The Role of Morphology in Learning to Read Written Words



A thesis submitted to the Department of Experimental Psychology, University of Oxford, in partial fulfilment for the degree of

Doctor of Philosophy

Hilary 2024

Acknowledgement

I would like to thank my supervisors, Professor Kate Nation and Dr Nicola Dawson. Kate and Nicky have been exceptionally supportive and insightful, guiding me from a morphology novice to a near-expert. Not only are they brilliant psychologists, they are also powerful, inspiring, caring, and wise women. I feel extraordinarily fortunate to have crossed paths with them, and I will always be honoured to have been their student.

Thank you to my wonderful colleagues for the stimulating discussions, shared cakes, friendly chats, and solid friendships in the office. Sean, Yuzhen, Jinyu, Nicole, Tonia, Yaling, Ellen, and Matthew, you've each carved out a special place in my heart and I'm glad to write your names on this page. If I had my way, I'd remain a DPhil student forever—if only to keep such good company.

I would like to thank Professor Gaia Scerif, whose guidance and encouragement kept my work on the right track. Seeing you in the office always made me feel encouraged!

Thanks to my friend Tongyu, who completed her dissertation at the same time, with whom I shared mutual encouragement during the final stages of writing.

I dedicate this thesis to my mentor, role model, and dear friend, Professor Kathy Sylva. Kathy, I wouldn't be the person I am today without your support and guidance. I miss our walks, chats, and bike rides dearly.

To my parents, XiaoLiang Zhang and JiRong Jiang, for their unconditional love and support.

To my dear husband, Yang Zheng, who is not only my best friend but also the love of my life.

Thank you to everyone who helped with data collection and to the schools and participants who took part in my research.

I am grateful for the valuable feedback from my examiners, Professor Elizabeth Wonnacott and Professor Jessie Ricketts, which made a difference to the final version of the thesis.

Lastly, I thank myself for persevering through the ups and downs. I was going to end this journey with an exclamation mark, but a period seemed more fitting. "I" standing above this full stop makes it into an exclamation. For now, my friend, if you are reading this, let us appreciate this simple, round and shining period together.

Abstract

This thesis investigates morphological processing in reading, examining morphology's role in linking print to meaning within alphabetic systems. It suggests that developing readers' systems become increasingly attuned to morphological regularities through accumulated reading experience. The findings underscore that the automatic activation of these structures is crucial in transitioning from novice to expert reader. Additionally, the efficacy of remote online testing for psycholinguistic research is evaluated, a vital aspect given the shifts towards online methodologies. A systematic review (Chapter 2) of 90 studies across Indo-European languages reveals the complexities of morphological processing, influenced by item-specific and reader-specific factors such as structure, frequency, age, grade, and reading ability. This overview sets the stage for the empirical studies that follow.

Experiment 1 (Chapter 3) tests the replication of morphological and pseudomorphological priming effects online, confirming that morphological effects are replicable remotely, while pseudomorphological effects are not, likely due to variabilities in online data collection. This outcome highlights the need to select appropriate paradigms for remote testing. Experiments 2 and 3 (Chapter 4) investigate the morpheme interference effect in both adults and adolescents, demonstrating that increased reading skills and greater suffix familiarity significantly enhance the morpheme interference effect, indicating that reading experience facilitates morpho-orthographic decomposition.

Experiment 4 (Chapter 5) explores whether induced suffix familiarity can enhance morphological interference, attempting to establish a causal link between suffix experience and morpho-orthographic processing. The findings suggest that one-session exposure is not as effective in enhancing suffix learning as prolonged reading experiences are.

Overall, this thesis expands our understanding of morphological processing and its developmental progression during an underexplored period, adolescence. It also offers insights into the methodological considerations of conducting psycholinguistic research remotely.

Contents

Acknowledgement **1**

Abstract **2**

Contents **4**

List of Figures **8**

List of Tables **10**

Chapter 1. Introduction **12**

1.1 The role of morphology in word reading 12

1.2 Theoretical models of morphological processing 15

1.3 Morphological processing in developing readers 19

1.4 Research aims and thesis overview 21

Chapter 2. Morphological processing and reading development — a systematic review **25**

2.1 Introduction 25

2.2 Method 30

2.2.1 Search overview 30

2.2.2 Search strategy and selection process 30

2.2.3 Study inclusion and exclusion 32

2.2.4 Coding 36

2.3 Results 37

2.3.1 Study Characteristics 37

2.3.2 Synthesis 40

2.3.3 Task-level Modulation 41

2.3.3.1 Reading aloud 41

2.3.3.2 Lexical decision 43

2.3.3.3 Masked priming 44

2.3.3.4 Other paradigms 48

2.3.4 Item-level factors 50

2.3.4.1 Morphological structure 50

2.3.4.2 Intactness vs. disruption 54
2.3.4.3 Stem vs. affix 57
2.3.4.4 Frequency 60
2.3.5 Individual-level factors 67
2.3.5.1 Age and grade: developmental milestones and years of schooling 68
2.3.5.2 Reading abilities: experience and individual variability 74
2.3.5.3 Reading difficulties 79
2.3.6 Cross-language differences 84
2.4 Summary and conclusion 87
2.5 Limitations and future directions 90
Chapter 3. Morphological effect in masked priming – an online replication study 93
3.1 Introduction 93
3.2 Method 95
3.2.1 Participants 95
3.2.2 Materials 96
3.2.3 Procedure 98
3.3 Results 99
3.3.1 Accuracy and RTs (above 200ms) 99
3.3.2 RTs between 200 – 2000ms 103
3.3.3 Posthoc power simulation 105
3.4 Discussion 107
Chapter 4. Effects of reading ability and suffix familiarity on morpheme interference effect in developing readers 111
4.1 Introduction 111
4.2 Experiment 2 115
4.2.1 Method 115
4.2.1.1 Participants 115
4.2.1.2 Materials and procedure 116
4.2.2 Results 119
4.2.2.1 Accuracy 120
4.2.2.2 RTs 121
4.2.3 Experiment 2 discussion 124

4.3 Experiment 3 125
4.3.1 Method 126
4.3.1.1 Participants 126
4.3.1.2 Material and procedure 126
4.3.2 Results 127
4.3.2.1 Accuracy 127
4.3.2.2 RTs 133
4.3.2.3 Exploratory analysis using age as a continuous predictor 137
4.3.3 Experiment 3 discussion 138
4.4 General discussion 141

Chapter 5: Influence of suffix exposure during reading on morphological processing 149

5.1 Introduction 149
5.2 Method 153
5.2.1 Sample size rationale 153
5.2.2 Participants 154
5.2.3 Materials 155
5.2.4 Procedure 160
5.3 Results 162
5.3.1 Accuracy 163
5.3.2 RTs 166
5.4 Discussion 169

Chapter 6. General discussion 176

6.1 Summary of findings 176
6.1.1 Reading ability and development of morphological processing 179
6.1.2 Revisiting the frequency effect 182
6.2 Implications 185
6.3 Limitations and future directions 189
6.4 Conclusion 191

References 192

Appendix for Chapter 2 218

Appendix 2A: Reading aloud 218
Appendix 2B: Lexical decision 222
Appendix 2C: Masked priming 225
Appendix 2D: Other paradigms 227

Appendix for Chapter 3 229

Appendix 3A. Experimental Stimuli 229
Appendix 3B: Contrast Comparisons for RT (200 – 2000ms) Analysis 232

Appendix for Chapter 4 233

Appendix 4A. Items of reading ability assessment 233
Appendix 4B. Lists of experimental stimuli for the lexical decision task 234
Appendix 4C. Model summary for accuracy analysis of Experiment 2 236
Appendix 4D. Model summary for RT Analysis of Experiment 2 237
Appendix 4E. (Experiment 2) 238
Appendix 4F. (Experiment 3) 239
Appendix 4G. Accuracy model output for Experiment 3 241
Appendix 4H. (Experiment 3) Two-way interaction between Age group × Suffix familiarity on Accuracy 242
Appendix 4I. RT Model output for Experiment 3 243
Appendix 4J. (Exploratory analysis) Accuracy model output of the exploratory analysis 244
Appendix 4K. (Exploratory analysis) 245
Appendix 4L. (Exploratory analysis) 247

Appendix for Chapter 5 248

Appendix 5A: Prevalence value of each word containing the target suffixes used in the stories during suffix exposure phase. 248
Appendix 5B: Stories Used across Three Exposure Conditions 249
Appendix 5C: Stimuli Lists (A and B) for the Pre- and Post- Lexical Decision Task. .. 263
Appendix 5D. Detailed Summary of Accuracy Model 264
Appendix 5E. Pairwise Comparisons of Accuracy (Bonferroni Corrected) 265
Appendix 5F. Detailed Summary of RT Model 266
Appendix 5G. Pairwise Comparisons of RTs (Bonferroni Corrected) 267

List of Figures

Figure 2.1: Flow Diagram of the Screening Process of Papers (* = Deviations from pre-registration) 34

Figure 2.2: Growth in Studies Investigating Morphological Processing in Developing Readers from 1981. Number of publications each year: 1981 (n=1), 1987 (n=1), 1989 (n=1), 1992 (n=2), 1995 (n=1), 1996 (n=2), 2000 (n=3), 2001 (n=1), 2002 (n=2), 2003 (n=3), 2004 (n=1), 2005 (n=1), 2006 (n=2), 2008 (n=3), 2009 (n=2), 2010 (n=1), 2011 (n=7), 2012 (n=3), 2013 (n=6), 2014 (n=5), 2015 (n=6), 2016 (n=3), 2017 (n=7), 2018 (n=6), 2019 (n=4), 2020 (n=3), 2021 (n=10), 2022 (n=3). 38

Figure 2.3: Counts of Language Investigated 39

Figure 2.4: Age (groups) Investigated in the Included Studies 40

Figure 3.1: Boxplot of Raw RTs (> 200ms) 101

Figure 3.2: Boxplot of Raw RTs (between 200-2000ms) 104

Figure 3.3: Power Curve from 500 Simulations 106

Figure 4.1: Two-way interactions between Condition × Suffix familiarity (i.e., log 10 transformed frequency of suffix in the OCC) for analysis of accuracy 121

Figure 4.2: Two-way interactions for RT analysis between: (a) Condition × Reading ability (i.e., ROAR score) and (b) Condition × Suffix familiarity (i.e., log 10 transformed frequency of suffix in the OCC). 123

Figure 4.3: Accuracy: three-way interaction of Condition × Suffix familiarity × Reading ability: (a) grouped by Reading ability; (b) grouped by Suffix familiarity (See Appendix 4F for the interaction plot grouped by individual suffixes). 130

Figure 4.4: Accuracy: three-way interaction of Condition × Suffix familiarity × Age group: (a) grouped by Age groups; (b) grouped by Suffix familiarity 131

Figure 4.5: RTs: three-way interaction of Condition × Reading ability × Suffix familiarity: (a) grouped by Reading ability: low, medium, and high reading ability; (b) grouped by Suffix familiarity: low, medium, and high suffix familiarity based on the ascending rank of suffix frequency (four suffixes per group; See Appendix 4F for the interaction plot grouped by each of the 12 individual suffixes). 135

Figure 4.6: Three-way interaction of Condition × Reading ability × Age in month 138

Figure 4.7: Reading ability across Age Ranges 138

Figure 5.1: Proportion Correct for Pseudomorphemic and Control Conditions Grouped by Each Suffix (Pre- and Post-tests Combined)..... 164

Figure 5.2: Proportion Correct for Pseudomorphemic and Control Nonwords in Pre- and Post-test (the lines represent the medians)..... 165

Figure 5.3: Proportion Correct in Different Exposure Conditions in Pre- and Post-test 166

Figure 5.4: Pre- and Post- test Performance across Exposure Conditions for Different Nonword Types..... 166

Figure 5.5: RTs (pre- and post-tests combined) for Pseudomorphemic and Control Nonwords Grouped by Suffix (the lines represent the means)..... 167

List of Tables

Table 2.1: Search terms derived from the three key components 31

Table 3.1: Means (SDs) for Lexical Characteristics of Primes and Targets across Conditions..... 97

Table 3.2: Means (SDs) of Raw RTs and Accuracy Across Conditions and Prime Types..... 99

Table 3.3: Model Output for Accuracy Analysis 100

Table 3.4: Model Output for RT (>200ms) Analysis 101

Table 3.5: Model Output for Analysis of RTs (200 – 2000ms, Log-transformed) 104

Table 3.6: Power for 'Prime type × Condition (pseudomorphological)' by Number of Participants 107

Table 4.1: Frequency (per million words) of Each Suffix in OCC 117

Table 4.2: Means (SD) for Linguistic Characteristics of Nonword Across Conditions..... 117

Table 4.3: Means (SDs) of RTs (untransformed), Accuracy, and Raw ROAR scores. 120

Table 4.4: Output for Accuracy Model 121

Table 4.5: Output for RT Model 122

Table 4.6: Means (SDs) of Accuracy (rate), RTs (untransformed), and Raw ROAR scores..... 128

Table 4.7: Output for Accuracy Model 128

Table 4.8: Effects of covariates on the morpheme interference effect for different reader and age groups and different suffix familiarity levels (*Contrast = Pseudomorphemic− Control*) 132

Table 4.9: Output for RT Model 134

Table 4.10: Effects of covariates on the morpheme interference effect for different reader groups and different suffix familiarity levels (*Contrast = Pseudomorphemic− Control*) 136

Table 5.1: Means (SDs) of Word Prevalence Values for Exposure Words containing Exposed Suffixes, across High- and Low-diversity Exposure Conditions. 156

Table 5.2: Examples of the First Three Fragments from Stories Featuring '-let' Suffix Exposure Words and Corresponding Comprehension Questions across Three Conditions 158

Table 5.3: Means (SDs) for Lexical Characteristics of Nonwords and Their Stems for Pseudomorphemic and Control Nonwords 160

Table 5.4: Means (SDs) of Accuracy (Proportion Correct) for Pseudomorphemic and Control nonwords. 163

Table 5.5: Accuracy Model Output for the Main Effects 165

Table 5.6: Means (SDs) of Raw RTs (ms) for Pseudomorphemic and Control Nonwords. 168

Table 5.7: RT Model Output for the Main Effects 168

Chapter 1. Introduction

1.1 The role of morphology in word reading

Reading plays a fundamental role in communication, literacy, and personal development, and is essential for knowledge acquisition, closely linked to academic, economic, and social success (Castles et al., 2018). Given that words are the most informative units of language, understanding the cognitive processes that underpin word reading is vital. The Simple View of Reading (Gough & Tunmer, 1986) posits that reading comprehension is the product of both word recognition and language comprehension abilities, highlighting the critical role of word reading in reading comprehension. Learning to activate meanings from symbols is the ultimate goal of reading acquisition (Rastle, 2019a). Various conceptual frameworks have attempted to explain how readers access meaning from print. For instance, the connectionist 'triangle model' (Harm & Seidenberg, 2004; Seidenberg, 2005) suggests that this process relies on the interplay between three key components: orthography, phonology, and semantics. Phonological decoding, which links orthography to phonology (O-P), is crucial in the initial stage of reading development, as noted by Share (1995). Having established the mapping between phonology and semantics (P-S) through spoken word recognition, phonological decoding (O-P) helps children indirectly activate meaning from written words via the phonological pathway (O-P-S). With increased reading experience, children can directly map orthography onto semantics (O-S). Reading development is thus explained as a transition from the O-P-S pathway to the more direct O-S pathway. Familiar written words can be rapidly recognised from their spellings, without the need for phonological decoding (see Nation, 2009, for a review). As morphemes are defined as the smallest meaningful constituents of words (Hamawand, 2011), morphology facilitates the establishment of direct mappings between written words and their meanings (Rastle, 2019b; Ulicheva et al., 2020).

The majority of words in Indo-European languages contain multiple morphemes—about 85% in English, 75% in French, and 80% in German (Marslen-Wilson & Tyler, 2007; Grainger & Ziegler,

2011; Segbers & Schroeder, 2017). Developing readers are more likely to be exposed to morphologically complex words as they get older. According to a large-scale analysis of texts targeted at English-speaking children and adolescents aged 5–14 years (Dawson et al., 2023), it was revealed that the proportion of morphologically complex words increases as a function of age. These words can combine stem morphemes, such as 'rain', with other stems to form compounds like 'raincoat', or with affix morphemes to create words like 'rainy' and 'rained' (Rastle, 2019a). There are inflectional and derivational morphemes—inflectional morphemes create different forms of the same word to make it fit various grammatical contexts (e.g., rain, rains, rained, raining), whereas derivational morphemes transform the word to express new concepts. For example, adding the suffix '-y' to 'rain' changes the stem from a noun into an adjective (Blevins, 2006). Recognition of morphological regularities allows for the easy recognition of multimorphemic words such as 'teacher' by analysing their components (e.g., teach + er). Besides aiding word recognition, morphology has the capability to create new words through morphological elements, a characteristic referred to as the productivity of morphological processes[1] (Bauer, 2001, pp.12-13). An example is the word 'netiquette', coined by combining 'net' and 'etiquette' to describe the proper conduct for online communication.

In English, morphological regularities tend to be far more salient and consistent in spelling than in spoken language, which may come at the cost of the spelling-sound irregularities at the grapheme-phoneme level (Berg & Aronoff, 2017; Rastle, 2019a). For instance, Carney (1994, p. 18) demonstrated that the pronunciations of the English past tense vary according to the surrounding phonological context (e.g., /id/ in 'started', /d/ in 'rained', and /t/ in 'watched'). Despite these phonemic variations, the regular form of past-tense morpheme has a constant spelling '-ed'.

[1] Productivity has been assigned to particular affixes, or morphological processes, or the complete module of grammar, or even the language system as a whole (see Bauer, 2001, pp.12-14, for a review). Here, the perspective that morphological processes are productive is adopted, but suffix productivity is also explored in other chapters of the thesis.

Conversely, homophonous word endings are often spelt differently (e.g., /əs/ in ‘marvellous’ and ‘cactus’). This allows some English suffixes to carry unique information that is not available phonologically (Berg & Aronoff, 2017). For example, adjectives that end in /əs/ are almost always spelt ‘-ous’, even though /əs/ is spelt in several other ways in different word types (e.g., ‘cactus’, ‘corpus’). In this way, suffix spellings are distinct from their homophones and can provide cues to lexical categories (Ulicheva et al., 2020). Thus, in written word recognition, morphology offers ‘islands of regularity’ amidst the irregularities of sound-spelling mapping (Rastle, 2019b), providing direct and predictable connections from text to meanings. As noted earlier, this further underscores the critical role of morphology in enabling orthography-to-semantics mapping, which aids in transitioning from novice to expert reader (Castles et al., 2018; Rastle, 2019b).

As unique linguistic units, morphemes introduce systematicity to the otherwise arbitrary connection between word form and its meaning, supporting word processing in two key ways: firstly, words sharing the same stem (morphological families) often possess similar meanings, and secondly, affixes systematically alter the meanings of words in highly predictable ways, providing cues to word class and meaning (Ulicheva et al., 2018). According to Perfetti (2007), effective word decoding relies on how lexical representations are stored in the mental lexicon, which is characterised by five features: orthography, phonology, morpho-syntax, meaning, and constituent binding—the extent to which these features are integrated. Morphemes, which possess phonological representations and carry grammatical cues, can strengthen the connections among these linguistic components, namely orthography, semantics, phonology and grammar. By doing so, morphology helps create more specified and precise mental representations of words, thus improving their lexical quality (the ‘lexical quality hypothesis’; Perfetti, 2007). Reichle and Perfetti (2003) outlined a computational model that suggests the familiarity and availability of the stem serve as a proxy for high lexical quality. The results demonstrated that derived word stems being more readily available when encountered with various affixes. Consequently, morphology supports word identification

processes not only through direct links from orthography to semantics but also by fostering the development of high-quality lexical representations through tightly linking orthographic, phonological, and semantic features among morphologically related words (e.g., Reichle & Perfetti, 2003). Therefore, morphological knowledge enhances literacy as higher-quality representations are readily retrieved during reading, freeing up cognitive resources for higher-level processes such as integration and comprehension (Bowers et al., 2010; Nagy et al., 2014; Nation, 2017).

Readers with greater reading expertise typically possess a higher average lexical quality across their vocabulary. Moreover, for each reader, some words are better known, indicating higher lexical quality, while others are relatively less familiar, reflecting lower lexical quality. Thus, lexical quality operates at both individual and item levels (Nation, 2017). By strengthening the connection between orthographic and semantic features of words in the mental lexicon, sensitivity to morphology may enhance the quality of lexical representations (see Castles et al., 2018; Nation, 2017, for a review). Such 'high-quality' lexical representations facilitate rapid and automatic word recognition with minimal conscious effort, thereby conserving cognitive resources for comprehension (Castles et al., 2018). While phonology and the mappings from orthography to phonology have traditionally received more focus in the literature on reading development, only in the past few decades has a growing body of research begun to pay more attention to the direct link between orthography to semantics (Rastle, 2022).Understanding how morphologically complex words are efficiently recognised during reading and how this ability develops is essential for a comprehensive understanding of the reading process.

1.2 Theoretical models of morphological processing

Morphological awareness tasks are commonly used to measure explicit morphological knowledge by asking participants to consciously generate, analyse, or manipulate morphological forms. Morphological awareness has been extensively studied (e.g., Kuo & Anderson, 2006). In contrast, morphological processing refers to the implicit activation of morphological knowledge

during real-time reading (Breadmore & Carroll, 2016). There has been a heated discussion concerning the theoretical accounts of morphological processing, i.e., how morphological information is represented and activated during visual word recognition (see Amenta & Crepaldi, 2012, for a review). The distributed connectionist approach views morphology as a reflection of statistical regularities in the activation across orthographic, phonological, and semantic units (Gonnerman et al., 2007; Rueckl, 2010). Distributed models, often informed by artificial neural networks, propose that explicit representations of morphological units (i.e., morphemes) are unnecessary. Instead, morphological sensitivity is acquired through statistical regularities in the form-to-meaning mappings of complex words (see Stevens & Plaut, 2022, for a review). This perspective suggests that morphological structures are processed in a graded manner, reflecting a varying degree of systematicity among semantic, phonological, and orthographic representations (Plaut & Gonnerman, 2000).

Unlike connectionist models, the localist approach advocates for modelling the lexicon with representations of individual morphological units. According to localist accounts, words are systematically broken down into their morphological components during processing (e.g., Rastle & Davis, 2008). Although all localist accounts suggest that discrete morphemic units are represented within the mental lexicon, they disagree on the stage at which morphological representations are activated. As a result, three main accounts have been proposed: supralexical (e.g., Giraudo & Grainger, 2001), sublexical (e.g., Rastle & Davis, 2008), and the dual-route accounts (e.g., Baayen & Schreuder, 2000). The supralexical account argues that morphological analysis happens after the activation of the whole-word representation (e.g., Giraudo & Grainger, 2001, Marslen-Wilson et al., 1994). The retrieval of the whole-word representation activates its morphemic components (stem and affixes), which then leads to the activation of all other words containing these components (Giraudo & Grainger, 2001). In contrast, the sublexical account suggests that morphological decomposition occurs prior to whole-word access. This account states that initially letter strings

with a surface morphological structure are automatically decomposed into their constituent morphemes based on their orthographic form, irrespective of the semantic relationship between the whole complex word and its stem (i.e., morpho-orthographic segmentation; Rastle et al., 2004). This obligatory decomposition process is commonly referred to as 'form-then-meaning' processing (Rastle & Davis, 2008). The dual-route account, as its name implies, postulates that morphologically complex forms are recognised through the interplay of supralexical and sublexical representations, whereby whole-word retrieval and morpho-orthographic decomposition occur simultaneously (e.g., Diependaele et al., 2009; Grainger & Ziegler, 2011). For example, the Augmented Addressed Morphology (AAM) account (Caramazza et al., 1988) suggests that whole-word lexical access is generally faster for known words, whereas decomposition is employed for unfamiliar words that have not been encountered previously and which cannot be directly activated via whole-word representations.

Research on morphological processing has widely adopted the masked priming paradigm. Here, a prime is displayed briefly (around 40-70ms) so as to be subconscious and is then replaced by a target. Rastle and colleagues (2004) asked participants to make lexical decisions about visual targets using different types of prime-target pairs. Compared to a nonmorphological form overlap condition (e.g., freeze-FREE), significant priming effects (mainly manifested by faster reaction times, RTs) were found in the truly morphological condition (e.g., bomber-BOMB) as well as in the pseudomorphological condition (e.g., corner-CORN). The absence of facilitation in the form-only overlap condition indicated that the priming effect could not simply be due to orthographic overlap. Instead, equivalent priming effects in the truly morphological and the pseudomorphological conditions suggested that words with apparent morphological structure were initially processed at the morpho-orthographic level, blind to semantics (Rastle et al., 2004, 2008). However, whether semantics play a role in the initial stage of processing is still under debate. Several studies (e.g., Beyersmann et al., 2012; Feldman et al., 2009; Diependaele et al., 2009) have reported larger

priming effects in the truly morphological condition than in the pseudomorphological condition. Furthermore, the morphological family size effect (e.g., Beyersmann & Grainger, 2018) showed that embedded target words with several morphologically related family members (e.g., 'serpentoche–SERPENT' [English: 'snakerel–SNAKE'], with family members like 'serpents', 'serpentin', etc.) exhibited a greater priming effect than those with fewer (e.g., dauphingri–DAUPHIN, which only includes the plural form as a family member). These findings indicate that semantic factors might be at play from an initial stage of processing. Despite this ongoing debate, the pseudomorphological effect provides evidence against the fully supralexical account, as priming for words with surface morphological structure (e.g., CORN+er) can only arise through sublexical decomposition.

To isolate the influence of whole-word semantics and further investigate morphological processing, consider evidence from lexical decision tasks that involve nonwords (e.g., Burani et al., 2002; Casalis et al., 2015; Dawson et al., 2018). Nonwords cannot activate whole-word lexical representations or undergo post-lexical analysis. For instance, Morris et al. (2011) demonstrated that recognition of the target stem was significantly primed by morphologically structured nonwords (e.g., flexify- FLEX) as well as by non-suffixed nonwords (e.g., flexint- FLEX). This observation challenges theoretical accounts that assert morphological processing requires initial whole-word activation, since nonwords are not represented in the mental lexicon. Moreover, the full decomposition mechanism also falls short of explaining the priming by nonwords like 'flexint' that lack an affix structure to decompose.

To account for these findings, Grainger and Beyersmann (2017) proposed an 'edge-aligned embedded stem activation' account, which posits that stems embedded at the beginning or end of letter strings can map directly onto whole-word representations and be activated independently of affix presence. The intensity of activation depends on the length of the embedded word: the orthographic representation of a word (e.g., 'teacher') receives the most activation, followed by the

longest embedded stem, e.g., ‘teach’ (see Grainger & Beyersmann, 2017, for a review). This account advocates for early morpho-semantic activation in morphological processing. For words, the morpho-semantic relationship between the whole word and the embedded stem (e.g., ‘teacher - TEACH’) enhances the co-activation of both the whole word and the stem. In the case of pseudomorphological priming (e.g., ‘corner - CORN’), when embedded word activation is inhibited due to different semantics between the whole word (e.g., ‘corner’) and the embedded stem (e.g., ‘corn’), the activation of affix representations (e.g. ‘-er’) can offset the inhibition and enhance the embedded stem activation. Such an effect is absent in non-suffixed words like ‘cashew’ where ‘-ew’ is not a suffix, or when affix representations are not well-established, as may often be the case in developing readers. For nonwords, as they lack whole-word representations in the mental lexicon, the embedded stems can be activated directly without inhibition.

The theoretical accounts of morphological processing discussed are largely based on studies with adult skilled readers. Despite varied perspectives, the accounts agree that skilled, automatic morphological processing of written words develops through extensive reading experience. Investigating whether developing readers follow similar processing patterns and understanding how they process morphologically complex words as a function of reading experience can provide deeper insights into the role of morphology in learning to read and the progression from novice to expert reader.

1.3 Morphological processing in developing readers

From an early age, preschool children already show sensitivity to morphological regularities in spoken language. For example, Berko (1958) found that 4- to 7-year-olds could orally generalise English morphological rules to novel words (e.g., ‘wug’) under varying phonological conditions (e.g., the plural ending ‘-s’ can be pronounced /s/ or /z/). The Berko task captures morphological awareness, which in young children often stems from implicit understanding or automatic use of morphological knowledge (see Carlisle, 2005, for a discussion), involving explicit production of

morphological inflections. While morphological awareness has been extensively considered in association with reading development (e.g., Kuo & Anderson, 2006), relatively little research has investigated implicit morphological processing as children read written words in real time. For instance, to understand how morphological knowledge is implicitly activated during online word reading, Carlisle and Stone (2005) found that from Grades 2 and 3 (7-9 years), children read aloud derived words like 'hilly' faster and more accurately than matched monomorphemic words like 'silly', indicating that children are sensitive to morphological structure during reading. Similarly, studies using masked priming and lexical decision paradigms with developing readers have found that developing readers show some sensitivity to morphological regularities during reading, but exhibit different patterns of morphological processing compared to adults (e.g., Beyersmann et al., 2012; Dawson et al., 2018; Hasenäcker et al., 2016). For instance, the relatively robust pseudomorphological priming observed in adults was absent in English preschool children (Beyersmann et al., 2012). Although there is no definitive answer as to when developing readers begin to exhibit adult-like automatic and rapid morphological processing, there is evidence to suggest that this develops late in reading development, potentially extending into middle or high school (e.g., Beyersmann et al., 2019; Dawson et al., 2018).

Although the literature is small, there is evidence to indicate that various individual- and item-level factors can influence children's morphological processing during reading. At the individual level, reading ability, for instance, has been shown to influence morphological processing, independent of age or grade level (e.g., Hasenäcker et al., 2020). At the item level, suffix frequency has been found to influence the magnitude of priming effects (e.g., Beyersmann et al., 2012) and cross-linguistic differences may account for some of the different patterns of findings reported across studies (e.g., Quémart et al., 2011). Furthermore, despite the limited number of studies conducted with children, task demands vary widely from study to study. For instance, some studies have used masked priming while others employed unmasked lexical decision. Different types of

stimuli are used across studies too, with some investigations using words (e.g., Beyersmann et al., 2012; Quémart et al., 2011), and others using nonwords or pseudowords (e.g., Burani et al., 2002; Casalis et al., 2015; Dawson et al., 2018). It is difficult to draw strong conclusions from this mixed evidence base. Rather than provide a narrative review, I decided to conduct a comprehensive systematic review on morphological processing and reading development, with the aim of clarifying how individual (child/reader)-, item-, and task- level factors influence children's reading of morphologically complex words. This review forms a substantial part of this thesis and is reported in Chapter 2.

1.4 Research aims and thesis overview

As briefly discussed above, morphology can provide a direct link between print and meaning in alphabetic writing systems, owing to the meaning-bearing and regular nature of morphemes (e.g., Rastle, 2019a). Skilled reading is characterised by the rapid, effortless, and automatic activation of these morphological regularities during reading, and the acquisition of sensitivity to morphological regularities marks the transition from novice to expert, as children learn to read (e.g., Castles et al., 2018; Nation, 2009). Although extensive research has been conducted on other aspects of reading development, such as phonological decoding and the relationship between phonological processing and learning to read (e.g., Melby-Lervåg et al., 2012), morphological processing remains relatively underexplored, with much still to be discovered about its development and the underlying mechanisms that drive it.

One notable feature of the existing evidence bases is the suggestion that morphological processing during visual word recognition continues to develop well into adolescence (e.g., Dawson et al., 2018). As research on reading development during adolescence is itself scarce (e.g., Ricketts et al., 2020), the questions addressed in this thesis aim to inform this topic.

Chapter 2 is a systematic review of morphological processing in developing readers across Indo-European languages. As previously discussed, the complex nature of morphological

processing, the variation across studies, and the various factors influencing it necessitate a comprehensive understanding of previous findings. This review includes 90 studies that cover a wide age range, from early childhood to late adolescence (6-18 years). Although the experiments in the thesis (Chapters 3-5) are conducted with British English-speaking readers, the review included all studies on Indo-European languages that met the systematic search criteria, as set out in Chapter 2. Those languages are similar in morphological paradigm (cf., Fortson, 2011). Therefore, our understanding of morphological processing in English can be informed by findings on other languages and cross-linguistic variations such as orthographic transparency within these languages contribute to a deeper and more comprehensive understanding of morphological processing (e.g., Casalis et al., 2015).

Due to the pandemic, the experiments in this thesis were conducted on an online experiment platform, Gorilla. Given the potential differences between remote online testing and traditional in-person lab-based testing (e.g., Rodd, 2024), it is crucial to validate this paradigm in an online environment first. Experiment 1 in Chapter 3 is a masked priming experiment conducted through remote online testing in adults and older adolescents aged 16 and 17 years. While it is well established that masked priming is an appropriate methodology to use with older children and adults, it requires precise timing in terms of both stimulus presentation and recording responses. An important step is therefore to ask whether the findings from in-person studies can be effectively replicated via remote testing before employing this approach to tackle further research questions.

Chapter 4 comprises two lexical decision experiments that probe the morpheme interference effect in both adults (Experiment 2) and adolescents (Experiment 3). This effect has been taken as evidence of morpho-orthographic decomposition via rapid analysis of morphological structure. In brief (see Chapters 2 and 4 for detailed reviews), skilled readers find it harder to reject pseudomorphemic nonwords (e.g., ‘earist’) in lexical decision compared to control nonwords without a morphological structure such as ‘earilt’ (Crepaldi et al., 2010). Dawson et al. (2018)

suggested that such rapid morphological decomposition occurred only until late adolescence. Presumably, the emergence of the morpheme interference effect is influenced by reading experience as extensive interaction with written suffixes might be needed to facilitate morpho-orthographic processing. To investigate how reading experience shapes morphological processing in visual word recognition, experiments in Chapter 2 examine the influence of reading experience on the morpheme interference effect. Reading ability and suffix familiarity (indexed by the frequency in children's reading material) are used as measures of reading experience at both individual and item levels. Experiment 3 also contributes to the understanding of reading development in an under-explored age group, adolescents. Additionally, the lexical decision paradigm does not require the strict timing of the masked priming paradigm, and this might make it more suitable for remote testing via Gorilla.

Experiment 4 reported in Chapter 5 further examines the role of suffix familiarity in morphological decomposition and attempts to establish a causal link between the frequency of suffix exposure and the implicit activation of morphological knowledge during word reading. Training studies that control for exposure frequency and manipulate context can provide valuable insights into the frequency effects on word learning (e.g., Mak et al., 2021). My experiment implements a training paradigm where suffixes appeared in varied words (e.g., '-ee' in 'employee') presented in short stories. The rationale was to induce variations in suffix familiarity, and measure is how this influences morphological processing by comparing the morpheme interference effect for nonwords containing the target suffixes pre- vs. post-exposure. In Experiment 4, suffixes possess even lower frequencies than those that did not elicit a morpheme interference in Experiment 3 in Chapter 4. The aim was to examine whether repeated exposure to these low-frequency suffixes within a single session can induce or enlarge a morpheme interference effect in nonwords containing these suffixes. Experiment 4 attempts to illuminate whether suffix experience underpins the development of skilled morphological decomposition, providing further insights into how word

reading is shaped by reading experience shapes and factors that might influence the learning of suffix representations.

Chapter 2. Morphological processing and reading development — a systematic review

2.1 Introduction

Reading entails deriving meaning from written symbols. Morphemes, the smallest meaning-bearing units in a language, facilitate this via systematic form-meaning correspondences, thereby creating a direct link between a written word and its meaning (Ulicheva et al., 2020). Beyond aiding in the comprehension of complex words, morphemic knowledge facilitates the generation of an 'unlimited' number of words from a finite set of morphological components, contributing to the productivity of a language. The vast majority of words in Indo-European languages (e.g., English, French, German, Italian) are constructed from multiple morphemes, frequently employing prefixes, suffixes, and infixes to convey meaning (Grainger & Ziegler, 2011; Segbers & Schroeder, 2017). These languages tend to be inflectional, standing in contrast to some non-Indo-European languages like Chinese and Korean. Despite their inflectional and concatenative nature, Indo-European languages generally contain fewer morphemes per word compared to agglutinative languages, such as Turkish or Finnish (Smith, 1995). While cross-linguistic differences exist within the Indo-European language family, this systematic review includes all Indo-European languages in its scope, due to their shared alphabetic writing system and morphological similarities.

Morphological regularities manifest in both spoken and written forms of language, but in English at least, are more consistent in the latter (Berg & Aronoff, 2017; Rastle, 2019). Despite phonemic variations, the spelling of suffixes is consistent across words (e.g., -ed pronounced as /id/ in ‘started’, /d/ in ‘rained’, and /t/ in ‘watched’) and is also distinguishable from their homophonic counterparts (e.g., adjectives ending in /əs/ are typically spelled ‘-ous’, despite alternative spellings like in ‘corpus’). These morphological patterns therefore provide ‘islands of regularity’ within the writing system (Rastle, 2019b). Acquiring sensitivity to morphological patterns is pivotal for efficient language processing and reading development, as morphology plays a key role in

establishing direct links between spelling and meaning and in doing, it is thought to facilitate the transition of children from novice to expert readers (Castles et al., 2018). The ability to explicitly recognise, consider, and manipulate morphemes, including base words and affixes, is usually referred to as morphological awareness (Apel, 2014; Kuo & Anderson, 2006). In contrast, morphological processing can be defined as the implicit processes that allow us to decipher morphologically complex words during the act of word reading (e.g., To et al., 2014). While morphological awareness can be broadly viewed as a subset of morphological processing (Deacon et al., 2008), tasks that measure awareness focus on explicit reflection. Thus, findings from these tasks cannot directly explain the processing of morphologically complex forms during reading. The significant role of morphological awareness in reading is well-established and reviewed elsewhere (e.g., Deacon et al., 2014; Kruk & Bergman, 2013) but our understanding of morphological processing during word recognition remains relatively scant, especially in children.

Studies have employed online methods, like masked priming using either word or nonword stimuli, to probe the implicit activation of morphological knowledge in adults. Yet, the mechanism for morphological processing has been a subject of extensive debate over the past few decades (cf. Amenta & Crepaldi, 2012; Rastle & Davis, 2008; Stevens & Plaut, 2022). The distributed connectionist approach advocates for morphology arising from learned sensitivities to statistical regularities traversing orthographic, phonological, and semantic dimensions (Gonnerman et al., 2007; Plaut & Gonnerman, 2000; Rueckl, 2010), treating morphological effects as graded and distributed. On the other hand, localist or decomposition accounts suggest that words undergo decomposition into morphemes during processing (Rastle & Davis, 2008). There are three main variations around decomposition. The supralexical account proposes morphological analysis post whole-word activation (Giraudo & Grainger, 2001); this has been challenged by studies revealing pseudomorphological priming effects (e.g., Rastle et al., 2004). The sublexical account states that decomposition precedes whole-word access (Rastle & Davis, 2008) and the dual-route account

suggests an interplay between both supralexical and sublexical representations (Grainger & Ziegler, 2011).

Like many complex cognitive abilities that develop over time, children exhibit differences in morphological processing compared with skilled readers (e.g., Dawson et al., 2018; Hasenäcker et al., 2016). While some knowledge comes to reading via spoken language experience (e.g., even preschoolers are sensitive to morphological patterns and can generalise them to new words, Carlisle, 1995), sensitivity to morphological patterns in written words takes time to develop. For example, the pseudomorphological priming effect (e.g., prime 'corner' facilitates a lexical decision response to 'CORN') has been consistently observed in adults, but the effect is absent in elementary school children (Beyersmann et al., 2012). The mechanisms that underly developmental changes remain largely unexplored, as does their developmental time course. Grainger and Beyersmann (2017) delineated a possible route towards achieving skilled adult-like morphological processing during reading. In brief, even before children start to read, pre-readers already possess morpho-semantic knowledge from exposure to morphologically complex words in spoken language. As they embark on formal reading instruction, they map newly learned orthographic forms onto their pre-existing morph-semantic representations. Free-standing stems are first learned because they are highly salient and carry the most semantic information of the complex words in which they are embedded. These mappings accelerate vocabulary growth as reading experience accumulates and pave the way for encoding affixes as morpho-orthographic units through embedded word activation, without requiring specialised morphological representations. Leveraging stem activation and the morpho-semantic information children have gathered via spoken language— such as understanding that a 'teacher' is somebody who teaches— children gradually establish affix representations for the remaining parts of complex words (e.g., -er in teacher) in the mental lexicon. As they encounter affixes across various lexical contexts (e.g., '-er' in 'teacher', 'singer', 'farmer', etc.), intensive reading experience helps fine-tune the representations of these affixes, eventually enabling these

affixes to be effortlessly activated during visual word recognition. Once affix representations are established, complex words can be processed through both the full-decomposition pathway (i.e., parsing complex words into their constituent morphemes) and the embedded stem activation pathway. As a result, pseudomorphological priming (e.g., corner activating CORN) emerges because the activation of the affix (-er) compensates for the inhibitory effect that stems from the lack of a semantic relationship between the embedded stem (e.g., 'CORN') and the whole word (e.g., 'corner').

While Grainger and Beyersmann's (2017) embedded word activation account offers a possible explanation, it is only one among several perspectives on the developmental trajectory of morphological processing. Other models, such as the connectionist accounts mentioned previously, also yield valuable insights. Despite their varied approaches, these models commonly agree on the pivotal role of reading experience in fostering rapid and automatic morphological processing skills. Reading experience is inherently multi-dimensional, involving a complex interplay between the reader and the reading materials (Castles et al., 2018). This multifaceted nature warrants an analysis of the role of reading experience in the development of morphological processing that considers both individual- and item-level factors. Collectively, these factors nurture a continuously evolving 'lexical legacy' that both shapes and is manifested by reading behaviours (Nation, 2017).

Morphological effects in developing readers are indeed shown to be influenced by various factors at the task, item, and individual child level. For instance, while individual reading ability is generally associated with grade level, the large inter-individual variance within each grade can result in diverse patterns of morphological processing (e.g., Hasenäcker et al., 2020). At the item level, some researchers use word stimuli (e.g., Quémart et al., 2011), whereas others employ nonwords (e.g., Burani et al., 2002; Casalis et al., 2015). Unlike words, nonwords lack lexical representations and are less associated with variables that could potentially affect the activation of morphological knowledge during reading, such as frequency, age of acquisition, or semantic

transparency (e.g., Carlisle & Fleming, 2003). Furthermore, task demands could be another influential factor. The lexical decision task, for instance, may be used with or without masked priming, influencing the interpretation of the time course of processing: Masked priming experiments capture the early stages of morphological processing, as words are read (e.g., Beyersmann et al., 2012), while non-priming lexical decision tasks may tap on a combination of both early and late stages (e.g., Dawson et al., 2018).

Several systematic meta-analyses and analytical narrative reviews have explored morphological processing in reading. None, however, have considered children. Instead, their focus has been with adult populations (Amenta & Crepaldi, 2012), specific languages (e.g., Spanish: Domínguez et al., 2000), distinguishing different theoretical approaches (Stevens & Plaut, 2022), and exploring neural mechanisms using certain methodologies including EEG, MEG, or fMRI (Bulut, 2022; Leminen et al., 2019). There are a reasonable number of individual studies with children, but as they use different paradigms, languages, stimuli, and age groups, it is difficult to draw any clear conclusions. It is striking that, to date, there has been no comprehensive systematic review of morphological processing in children learning to read Indo-European languages. This systematic review aims to bridge this gap. We focused on four preregistered research questions, namely:

1. How does sensitivity to morphological structure in written words develop and change through childhood and adolescence?
2. Do individual differences within an age group influence the morphological processing of written words?
3. Which stimulus-level factors modulate morphological processing of written words by children and adolescents?
4. Are patterns of morphological processing and its development task-dependent?

Having used these research questions to guide our literature search, if data were available, we also aimed to explore: (1) possible cross-linguistic differences (within Indo-European languages) in

morphological processing; (2) similarities and differences in processing different types of morphemes (e.g., stems vs. affixes; prefixes vs. suffixes) and (3) which skills early in development predict variance in the development of children's morphological processing.

2.2 Method

2.2.1 Search overview

The review was conducted following the NIRO-SR framework which sets guidelines for planning and conducting non-interventional, reproducible, and open systematic reviews (Topor et al., 2020). To ensure transparency, the protocol was preregistered on the PROSPERO database (Registration number: CRD42020202170; accessible at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020202170). According to the aims of this review, the inclusion and exclusion criteria were pre-specified to guide the subsequent literature search and screen.

2.2.2 Search strategy and selection process

The original search covered both published and unpublished grey literature up until November 2020. We did not restrict the earliest publication date to capture a comprehensive range of relevant literature. The following databases were searched for published literature: PsycINFO, Web of Science (both the core collection as well as the Social Science Citation Index (SSCI)), Scopus, MEDLINE, ProQuest Education Collection (incl. Education Database and ERIC), ASSIA Applied Social Sciences Index and Abstracts, British Education Index, ProQuest Linguistic collection (incl. Linguistics and Language Behavior Abstract (LLBA), and CINAHL (EBSCO), with no restrictions as to the earliest publication date (but up to November 2020). Grey literature, such as dissertations and conference proceedings, was searched using ProQuest Dissertation & Thesis Global, OpenGrey, and Conference Proceedings Citation Index-social Science& Humanities (via Web of Science).

A comprehensive search was conducted using combinations of search terms related to three key constructs: morphology, processing, and reading. Table 2.1 lists the search terms used for each construct, alongside the truncation and wildcard operators applied. To define the age range, age-specific filters in databases were employed where available. Articles were restricted to those published in English. An example of the Boolean search string used on Web of Science is: " TI ((morpholog* OR morphem* OR deriv* OR inflect* OR affix* OR suffix* OR prefix*) AND (process* OR prime OR priming OR segment* OR pars* OR decompos* OR decod* OR represent* OR identif* OR recogni*) AND (read* OR litera* OR language OR linguistic*)) OR AB((morpholog* OR morphem* OR deriv* OR inflect* OR affix* OR suffix* OR prefix*) AND (process* OR prime OR priming OR segment* OR pars* OR decompos* OR decod* OR represent* OR identif* OR recogni*) AND (read* OR litera* OR language OR linguistic*)) (Language: English; Age level: Children, Teenagers)".

Table 2.1: Search terms derived from the three key components

Concepts	Alternative Terms	Truncations
Morphology	morphology, morphological, morphemic, morpheme, derivation, derivational, derived, inflection, inflected, inflectional, affix(es), suffix(es), prefix(es), affixed, suffixed, prefixed	morpho*, morphe*, suffix*, affix*, prefix* deriv*, inflect*
Processing	processing, process, prime, priming, segment, segmentation, parse, parsing, decompose, decomposition, decode, decoding, represent, representation, identify, identification, recognise, recognition	process*, prim*, segment*, pars*, decompos*, decod*, represent*, identif*, recogni*
Reading and language	read, reading, reader, literacy, language, linguistic, linguistics	read*, litera*, language, linguistic*

In addition to database searches, backward citation searching was conducted, whereby the reference lists of included articles were manually scanned for relevant papers that was not extracted from the initial search. Furthermore, given the time that had elapsed since our initial search and data extraction, a forward citation search was undertaken in January 2022. This search targeted

publications from 2020 to the end of 2021 that cited the eligible articles from our original search, thereby reducing the possibility of overlooking more recent contributions. Moreover, an alert on Web of Science was set up to continuously updated the reviewers on relevant publications spanning from 2020 through December 2021. Both the backward and the forward citation searching were conducted by the primary reviewer (MZ).

2.2.3 Study inclusion and exclusion

To avoid over-specificity, the search was not limited by employing age-specific search terms in databases without age filter or using additional search terms related to Indo-European language. These decisions led to the large number of 12,397 studies identified. After duplicates were removed, the search yielded a total of 5508 records, which were then screened for eligibility, as depicted in Figure 2.1. The following inclusion criteria were applied: (1) studies published in English, (2) research focusing on Indo-European languages, (3) research focusing on children or adolescents aged between 6 and 17 years, (4) participants completed at least part of the study in their native language, and (5) research investigating the online morphological processing of written stimuli.

Due to this large volume, a deviation from the pre-registered protocol was made for practicality: Instead of two reviewers independently screening 25% of all the titles and abstract, a two-stage approach was adopted. In the first stage, only the primary reviewer screened titles and abstracts based on the first three, more clear-cut criteria. This process excluded studies that were clearly ineligible, such as those that were not empirical research, were published in languages other than English, fell outside the target age range, or focused on non-Indo-European languages like Turkish, Finnish, or Hebrew. The second reviewer then briefly reviewed the titles and abstracts of the excluded studies. In the second stage, both reviewers independently screened random 25% (n = 102) subset of the titles and abstracts of the remaining 407 records according to all the inclusion criteria. The inter-reviewer reliability was r =.70. This represents the second deviation from the pre-

registered protocol, which originally proposed an agreement rate of 80% between reviewers before proceeding to a single screening by the primary reviewer. However, despite this lower-than-planned inter-reviewer reliability, all disagreements were comprehensively addressed through discussions between the two reviewers. In instances where consensus could not be reached, a third expert in the field was consulted, ensuring all decisions were well-informed and mutually agreed on.

Full-texts were retrieved and where necessary, full-text book chapters and theses were acquired in hard copy or electronic form via inter-library loans or directly from the authors. However, one eligible thesis (Condry, 1979) could not be obtained and was consequently excluded. For the subsequent full-text screening, two reviewers independently evaluated 25% of the full-texts (n = 48), achieving an inter-reviewer reliability of r = .81. Disagreements were resolved through discussions between the two reviewers, with input from a third expert when necessary. Based on the discussions, as an intermediate step, we further refined our definition of online visual morphological processing to: “tacit, non-strategic activation of morphological knowledge while reading morphologically complex written stimuli, involving timed components such as reaction times (RTs) during quick lexical decision, eye movements during reading, or ERP waveforms.”

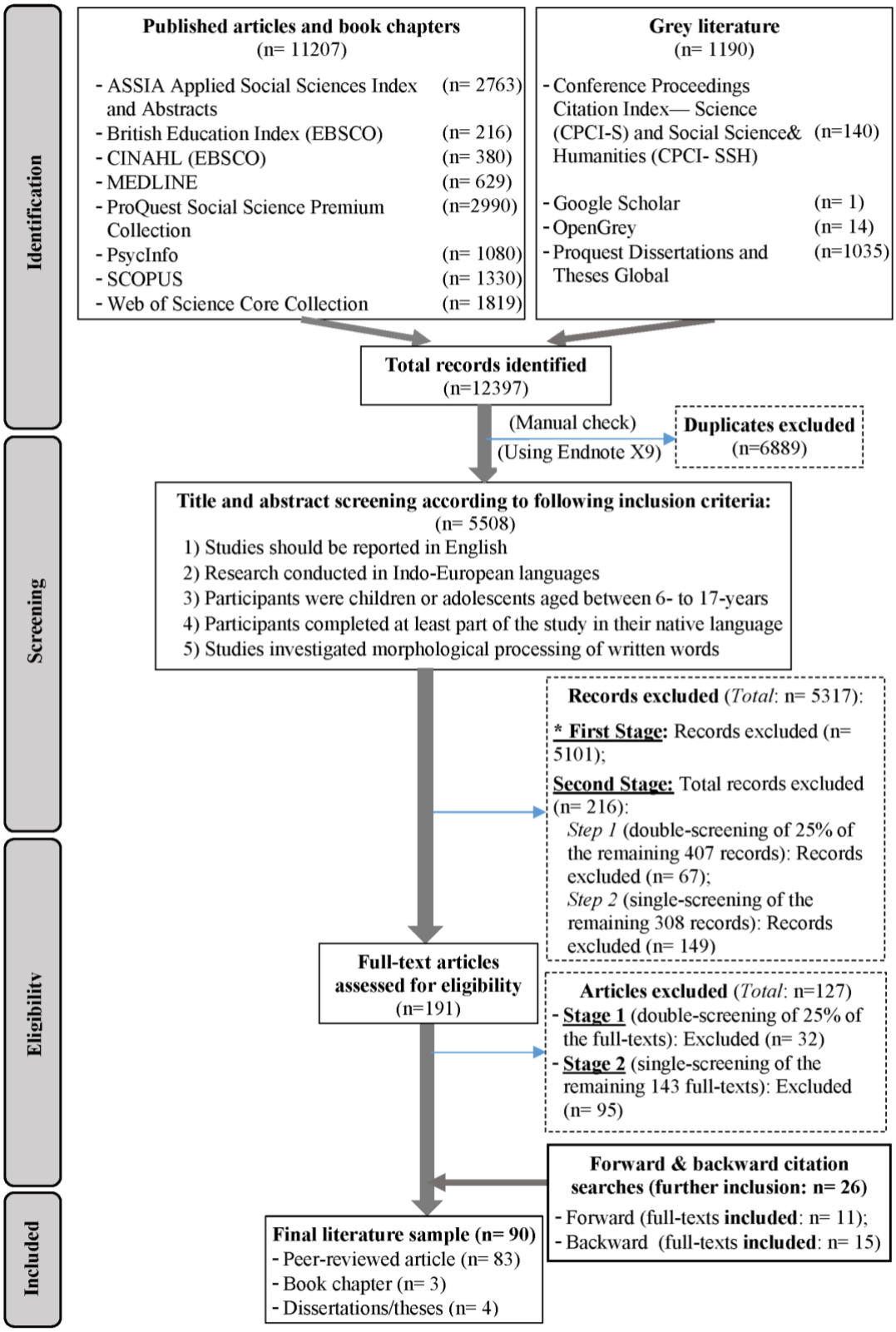


Figure 2.1: Flow Diagram of the Screening Process of Papers (* = Deviations from pre-registration)

Therefore, studies that used only monomorphemic written stimuli or only required explicit morphological knowledge were excluded. For instance, some studies used the cross-modal priming paradigm to present morphologically complex primes auditorily. Even though written target stimuli were used, they were in morphologically simple forms. Hence, any morphological processing was primarily associated with the auditory prime rather than the written target (e.g., Clahsen & Fleischhauer, 2013[2]; Quémart et al., 2017). Intervention studies were not the main focus of this review, meaning that intervention studies were only included if data could be extracted from control groups or from a baseline condition, provided they met the inclusion criteria. In addition, before data extraction, the primary reviewer also cross-referenced any eligible, unpublished thesis chapters with their published versions— upon publication, the unpublished chapters were then excluded as duplicates, and only the corresponding peer-reviewed journal articles were included, categorised under forward citation searches. The primary reviewer then single-screened the remaining full texts (n= 147). Articles that the primary reviewer could not clearly categorise (n = 25) were further discussed among the two reviewers and an expert.

In total, 127 full-text articles were further excluded for the following reasons: (1) Inability to access the full-text (n= 1); (2) Articles not available in English (n= 1); (3) Duplicates, such as journal articles that were also published as thesis chapters (n= 4); (4) Studies focusing on non-Indo-European languages (n= 3); (5) Research conducted with non-native speakers (n= 2); (6) Studies with participant age ranges either outside of 6-17 years or too broad, making it impossible to isolate data specific to the target age group (n= 9; e.g., age range: 9-87 years, as in Savinova & Malyutina, 2021); (7) Studies not focusing on online processing of morphologically complex written stimuli (n= 101), including those not utilising complex written stimuli (e.g., Quémart et al., 2017; Nation & Cocksey, 2009), presenting stimuli auditorily (e.g., Montgomery & Leonard, 1998; Quémart et al.,

[2] The visually presented targets in Clahsen & Fleischhauer (2013) were first person singular present tense forms, e.g., *tanze* (primed by *getanzt*) ‘dance (primed by danced)’. Although these forms involve morphological inflection, they can be seen as morphologically simple compared to other verb forms in German.

2017), or requiring explicit morphological analysis (e.g., Apel & Henbest, 2020; Carlisle & Fleming, 2003); and (8) Studies employing morphological processing tasks but failing to provide sufficient data (n= 6), such as intervention studies without comprehensive baseline or control-group data (e.g., Jöbstl et al., 2021), or studies where morphological processing was merely a predictor for another outcome, without detailed information about morphological processing itself (e.g., Goodwin et al., 2017). No studies were excluded due to critical methodological flaws. Any methodological issues that might affect the reliability and generalisability of the results were carefully noted during the following coding process. The final sample included 90 studies. The included studies are highlighted in bold and marked by '*' in the reference list.

2.2.4 Coding

Studies included in the review were coded for (1) study details (e.g., year of publication, authors), (2) participant demographics (e.g., type, age), (3) study type (cross-sectional vs. longitudinal, experimental vs. observational), (4) research methodology (e.g., target languages, paradigms, tasks, stimuli), (5) results of interest (analytical approach, statistical outcomes, key findings), and (6) implications and limitations (e.g., critical methodological issues such as small sample size). Importantly, in line with the research questions this review aimed to address, individual, item, and task level factors influencing morphological processing were also coded for each study when applicable. The primary reviewer coded all the eligible studies. The comprehensive data extraction and detailed results of the coding for individual studies are accessible in the Supplementary Material (Data extraction form) that is available online on OSF via: https://osf.io/srdmq/?view_only=ffbd7b2d3d8e41c9b27effde99e2c49b.

2.3 Results

Appendices 2A-2D provide an overview of the 90 studies included in the review, organised by the paradigms employed. Notably, the inherent conceptual complexity of morphological processing, coupled with the varied designs of the studies, renders the distinction between online and offline processing less clear-cut than envisioned, even with our 'timed component' inclusion criterion. It could be argued that online and offline processing should be viewed more as a 'continuum' than as a strict dichotomy, necessitating a degree of interpretive flexibility. In the spirit of openness, thoroughness, and relevance, we included studies that utilised tasks with less stringent timing constraints, such as copying (e.g., Hess et al., 2021) and fragment completion (e.g., Feldman et al., 2002), alongside those that are obviously tapping online word recognition. Taken together, we hoped that the studies would provide important insights into implicit morphological processing from different perspectives.

2.3.1 Study Characteristics

The articles included in this review, published between 1981 and 2021[3], reflect an incremental trend in research focusing on morphological processing as children read words over the past few decades. This trend peaked in 2021, with 10 studies published that year, the initial cut-off year for the literature search, as shown in Figure 2.2. Besides the included book chapters, unpublished thesis chapters and dissertations, included studies were published in 39 different journals.

[3] Three articles, found during the forward citation search, were published in 2022 but obtained online in 2021.

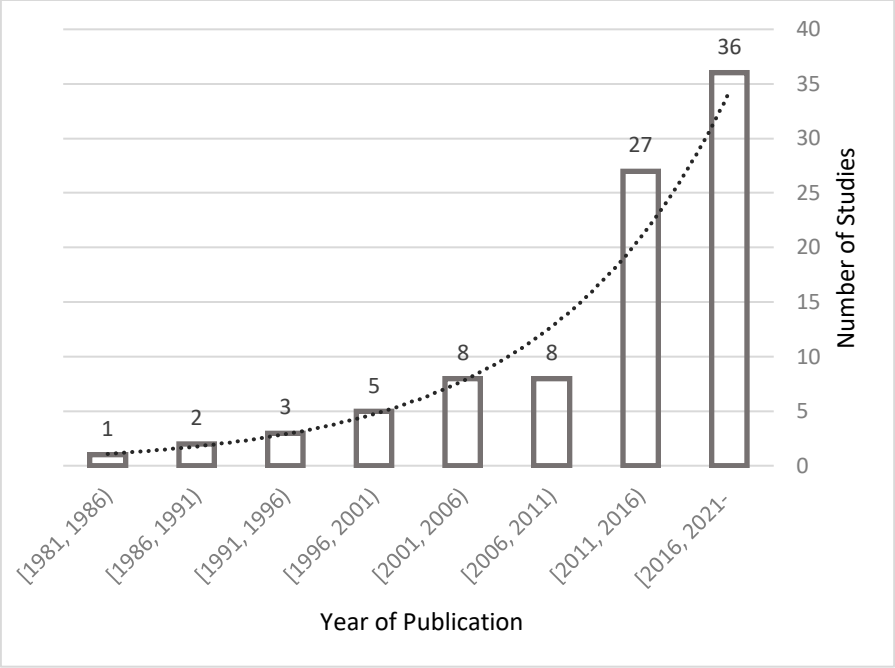


Figure 2.2: Growth in Studies Investigating Morphological Processing in Developing Readers from 1981. Number of publications each year: 1981 (n=1), 1987 (n=1), 1989 (n=1), 1992 (n=2), 1995 (n=1), 1996 (n=2), 2000 (n=3), 2001 (n=1), 2002 (n=2), 2003 (n=3), 2004 (n=1), 2005 (n=1), 2006 (n=2), 2008 (n=3), 2009 (n=2), 2010 (n=1), 2011 (n=7), 2012 (n=3), 2013 (n=6), 2014 (n=5), 2015 (n=6), 2016 (n=3), 2017 (n=7), 2018 (n=6), 2019 (n=4), 2020 (n=3), 2021 (n=10), 2022 (n=3).

Due to the inclusion criteria, all reviewed articles focused on Indo-European languages, covering nine different languages: English, Italian, French, Dutch, German, Spanish, Portuguese, Bosnian, and Serbo-Croatian, with all participants being native speakers of their respective languages (see Figure 2.3). It is notable that the research meeting our criteria has predominantly focused on a limited subset of Indo-European languages. Considering the Indo-European language family comprises over 400 languages spoken by nearly 3 billion people globally (Kapović et al., 2017, p. 1), the research field has so far explored just a fraction of its linguistic diversity. English emerged as the most studied language, comprising 36 studies. Among the 90 studies identified in our review, three investigated multiple languages (Beyersmann et al., 2021; Casalis et al., 2015; Mousikou et al., 2020), and the rest focused on a single language.

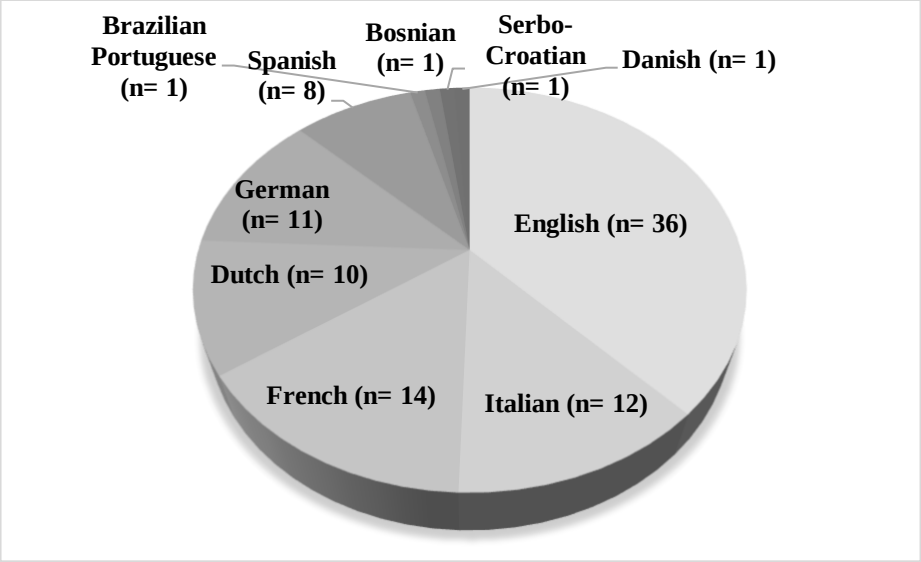


Figure 2.3: Counts of Language Investigated

Note: In studies investigating multiple languages, each language was counted separately; thus, the total counts exceeded the number of studies included (i.e., 90).

As illustrated in Figure 2.4, most studies focused on children and young adolescents (9-12 years). There is a notable gap in research through adolescence (13-17 years). This is unfortunate as this developmental stage marks a significant transition characterised by various cognitive, biological, and social changes (e.g., Ben-Shachar et al., 2011; Dumontheil et al., 2010; Yurgelun-Todd, 2007). During this period, the maturation of brain structures associated with language processing continues. Neuroimaging data indicate that the ventral pathway for reading, which is associated with direct mappings between orthography and semantics as well as morphological processing, continues to develop into mid-adolescence (Ben-Shachar et al., 2011; Yablonski et al., 2021). Furthermore, adolescents are increasingly exposed to complex linguistic environments and morphologically complex items in their reading materials (Dawson et al., 2023), making this period pivotal for the acquisition of morphology from linguistic input. While there is considerable variation in language abilities within each age group, the developmental trajectory of core language skills exhibits a high level of stability from childhood into early adolescence, suggesting that

relative language skill levels are generally maintained over time (Bornstein et al., 2016). This stability provides a theoretical basis for investigating the continuous development of language and reading abilities. Therefore, while existing research often compares children with adults, a more detailed examination of the developmental changes during the intermediate adolescent period has the potential to reveal the full developmental trajectory of morphological processing in reading.

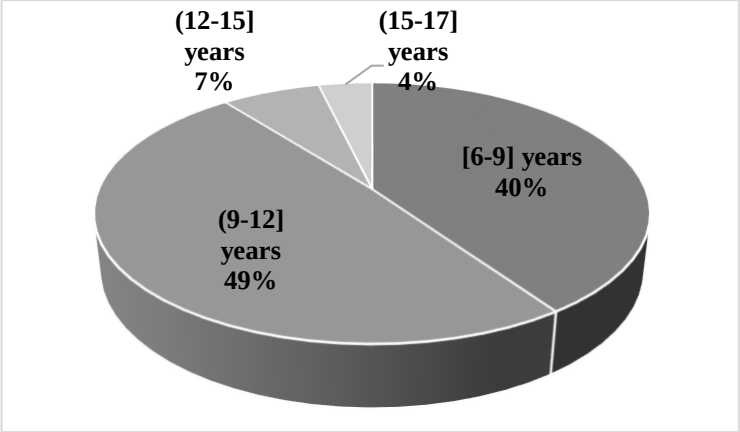


Figure 2.4: Age (groups) Investigated in the Included Studies

Note: Categorisation based on chronological age. The categorisation takes into account both the stated age range and the mean age (when available), providing a rough estimate in certain cases. In articles containing multiple experiments, each distinct age group was counted only once.

2.3.2 Synthesis

The narrative synthesis is organised around three central themes that correspond to our research questions. The first theme focuses on individual-level factors, which encompass characteristics related to the subject of reading behaviour, namely the child, including factors such as age, grade, and reading ability. The second theme addresses item-level factors, pertaining to the characteristics of the object of reading, the reading material (e.g., complex written words). This includes elements like the distributional properties of words, including word frequency and family size. The third theme focuses on task-level factors, stemming from the methodological diversity observed across different studies. In addition to contributing to the evaluation of study design and quality, these

factors also potentially illuminate the varied dynamics of morphological processing. Given that each task carries its own cognitive demands and temporal constraints, morphological processing may be task dependent. Therefore, different tasks might reveal distinct aspects and varying time-courses of morphological processing. We begin our discussion of findings by considering task-level factors first. This is to establish a methodological foundation for the subsequent discussion of item-level factors, and then individual-level factors.

2.3.3 Task-level Modulation

Research papers included in this review predominantly adopted three paradigms: Reading Aloud (Naming) (see Appendix 2A), Lexical Decision[4] (Appendix 2B), and Masked Priming (Appendix 2C). We consider each in turn before considering other paradigms used (Appendix 2D).

2.3.3.1 Reading aloud

This paradigm was used in 32 studies. It requires participants to read items aloud and accuracy and RTs are typically recorded. Reading aloud is suitable for young children, and it is attractive in that it requires low task demands and aligns with their routine school practices (Clahsen & Fleischhauer, 2013). The nature of the items has varied from study to study and has included words (e.g., Carlisle & Stone, 2003, 2005; Deacon et al., 2011, 2017; Marcolini et al., 2011), pseudowords (e.g., Burani et al., 2002; Traficante et al., 2011; Verhoeven et al., 2003), or both (e.g., Angelelli et al., 2014, 2017; Cole et al., 2012; Laxon et al., 1992), each varying in aspects like morphological structure, whole-word and constituent morpheme frequency, and stem stability. Reading aloud sheds light on the implicit activation of morphological knowledge, such as how constituent morphemes influence the pronunciation of complex stimuli. In most studies, RTs are generally defined as the time taken to initiate a response from the presentation of the written word to the onset of naming. The RTs might not fully capture the processing of the entire stimulus, as

[4] In this review, we use the term 'Lexical decision' specifically to refer to the 'word/nonword' decision task without priming, distinct from other paradigms like masked priming that also involve lexical decisions.

readers could begin pronouncing without fully processing the whole stimulus (Marcolini et al., 2011; Suarez-Coalla & Cuetos, 2013). As a result, RTs might naturally tend to be sensitive to elements on the left part of the stimulus, such as prefixes (Verhoeven et al., 2006) or base frequency (Angelelli et al., 2014), due to the nature of the left-to-right reading system. In one study, Suarez-Coalla et al. (2017) examined the online time course of reading aloud more closely. Alongside RT, they evaluated reading durations (total articulation time) and critical segment durations (time to read the first three phonemes, often corresponding to the root in complex words). These additional measures enabled the researchers to investigate the role of the base word in initiating reading, specifically its potential to offer a head-start in the early stages of processing. This demonstrates how reading aloud has the potential to offer detailed insights into the intricacies of the naming process.

While almost all studies asked participants to name words in isolation, notable exceptions include Deilen et al. (2022, Experiment 2), Traficante et al. (2018), and Whylie (2021), where target stimuli were presented within sentence contexts. For instance, in Traficante et al. (2018), nouns derived from either noun or verb bases were embedded in sentences (e.g., 'Un amico mi fornisce **assistenza** nei compiti di latino.'; ['A friend gives me his **assistance** with my Latin homework']). Participants read full sentences aloud, and eye movements to the target words were recorded. This approach allowed an investigation into the role of syntactic expectations in recognising both the base and the whole word. For example, the presence of a verb-derived noun in a sentence that requires a noun in the target-word position demonstrated a longer inhibitory base frequency effect in the first fixation duration, as the early detected verb base and the expected noun are likely to cause difficulty in processing at the early stage. This experimental setup was contrasted with Traficante et al. (2014), where participants named the same set of words but presented in isolation, thus leading to immediate naming upon base recognition without the need for considering the whole word. Embedding words in sentences provided a more complex picture of the role of semantic and

syntactic components in morphological processing. Additionally, sentence context offers a more dynamic and natural reading experience, and findings that are potentially more generalisable to normal reading; such studies may also highlight the interaction between morphological processing and syntactic or grammatical constraints (Whylie, 2021). The interplay is also echoed in studies using grammatical priming (Bowey, 1996) and silent reading of sentences with embedded target words (e.g., Deilen et al., 2022). As morphological processing unfolds naturally during the reading of morphologically complex words within sentence contexts, further research should consider embedding target words within sentences. However, sensitive measurement (e.g., eye-tracking) is essential to reliably capture responses to the target words.

2.3.3.2 Lexical decision

Lexical decision was used in 26 studies. The classic task requires participants to make a yes-no judgement as to whether a written string is a word or not. While both lexical decision and reading aloud are widely used, they differ in processing demands (Carlisle et al., 2001). Lexical decision tasks, typically performed under time constraints, might be quite demanding for young children or those with poor reading (Lazaro et al., 2013) compared to tasks like naming. Carlisle et al. (2001) directly compared children's reading aloud and lexical decision for the same stimuli and found a stronger phonological transparency effect in reading aloud, with derived words having stable stems being read faster and more accurately than those with phonological shifts. In contrast, the phonological transparency effect was more subtle in lexical decision, though still evident. This difference could be attributed to task factors, with naming involving subsequent production processing, leading children to prepare to pronounce the base word before fully processing the affix or stress pattern of the derived word. This reliance on the base might interfere with the retrieval of the phonological representation of the whole word. Conversely, lexical decision does not involve pronunciation and so might be comparatively less sensitive to phonological shifts (Carlisle et al., 2001). Therefore, it might be that reading aloud and lexical decision are sensitive to different

aspects of morphological processing. What follows is that the outcomes obtained from the two tasks might not be directly equivalent or comparable.

The go/no-go version of lexical decision asks participants to press a button only for 'yes' responses and refrain from pressing any key for 'no' responses. This is considered to place fewer task demands on younger children compared to the standard version resulting in lower error rates, and faster and less variable RTs (Moret-Tatay & Perea, 2011). Building on these findings, Lázaro et al. (2015, 2016) used go/no-go lexical decision tasks to probe complex word processing in Spanish-speaking children with reading difficulties and younger reading-level controls. Different from their earlier study (Lázaro et al., 2013), they found that the suffix frequency effect emerged as early as 2nd grade (Lázaro et al., 2016). High-frequency base words facilitated word recognition in 8-year-olds with reading difficulties and in both reading- and age-matched controls (Lázaro et al., 2015).

2.3.3.3 Masked priming

Masked priming (N = 17 studies) requires lexical decisions to a target word, following the brief presentation of a prime, with primes and targets being related to each other in some way including for example morphological (e.g., 'teacher-TEACH'), pseudomorphological ('mother-MOTH'), or orthographic (e.g., 'cashew-CASH') overlap. As the prime duration is brief, masked priming is thought to capture the very early stages of the word recognition process. As such, it sheds light on automatic, nonstrategic processes as reading happens (Beyersmann et al., 2012; Lazaro et al., 2018).

Most of the masked priming studies with children used the classic version of lexical decision, but three studies opted for the go/no-go version (de Oliveira & Justi, 2017; Lazaro et al., 2018) or reading aloud of the target word (Duranovic et al., 2020), either to accommodate younger participants, (i.e., the 2nd graders in de Oliveira & Justi, 2017) or to foster a more natural reading situation (Duranovic et al., 2020). Alongside these format adaptations, prime duration, or Stimulus Onset Asynchrony (SOA), also varies across studies and might serve to influence the morphological priming effect. Beyersmann et al. (2012) were pioneers in applying this paradigm with children and

observed that 3rd and 5th graders responded faster and more accurately to a target word preceded by a morphologically related prime than to one preceded by an unrelated prime. They set the prime duration at 50ms, and across subsequent studies with children, prime duration has ranged between 50-75ms. Although these durations are often longer than those typically used in adult studies, they are considered short enough to remain undetected by children. This is important as at such short durations, semantics processing is presumed to be minimal, or at least not as dominant as morpho-orthographic processing. Beyersamm et al. (2012) observed morphological priming but did not find pseudomorphological priming, which seems to suggest that semantic information might exert some influence. This initial stage is generally argued to be pre-semantic[5], with processing that is not predominantly reliant on meaning activation. This is supported by studies like Quémart et al. (2011), which observed both morphological and pseudomorphological priming in children, indicating possible semantic-independent morphological processing at the initial stage.

To ensure detection of the prime, researchers generally employ durations exceeding 200ms (de Oliveira & Justi, 2017) and at these longer durations, semantic information of the prime word can be actively processed (Quémart et al., 2011). Five studies (Casalis et al., 2009; de Oliveira & Justi, 2017; Duranovic et al., 2020; Quemart & Casalis, 2014; Quemart et al., 2011) examined the time course of the morphological priming effect more closely by manipulating prime presentation durations. In French 4th-grade children, Casalis et al. (2009) found significant priming for morphologically (e.g., 'LAVEUR-lavage', equivalent to 'TEACHER-teaching' in English) and orthographically related primes (e.g., 'LAVANDE-lavage', similar to 'DECIDE-decimal' in English) at a relatively short 75ms prime duration. However, at 250ms prime duration, priming was only seen in the morphologically related condition. In a parallel finding with Bosnian-speaking children, Duranovic et al. (2020) observed both types of priming at 50ms, with priming from

[5] As mentioned in Chapter 1, the role of semantics at the initial stage of processing is still under debate across studies with adults (see Amenta & Crepaldi, 2012, for a review; cf. Rastle et al., 2004, 2008, Beyersmann et al., 2012; Beyersmann & Grainger, 2018; Diependaele et al., 2009; Feldman et al., 2009).

orthographic overlap becoming non-significant at 250ms. In a study including a broader range of age and prime durations, Quémart et al. (2011) observed priming in both morphological and pseudo-derivative (e.g., 'baguette–BAGUE', equivalent to 'corner-CORN' in English) conditions, when prime duration was 60ms or 250ms. Notably, pseudo-derivative priming was significantly weaker than the morphological priming at the longer 250ms duration, and it disappeared when the prime duration was extended to 800ms. In contrast, morphological priming remained robust across all durations. Moreover, at the 800ms duration, significant priming effects emerged for semantically related prime-target pairs (e.g., 'tulipe–FLEUR', English equivalent: 'tulip-FLOWER'), indicating that semantic factors are active when the prime is actively available. The time-course revealed by manipulating prime durations suggests that sensitivity to morphological structure is evident throughout the word recognition process.

Despite robust morphological priming with short prime duration being observed in most studies with children, there are some exceptions. For example, de Oliveira & Justi (2017) found that while Brazilian Portuguese-speaking 5th graders showed morphological priming with a 60ms duration, 3rd-4th graders only showed the effect when the prime duration was 250ms. This pattern of priming might suggest that morpho-orthographic processing at a short prime duration requires abstract orthographic patterns of morphemes, the acquisition of which depends on reading experience. In contrast, morpho-semantic processing is less dependent on reading experience but is only available in processing when the prime duration is longer. Caution is warranted when interpreting the absence of morphological priming among younger children as this discrepancy may partly be attributed to several statistical decisions made in their analysis. Notably, the lack of a by-items (F2) analysis[6] in de Oliveira & Justi (2017) could allow certain items to disproportionately influence the results. Additionally, posthoc pairwise comparisons (splitting by SOA group) were conducted even though

[6] The lack of by-items (F2) analysis was noted in several studies included in this review. This issue is coded as 'critical methodological issues' in the Data extraction form in the supplementary materials on OSF and discussed in Section 2.5 Limitations and Future Directions.

the interaction between prime types and SOA was not significant in 3rd grade, and no corrections were applied for multiple comparisons (morphological vs. orthographic vs. unrelated across two SOAs). Both issues could increase the likelihood of Type I errors, raising the risk of overstating or misinterpreting the priming effect and the changes in priming based on the SOA. Therefore, while the developmental trend of priming effects emerging at longer SOAs seems reasonable, the lack of a significant interaction suggests that the effect of prime type does not robustly vary with SOA, especially in the 3rd graders. However, as this study is the only representation of Portuguese-speaking children in the review, it is challenging to determine whether the differing results were solely due to statistical issues or if cross-linguistic differences, such as specifics of Portuguese or the Brazilian educational system, also played a role. Further research on morphological processing in Brazilian-Portuguese-speaking children is essential, as discussed in Sections 2.3.6 on cross-language differences and 2.5 on limitations and future directions.

Sensitivity to morphological structures at short prime durations also hinges on the recognition of base words, as demonstrated by Quémart and Casalis (2014). They explored the effects of morphological primes with or without base modifications in form (e.g., 'bergerie [sheepfold]-BERGER [shepherd]', English equivalent: 'natural-NATURE') and noted that significant priming at a brief 60ms prime duration occurred solely for primes with unmodified bases (e.g., 'nuageux [cloudy]-NUAGE [cloud]'). Yet, when primes were presented for 250ms, priming extended to all morphologically related prime-target pairs, including those with phonological or/and orthographic modifications to the base. Evidence across the five studies that manipulated prime durations indicates that morphological priming becomes stronger as prime durations increase and in doing so, its signature is different from that of pseudomorphological or orthographic priming which diminishes as prime duration increases. This is likely due to the emergence of stronger semantic effects in later processing stages. This modulation by prime durations suggests a nuanced interplay between morpho-orthographic and morpho-semantic processing as visual word recognition unfolds.

2.3.3.4 Other paradigms

Together, these three paradigms – reading aloud, lexical decision and masked priming – have provided insights into morphological processing as children read and process morphologically complex stimuli. Studies have used alternative approaches as well. For instance, in written production or copying (Hess et al., 2021; Quemart & Lambert, 2019), children are asked to write stimuli presented on a screen using a digitiser. This setup enables tracking of total writing speed and letter writing durations, with a specific focus on critical elements such as morphemic boundaries (Quemart & Lambert, 2019). The findings will be discussed in detail later; however, it is worth noting that these studies offer evidence concerning morphological processing during written production, which is claimed to be much like reading aloud as they lack the decision-making element of lexical decision (Quemart & Lambert, 2019). However, careful interpretation of copying task results is needed— while timed and occurring in real-time, these tasks engage offline processing as well and are dependent on motor skills.

Similarly, semantic categorisation (Hasenäcker, Olga et al., 2021), fragment completion (Deacon et al., 2010; Feldman et al., 2002; Rabin & Deacon, 2008), and probe detection (Breadmore & Carroll, 2016) do not require explicit lexicality judgments about a target word itself. The latter two approaches focus on the memory of complex words recently experienced. Memory is also targeted in a novel paradigm by Pescuma et al. (2022) who employed a Fast Periodic Visual Stimulation Oddball paradigm, combined with MEG to examine brain responses to different morphemes. Children were presented with pseudowords featuring various stem and suffix combinations. For example, the ‘oddball’(target) pseudoword ‘softity’, comprising a real stem and suffix, was placed in a sequence of nonstem + real suffix stimuli (e.g., 'trumess') to examine stem identification, and among real stem + nonsuffix stimuli (e.g., 'stopust') for suffix identification. Stimulus sequences were constantly overlaid by a fixation cross that changed colour randomly. Participants were simply required to tap a button when detecting a colour change in the fixation

cross, ensuring engagement with the area without drawing attention to the pseudowords. This approach is thought to be particularly suitable for measuring automatic neural responses to words in children, given minimal task demands. However, it is crucial to acknowledge the downside of this, in that the task is quite different to natural reading. This distinction highlights the need to carefully consider how differences between experimental settings and real-world reading scenarios could impact the generalisability of findings to actual reading contexts (e.g., McCutchen et al., 2009).

McCutchen et al. (2009) used a continuous lexical decision paradigm. This is still a contrived task but is arguably closer to normal reading than a measure like fragment completion or masked priming. Children were tasked with making lexical decisions on a continuous list of stimuli, where each item 'n' acted as a prime for the following 'n+1' item. A similar repetition priming approach was reported by Leong and Parkinson (1995). Diverging from serial stimulus presentation, Cauchi et al. (2022) used a flankers paradigm to explore morphological processing in developing readers. This involved simultaneously presenting the target word and its flankers (e.g., 'corn corner corn') and asking children to make lexical decisions about the central words. This approach allowed for parallel processing in a multiword context, enabling the researchers to investigate the spatial integration of adjacent words, which could be seen as a step closer to natural sentence reading than single-word recognition tasks. As highlighted earlier, eye movement studies that capture reading behaviour as children read morphologically complex words within sentences are useful (Breadmore & Carroll, 2018; Deilen et al., 2022; Whylie, 2021). The eye movement record provides high temporal precision and reflects the interconnected nature of morphology, syntax, semantics, and orthography (Carlisle, 2000; Joseph et al., 2013), thus making the findings more generalisable to natural, self-paced silent reading.

In an early study, Drewnowski (1981) used self-paced paragraph reading combined with a letter search task to explore how suffixes ('-ing' or '-ent') were processed within textual passages. The observation of a 'missing letter' effect predominantly with '-ing', but not as pronounced with '-ent',

was interpreted as evidence for ‘-ing' being a discrete processing unit. However, a recent study (Hasenäcker, Ktori et al., 2021) adopting a letter search task with pseudowords containing stem and suffix morphemes detected no effect in letter search related to suffix status. Despite the difference in stimulus presentation and lexicality, the absence of significant findings in Hasenäcker, Ktori et al. (2021), backed by additional Bayes Factor analyses, led the authors to question the suitability of letter search tasks for studying morphological processing, especially in light of consistent findings of morphological effects across other tasks.

In summary, different paradigms have been used to investigate morphological processing in children’s reading. These might yield varying patterns of results (Berthiaume & Daigle, 2014) and therefore, it is important to attend to task demands. Reading aloud relies on speech production as well as word recognition. Making lexical decision also requires word reading as well as decision-making under time constraints, which can be challenging for young children. Neither paradigm can fully capture the very early stages of word recognition in the way that masked priming is claimed to. Masked priming itself is a complex paradigm, and we need to remain mindful of factors such as prime durations. Hasenäcker, Ktori et al. (2021) emphasised the importance of interpreting results within the context of the methods used and the limitations associated with different methods. They highlighted the importance of their null results, which contributed to the ongoing discussion about the replication crisis. As the field advances, critically evaluating and adapting research methods becomes increasingly essential to ensure that results are reliable and accurately reflect the nature of morphological processing. Having reviewed the main methods used to date, I now turn to what has been learned about children’s morphological processing from these methods.

2.3.4 Item-level factors

2.3.4.1 Morphological structure

Research has largely shown that from the elementary school years onwards, individuals are better able to read morphologically complex words compared to simple words (e.g., D'Alessio et al.,

2018; Marcolini et al., 2011; Suarez-Coalla & Cuetos, 2013), and compared to words with pseudo-affixes (e.g., 'dinner', 'naked'; Carlisle & Stone, 2003, 2005; Laxon et al., 1992). There are however some exceptions to this general finding. For example, Quemart & Lambert (2019) found no significant impact of morphological structure on word naming. One possible explanation is that in certain cases, especially in more transparent languages like French, the whole word is more readily accessible and familiar than its constituent morphemes, potentially diminishing the facilitative effect of morphemic structure (Angelelli et al., 2014). Contrary to this idea, Italian children appear to use morphemes as reading units, despite Italian being a transparent orthography: Traficante (2014, 2018) observed that the grammatical class of the base influenced children's reading aloud of derived words, both in isolation and in sentence contexts. While the extent of the influence of morphological structure on word naming can vary, reading aloud words is generally influenced by morphemic structure, challenging the exclusive role of syllables in the naming process (Carlisle & Stone, 2005). The effects of morphological structure may be particularly evident when naming pseudowords which lack established whole-word representations. Research (Angelelli et al., 2014; 2017; Burani et al., 2008) has demonstrated that morphemic pseudowords composed of an existing root and a suffix were more effectively named than those without identifiable morphemes. Angelelli et al. (2014) argued that facilitation in naming morphologically structured pseudowords is due to morphological processing, rather than merely being a 'wordlikeness' effect. They were able to conclude this as they matched the frequency of non-suffix endings in nonmorphemic pseudowords with suffixes in morphological ones (similarly in Burani et al., 2008), ensuring that suffixes were not simply more familiar. Additionally, the presence of morphemic units, whether roots or suffixes, contributed to improved naming performance in other studies (Mousikou et al., 2020; Traficante et al., 2011).

A different line of work has examined the pronunciation of prefixes. Children tend to pronounce the first part of a pseudoword in the manner typical for prefixes, indicating that the

initial segment of word is treated as if it were a prefix (Verhoeven et al., 2006). Cole et al. (2012) further reported that the quantity of morphemes in a pseudoword affected reading speed and accuracy, with pseudowords encompassing two morphemes (a real stem and a real suffix) being processed more efficiently than those with one morpheme pseudostem + real suffix) or no morpheme (pseudostem + pseudosuffix). Verhoeven et al. (2006) also found that children recognise words with real or phonological prefixes (which share the same pronunciation as real prefixes) more readily in lexical decision, compared to those with pseudoprefixes (i.e., those pronounced differently from real prefixes).

Other studies using lexical decision generally report compatible findings with morphological structure being associated with better performance (Casalis et al., 2015; D'Alessio et al., 2019; van Hoogmoed et al., 2011). However, van Hoogmoed et al. (2013) observed that responses to monomorphemic words were faster compared to derived words, possibly because the monomorphemic words were shorter than the derived words in their study. Additionally, Duncan et al. (2011) reported no advantage in lexical decision for words with both a stem and a suffix (e.g., 'farmer') over those with only a stem (e.g., 'window') in children with good reading ability, possibly due to the direct activation of whole-word representations. In contrast, they found a different pattern emerged for pseudowords — children found pseudowords containing both a stem and a suffix (e.g., 'gifter') more challenging to reject than those with a stem and a non-suffix ending (e.g., 'puffow'). This 'morpheme interference effect', consistently observed in adults since Taft and Forster (1975), has more recently been investigated and observed in developing readers (e.g., Assink et al., 2000; Casalis et al., 2015; Dawson et al., 2018). The effect remains even after controlling for factors like orthographic neighbours (Dawson et al., 2018) or letter cluster frequencies of affixes and non-affixes (Assink et al., 2000), which could otherwise contribute to the 'wordlikeness' of morphemic pseudowords. Therefore, the challenge in rejecting morphemic pseudowords suggests that morphemic representations were activated through morphological

decomposition, thereby enabling these words to be licensed and have semantic interpretability (Duncan et al., 2011). Consistent with this suggestion, in a post-hoc analysis Dawson et al. (2018) observed that the semantic interpretability of pseudowords influenced outcomes. However, they noted that this influence still depended on the prior decomposition of morphological structure. This echoes Berthiaume and Daigle's (2014a, 2014b) finding that affixed pseudowords adhering to language-specific formation rules were perceived as more plausible. One exception was reported by Burani et al. (2002) who found that children were quicker to reject morphological pseudowords relative to nonmorphological pseudowords. This may be because the nonmorphological pseudowords used in their study contained no morpheme (i.e., non-root + nonsuffix), unlike the control nonwords used in other studies, which contained embedded roots. Therefore, as the nonmorphological pseudowords in Burani et al. (2002) lacked embedded words for activation, they were processed as whole words, whereas their morphological counterparts could be broken down into smaller morphemic units. This could potentially account for the unexpected findings.

As previously reviewed, robust priming for morphologically related primes (e.g., 'bomber-BOMB') under masked conditions has been consistently demonstrated in developing readers. Even pseudomorphological primes with surface morphological structure (e.g., 'corner-CORN') show priming effects in some studies with children (e.g., Quémart et al., 2011). This would be consistent with the type of automatic morpho-orthographic processing that characterises skilled processing (see Amenta & Crepaldi, 2012 for a review). However, the pseudomorphological effect appears to vary across languages and developmental stages (cf., Beyersmann et al., 2012; Dawson et al. 2021; Quémart et al., 2011), something I will return to discuss in Sections 2.3.5 and 2.3.6. Nevertheless, it has also been demonstrated that morphological structure is not always required for priming effects. For instance, research like Beyersmann et al. (2015) and Hasenäcker et al. (2016) found that young elementary school children exhibited significant priming when the target (e.g., TRISTE [SAD]) was primed by a suffixed nonword (e.g., tristerie [sadation]) or a non-suffixed nonword (e.g., tristald

[sadald]), with no significant difference in priming between the two conditions. Similarly, a study using a semantic categorization task (Hasenäcker, Olga et al., 2021) further supported the idea that stem activation could be independent of morphological structure: Words with category-congruent embedded stems were harder to reject as a category member, and this effect on RTs was not modulated by morphological structure (that is, they observed similar results for categorising 'pea-ce' or 'corn-er' as food). These findings indicated that the non-morphological activation of orthographic stems extended beyond being recognised as familiar letter strings, also feeding forward to the semantic level. Thus, the activation of embedded stem semantics affected the speed of whole word categorisation, irrespective of morphological structure. The early-emerging non-morphological activation of the embedded stem could potentially be explained by the 'edge-aligned embedded stem activation' account by Grainger and Beyersmann (2017) which discusses different impacts, distinct roles, and varying developmental trajectories of stems and affixes on morphological processing. This is discussed in more detail in Section 2.3.4.3 'Stem vs. Affix'.

2.3.4.2 Intactness vs. disruption

The impact of morphological structure on processing varies depending on the integrity of both the structure and its constituents. Hasenäcker, Ktori et al. (2021) observed that the presence of a suffix influenced the search for letters embedded within the suffix when these suffixes appeared in typical word-ending positions, but not when the stem-suffix order was reversed. However, this position-specific effect is small and was not consistently replicated. This might be due to the limitations of the letter search task, as discussed in Section 2.3.3.4. Also, disrupting morphemic boundaries by transposing letters in prime words (e.g., 'cor**en**r') eliminates the morphological priming effect (Rosenberg & Kruk, 2021). This has been interpreted to suggest that children employ both fine-grained and coarse-grained (i.e., morpho-semantic) processing strategies in these cases.

Non-lexical visual disruptions as segmentation cues have also been shown to affect the processing of morphological structures. Egan and Pring (2004) demonstrated that children struggled

to read words containing mixed-case letters that crossed morpheme boundaries (e.g., 'walKEd'), emphasising the significance of intact morphemic units. Cole et al. (2012, Experiment 2) observed that children performed similarly in word naming in morphological ('mal ade') and syllabic ('ma lade') segmentation formats. However, a decrease in speed and accuracy was found in the 'morphological + 1' format (i.e., with an additional letter added to the stem, e.g., 'mala de'), suggesting that children are sensitive to both syllables and morphemes as they read complex words and that non-linguistic segmentation patterns impede this process. Cole et al. (2012) concluded the linguistic nature of the units, not the segment length, was crucial. Along similar lines, Deilen et al. (2022) examined eye movements of hearing-impaired children when reading words in isolation and in sentence context. They found that compound words were easier for less-skilled hearing-impaired children to process when segmented by a 'medipoint' segmentation marker between morphemes (e.g., 'Regen·chirm' [umbrella]) compared to unsegmented and hyphenated formats (e.g., 'Regen-chirm'). However, research using a copying task (Hess et al., 2021) found that colour changes aligned with morpheme boundaries led to longer durations before the onset of word writing. This suggests that visual cues between morpheme boundaries might help trigger the activation of both stem and suffix, resulting in a delay before children began to write.

Research has also examined the interplay between syllabic and morphemic units. Using lexical decision, Hasenacker and Schroeder (2017) found that syllable-congruent disruptions (e.g., FAH:RER) facilitated word processing in both 2nd and 4th graders, whereas the effects of morpheme-congruent disruptions (e.g., FAHR:ER) on pseudoword rejection was only seen in 4th graders. This suggests a developmental change in morpheme-based processing, where syllables initially serve as functional units before morphemic units become salient later in development; we will return to discuss this later when we review child-level factors in section 2.3.5.1. Conversely, earlier studies by Leong (1989) and Leong and Parkinson (1992) reported improved lexical decision performance with morpheme-congruent disruptions, indicating a preference for morphemic units

(e.g., tractOR) over syllabic units (e.g., tracTOR). These discrepancies may stem from cross-linguistic (i.e., German in Hasenacker & Schroeder, 2017; English in Leong, 1989; Leong & Parkinson, 1992) and age-related differences, as participants in the latter studies were older and likely had more developed morpheme representations.

Besides structural integrity, the morpheme stability in the sense of unaltered constituent morphemes, either phonologically, orthographically, or both, particularly stem transparency, influences morphological effects. Studies have demonstrated that children encounter more difficulty when naming words with phonological and/or orthographic modifications in the stem (e.g., phonological shifts as in 'natural') compared to transparent words with stable stems (e.g., phonologically stable stems as in 'cultural') (Carlisle, 2000; Carlisle et al., 2001; Carlisle & Stone, 2003, 2005; Deacon et al., 2011; Mann & Singson, 2003). Leong and Chen (2000) also found that it was more difficult for children to produce the base form of a word when presented with its derived form if that involved phonological or orthographic changes, compared to the no-change condition. The influence of stem stability or transparency is also evident in lexical decision, though in a more subtle manner than in reading aloud (Carlisle et al., 2001). Developing readers tend to be faster to accept stable derived words as opposed to those with phonological shifts in stems, e.g., natural (Carlisle et al., 2001; Carlisle & Stone, 2005). In fragment completion, Feldman et al. (2002) observed that children more readily completed the fragment as per the prime when it was morphological transparent, compared to when there was a shift in form (e.g., target 'b_ _d' ('bend') primed by 'bent'). McCutchen et al. (2009) observed a similar trend in their continuous lexical decision priming study, with children faster to respond to targets that shared complete orthographic and phonological forms with the prime, compared to the primes with phonological shifts. However, the impact of transparent or stable stem was less pronounced in priming than reading aloud tasks. This may be attributed to phonological transparency effect being highly sensitive to task characteristics, such as response type (e.g., lexical decision vs. naming). Despite this, the stem

transparency effect has been observed in masked priming studies, where elementary school children respond faster to targets preceded by morphologically related primes without base modifications, but do not show significant priming there is a phonological shift between prime and target (e.g., bergerie [sheepfold]-BERGER [shepherd]; English equivalent: direction-DIRECT) (Quemart & Casalis, 2014, 2015). This suggests that phonological codes might be activated at the early stage of morphological processing. Quemart and Casalis (2015) found priming effects for phonologically shifted primes as well, but only in adolescent readers aged 12 years (including both dyslexic adolescents and age-matched skilled readers), but not in younger 8-year-old typically developing children, highlighting developmental progression in morpho-orthographic decomposition. However, even among skilled adolescent readers, priming was still absent with primes involving both phonological and orthographic shifts (e.g., soigneux [careful]-SOIN [care]; English equivalent: fifth–FIVE), which only dyslexic adolescents exhibiting priming across all conditions. These findings underscore the importance of an intact base word for children to effectively locate it during morphological decomposition (Quemart & Casalis, 2014). This study will be revisited in Section 2.3.5.3 on reading difficulties.

2.3.4.3 Stem vs. affix

As reviewed above, stems and affixes both uniquely and interdependently influence morphological processing, with stems or base words appearing particularly crucial as the identification of an intact base word is essential for effective morphological decomposition (e.g., Carlisle & Stone, 2003). The semantic categorisation study by Hasenäcker, Olga et al. (2021) revealed automatic activation of embedded stem semantics from orthography in children in Grade 3-5, demonstrating that stems were recognised beyond merely being familiar letter strings. Similarly, embedded stems in nonwords were activated by Grade 2 children, regardless of affix status (e.g., Beyersmann et al., 2015, 2021; Hasenäcker et al., 2016, 2020). Beyersmann et al. (2021) found that although suffix processing is slower compared to prefix processing due to the

‘left-to-right’ reading system, embedded words were processed rapidly in both initial and final string positions. The morphological-structure-independent activation of embedded stems in nonwords suggests that morpho-orthographic processing might not be employed. This corresponds with the relatively late emergence in mid-adolescence of pseudomorphological priming and the ‘morpheme interference effect’ in pseudoword rejection (e.g., Dawson et al., 2019; Beyersmann et al., 2012), both of which are considered dependent on automatic morpho-orthographic processing.

The prominence of word-based, non-morphological processing in early reading development suggests that recognising embedded stems might be a precursor to acquiring detailed morpho-orthographic representations (Hasenäcker et al., 2016). This supports the ‘semantic view’, which posits that children initially identify meaningful parts within words, such as the stem 'teach' in 'teacher', and might start treating the remaining orthographic portion, like the affix '-er', as a separate morphological unit after recognising the stem morpheme. Grainger and Beyersmann (2017) proposed the ‘edge-aligned embedded stem activation’ model, explaining how embedded stems map directly onto whole-word representations, without the need for complete decomposition of the stimuli or stripping off affixes. In cases like 'teacher', where there is a morpho-semantic relationship between the stem ‘teach’ and the whole word, their co-activation facilitates stem recognition. However, in cases where whole words and embedded stems are semantically unrelated, inhibition can occur. This inhibition does not occur in nonword primes, due to their lack of established lexical representations. In pseudomorphological primes like ‘corner’, the inhibition between the whole word and embedded word ‘corn’ can be compensated by the presence of the suffix-like unit ‘-er’, thus boosting the activation of ‘corn’. Yet, this process relies on well-established affix representations in the mental lexicon, which gradually develop over time. Hence, the compensatory effect is not observed in non-suffixed form-related primes like ‘cashew’ or in developing readers with less well-established affix representations. Therefore, it is proposed that

morpho-orthographic processing is initiated by edge-aligned embedded word activation, with affixes playing a secondary role (Grainger & Beyersmann, 2017).

Some studies have directly explored the interplay between stems and suffixes by orthogonally manipulating their presence or absence in various conditions (e.g., stem + suffix, stem + nonsuffix, nonstem + suffix, and nonstem + nonsuffix). Research (Casalis et al., 2015; Duncan et al., 2011; Quémart et al., 2012) found that with pseudowords, the combination of a root and a suffix hindered rejection, confirming the 'morpheme interference effect', whereas no suffix effect was observed for pseudowords without an embedded word. However, patterns varied for familiar words where suffixes played a more crucial role: children performed similarly in word recognition when there was a stem (e.g., similar performance to 'farmer' or 'window'); and only for in the absence of a stem did suffix presence (e.g., in 'murder') lead to more accurate word acceptance compared to words without suffixes (e.g., 'narrow'). This might be attributed to the fact that stems often standing alone as words, whereas affixes, which typically cannot stand alone as independent words, play a secondary role in enhance word recognition. In tasks like reading aloud, both stem and suffix facilitation effects are seen: readers benefited from these familiar morphemes, resulting in faster and more accurate naming when either a stem or a suffix was present compared to when there was no stem or suffix present, though the suffix effect was more dependent on reading skill (Mousikou et al., 2020). Furthermore, Pescuma et al. (2022) outlined neurological evidence to support the view that stems are more salient, primary units in the developing reading system. They used an 'oddball-MEG' paradigm to investigate brain activation during stem versus suffix identification. They found significant brain responses in children during stem identification (e.g., oddball with stem+ suffix like 'softity' among nonstem+ suffix base stimuli, like 'joskive') but not in suffix identification conditions (i.e., the same oddball among stem+ nonsuffix base stimuli, like 'helpune').

Compound words contain multiple stem words (e.g., 'sunflower). These offer valuable insights into the roles of stems in word recognition. Studies have shown that the compound written word

recognition is influenced by the properties of their first constituent, suggesting a left-to-right decompositional processing (de Zeeuw et al., 2015; Hasenäcker & Schroeder, 2019). Beyersmann et al. (2019) further supported the notion of full decomposition by demonstrating no priming effect in noncompound conditions (e.g., sandwich-SAND). This occurs even for semantically opaque compounds (e.g., butterfly) and is seen in children as young as Grade 3. This is an interesting finding, given pseudosuffix priming is not typically observed at this age. The automatic segmentation of compound words implies that embedded stems might act as a bootstrapping mechanism for morphological parsing. While children quickly parse stems in compound words, they may develop the processing mechanisms for affixed words more slowly. This hypothesis is corroborated by Hasenäcker et al. (2017), who found that the compound effect emerged earlier than prefix and suffix effects in both word recognition and pseudoword rejection.

Taken together, these findings suggest that stems and affixes play integral yet distinct roles in morphological processing. While children readily activate embedded words and parse stems in compound words, they might be slower to develop the mechanisms needed to process affixed words, indicating that stems act as primary anchors in children's processing of written words.

2.3.4.4 Frequency

Manipulation of frequency is common in morphological processing research, particularly in studies examining the impact of linguistic experience. Like other characteristics such as word length, imaginability, and age of acquisition, frequency – as a crucial distributional property – is often matched across conditions or included as a covariate. This approach could be considered 'standard practice' in research that either does not primarily focus on frequency effects or concentrates on one type of frequency but not the other, thereby controlling for potential confounds. Often, this includes controlling for the frequency of both whole-word stimuli (e.g., de Zeeuw, 2013; Fleischhauer et al., 2021; Leong & Chen, 2000) and sublexical units like bigram frequency (e.g., Angelelli et al., 2014; D'Alessio et al., 2018). There are two primary types of frequency: token and

type. Token frequency refers to the total occurrences of a lexical unit in a corpus, whereas type frequency is concerned with the number of distinct word forms containing that unit (Diependaele et al., 2012, pp. 316-319).

Token frequency significantly influences language processing, as increased exposure to words leads to greater familiarity. This familiarity builds lexical quality (Reichle & Perfetti, 2003) and is associated with better reading of all word types including complex words, as revealed by studies of reading aloud and lexical decision (e.g., de Zeeuw et al., 2013; Fleischhauer, 2013; Kearns, 2015). Additionally, surface frequency has been found to influence the utilisation of morphological structures during reading. For instance, Marcolini et al. (2011) found that typically developing children named complex words more quickly than simple ones, but this pattern was exclusive to low-frequency words. However, D'Alessio et al. (2018) did not replicate the findings, observing that for both low- and high-frequency words, morphological structure affected the naming. They attributed this discrepancy in findings to different data analysis approaches and control variables. Moreover, the high-frequency morphologically complex words used by D'Alessio et al. (2018) were less common than those in Marcolini et al. (2011); it might be that their frequency was not sufficient to support processing as whole words, resulting in reliance on morpheme-based reading strategies. Subsequent research by D'Alessio et al. (2019) found that the positive effects of morphological structure on lexical decision accuracy were most apparent in low-frequency words, supporting the notion that while frequent words are processed rapidly as whole entities, morphology plays a role in recognising infrequent words. The eye movement data from Traficante et al. (2018)'s study provided additional evidence for differing processing strategies in low- versus high-frequency words, showing that the inhibitory effect of word length on gaze duration was stronger for low-frequency words. The reduced impact of word length on high-frequency words might stem from their processing through coarse-grained orthographic representations, in contrast to the fine-grained representations used for low-frequency words.

While high surface frequency benefits the processing of both simple and complex words, the latter are characterised by various distributional factors including distinct frequencies for bases (roots or stems) and affixes, each playing a unique role in the morphological processing of complex words. For instance, Lazaro et al. (2013) demonstrated that children made faster lexical decisions for stimuli composed of frequent bases. The frequency of the base, the main meaning-bearing unit of a word, influences word processing, even after accounting for related factors such as family size. Through regression analysis, Deacon and Francis (2017) illustrated that base frequency significantly impacts word naming accuracy, after factoring in variables such as family size and whole-word frequency. Importantly, Traficante et al. (2011) suggested that the effects of base frequency were not solely attributable to familiarity with grapheme sequences, as demonstrated in studies using pseudoword stimuli. Higher base frequency emerged as the most reliable predictor of enhanced speed and accuracy in naming, even when the bigram frequency of the left part of the pseudowords (corresponding to the same root/nonroot unit in various conditions) was taken into account (Traficante et al., 2011). This suggests that the impact of root frequency on word reading extends beyond the influence of its orthographic patterns.

Studies have also investigated affix frequency effects in children's reading. Angelelli et al. (2014, 2017) found that high-frequency suffixes facilitated naming accuracy. However, it might be difficult to distinguish the suffix frequency effect from base frequency effect because the stimuli used in Angelelli et al. (2014, 2017) consisted of base + suffix combinations where either both the base and the suffix were high-frequency, or both were low-frequency. Traficante et al. (2011) observed that although high-frequency suffixes did not facilitate the onset of pseudoword pronunciation, likely due to their position at the end of words, higher suffix frequency was associated with more accurate naming. This effect contrasted with the effects of the bigram frequency of the right part of the stimuli (corresponding to the suffixes and nonsuffix endings), where higher bigram frequency of the right part letter string was associated with lower accuracy.

These findings further underscore the morphemic nature of suffix frequency effects, distinguishing them from purely orthographic influences. Burani et al. (2018) also found that frequent suffixes were associated with higher naming accuracy in both dyslexic and typically developing children, with faster responses to words with high-frequency suffixes also observed in the latter group. Similarly, the magnitude of the morphological priming was found to increase with higher suffix frequency (Beyersmann et al., 2012). Lázaro et al. (2016) focused primarily on the suffix frequency effect in lexical decision, finding that words with frequent suffixes were processed more quickly and accurately than those with infrequent ones, a trend consistent across Grade 2-4 children. The early sensitivity to suffix frequency among Spanish-speaking children may stem from the frequent use of derivation in Spanish, resulting in regular exposure to derived structures and increased familiarity with them.

According to the dual-route reading model (Grainger & Beyersmann, 2017), both whole words and their morphemic constituents are concurrently activated during processing. For instance, Lázaro et al. (2015) found that increased frequencies in both surface and base led to fewer errors in lexical decision tasks. Burani et al. (2018), however, noted that base and surface frequencies influenced different outcomes of word processing: a higher root frequency sped up naming, while a higher surface frequency increased accuracy. This suggested that base frequency likely affected pronunciation onset due to left-to-right reading, and surface frequency is key for overall reading accuracy, as previously discussed in the task modulation section. Similarly, eye movement data from Whylie (2021) demonstrated that in silent sentence reading, while target word surface frequency affects later processing measures, higher base frequencies specifically reduced go-past duration. This, they argued, reflected greater reliance on morphological processing for words with higher base frequencies, particularly during re-reading for contextual integration. In a similar sentence reading study that also had children read aloud, Traficante et al. (2018) found that base and derived frequencies influenced reading aloud derived words differently, based on the grammatical

class of base. Deverbal nouns with higher base frequencies exhibited longer first fixation durations due to conflicts between the verb base and syntactic constraints. Conversely, for nouns with noun bases, high-frequency bases accelerated first fixation whereas higher derived frequencies caused delays, suggesting that although children may gain a head-start from a high-frequency base, they may also face processing conflicts between the base and the high-frequency whole word when both are grammatically appropriate. This is in line with earlier findings reported by Traficante et al. (2014), who observed that frequency effects in single-word reading were modulated by base grammatical class. Besides this, they noted an effect of suffix length, where longer suffixes increased RTs for both types of derived words, a factor not commonly explored in other studies investigating frequency effects. In addition, root frequency was found to influence lexical decisions in compounds that usually had two or more roots. Studies by de Zeeuw et al. (2015) and Hasenäcker & Schroeder (2019) revealed that higher frequencies of both the compounds and their first constituents were associated with faster lexical decisions, while the second constituent root primarily influenced reading accuracy.

These studies imply that examining the frequency of either whole words or morphemes in isolation might offer only a partial understanding. Research has delved into how the independent and interactive roles of surface(whole-word) and base frequencies contribute to morphological processing. Carlisle (2000) suggested that surface frequency played a major role in children's reading of complex words, as the advantage of high-surface-frequency words over their low-surface-frequency counterparts was not offset by their comparable base frequencies. In this study, Carlisle (2000) analysed both high- and low-surface frequency words collectively, using an outcome measure that blended speed and accuracy (e.g., awarding two points for correctly reading a word within two seconds). When focusing specifically on words with low surface frequency, as in the study by Carlisle and Stone (2003), or when RTs and accuracy were analysed separately, as in Carlisle and Stone (2005), regression analyses indicated that base frequency accounted for a

significant proportion of the variance in reading rare low-frequency, words, alongside the robust surface frequency effect. Similarly, Deacon and Francis (2017) observed that complex words with a low surface frequency were named more accurately when they had a higher frequency base. Additionally, Verhoeven and Schreuder (2011) found in their lexical decision study of plural words that, while higher surface and base frequencies generally facilitated accuracy and speed, the root frequency effect was more pronounced in low surface frequency words than in those with high surface frequency.

While token frequency measures the individual occurrences of words, bases, and affixes in texts, type frequency counts the variety of distinct lexical elements. This interrelation between token and type frequencies is well-recognised, as evidenced by studies controlling for type frequency-related factors such as family size, family frequency, and suffix productivity when investigating token frequency effects in children's reading (e.g., Deacon & Francis, 2017; Lazaro et al., 2013, 2015; Traficante et al., 2014). The significance of type frequency by itself in morphological processing has also been investigated. Within this area, a few studies have focused on the family size effect. Carlisle and Katz (2006) found that children more effectively named words belonging to larger families with higher Average Family Frequencies (AFF; a metric calculated by dividing the total frequency of all family members by the family size). However, the results should be interpreted with caution as words with different family sizes also varied in base word frequency. In contrast, Burani et al. (2018) reported an inhibitory effect of family size in their study focusing on reading low surface frequency derived words, where larger root family sizes delayed naming onset as dyslexic children read words aloud. This discrepancy could stem from differing methodologies, with Carlisle and Katz (2006) using a combined score of accuracy and whole-word pronunciation speed, and Burani et al. (2018) measuring naming onset speed. The latter study might suggest potential serial processing as larger root families could lead to multiple base candidates, causing uncertainty in planning pronunciation, particularly for low-frequency words and in children with

limited reading abilities who might struggle to process the entire word at onset. There could potentially be cross-linguistic differences as Carlisle and Katz (2006) studied English, whereas Burani et al. (2018) investigated Italian, a more transparent and morphologically rich language, compared to English.

The impact of affix type frequency, linked to stem diversity and affix productivity, is another key aspect of type frequency. Verhoeven et al. (2006) and van Hoogmoed et al. (2011) found that Dutch-speaking children recognised words with real prefixes ('GE-' and 'BE-') or phonological prefixes with the same sound patterns (i.e., /gə/ and /bə/) more effectively than those with different-sounding pseudoprefixes (i.e., /ge/ and /be/). The prefix status effect [7] was more pronounced in words with 'GE-', potentially due to its higher productivity and type frequency. However, this interpretation requires caution, as GE- also differs from BE- in token counts and other prefix characteristics. For instance, Van Hoogmoed et al. (2011) noted that GE- is more frequently used as a true prefix /gə/ compared to its pseudoprefix (/ge/), exhibiting a higher true-to-pseudo usage ratio than BE-. The relative frequency of a letter string as a real versus pseudo prefix might influence word identification, intertwining with type frequency and other frequency-related factors of affixes (Verhoeven et al., 2006). This phenomenon was investigated by Assink et al. (2000), who found a stronger morpheme interference effect in nonwords with higher type frequency prefixes, an effect further modulated by the ‘corruptness’ of the prefixes (defined by how commonly they occur as true prefixes, with incorrupt prefixes having a high true/pseudo ratio). In cases with high type-frequency prefixes, the largest error rate difference was seen in the 'incorrupt' prefixed nonwords, whereas in the low prefix frequency condition, a contrasting pattern was observed. It can be inferred from the results that the impact of prefix type frequency on morpheme interference was more salient with incorrupt than corrupt prefixes. One explanation could be that incorrupt prefixes, often functioning

[7] Yet, it might be challenging to attribute the prefix status effect solely to morphology or to fully disentangle the roles of morphology and phonology, as neither study distinguished between true and phonological prefixes.

as true prefixes, and thus, when they have higher type-frequency, could generate stronger expectations for legitimate words compared to their corrupt counterparts, making it more difficult to reject nonwords containing them. The likelihood of prefix status, which is reminiscent of Ulicheva et al. (2020)'s ‘Diagnosticity’ construct, pertains to the ‘cue validity’, or the amount of meaningful information an affix provides in word recognition. Consequently, the recognition of a sublexical unit in morphological processing is shaped not just by its frequency of occurrence as a token, but also by other statistical regularities influencing the encoding and activation of a prefix when words are experienced across a child’s reading history.

Further research in this domain could yield valuable insights into the interplay between type frequency and other affix characteristics in morphological processing. Computational modelling is also helpful, and these frequency effects align with the word-experience model simulated by Reichle and Perfetti (2003). This posits that frequent exposure to words and their morphemes boosts familiarity, thus aiding the formation of high-quality lexical representations of words and morphemic units in the mental lexicon. As effective reading depends on using orthographic forms as retrieval cues, frequency can be seen as a key indicator of both how well these forms are established and how easily they can be activated, thereby affecting their morphological processing.

2.3.5 Individual-level factors

Knowledge about words is partially influenced by item-level factors such as frequency. This type of factor depends not only on the inherent frequency of these units but also on the experiences readers have with them. This leads naturally to a discussion of individual- or child-level factors. Both age and reading ability influence morphological processing (e.g., Carlisle & Katz, 2006). Increases in proficiency and accumulated reading experience with age lead to higher-quality lexical representations (Perfetti & Hart, 2001; Reichle & Perfetti, 2003), thereby facilitating the use of stems and affixes as processing units. While chronological age or years of schooling are generally correlated with accumulative reading experience, they do not always serve as reliable indicators of

reading expertise as there is significant individual variability in reading ability within an age or grade band. Therefore, morphological processing is influenced by both age and varying reading abilities. Studies included in this review consistently demonstrate the significant impact of these individual-level factors. Studies involving different age groups (including longitudinal research) or levels of reading proficiency typically show expected improvements in morphological processing as age and ability increase (e.g., Deacon et al., 2017; Leong & Parkinson, 1995; McCutchen et al., 2009; Quémart et al., 2011). Moreover, in addition to general performance trends, these factors can modulate morphological processing in various ways.

2.3.5.1 Age and grade: developmental milestones and years of schooling

Developmental progression is evident in studies such as Carlisle and Katz (2006), where 4th graders exhibited a greater discrepancy in naming derived words from large vs. small families compared to 6th graders, illustrating that proficiency in reading complex words advances with grade. Similarly, de Zeeuw et al. (2013) documented improvement in lexical decisions to past tense verbs, with 3rd graders struggling more with irregular verbs than with regular ones, a gap that was mitigated by Grade 6. These patterns suggest that younger readers tend to rely on familiar patterns, while older students show increased adaptability in processing derived words, whether from smaller families or in irregular forms, indicating more advanced morphological processing abilities. Further evidence of this developmental progression comes from Carlisle and Stone (2005), who observed a narrowing performance difference in recognising high and low derived-frequency words — both with high-frequency bases — among 5^{th} -6^{th} graders compared to 2^{nd} -3^{rd} graders. This suggests that enhanced morphemic processing allows upper elementary students to benefit more from identifying high-frequency bases. Additionally, Angelelli et al. (2014) found that 5th graders, unlike 3rd graders, were more prone to mispronounce derived words with low-frequency morphemes than non-derived words. This indicates that older readers, upon encountering a root, might face uncertainty when they expected a commonly associated suffix. These findings highlight the

development of morphological processing with grade and reading experience, and how this leads to improved morphological parsing and potentially, an increase in reliance on morphology in visual word recognition.

The growing dependency on morphological decomposition relies on refined morphemic representations, achieved through accumulated experience with complex words. Although children show sensitivity to morphological structures early on—for example, 2nd and 3rd graders more accurately read derived words (e.g., 'winner') over pseudo-derived ones (e.g., 'dinner') (Carlisle & Stone, 2005)—morphology does not play an exclusive role in word processing, especially before children can efficiently use morphemes as processing units. At this stage, they may primarily rely on other lexical units, such as syllables, regardless of whether the task requires oral output. Hasenacker and Schroeder (2017) found that when making lexical decisions on stimuli disrupted at either a syllable-incongruent morphemic boundary (e.g., FAHR:ER) or a syllable-congruent but morphology-incongruent boundary (e.g., FAH:RER) both 2nd and 4th graders were facilitated by syllable-congruent disruptions in word reading; only 4th graders, however, exhibited a selective delay when rejecting multimorphemic pseudowords disrupted at the morpheme boundary. This distinction might indicate a transitional phase in reading development where morphemes start to become more relevant. From 4th onwards (through 6th grade), children might process words defined by morphology more efficiently than those defined by phonological principles, as suggested by an earlier study using a similar lexical decision paradigm (Leong, 1989). Also, Quemart and Lambert (2019) demonstrated that when copying derived words, 4th graders showed shorter writing durations for letters at syllabic boundaries in complex words than in simple ones, whereas 6th graders showed longer durations for letters preceding the morphemic boundary in complex words with no significant differences at the syllabic boundary. Therefore, with increasing reading experience, the mental lexicon might become more organised around morphemes, potentially due to the larger grain sizes of morphemes and their effectiveness in directly retrieving meanings from

orthography (see Mousikou et al., 2020, for a discussion). This possible transition from early, laborious phonological decoding to more efficient morpheme-based reading is further suggested by studies showing a decreasing gap between reading complex words with and without phonological alternations (e.g., Carlisle & Stone, 2005, Study 2). Yet, more research is needed to examine whether and under what circumstances morphology overtakes syllabic processing, and the individual(child)- and item-level characteristics that enable this to happen.

Understanding these developmental milestones and the mechanisms underlying their acquisition is essential. As mentioned previously regarding task level modulations, distinct pathways might be involved in processing. For example, according to Beyersmann and colleagues, embedded words are processed through the whole-word pathway based on nonmorphological processes, while affixes are processed via affix activation and morpho-orthographic full decomposition (e.g., Beyersmann et al., 2021; cf. 'word and affix' account; Grainger & Beyersmann, 2017). A robust embedded word activation effect has been observed from early elementary school years onwards. For instance, masked priming studies (e.g., Beyersmann et al., 2015, 2019; Hasenacker et al., 2016, 2020) have demonstrated that from the second grade onwards, priming of a target stem word (e.g., 'KLEID', [English equivalent: 'FARM']) occurs when it is preceded by suffixed words (e.g., 'kleidchen', [English equivalent: 'farmer']), suffixed nonwords (e.g.,'kleidtum', [English equivalent: 'farmation']), or nonsuffixed nonwords (e.g., 'kleidekt', [English equivalent: 'farmald']) containing it, all showing a similar magnitude of priming. This provides evidence that children automatically activate embedded words from this age, indicating an initial reliance on whole-word representations without morpho-orthographic decomposition. Using semantic categorisation, Hasenäcker, Olga, et al., (2021) found that embedded word activation might strengthen with grade, as the influence of the embedded word became stronger from 3rd to 4th graders. However, this association between embedded word activation and grade may stem more from increasing reading abilities than from grade or age itself, a topic we come back to later in this review.

Although embedded word effects are evident early on in reading acquisition, rapid and automated morphological processing, including affix activation during the full decomposition of morphologically structured words, unfolds gradually through the primary years, though the exact timeline varies across studies. While early morphological priming has been observed with brief primes in several studies (e.g., Beyersmann et al., 2012; 2015; Hasenacker et al., 2016), de Oliveira & Justi (2017) reported later emergence of this effect, as noted earlier. Research on the time course of morphological processing (e.g., Quémart et al., 2011) delineates two stages: initial morpho-orthographic processing marked by rapid, automatic form-based processing, and a subsequent morpho-semantic phase where semantic processing becomes relevant. Thus, the findings of de Oliveira and Justi (2017) might reflect younger children depend more on morpho-semantic information when processing, whereas by 5th grade, children have acquired sensitivity to morpho-orthographic regularities and can decompose complex words rapidly and automatically. This shift suggests that with sufficient experience of morphemes, children can gradually access word meanings based on morphological regularities in a morpho-orthographic item-based manner, rather than mapping each written word onto its meaning.

The embedded word activation account (Grainger & Beyersmann, 2017), as detailed in the 'stem and affixes' section earlier in this review, suggests that activating embedded stems may initiate the acquisition of morphological decomposition. This process aids children in gradually encoding the orthographic representations of morphemic constituents, such as affixes (e.g., '-er'in 'teacher') and from this, develop full morpho-orthographic decomposition. As mentioned previously, Hasenacker et al. (2017) demonstrated that the use of stems and affixes as processing units in visual word recognition depends on experience with these constituent morphemes. In their study, stem representations emerged first, resulting in earlier facilitation effects for complex words over monomorphemic words—specifically, the facilitation effect for complex words was observed in compounds first, from the end of 2nd grade onwards, in suffixed words from 3rd grade onwards,

and emerged in prefixed words from 4th grade onwards, highlighting the emergent status of morphemic representations in the developing reading system (Hasenacker et al., 2017). Supporting this shift in morpheme recognition, Hasenäcker et al. (2020) found that from Grades 2 to 4, with increasing years of reading experience, there was an increase in priming from suffixed words and suffixed nonwords, while priming from non-suffixed nonwords decreased. This pattern indicates that sensitivity to affixes is acquired over grades, and with continued schooling, children, irrespective of their early differences in embedded word activation, begin to employ an affix-stripping mechanism, transitioning from primary reliance on embedded word activation to a more efficient affix-stripping approach, as their affix representations become more refined[8]. Furthermore, Lazaro et al. (2018) highlighted progressive sensitivity to affixes with age: only 6th graders, but not the 4th and 5th graders, exhibited significant facilitation when responding to complex words primed by other words sharing the same suffix. This suggested that the facilitative effect of suffixes in lexical access is associated with grade level and reading experience. It also supports suffixes serving as meaningful processing units in visual word recognition, as children develop expertise. Additionally, Quémart et al. (2012) found in a lexical decision task that the presence of suffixes sped up word recognition in 5th graders but not in 3rd graders, implying that the younger children had not yet fully developed the ability to leverage morphological structures like suffixes, effectively.

Pseudomorphological priming (e.g., corner-CORN), depends on the rapid activation of affix representations based on the surface morphological structure, and serves as an indicator of automatic morpho-orthographic processing. For example, Quémart et al. (2011) found priming effects in both truly morphological and pseudomorphological conditions in French children from 3rd grade, but not in the form-overlap condition (e.g., brother-BROTH). Furthermore, Flischhauer

[8] This does not imply that the embedded stem activation is replaced by the later acquired morpho-orthographic mechanisms. Instead, research has found that both mechanisms coexist (e.g., Hasenäcker et al., 2020).

et al. (2021) found significant pseudomorphological priming beyond the baseline form condition in German 4th graders, indicating a rapid segmentation into morpheme-like elements, whereas 3rd graders still relied primarily on truly morpho-semantic relationships (see also Casalis et al., 2009). Notably, neither type of morphological priming was observed in 1st and 2nd graders. Conversely, Beyersmann et al. (2012) reported that even in Year 5, English children did not show distinguishable pseudomorphological priming from the form-related baseline. Advancing this research, Dawson et al. (2021) adopted the stimuli from Beyersmann et al. (2012) to study participants across a wider age range, from 9-10 years (Year 5) to 16-18-year-old adolescents. The results showed an age-related (analysing age in months as a continuous variable) increase in the magnitude of priming, with older participants showing pseudomorphological priming. This points to a gradual development of morpho-orthography processing throughout the elementary school years and may extend well into adolescence. This fits with Dawson et al.'s (2018) observations from morpheme interference across the same age range. Across age, children made more mistakes when trying to reject pseudomorphemic nonwords (e.g., 'earist') in lexical decision, compared to control nonwords (e.g., 'earilt'). The interference effect was significantly stronger among older adolescents (16- to 17-years) compared to children (7-9 years) and younger adolescents (12-13 years). Notably, the morpheme interference effect was only seen in the RTs in older adolescents. These findings indicate that adolescence could be a crucial phase for the development of rapid online morphological decomposition.

The studies I have reviewed illustrate the path towards rapid, automatic morphological processing in reading development. Yet, the variability across languages, samples, and paradigms used makes it difficult to pinpoint exact ages or grade-levels. The progression observed through years of schooling emphasises the critical role of reading experience. This might reflect a growing sensitivity to morphological structure as a function of reading more complex words. According to a large-scale corpus analysis by Dawson et al., (2023), complex words constitute a large proportion of

the vocabulary in reading materials targeting English children aged 5-14 years, and this proportion increases with age. It might also reflect the importance of explicit word reading instruction in fostering sensitivity to morphological structures.

Note that our discussion has not clearly distinguished between age and grade—although the relationship between age and grade is generally straightforward, cross-linguistic studies (e.g., Casalis et al., 2015) show that children of different ages can be in the same grade level. Age reflects accumulated reading experience and cognitive maturation, whereas grade corresponds to years of schooling and exposure to reading instruction, a topic I return to later. As highlighted earlier in the '2.3.1 Study Characteristics' section, relatively few studies have captured adolescence resulting in an evidence gap. Future studies might thus benefit from broadening the age range. It is also possible that certain tasks might be too simple for older readers, causing a ceiling effect (e.g., Deacon et al., 2011) and thus masking potential variations. Importantly, other factors such as reading proficiency, which typically develops with age and grade level, also have considerable influence on morphological processing, potentially exceeding the impact of age alone, which underscores the importance of accounting for both proficiency as well as age and grade.

2.3.5.2 Reading abilities: experience and individual variability

Reading ability, developing from cumulative exposure to diverse texts (Dawson et al., 2021; Nation, 2017), does not uniformly increase with years of schooling. Traficante et al. (2018) posited that grade pertains to literacy exposure and explicit learning whereas reading proficiency encompasses reading-related skills and implicit learning. To disentangle these effects, they asked children to read suffixed words and tested whether proficiency (calculated using factor analysis on reading speed) influenced performance above the influence of grade. They found that incorporating proficiency scores improved model fit and rendered the influence of grade insignificant. This indicates that the processing of morphologically complex words was modulated by individual differences in reading skills, not by grade itself. Additionally, Dawson et al., (2021) hypothesised

that orthographic knowledge (as tapped by reading ability) could be a better predictor of morpho-orthographic processing than age, as processing relies on orthographic representations of morphemes. Their exploratory analysis, which replaced age with a standardised measure of word reading fluency revealed that pseudomorphological priming (‘corner-CORN’) showed a significantly larger increase than form priming (‘mother -MOTH’) as a function of reading ability. Notably, this pattern was not seen in morphological priming (‘teacher-TEACH’).

In a longitudinal study of masked priming, Hasenäcker et al. (2020) aimed to distinguish the effects of grade and reading ability. Accordingly, they analysed word reading scores standardised within each grade. They found that both grade and reading skill influenced priming effects, but in distinct ways: increasing grade correlated with increased sensitivity to affixes, as discussed above, whereas higher levels of reading skill were associated with embedded stem activation. Good and average readers exhibited stem activation by Grade 2, as evidenced by priming effects, not only with suffixed words but also nonwords containing the target, regardless of the affix status. In contrast, poor readers were less efficient in activating embedded words, even by Grade 4. This might be attributed to differences in lexical quality (Reichle & Perfetti, 2003) between good and poor readers. On this view, good readers possess more stable morpheme representations, and this enables them to directly map embedded stems onto whole-word representations. In contrast, poor readers, with fewer and weaker orthographic representations, are more dependent on meaningful morphological relationships. Consistent with this, Beyersmann et al. (2015) found that priming effects from both suffixed and non-suffixed nonwords were modulated by reading proficiency (a composite of spelling, reading proficiency, and morphological awareness) in Grade 2-5 French children. Proficient readers exhibited facilitatory priming, whereas less proficient readers encountered inhibitory priming. The authors posited that low-proficiency children made slower responses as the nonword primes competed with orthographically similar known words. This was less evident in high-proficiency children as embedded words were activated more strongly. These

findings underscore the impact of reading proficiency on embedded stem activation during visual word recognition.

Using traditional (unprimed) lexical decision, Duncan et al (2011) systematically manipulated the presence of embedded stems and suffixes and found that good readers were more accurate with words containing orthographic suffixes (e.g., 'murder') than those without (e.g., 'narrow'); they also showed the morpheme interference effect and found it hard to reject pseudowords with morphological structure (e.g., 'gifter'). Notably, the presence or absence of a suffix did not significantly affect responses when a stem was present in words (e.g., comparing 'farmer' with a suffix to 'window' without a suffix), or absent in pseudowords (e.g., comparing 'gopter' = 'gopt' + '-er' to 'ferbow' = 'ferb'+ 'ow'). This highlights the primary role of stems in morphological processing. In contrast, poor readers were not influenced by suffix presence, with stem presentation mainly playing the role of facilitating word recognition and impeding pseudoword rejection. Similarly, Hasenäcker et al. (2017) reported that individual differences in vocabulary knowledge moderated the effects of morphology on reading in German children in Grades 2 to 6. Children with stronger vocabulary knowledge demonstrated early sensitivity to constituent morphemes, benefiting from morphology in compounds and derived words over simple words since Grade 2. In contrast, those with weaker vocabulary experience difficulties in processing derived words and exhibited the interference effect with prefixed pseudowords at an older age, compared to their more proficient peers. This performance gap between good and poor readers demonstrates that strong word reading and vocabulary contribute to the development of stable morpheme representations, which in turn are crucial to support their rapid activation, and further enhance effective decomposition and processing of complex words. Not surprisingly then, reading ability is closely associated with morphological processing. For instance, Lázaro et al. (2016) found significant correlations between reading, vocabulary size and the size of the suffix frequency effect (i.e., facilitation of words containing frequent suffixes). Leong and Parkinson (1992) observed that poor readers showed less sensitivity

to morphological boundaries than better readers. The researchers applied discriminant function analysis to their latency data and found that this effectively predicted group membership (i.e., whether a child was a good or poor reader, based on word reading and spelling skills.

In reading aloud, phonological components of reading ability are naturally pertinent but morphological aspects have also been implicated when reading complex words. Saiegh-Haddad and Geva (2008) demonstrated that both phonological awareness and morphological awareness independently contributed to reading derived words. Importantly, morphological awareness remained a significant predictor of reading fluency in their hierarchical regression model, even after accounting for phonological awareness. These findings suggest that readers might draw on morphological knowledge when naming morphologically complex words. Specifically, developing readers might seek out morphology more for decoding certain types of words. For example, morphological awareness, beyond basic phonological processing abilities, particularly enhances the decoding of polysyllabic polymorphemic words that feature phonological shifts (Kearns, 2015), as well as opaque and low-frequency words (Mann & Singson, 2003) that are not readily decoded according to phonological principles. Furthermore, in a study of Italian children, Marcolini et al. (2011) found that 6th-grade poor readers named derived words more quickly than simple ones, irrespective of frequency, whereas their more skilled peers only benefitted from morphological structure when reading low-frequency words. This pattern suggests that in transparent orthographies, activating whole-word units could be more efficient than morphemes for skilled readers, who might only lean on morphological knowledge when dealing with unfamiliar words. However, as morphemes provide shorter reading units, morpheme-based reading could be particularly beneficial for children with limited reading abilities who have yet to fully develop whole-word processing. Indeed, for unfamiliar items such as nonwords, it has been shown that the advantage of suffix presence in reading increases with higher word reading fluency and vocabulary skills (Mousikou et al., 2020). Similarly, the speed to produce a stem after being primed by its

derived forms is associated with spelling performance and vocabulary size (Leong & Chen, 2000). Laxon et al. (1992) found that children named affixed words more accurately than nonwords created by adding an affix to a stem, but poor readers exhibited a larger disparity between word and nonword reading accuracy. This suggests that they benefitted less from recognising the word stem in nonwords, compared to good readers who were more adept at utilising morphological structures, presumably due to higher levels of morphological sensitivity.

In summary, individual differences in reading ability appear to modulate morphological processing, potentially accounting for at least part of the effects of age and grade. An increase in reading ability could be associated with higher lexical quality and more stable orthographic representations of morphemes, leading to a greater sensitivity to morphological structure and therefore greater opportunity to utilise this during reading. There are several issues worth considering. The relationship between reading and morphological processing might appear circular, as morphological processing ability can be seen as an indicator of reading ability and more generally, it may be both a driver and consequence of reading expertise (Nation, 2017). The question remains whether efficient morphology use by proficient readers stems from their robust word representations or if other factors contribute to both reading and morphological processing skills. For example, given that children show sensitivity to morphological structure before formal reading instruction begins (e.g., Berko, 1958), external factors such as the home learning environment or print exposure could contribute to both early sensitivity to morphology and higher reading ability. However, these factors have rarely been studied in the existing literature. Furthermore, reading ability is multifaceted and can be measured in different ways. However, given the often-correlated nature of these various aspects, it is crucial to clarify why specific facets are chosen because some abilities directly affect morphological processing, whereas others may seem influential due to correlations with other components of reading ability. Furthermore, definitions of good and poor readers vary, as do grade boundaries and the age of initial literacy instruction.

2.3.5.3 Reading difficulties

Some studies have included children with reading disorders. While these children may form part of the continuous variation associated with individual differences in reading ability and discussed above, we discuss the studies in a separate section, given there is a reasonable number of them. Many have included both reading ability (RA) as well as a chronological age (CA) comparison group (Angelelli et al., 2017; Burani et al., 2018; Carlisle et al., 2001; Law & Ghesquière, 2021; Suarez-Coalla & Cuetos, 2013; Suarez-Coalla et al., 2017; Traficante et al., 2011; van Hoogmoed et al., 2013) and some included RA controls only (Elbro & Arnbak (1996). We note too that the nature of the reading impairment might well vary from study to study.

Reading-impaired children might not be as sensitive to orthographic patterns at an early stage of processing as their skilled peers (CA), as evidenced by an unpublished thesis using masked priming (Madziak, 2021). Although Grade 5-6 reading-impaired children and their younger RA (Grade 2) controls both responded more quickly to targets preceded by related versus unrelated primes, like the older skilled CA readers (Grade 5-6), only the CA group exhibited a reduction in the N250 component (reflecting the processing of phonological and orthographic information) upon encountering related primes. This indicated that reading-impaired children and their younger RA controls did not exhibit early processing of sublexical orthographic information, whereas only the older skilled readers processed the orthographic information at an early time point. In addition, Lazaro et al. (2013) observed that the older skilled CA readers made faster lexical decisions to stimuli with frequent bases, a facilitation effect not seen in children with reading disorders and younger RA children matched for vocabulary size. This discrepancy could be attributed to a deficiency in processing and an insufficiently large vocabulary, limiting the benefit of frequent bases in the latter groups Comparatively, Lazaro et al. (2013) highlighted a contrast with studies like Marcolini et al. (2011), where ‘poor readers’ benefitted from morphological structure in word naming tasks. Lazaro et al. (2013) posited that task demands might be a potential explanation,

suggesting that while children with reading disabilities did not benefit from frequent bases in their lexical decision task, they could exploit the bases as shorter reading units in tasks such as reading aloud. In a 'go/no-go' lexical decision task, Lazaro et al. (2015) found that children with reading deficits, though underperforming compared to their skilled counterparts, were nonetheless influenced by orthographic alterations to the stems and benefited from high-frequency bases, similarly to the control groups. This indicated that children with reading disabilities could parse morphologically complex words into morphemes when the stem is familiar and intact. The variations observed when compared to Lazaro et al. (2013) could be attributed to the children in Lazaro et al. (2015) being a few months older and, more critically, the 'go/no-go' procedure requiring less cognitive demand, thereby particularly benefiting children with reading difficulties.

Some studies have included children with dyslexia and have asked about how children who are poor at phonological processing and mapping between orthography and phonology deal with complex words. Some studies, like Egan and Pring (2004) suggested that dyslexic children are insensitive to the internal structures of words, including both phonemes and morphemes, as they were not affected by disruption to either phoneme (e.g., 'clOCk') or morpheme (e.g., 'walKEd') boundaries. In contrast, other studies suggest that dyslexic readers, though underperforming compared to typically developing readers, are able to process morphemic units or and sensitive to morphemic properties like root frequency, mirroring the performance pattern of typical readers (e.g., Berthiaume & Daigle. 2014a; Suarez-Coalla et al., 2017; Traficante et al., 2011; Verhoeven & Schreuder, 2011). The majority of research indicates that dyslexic readers show sensitivity to morphemic units but might process them differently due to processing limitations. For instance, studies suggest dyslexic children predominantly depend on root representations that lack morphological specificity. In an eye movement study, Breadmore and Carroll (2018) observed that, unlike typically developing peers, dyslexic children did not show an extended gaze duration for misspelt target verbs due to morphological over-regularisation (e.g., 'growed'); this might suggest

diminished sensitivity to suffixes while relying predominantly on root morphemes. This reliance on roots was further supported by Breadmore and Carroll's (2016) previous study using a probe detection task. This found that after reading a list of 'lure' words (e.g., 'post'), dyslexic children struggled more to decide whether a 'probe' word (e.g., 'postal') was on the list or not when the lure and probe had morphological overlap (e.g., 'post-postal'), compared to when there was pseudomorphological or semantic lure-probe overlap. In addition, Burani et al. (2018) observed that unlike typically developing children who benefitted from longer roots when reading aloud derived words, dyslexic children lacked this advantage and were further hindered by larger root family sizes. The differences could be attributed to perceptual limitations, and a tendency towards serial processing among dyslexic children— for example, they might be less able to consider the whole word at pronunciation onset, and word roots with larger family sizes could lead to multiple potential word outcomes, creating uncertainty in pronunciation planning.

A different line of emerging evidence suggests that dyslexic readers are capable of morphological processing and furthermore, that they might adopt morpheme-based reading as their preferred strategy. Angelelli et al. (2017) found that dyslexic children exhibited lower accuracy when reading words containing low-frequency morphemes compared to non-derived words, unlike typically developing CA control children, who showed similar accuracy for both types of words. This might indicate distinct morphological parsing approaches: CA control children did not engage in morphological parsing for low-frequency morphemes, treating those complex words as simple ones. In contrast, dyslexic children were inclined to systematically parse stimuli into smaller units due to their challenges in processing words as wholes, aiming to maximally employ root representations to facilitate overall word processing. Differences are also evident in studies showing that while both typically developing and dyslexic children pronounced morphological pseudowords better than non-morphological pseudowords, only the dyslexic children benefitted more from morphological structure when reading words aloud. In contrast, skilled readers read derived and

simple words with comparable proficiency due to their direct activation of whole-word lexical units. This pattern has been observed in children (e.g., Burani et al., 2008; Suarez-Coalla & Cuetos, 2013) and adolescents (e.g., Elbro & Arnbak, 1996, Experiment 1). Elbro and Arnbak (1996) argued that morphological structure benefitted dyslexic individuals in both word decoding and comprehension, as evidenced by a significant correlation between sensitivity to morphological structure and enhanced reading comprehension in the dyslexic group, but not the RA controls. Dyslexic adolescents also demonstrated improved performance when reading texts morpheme-by-morpheme compared to syllable-by-syllable[9] (Elbro & Arnbak, 1996). In addition, developing readers with dyslexia were found to rely more heavily on morpho-semantic information, as evident in masked priming studies (Law & Ghesquière, 2021; Quemart & Casalis, 2015), showing that while RA and CA control groups exhibited priming in both morphological and pseudomorphological conditions, the dyslexic group exhibited priming only in the morphological condition. Quemart and Casalis (2015) observed that such morphological priming remained robust in dyslexic children across conditions where there were phonological or orthographic shifts to the base, in contrast to the performance in their CA or RA controls who exhibited diminished morphological priming when the base underwent phonological or orthographic shifts. Additionally, there was a negative correlation between morphological priming effects in fifth grade and phonological skills in second grade among dyslexic children—a relationship not seen in controls (Quemart & Casalis, 2015). This implies that poorer early phonological processing correlates with stronger morphological priming later on, suggesting a potential cognitive shift to compensate for phonological challenges (Law & Ghesquière, 2021). These 'morphological advantages' findings have been used to suggest that morphological processing might serve as a compensatory strategy, helping dyslexic children to overcome their phonological decoding challenges.

[9] This finding of Experiment 2 in Elbro & Arnbak (1996) is discussed for its theoretical relevance; however, it was excluded from data extraction because it involved forcing children to read sentences morpheme-by-morpheme, contrasting with our criteria for 'tacit, non-strategic morphological knowledge activation in complex written stimuli'.

Hearing problems constitute a unique category, distinct from reading impairment and dyslexia. However, children with hearing impairments tend to have difficulties with phonological processing and with reading. Research has demonstrated that deaf or hearing-impaired children performed comparably to typically developing controls in tasks such as probe detection (Breadmore & Carroll, 2016) and plausibility judgement (Berthiaume & Daigle, 2014b), indicating that morphological processing develops effectively, despite phonological challenges. In Berthiaume and Daigle (2014b)'s study, although deaf children and younger RA controls showed similar performance, both groups were outperformed by CA controls, implying that reading ability, rather than hearing status might be a more crucial factor contributing to the variation in morphological processing. This view is supported by Deilen et al. (2022) who found that the impact of visual segmentation at morpheme boundaries by hearing-impaired children varied according to their literacy skills, with lower-literacy children benefiting from visual segmentation, whereas those with higher levels of literacy skills did not. In addition, van Hoogmoed et al. (2013) observed that deaf children and their hearing peers showed similar RT patterns when making lexical decisions about derivational words and compounds. However, despite both groups responding more accurately to compounds, deaf children made more errors. This difference could stem from deaf children's reduced exposure to morphological structures in spoken language, echoing Van Hoogmoed et al. (2011), who noted that both deaf and hearing children utilised morphology for visual word recognition but only with the more frequent prefix; the deaf children were less sensitive to low-frequency prefix compared to their hearing peers. These findings imply that the development of morphological processing is not inherently constrained by phonological difficulties. Instead, the performance gap between deaf and hearing children might be largely due to differences in reading ability and reading experience, rather than phonological difficulties themselves.

In summary, reading difficulties have been found to relate to morphological processing in distinct ways and findings have been mixed. While some researchers identify a shortfall in

morphological knowledge among dyslexic children, others argue that these individuals show relative morphological strengths that could serve as a compensatory mechanism for their phonological deficits. Together with the research findings on hearing-impaired children, the results indicate that morphological sensitivity can develop in the context of phonological deficits. However, the heterogeneity of these deficits complicates drawing clear conclusions about their specific impact. Moreover, such deficits may overlap with other disorders, as Pennington and Bishop (2009) suggest, further complicating the analysis and making consensus difficult to achieve. Therefore, further research is needed in both morphological processing and language deficit domains to better understand how poor readers become more proficient at morphological processing and to inform the development of evidence-based teaching strategies.

2.3.6 Cross-language differences

Language serves as an influential factor at both the item and individual levels as the development of morphological processing is shaped by the specific language and writing systems that children encounter. The review included studies spanning 10 Indo-European languages that differ in several aspects, such as orthographic consistency, morphological richness, and complexity. These linguistic properties potentially influence morphological processing, and it might well be that some discrepancies across studies reflect cross-linguistic differences. For instance, as previously mentioned, morphemic parsing or activating morphemic knowledge does not necessarily facilitate reading, and this may be particularly true in transparent languages where the whole word is readily available and can be easily activated (e.g., Angelelli et al., 2014). Additionally, the pseudomorphological masked priming effects observed in 7-year-old French children were not seen in English children (e.g., Beyersmann et al., 2012; Quémart et al., 2011). This might, in part, reflect the distinct nature of the two languages. French, with its richer derivational morphology than English, might enable French children to develop morphological parsing mechanisms more rapidly than children learning to read in English. This is corroborated by Casalis et al. (2015), who explored

lexical decisions in English and French 7- to 10-year-olds. Overall, they found that the presence of morphemes facilitated word recognition and impeded nonword rejection. However, these effects were noticeable in both accuracy and RTs for French children, whereas for English children, the effect was observed in accuracy only. Casalis et al. (2015) suggested that morphology plays a key role in reading across languages, regardless of orthographic transparency but that a rich morphological language context (i.e., French) could accelerate the development of morphological knowledge and enhance sensitivity to written morphology.

In contrast, Mousikou et al. (2020) suggested that it was orthographic depth/consistency, rather than morphological complexity, that modulated how extensively morphology was employed in word reading. This was evidenced by their finding that Grade 3 English readers employed morphology in reading aloud to a greater extent than their French, German, and Italian counterparts, even though the latter languages have higher degrees of morphological complexity and orthographic consistency (i.e., lower orthographic depth). Mousikou et al. (2020) pointed out that Casalis et al. (2015) did not adequately match their stimuli on psycholinguistic variables across conditions; they also noted that some stems were modified, which might have impacted word processing and limited the generalisability of their findings (as also indicated by Dawson et al., 2018). Furthermore, the contrasting conclusions between the studies could stem from the different methodologies used. The influence of orthographic transparency is more obvious in reading aloud, as it allows children from languages with high orthographic transparency to utilize direct letter-sound mappings for accurate pronunciation; this is less crucial in making lexical decisions, as required in Casalis et al. (2015). The view that orthographic depth outweighs the influence of morphological richness on morphological processing is echoed in Beyersmann et al.'s (2021) investigation of masked priming in French and German 3rd-4th grade children. French is relatively more orthographically complex whereas German is more morphologically complex and productive, due to abundant compounds. The results showed a robust embedded word effect (contrasting nonaffixed nonwords vs. unrelated

control conditions, e.g., ‘farmity–FARM’ vs. ‘workald–FARM’) across languages, grade, and affix-type. However, only the French children exhibited both morpho-orthographic processing (contrasting affixed nonwords vs. nonaffixed nonwords, e.g., ‘farmity–FARM’ vs. ‘farmald–FARM’) and morpho-semantic processing (contrasting affixed word vs. affixed nonword, e.g., ‘farmer–FARM’ vs. ‘farmity–FARM’), while neither effect was shown in the German cohort. Although challenges exist in delineating how specific linguistic properties impact morphological processing, given the rich and diverse nature of languages (Beyersmann et al., 2021), these findings indicate that morphological processing in children from different linguistic backgrounds is likely to follow varied developmental trajectories, potentially influenced by the intrinsic linguistic properties of their respective languages and writing systems.

Aside from the language characteristics, another key point is that educational contexts also vary. For instance, in the UK, schooling begins at the age of five, while in France, it starts at six, making French children typically a year older than their English counterparts in the same grade (Casalis et al., 2015), Furthermore, differences in the school curriculum, especially in how morphology is taught, exist across languages. For example, according to Casalis et al. (2015), morphological structures are explicitly taught in Grade 4 in France, whereas in the UK, intensive instruction about derivational affixes begins in Grades 3 or 4 against a background of systematic phonics instruction. In addition, Hasenäcker et al. (2020) partly attributed the discrepancy between their results—where affix stripping in priming varied by grade among German children—and Beyersmann et al. (2015), who found no such grade-dependent priming effect in French children, to instructional differences since German schools typically begin explicit instruction in morphology from Grade 3, earlier than that in France.

In summary, there are cross-linguistic differences in children’s morphological processing and this can be attributed to both the intrinsic linguistic structures and differences in educational systems. However, achieving a perfect match between children from different linguistic

backgrounds is challenging. Regarding methodological challenges related to educational systems, separate analyses can be conducted, based on either grade-level or age-matched groups, as demonstrated by Casalis et al. (2015). Beyersmann et al. (2021) suggested that to clarify how cross-linguistic differences influence morphological processing, future studies should explore morphological processing within a 'broader linguistic context', including phonology, orthography, morphology, and semantics. Undoubtedly, cross-linguistic studies are invaluable in exploring the universality and common mechanisms of the development of morphological processing. However, this is not the primary focus of the current review, and with only three included studies involving cross-linguistic comparisons employing different paradigms (Beyersmann et al., 2021: masked priming; Casalis et al., 2015: lexical decision; Mousikou et al., 2020: read aloud), it is difficult to draw firm conclusions. Given the limited number of studies on this topic, disentangling the specific linguistic prerequisites remains challenging, highlighting the need for further research.

2.4 Summary and conclusion

This systematic literature review aimed to illuminate the factors that influence morphological processing as children learn to read Indo-European languages. Thus, it is worth noting that the interpretation and generalisation of the findings should be limited to Indo-European languages, where the vast majority of complex words contain two morphemes, exhibit a significant amount of concatenative morphology, and feature spaces between words. For example, findings such as embedded word activation may not be applicable to languages with different morphological structures, like Hebrew (see Stevens & Plaut, 2022). Specifically, we explored how task-, item-, and individual (child) -level factors modulate the implicit activation of morphological information as children read and process written words. Given the inconsistent patterns of results, providing a unified overview of morphological processing modulations is challenging. However, despite the heterogeneity present in paradigms, designs, participant and stimulus characteristics, covariates, and

outcome measures, the findings collectively highlight the importance of morphology in word reading and its pivotal role in reading development.

At the task level, reading aloud, lexical decision, and masked priming are the primary paradigms. The choice of paradigm influences the insights gained, as each method varies in sensitivity and processing demands. Understanding these nuances is essential for interpreting results and advancing research in this field, particularly when working with young readers. For instance, reading aloud, which relies on speech production, can be used to measure phonological factors influencing morphological processing but may be challenging for children with dyslexia, depending on the research question. Lexical decision tasks, which require decision-making under time constraints, can also be difficult for young children. Although masked priming is valued for capturing the early stages of word recognition, factors like prime duration can significantly impact results. Alternative methods, such as copying tasks, flankers tasks, or sentence reading tasks, offer promising avenues for exploring different aspects of morphological processing. For example, collecting eye movement data during sentence reading can provide insights into morphological processing in more natural, self-paced silent reading. Similarly, flankers tasks, which measure processing in a multiword context, serve as a closer approximation to sentence reading than single-word recognition tasks. However, since these alternative paradigms have only been adopted in limited number of studies with mixed results, it is crucial to replicate these findings and interpret them carefully, taking into account the specific demands and capabilities of each method.

The complexity of morphological processing is also influenced by various characteristics of the lexical items being processed. Whether a morpheme functions as a stem or an affix significantly impacts its processing, likely due to the central role of stems in word recognition and formation. Additionally, variations in morphemes or morphological structures, including phonological or orthographic alterations, can also affect the ease with which they are activated. Furthermore, Frequency is a key distributional property of lexical units, encompassing more than just the number

of times a word is encountered. It serves as an umbrella factor that captures the co-occurrence of diverse linguistic features, such as semantic diversity, productivity, and genre, across different reading experiences. Consequently, frequency both quantifies and qualifies the varied interactions with lexical items, refining their mental representations over time. These nuanced aspects of frequency are often overlooked by studies.

Word recognition relies on both inherent item-level factors and reader-specific factors. Morphological processing becomes more automatic and rapid mainly due to extensive experience with the (sub)lexical units and the more refined representations of them in the mental lexicon. Therefore, the development of morphological processing is not solely a function of chronological age or grade level but is strongly mediated by inter-individual differences in reading skills (Hasenäcker et al., 2020). Given the long-term stability observed in developing readers with different levels of language abilities (Bornstein et al., 2016), it is important to consider both individual variances in reading ability and developmental trajectories. On the other hand, morphological sensitivity can develop in the context of reading difficulties like dyslexia. Children with reading difficulties (including those caused by hearing impairments, although this is not technically a 'reading' difficulty) share many similarities with their typically developing peers but exhibit varied patterns in morphological processing. Evidence on the influence of reading difficulties on morphological processing is mixed, and the heterogeneity of these deficits makes it challenging to draw firm conclusions. However, understanding morphological processing within the context of reading deficits can help in developing tailored interventions and in fostering more effective learning strategies to support children with diverse educational needs.

In summary, in addressing the three major questions proposed, factors at the individual child level—such as accumulated reading experience, reading ability, and reading-related deficits—significantly influence the formation of mental representations of morphemes and the activation of morphological information during reading. At the item level, morpheme attributes are key,

including whether a morpheme functions as a stem or an affix, their distributional properties (e.g., frequency), and their structural form (e.g., how stimuli are presented—whether intact or disrupted as well as the integrity of the morphological structure such as whether phonological or orthographic alterations occurred).

Regarding the third question, 'Is morphological processing task-dependent?' The answer is both yes and no. Morphological processing, as the term implies, encompasses the implicit activation of morphemes and morphological structures encountered during reading. This process unfolds naturally as readers interact with lexical stimuli in real time. While variability in morphological processing can be influenced by the nature of the reading material and reader characteristics, the fundamental mechanisms of this cognitive process should, theoretically, remain unaffected by how it is measured. However, given that morphological processing is multifaceted and involves various stages, it could be metaphorically compared to observing the same object through either a microscope or a telescope. Different tasks might tap specific aspects or different stages of morphological processing or require different information to be activated during processing, depending on the tasks or measurements employed. Therefore, gaining a holistic and comprehensive understanding of morphological processing necessitates examining it through varied 'analytical lenses'.

2.5 Limitations and future directions

Despite our efforts to conduct a thorough, rigorous, and unbiased analysis, certain limitations of this review and the studies it encompasses should be noted. First, our systematic literature search did not bring up studies focused on the reading of irregular complex forms (e.g., 'fell', 'sought'). One possibility of the absence of such studies could be the use of search terms like 'decomposition' and 'parsing', which might have unintentionally suggested a focus on 'isolable morphological constituents'. It would be interesting to see this investigated in children, as per a study by Crepaldi et al. (2010) on irregularly-inflected primes (e.g., 'brought-buy') in adults. The processing of these

irregular forms may be better explained by distributed models, which argue that explicit morpheme representation is unnecessary (for a review, see Stevens & Plaut, 2022). This contrasts with the decomposition models that were used for theorisation in our review (e.g., Grainger and Beyersmann, 2017). Note however that this review does not aim to assess any particular theoretical account: this is beyond our intended scope, and in general, the available data can be accommodated in different theoretical frameworks. Future research could benefit from expanding the linguistic scope to include languages beyond the Indo-European family, potentially shedding light on more universal mechanisms of morphological processing.

Second, this review did not exclude studies or weigh their significance based on the analytical methods employed. It is worth noting that only in recent decades have linear mixed-effects models, treating participants and items as random effects, emerged as the standard for data analysis in this field. Prior to this, ANOVA analyses with F1 (by-participants) and F2 (by-items) were the predominant approach. However, several studies, especially the early ones, relied on a single ANOVA approach without conducting the F2 analyses (e.g., de Oliveira & Justi, 2017). This raises questions regarding the statistical power and generalisability of their findings, a concern highlighted in the 'critical methodological issues' column of the data extraction form shared on OSF. Conducting a meta-analysis would be helpful to the issues related to varying analytical approaches that this narrative synthesis could not address. However, due to the substantial differences in factors like populations, languages, and research paradigms, it was not feasible to perform a meta-analysis.

Finally, our search and screening process resulted in the inclusion of studies focusing only on a subset of Indo-European languages. While there is no compelling evidence to predict significant differences in results from other languages within this family, cross-linguistic variations among languages from the same family are recognised. Future research endeavours should both extend the language spectrum beyond those currently examined and conduct more cross-linguistic analyses. Importantly, as discussed above, to mitigate potential confounds such as stimuli or age differences,

effective cross-linguistic comparisons should employ similar stimuli and paradigms, ensuring participants are matched by grade and/or age. Moreover, the relative scarcity of studies employing longitudinal design, focusing on adolescent populations, exploring additional modulators, such as print exposure, and putting target words within sentence contexts to create a more natural reading setting, might suggest promising avenues for future research.

Chapter 3. Morphological effect in masked priming – an online replication study

3.1 Introduction

As highlighted in the systematic review (Chapter 2), the detection of morphological effects may vary based on the chosen experimental paradigm. To shed light on the real-time activation of morphological representations, research has adopted online techniques, such as the masked priming paradigm. This paradigm, with its brief prime durations and precise time measures, offers a promising approach for investigating the early stages of nonstrategic morphological processing as children read words. By varying prime durations and types of prime-target overlap, this paradigm could allow for an examination of various factors influencing morphological processing. Changes in the size of the priming effect might also act as an outcome measure in learning experiments, for example, children could be exposed to new suffixes and priming could be used to identify changes in their consolidation.

Previous research on masked priming with children as well as adults has predominantly involved in-person studies (e.g., Rastle et al., 2004; Beyersmann et al., 2012). Due to the pandemic, the present research needed to be conducted online. A critical first step is to establish the feasibility of this. I used the Gorilla Experiment Builder platform (www.gorilla.sc) and aimed to replicate a masked priming experiment with adults. This allowed me to establish its reliability and to consider the validity of remote testing; if the paradigm turned out to be reliable and valid, it could then be extended to children and used to explore a range of theoretical questions.

Skilled readers are sensitive to the morphological structures of complex words and can attend to these structures in a rapid and highly automatic manner during visual word recognition (for a review, see Amenta & Crepaldi, 2012). Masked priming (e.g., Beyersmann et al., 2016; Rastle et al., 2004, 2008), provides evidence to suggest that complex words are decomposed based on their

apparent morphological structure at the initial stage of processing, with lexical decision responses to targets being faster and more accurate when preceded by primes that share a morphological relationship (form and meaning, e.g., 'fuller - FULL') or pseudomorphological relationship (form due to a stem overlap but no semantic relationship, e.g., 'corner-CORN'), compared to unrelated primes. Priming is typically not observed in cases where the prime and target share orthographic overlap without an apparent suffix structure (e.g., 'brothel-BROTH'; Rastle et al., 2004). This indicates that priming depends on morpho-orthographic relationship, rather than mere orthographic overlap between the target and its prime. Form-based morphological decomposition is well-documented in adults. Additionally, Dawson et al. (2021) also found that among adolescents, the magnitude of priming increased with age, with older adolescents aged 16-18 years exhibiting morphological and pseudomorphological priming effects, comparable to the patterns observed in adult readers. As reviewed in Chapter 2, a different pattern of results in seen in younger children.

Internet-based platforms like Gorilla are able to present stimuli and record responses. Online experiments offer the advantage of being conducted in a variety of locations, free from the constraints of traditional lab settings. This flexibility significantly aids in recruiting participants from diverse backgrounds and in collecting large samples (Anwyl-Irvine et al., 2021; Majid, 2023). However, despite these advantages and the growing shift towards remote testing—a trend propelled by the pandemic—concerns about data quality persist, mainly stemming from two main sources: (1) Participant-related factors, including behavioural differences from those observed in lab settings, such as distracting environments, attention lapses, multitasking or potential 'cheating' (Rodd, 2019), and (2) Variability in testing conditions, such as using devices with different operating systems and experiencing varied internet speed (Anwyl-Irvine et al., 2020; Rodd, 2019). Both factors could affect data quality, and this might be especially problematic in priming experiments that depend on precise timing for effective masking and presentation of primes for the intended durations. Anwyl-Irvine et al. (2021) highlighted that the impact of timing errors varied with the

required display duration or expected RT differences. Thus, when displaying stimuli for a few frames (with one frame equalling 1/60th of a second), missing one frame could omit the stimulus presentation entirely, while an additional frame could substantially extend its display time. In studies that use masked priming to explore morphological processing, prime duration is brief, and effects are small (e.g., for adults in Beyersmann et al., 2012 (Experiment 1): ΔRT in morphological condition = 44ms; in pseudomorphological condition = 11ms; and for older adolescents in Dawson et al. (2021): ΔRT in morphological condition = 34ms; in pseudomorphological condition = 18ms). Noise in the data could pose a challenge to accurately detecting these effects.

The objective of Experiment 1 reported in this chapter was to replicate the study by Beyersmann et al. (2012) using Gorilla to examine the feasibility of conducting masked priming research through the web-based platform. Demonstrating the efficacy of this paradigm is crucial for facilitating sufficiently powered masked priming studies online, thereby enabling large-scale data collection. Given that online testing could be subject to timing accuracy variations due to hardware and software issues as well as participant behaviours (Anwyl-Irvine et al., 2020; Rodd, 2019), effects were made to minimise potential variability in data. The design of testing and data analysis was informed by best-practice guides for online data collection (e.g., Rodd, 2019; Sauter et al., 2020), such as including within-experiment safeguards like a time limit for each trial (detailed in the Methods section) and excluding participants who might not engage fully with the task (as detailed in the Results section).

3.2 Method

3.2.1 Participants

Forty-seven undergraduate students (aged 18-22 years) and 114 adolescents (aged 16-17 years) were recruited. The undergraduate students were recruited via a departmental participant recruitment system, and the adolescents were recruited from a Sixth Form College in a city in the United Kingdom. Two adolescents were excluded because they did not complete the experimental

task. Five undergraduate and seven adolescent participants were excluded because they self-reported as non-native speakers of British English. Additionally, one undergraduate participant and four adolescents were further excluded due to self-reported diagnosed reading disorders. A threshold was established to exclude participants with less than 75% accuracy (for both word and nonword targets) as a measure of potential inattention, but no exclusions occurred based on this criterion. The final sample consisted of 142 participants: 41 undergraduate students (M_{age} = 18.46, SD = .90; 37 females) and 101 adolescents (M_{age}= 16.18 years, SD = .38; 69 females, 1 other, 1 prefer not to say). All remaining participants reported that they were native speakers of British English, with 14 adolescents additionally reporting speaking another language.

Undergraduate participants were awarded course credits for their participation, while adolescents recruited through Sixth Form College were entered into a prize draw to win a £30 Amazon voucher. The study was approved by the University of Oxford Research Ethics Committee.

3.2.2 Materials

The same stimuli and manipulations as in Beyersmann et al. (2012) were employed, with morphological condition (morphological vs. pseudomorphological vs. form) and prime-target relatedness (related vs. unrelated) being manipulated within participants. In each related condition, there were 34 prime-target pairs, resulting in a total of 102 trials: 1) the morphological condition, in which the prime was a true derivative of the target (e.g., 'golden – GOLD'), 2) the pseudomorphological condition, where the prime shared an apparent morpho-orthographic relationship with the target but lacked a semantic overlap (e.g., 'corner – CORN'), and 3) the form-related control condition, where the prime and target shared only letter overlaps without morpho-orthographic relatedness (e.g., 'freeze – FREE). An unrelated prime was also selected for each of the target words (e.g., 'frosty – GOLD', 'sticky– CORN', 'tender – FREE'), resulting in each set of 34 targets being divided into 17 with a related prime and 17 with an unrelated prime. To ensure participants encountered each target only once, two lists were created: a target paired with a related

prime in one list was paired with an unrelated prime in the other. Participants were randomly assigned to one of the two lists. Stimulus lists are provided in the Appendix 3A.

According to Beyersmann et al., (2012), experimental stimuli across the three conditions were matched on orthographic overlap (defined as prime length divided by target length in letters), neighbourhood size (N) of both primes and targets, length and frequency of both primes and targets. Stimulus characteristics were recalculated using CELEX in N-Watch (Davis, 2005) and shown in Table 3.1.

Table 3.1: Means (SDs) for Lexical Characteristics of Primes and Targets across Conditions

Characteristics	**Morphological**	**Pseudomorphological**	**Form**
Prime-target orthographic overlap	1.61 (.26)	1.48 (.20)	1.50 (.27)
Target length (in letters)	3.62 (.65)	3.94 (.74)	3.71 (.58)
Target log10 frequency	1.80 (.46)	1.86 (.84)	1.86 (.82)
Target neighbourhood size	8.82 (4.69)	8.82 (5.43)	9.97 (5.29)
Prime length (in letters)	5.71 (.63)	5.74 (.96)	5.47 (.90)
Prime log10 frequency	1.38 (.60)	1.61 (.73)	1.32 (.73)
Prime neighbourhood size	*2.85 (1.97)*	*2.24 (2.16)*	*2.74 (3.17)*

In addition to the experimental stimuli, another set of unrelated word pairs (N= 34) was created to reduce the proportion of related prime-target pairs. To make the lexical decision task, 136 orthographically legal and pronounceable nonword targets (e.g., ‘CEBB’) were created, with each being also preceded by a prime that had not been paired with any target in the experimental trials (e.g., ‘jumping- CEBB’). Finally, diverging from Beyersmann et al. (2012), three 'fake trials' were inserted at the start of each block as a buffer to help improve data quality by allowing participants to refocus on the task after taking the short break following each experiment block. These trials (e.g.,

‘fror- LIBERTY’) totalled 12 across four blocks, with six word and six nonword targets. All primes for these dummy trials were nonwords, and the order of their presentation was fully randomised.

3.2.3 Procedure

Participants undertook the task remotely via Gorilla, accessing the experiment through a URL. They were required to complete the task on a PC or laptop. The university students participated individually at locations of their choosing, while the 16- and 17-year-old sixth-form college participants completed the task in computer rooms at their institutions in groups of different sizes. Instruction and questions on demographic information were displayed before the experiment task began, with all participants independently reading the instructions on their own computer and completing eight practice trials with feedback on accuracy when answered incorrectly. Feedback was not provided for experimental trials.

The experimental trials were divided into four blocks, with each starting with a 5000ms fixation cross and a 'Get ready' message displayed simultaneously at the top of the screen alongside a 5s countdown. Each trial began with a 500ms forward mask of hash keys ‘########’, followed by a prime in lower case presented for 50ms, which was then immediately replaced by the target in upper case. The prime duration was chosen not only to match Beyersmann et al., (2012) but also in light of the observations in in Anwyl-Irvine et al. (2020) that with a common display rate for monitors and screens being 60 frames per 1000ms, presenting stimuli for a duration that is an integer multiple of a frame could potentially improve the display accuracy. Thus, a 50ms duration corresponding to three frames, was selected. Participants were instructed to place the index fingers of each hand on the ‘F’ and ‘J’ keys on the keyboard, and to decide as quickly and accurately as possible whether the target on the screen was a word or not by pressing the corresponding key (‘J’ for word and ‘F’ for nonwords). The target was presented on the screen until a response was made or until a 5s time limit had passed. Trial order was randomised order, both within and across blocks.

3.3 Results

3.3.1 Accuracy and RTs (above 200ms)

Since age difference was not the focus of this replication study, and older adolescents have been shown to exhibit similar performance patterns in masked priming tasks as their adult counterparts (e.g., Dawson et al., 2021), data obtained from university students and sixth form college participants were combined and analysed together. Only data for word targets were analysed using the 'lmerTest' package (Kuznetsova et al., 2017) in RStudio (RStudio Team, 2020). Post-hoc comparisons were performed using the 'emmeans' package (Lenth et al., 2019). Data wrangling was conducted using the 'tidyverse' package (Wickham et al., 2019). The two categorical variables were recoded as follows: 'Prime type (related vs. unrelated)' was coded using deviation contrast coding (with related = 0.5, unrelated = -0.5), while 'Condition (morphological vs. pseudomorphological vs. form)' was coded as a categorical factor with three levels. In this setup, the 'form' condition was set as the reference level, with the other two conditions being compared to it.

Accuracy across all the three conditions was high (see Table 3.2). The accuracy model yielded no significant effects. Table 3.3 shows the model output for all fixed effects.

Table 3.2: Means (SDs) of Raw RTs and Accuracy Across Conditions and Prime Types

Condition	Prime type	Accuracy	RTs (above 200ms)	*RTs (between 200-2000ms)
Form	*Related*	0.97 (0.18)	618 (175)	615 (164)
	Unrelated	0.96 (0.19)	628 (176)	624 (160)
Pseudomorphological	*Related*	0.94 (0.23)	621 (187)	617 (177)
	Unrelated	0.94 (0.24)	638 (187)	634 (173)
Morphological	*Related*	0.98 (0.15)	595 (172)	591 (160)
	Unrelated	0.97 (0.16)	615 (171)	612 (163)

Note. RTs for correct responses were calculated after post log-transformation data trimming. The '*RTs' column reports the second analysis of RTs described in Section 3.3.2, excluding RTs shorter than < 200ms as well as those longer than 2000ms.

Table 3.3: Model Output for Accuracy Analysis

Predictors	**Odds Ratios**	**CI**	***z* value**	***p***
(Intercept)	41.11	29.19 – 57.89	21.277	<0.001
Prime type	1.17	0.87 – 1.59	1.027	0.304
Condition [Pseudomorphological]	0.69	0.44 – 1.09	-1.595	0.111
Condition [Morphological]	1.50	0.94 – 2.41	1.689	0.091
Prime type × Condition [Pseudomorphological]	0.88	0.59 – 1.31	-0.622	0.534
Prime type × Condition [Morphological]	0.85	0.53 – 1.36	-0.664	0.507

Note. the final model is Model_Accuracy <- glmer (Correct ~ prime.type*condition + (1|participant) + (1|item), family = binomial). In this model, the reference level for the 'Condition' is the '**Form' condition**.

In the RT analysis (Table 3.2 for descriptive statistics), incorrect trial RTs were excluded (570 trials, 3.9%). RTs under 200ms, viewed as premature button presses, and those over 5000ms, marked as timeouts, were also removed (nine trials, bringing the total to 579 trials removed, 4.0% of all trials). To reduce the impact of extreme RTs, a cutoff was applied at 3 standard deviations from the individual mean RT of each participant. This excluded an additional 95 trials (totalling 674 trials, 4.7% of all trials). The analysis showed a significant effect of Prime type in each of the three conditions[10] (Form: β = -0.016, t = -2.52, p = .01; Pseudomorphological: β = - 0.028, t = -4.31, p < .001; Morphological: β = -0.032, t = -4.88, p < .001), indicating that targets preceded by related primes were responded to faster than those preceded by unrelated primes. This effect was most pronounced in the morphological condition, followed by the pseudomorphological and form conditions. Figure 3.1 displays the mean RTs for related and unrelated primes across the three conditions.

[10] The output in Table 3.4 shows the effect of Prime type at the reference level for the condition factor (i.e., the Form condition). To obtain the estimates and p values of Prime type in the other two conditions, the model was refitted twice, each time changing the reference level of Condition so that each of the other two conditions served as the reference level in separate models.

There was also a significant effect of Condition when comparing the morphological condition with the form condition (β = -0.03, t = -2.33, p = .02), indicating that responses were faster in the morphological condition compared to the form condition. However, there was no significant interaction between condition (morphological vs. form) and prime type, suggesting that the difference in RTs in the morphological condition compared to the form condition cannot be attributed to a priming effect. The model output for all fixed effects is shown in Table 3.4.

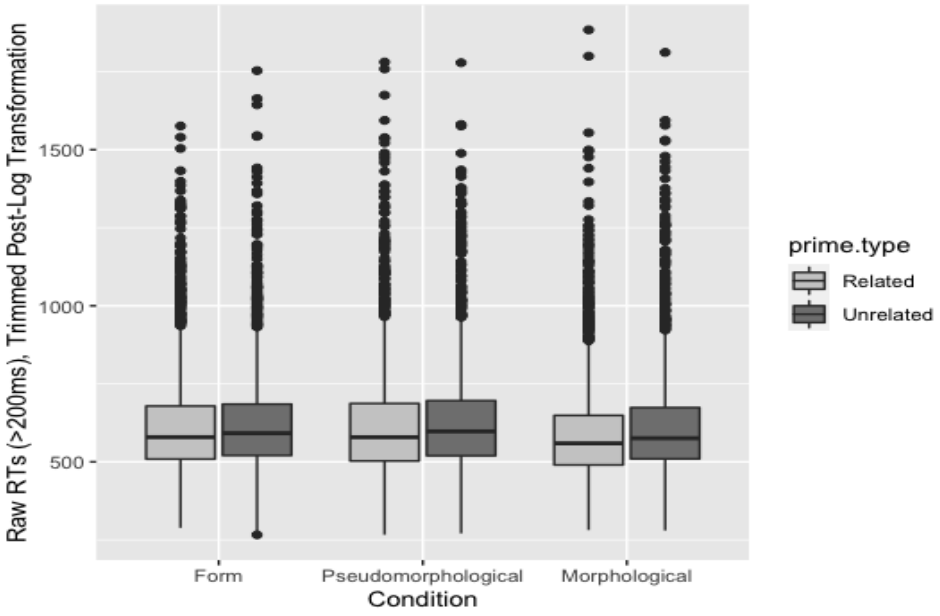


Figure 3.1: Boxplot of Raw RTs (> 200ms)

Table 3.4: Model Output for RT (>200ms) Analysis

Predictors	***β***	**CI**	***t* value**	***p***
(Intercept)	6.40	6.38 – 6.43	473.287	<0.001
Prime type	-0.02	-0.03 – -0.00	-2.515	**0.012**
Condition [Pseudomorphological]	0.01	-0.02 – 0.03	0.664	0.507
Condition [Morphological]	-0.03	-0.05 – -0.00	-2.329	**0.020**
Prime type × Condition [Pseudomorphological]	-0.01	-0.03 – 0.01	-1.299	0.194
Prime type × Condition [Morphological]	-0.02	-0.03 – 0.00	-1.656	0.098

Note. the final model is Model_RT<- lmer(RT_trimmed ~ prime.type*condition + (1|participant) + (1+prime.type|item)). In this model, the reference level for the 'Condition' is the '**Form' condition**.

Besides the linear mixed-effects model, Bayes Factor analyses were conducted to probe the potential existence of a Condition (reference level: form condition) × Prime type interaction. This interaction represents the morphological and pseudomorphological priming (compared to the form overlap only condition) I aimed to replicate in this study; however, it was not significant in the linear mixed-effects model analysis. This means that there is no evidence for the interaction, but it does not mean that there is evidence for no interaction. In order to evaluate this, I computed a Bayes Factor using the method advocated by Dienes (2008) and following the instructions outlined by Silvey et al. (2021). The Bayes Factors calculated using the R code adaptation of the Dienes calculator (Dienes, 2008) by Bence Palfi, originally implemented by Baguley & Kaye (2010)[11]. This requires three numbers- an estimate of (i) the effect size and (ii) the SE associated with the effect and (iii) an estimate of the predicted effect. (i) and (ii) were the relevant beta value and SE from the coefficients of the linear mixed effect model above and (iii) was taken from an equivalent model run over the data from Dawson et al. (2021**)** [12] (recall that this study used the same stimuli and similar procedures as in Beyersmann et al. (2012) but with a sample including adolescents). H1 was modelled as a half normal with a mean of 0 and the SD set to the size of the predicted effect (i.e. iii) which is appropriate when the prediction is directional and smaller values are more likely than larger values (which is generally true in psychology) and H0 was modelled as a point null. The calculator computes whether the data is more likely under H0 than H1, the resulting BF is a ratio with values above 1 suggesting more evidence for H1 and values below 1 suggesting more evidence for H0. The BF was 1.03 for the Condition (pseudomorphological vs. form) × Prime type interaction and 0.98 for the Condition (morphological vs. form) × Prime type interaction. According to the criteria laid out in Jeffreys (1939) and recommended in Dienes (2008) BFs between 1/3 and 3

[11] The function and detailed implementation can be found at https: //users.sussex.ac.uk/~dienes/inference/Bence%20Bayes%20factor%20calculator.html

[12] The original data from Dawson et al. (2021) were obtained from the Open Science Framework (OSF: https://osf.io/3g7vx/) and reanalysed using the same method as in the current study, resulting in the interaction effect of Prime_type × Condition [Pseudomorphological]: $\beta = 0.034$; Prime_type × Condition [Morphological]: $\beta = 0.067$.

indicate ambiguous evidence, i.e. there is no evidence for either the null or the alternative hypothesis for either of these contrasts in the current study. These results suggest that the paradigm with this age group is not sufficiently sensitive to assess this hypothesis with this number of participants through online testing.

3.3.2 RTs between 200 – 2000ms

Although the Bayes Factors indicated ambiguous evidence for or against the presence of priming effects, a closer inspection of the boxplots presented in Figure 3.1 revealed some potential reasons for this outcome. It was obvious that many data points lay outside the 1.5 times the interquartile range (IQR) from the third quartile (Q3) of the RT data, which could be considered outliers, even though they were not removed by the 3SD from the mean threshold. Those values could have led to extreme outcomes in the data analysis and potentially resulted in nonsignificant findings. Therefore, to mitigate this possibility, in an explorative manner, RTs for correct answers that were both shorter than 200ms and longer than 2000ms were removed (654 trials, 4.5%), thus ensuring RTs were within the 200-2000ms range, rather than the <5000ms timeout initially chosen to align with Beyersmann et al. (2012). The 2000ms cutoff point has been employed in many masked priming studies (e.g., Adelman et al., 2014; Gil-López et al., 2011; Massol, et al., 2010). As shown in Table 2, both means and SDs of RTs were reduced after excluding trials beyond the 200–2000ms range, compared to the statistics of RTs when only those below 200ms were removed. Figure 3.2 illustrates the new mean RTs for related and unrelated primes across the three conditions.

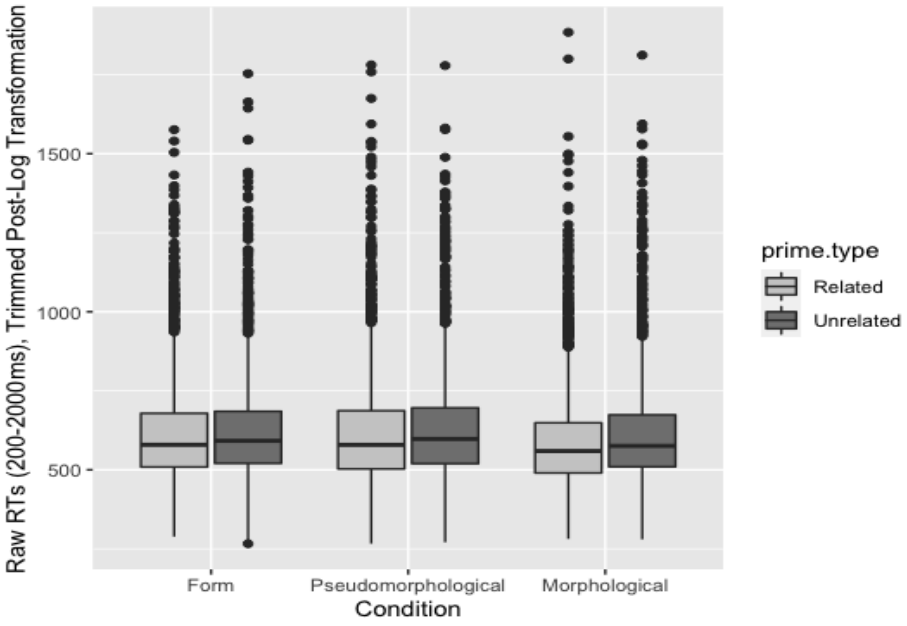


Figure 3.2: Boxplot of Raw RTs (between 200-2000ms)

The same analyses were run, and model output is in Table 3.5. Besides significant effects of Prime type in form (β = -0.015, t = -2.43, p = .017), pseudomorphological condition (β = -0.029, t = -4.62, p < .001), and morphological (β = -0.033, t = -5.39, p < .001) conditions, and a significant effect of Condition (morphological vs. form, β = -0.03, t = -2.35, p = .019), there was a significant interaction between Condition (morphological vs. form) and Prime type, β = -0.02, t = -2.08, p = .038, indicating that the priming effect obtained in the morphological condition was significantly larger than that in the form condition (i.e., the 'morphological priming effect').

Table 3.5: Model Output for Analysis of RTs (200 – 2000ms, Log-transformed)

Predictors	**β**	**CI**	***t* values**	***p***
(Intercept)	6.40	6.37 – 6.43	480.72	**<0.001**
Prime type	-0.01	-0.03 – -0.00	-2.428	**0.015**
Condition [Pseudomorphological]	0.01	-0.02 – 0.03	0.617	0.537
Condition [Morphological]	-0.03	-0.05 – -0.00	-2.345	**0.019**

Prime type × Condition [Pseudomorphological]	-0.01	-0.03 – 0.00	-1.586	0.113
Prime type × Condition [Morphological]	-0.02	-0.03 – -0.00	-2.075	**0.038**

Note. In this model, the reference level for the 'Condition' is the '**Form' condition**.

Follow-up posthoc analysis (see Appendix 3B for the output from all contrasts of the posthoc analysis) demonstrated that RTs for targets preceded by related primes were significantly faster than those preceded by unrelated primes in the morphological condition (ΔRT = 21ms, $Estimate$ = -0.033, z = -5.39, p < .001). In addition, although there was no significant effect or interaction for the pseudomorphological condition compared to the form condition, posthoc pairwise comparisons showed faster responses to targets preceded by related vs. unrelated primes in pseudomorphological condition as well (ΔRT = 17ms, $Estimate$ = -0.028, z = -4.62, p < .0001). However, no such significant prime-type effect was observed in the form condition ($Estimate$ = -0.015, z = -2.43, p = .15).

3.3.3 Posthoc power simulation

The pseudomorphological priming (compared to the form condition) was not found in our online experiment. This might be due to increased noise in the data due to its online format, which might have rendered the sample size too small to detect a pseudomorphological priming effect. To evaluate this possibility, I undertook data simulation, using the Monte Carlo method and the powerSim function from the 'SIMR' package in R (Green & MacLeod, 2016). Using the effect size of 'Prime type × Condition [Pseudomorphological]' obtained from the current study, with 500 simulations, it was revealed that to reliably detect an effect size equivalent to that of the Prime type × Condition (pseudomorphological vs. form) interaction with 80% power and a < .05 alpha level, at least around 560 participants would be necessary. See Figure 3.3 for the simulated power curve and Table 3.6 for the power simulation output. It is important to note that while post-hoc power simulations can provide insight into the required sample size for detecting the interaction effect that

was not observed in this study, they have several limitations. Inappropriate use of post-experiment power calculations can lead to overestimation of effect sizes when calculated based on significant results and could also be misused to explain away statistically nonsignificant findings (Gelman & Carlin, 2014; Hoenig & Heisey, 2001). Unlike pre-experiment power calculations which are valuable for planning experiments, post-hoc power calculations rely on the observed effect size. This means that the results can only inform us about the sample size and power needed for future studies but do not provide evidence supporting the effect in the current study (Hoenig & Heisey, 2001). Therefore, while the power simulation suggests that a larger sample size would be necessary to reliably detect the effect observed in my study, the post-hoc nature of this analysis means the results should be interpreted with caution. Future research with adequately powered designs is crucial to more confidently determine the presence or absence of pseudomorphological priming effects.

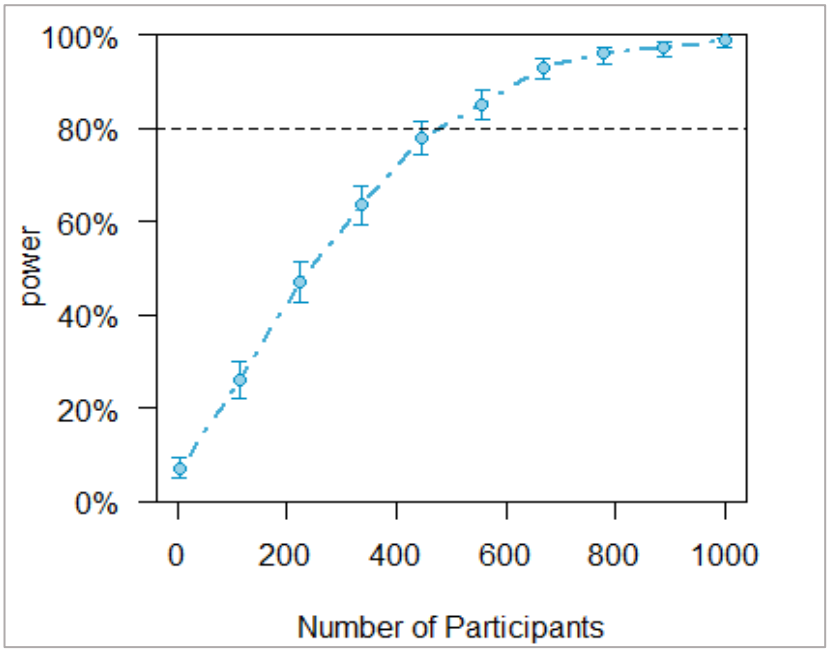


Figure 3.3: Power Curve from 500 Simulations

Table 3.6: Power for 'Prime type × Condition (pseudomorphological)' by Number of Participants

Number of Participants	Power	95% confidence interval
3	6.80%	(4.75, 9.37)
114	25.80%	(22.02, 29.87)
225	47.00%	(42.55, 51.48)
335	63.60%	(59.21, 67.83)
446	78.00%	(74.11, 81.56)
557	85.20%	(81.78, 88.20)
668	93.00%	(90.40, 95.08)
778	96.00%	(93.89, 97.54)
889	97.20%	(95.35, 98.46)
1000	98.80%	(97.41, 99.56)

3.4 Discussion

Experiment 1 used online testing to explore its potential utility for masked priming experiments that tap individuals' sensitivity to morphological structure as they process words in lexical decision. Specifically, the experiment aimed to replicate the morphological and pseudomorphological priming effects reported by Beyersmann et al. (2012). While the accuracy analysis yielded no significant results, there were some effects on the latency data. However, in the initial analysis of RTs between 200-5000ms, neither the morphological (e.g., 'teacher-TEACH') nor the pseudomorphological ('corner- CORN') priming effects differed significantly from those in the form condition (e.g., 'mother-MOTH'). Moreover, the Bayes Factor analyses also did not provide clear evidence for or against the presence of the morphological or pseudomorphological priming effect. A second RT analysis further removed RT points that were slower than 2000ms (instead of 5000ms), and the effects of interest started to emerge, mirroring those reported by Beyersmann et al. (2012) to some extent. However, while there was evidence of morphological priming effect, the pseudomorphological priming effect (compared to the form condition) remained elusive. The RT difference in the morphological condition was also smaller compared to the original study (ΔRT =

21ms in Experiment 1 versus ΔRT = 44ms in Beyersmann et al., 2012). Interestingly, although the data in Experiment 1 did not replicate the pseudomorphological priming reported by Beyersmann et al. (2012), the mean ΔRT in the pseudomorphological condition in Experiment 1 was larger than that in Beyersmann (Experiment 1: 17ms vs. Beyersmann et al., 2012: 11ms). This counterintuitive result might be due to the possible pseudomorphological priming in Experiment 1 was 'masked' by the greater noise and variance in data (as can be seen from the large SDs), which might be particularly impactful in the pseudomorphological condition—a point that will be discussed later. Based on the data obtained, the critical priming effects were weak and inconsistent, and they provided no conclusive evidence for morpho-orthographic processing during the early stages of morphological processing. The power simulation indicated that a large sample (N = 557) would be required to detect the pseudomorphological priming effect.

Pseudomorphological priming effects are frequently observed in skilled readers in controlled lab settings (see Amenta & Crepaldi, 2012, for a review). The much weaker effect in my study might be due to the masked priming paradigm requiring precise timing for accurate stimulus presentation and for measuring rapid responses. These might be vulnerable when experiments are remote. Variability and noisy data might stem from hardware and software variations, or participant inattention, particularly compared to performance in lab settings. Online testing remains not quite as precise as the lab-based equivalents (Anwyl-Irvine et al., 2020). For example, a consistent error in timing, often referred to as a 'lag', has been widely identified in online studies (Bridges et al., 2020). Moreover, controlling for participant inattention is more challenging due to factors like distractions in their environment, though there are things the researcher can do to aid data quality (Rodd, 2024).

That said, this does not imply that online testing is not robust or useful for testing morphological processing. Bridges et al. (2020) and Anwyl-Irvine et al. (2020) conducted timing validation of online testing by comparing data across different software, operating systems, web

browsers, and devices. They found that various combinations were reasonably precise in presenting visual stimuli and recording responses. For instance, visual lagging was found to be less than 2 frames for most combinations, and in theory, all platforms could provide reliable results for studies not requiring less than 100ms RT accuracy or less than two-frame presentation accuracy (Anwyl-Irvine et al., 2020; Bridges et al., 2020). However, such variations might still affect masked priming that presents stimuli for 3 frames (i.e., 50ms) or so, and thus, occasional under-presentation or over-presentation could substantially impact the duration of stimuli presentation, and such noise might mask the already small priming effects. Pseudomorphological priming might be particularly sensitive to timing inaccuracies in prime durations. For instance, extending the display time might result in the pseudomorphological prime being consciously perceived by participants, which would undermine any pseudomorphological priming effect due to semantic differences in the prime-target pairs (e.g., 'corner-CORN') as discussed in Chapter 2. However, morphological priming is less likely to be affected by over-presentation. The disproportionate impact of variations on morphological and pseudomorphological conditions could further explain why it was more difficult to replicate the pseudomorphological priming via online testing.

Nevertheless, these findings are encouraging, suggesting that online testing can be reliable for paradigms that are less sensitive to these variations because they do not require high-precision timing for stimuli presentation. Therefore, the failure to replicate the effects does not imply the disproof of them. Rather, it suggests that the data quality obtained online might not be reliable enough to detect the priming effect using the masked priming paradigm, thereby raising concerns about the suitability of conducting masked priming online. Therefore, the key conclusion of this chapter is that masked priming may not be the best option to test morphological processing via online remote testing, especially given my aim was to explore processing to track development across children and young people. A potential alternative might be the lexical decision paradigm, which employs overt stimuli presentation and is likely should be relatively more robust to be

conducted via remote online testing. Informed by the results of Experiment 1, I decided to use lexical decision with morpheme interference paradigm to test my hypotheses on morphological processing in developing readers.

Chapter 4. Effects of reading ability and suffix familiarity on morpheme interference effect in developing readers

4.1 Introduction

As reviewed in Chapter 2, cross-linguistic research emphasises the critical role of morphological processing in the early stages of visual word recognition, though the exact mechanisms are still debated (Amenta & Crepaldi, 2012; Giraudo & Voga, 2014; Stevens & Plaut, 2022). Skilled adult readers rapidly and automatically utilise morphological structures in lexical processing (Rastle, 2022). For instance, it has been found that nonwords possessing a surface morphological structure (e.g., gasful) were harder to reject than their matched control nonwords lacking such morphological structure (e.g., gasfil), a phenomenon termed the 'morpheme interference effect' (e.g., Crepaldi et al., 2010; Taft & Forster, 1975). Since Taft and Forster's (1975) seminal study, this morpheme interference effect has been replicated across languages (e.g., English: Crepaldi et al., 2010; Italian: Caramazza et al., 1988; Swedish: Jarvella & Wennstedt, 1993). Moreover, cognitive processes for pseudomorphemic nonwords resemble those for real words, as evidenced by similar ERP responses between nonwords containing bound stems with prefixes and real words (McKinnon et al., 2003). Decomposition accounts suggest that the morpheme interference occurs because letter strings with an apparent morphological structure are initially decomposed into their constituent morphemes based on their orthographic form, irrespective of semantics or lexical status (i.e., 'automatic' morpho-orthographic segmentation; Rastle et al., 2004).

Empirical evidence from developing readers supports the presence of some morphological ability from a young age, but the rapid and automatic decomposition of morphological structures takes time to develop (e.g., Beyersmann et al., 2012; Dawson et al., 2018). Exposure to complex words and their morphemes can facilitate the development of morphological processing (Dawson et al., 2021; Nagy et al., 2014; Verhoeven & Perfetti, 2011). Such reading experience enables

automatic morphological decomposition by refining affix representations within the mental lexicon, a process typically slower than developing root representations because affixes cannot function as stand-alone semantic units (e.g., Chapter 2; Grainger & Beyersmann, 2017). For instance, Dawson et al. (2018) found that the morpheme interference effect on RTs was absent in children aged 7-9 years and younger adolescents aged 12-13 years; however, it was only in late adolescence (16-17 years) that adult-like automatic morphological decomposition started to emerge. While reading experience generally increases with years of reading instruction, chronological age or school grade only provides a rough estimate of actual reading experience and tends to overlook individual differences in reading ability within each age group (Hasenäcker et al., 2020).

Word reading ability, as a product of both years of reading instruction and individual differences in reading expertise (e.g., Nation, 2017), might serve as a more robust individual-level indicator of reading experience. The influence of reading ability on morphological processing is evident in both skilled adult readers (e.g., Beyersmann, Casalis, et al. 2015; Andrews & Lo, 2013) and developing readers (e.g., Beyersmann, Grainger, et al., 2015; Dawson et al., 2021; Hasenäcker et al., 2020). For instance, Andrews and Lo (2013) demonstrated systematic individual differences in morphological processing among adults, with those having a higher vocabulary-than-spelling ratio showing stronger morphological priming, whereas those with a reverse profile exhibited equivalent levels of morphological and pseudomorphological priming. Similarly, Beyersmann, Casalis, et al. (2015) found that adults with higher language proficiency (reflected by a composite score of spelling and vocabulary) demonstrated stronger embedded stem activation.

In developing readers, individual differences in reading ability assume an even more pivotal role in morphological processing, potentially overshadowing the effects of age and grade (e.g., Traficante et al., 2018). Children with higher vocabulary levels were able to use morphological structures to aid visual word recognition from a younger age compared to their peers with lower reading abilities (Hasenäcker et al., 2017). In a study aiming to disentangle the effects of reading

ability and grade, Hasenäcker et al. (2020) found that stem activation was strongly influenced by individual differences in reading ability, while sensitivity to affixes developed with grade (see Chapter 2 for a detailed discussion).

There is still debate about when adult-like morphological processing emerges during reading development. For instance, studies using the morpheme interference paradigm found that the presence of a stem and/or a suffix in a nonword increased both the response error rate and RTs in French 3rd and 5th graders (8- to 10-year-olds; Quémart et al., 2012; Casalis et al., 2015). In contrast, the morpheme interference effect was restricted to accuracy in children aged 7-9 and younger adolescents aged 12-13, and extended to RTs only in later adolescence, aged 16-17 years (Dawson et al., 2018). The discrepancies might stem from cross-linguistic differences in orthographic depth (i.e., English is deeper than French) and morphological productivity (i.e., English is less morphologically productive than French) (Casalis et al., 2015; Mousikou et al., 2020). This suggests that specific the nature of the linguistic input shapes morphological processing, underscoring the need to consider item-level factor characterising reading experience, as discussed in Chapter 2. The acquisition of morphological representations involves integrating frequently encountered letter chunks into meaningful linguistic units (Rastle & Davis, 2008). As such, beyond individual-level reading ability, affix frequency emerges as a key item-level factor characterising reading experience. The accumulation of affix-related experience contributes to the refinement of affix representations into fine-grained morpho-orthographic units, potentially enhancing the activation of these sublexical units during reading (e.g., Castles et al., 2018; Nation, 2009).

Using secondary data from the English Lexicon Project (ELP; Balota et al., 2007), researchers (Muncer et al., 2013a; Muncer et al., 2013b; Yap et al., 2015) found that nonwords with affix morphemes were named faster and were harder to reject in lexical decision tasks than those without affixes, indicating that suffixes as processing units enhance word reading. Studies on affix

frequency have also shown its significant impact on morphological processing. For instance, Suffix frequency was found to modulate responses in both lexical decision and naming tasks in Italian-speaking adults (Burani et al., 1997): Compared to control nonwords, longer RTs and higher error rates were observed only when rejecting pseudomorphemic nonwords with high-frequency suffixes. Additionally, while pseudomorphemic nonwords were generally named faster, this facilitation was more pronounced in those with high-frequency suffixes. Notably, Burani et al. (1997) matched the nonsuffix letter strings that attached to stems in the control condition with suffixes used in the pseudomorphemic condition on several lexical attributes, including orthographic, phonological, and syllabic similarities, and crucially, position-specific (word-final) orthographic frequencies. This design ensured that the suffix frequency effects observed were not merely due to the high-frequency co-occurrence of letter chunks (cf. Crepaldi et al., 2010; De Rosa & Crepaldi, 2022). Similarly, in Burani & Thornton (2003, Experiment 1), where suffixes with varied frequencies were combined with pseudostems, the morpheme interference effect was exclusively seen in nonwords with high-frequency suffixes. Beyersmann et al. (2012) also found that in masked priming, increases in suffix frequency were associated with an increase in priming magnitude, as evidenced by RTs in Year 3 and Year 5 English-speaking children. In a go/no-go lexical decision task (Lázaro et al., 2016), Spanish-speaking children from 2nd, 4th, and 6th grades responded faster to words with high-frequency suffixes than to those with low-frequency suffixes, and this suffix frequency effect was not modulated by year group. Studying the affix frequency effect is crucial, as increased exposure to morphemes enhances their familiarity, fostering high-quality lexical representations that facilitate rapid processing during visual word recognition (Reichle & Perfetti, 2003).

Chapter 4 examines the impact of reading experience on morphological decomposition. The primary goal was to validate and extend the morpheme interference effect identified in earlier studies (e.g., Crepaldi et al., 2010; Dawson et al., 2018) through remote online testing in adults (Experiment 2). The role of reading experience, characterised by reading ability (individual-level)

and suffix familiarity (item-level), on the morpheme interference effect, was investigated. While extensive research has explored morphological processing differences between children and adults, a significant gap remains concerning adolescent readers (cf. Chapter 2; Ricketts et al., 2020). Experiment 3 involved participants aged 11-17 years, covering a wide age range during the adolescent phase and capturing a wide range of reading abilities.

4.2 Experiment 2

Experiment 2 was designed to test the morpheme interference effect in a group of skilled readers by comparing their responses to morphologically-structured nonwords (pseudomorphemic nonwords, e.g., birdly) and non-morphologically-structured nonwords (control nonwords, e.g., birdla). Since the difference between online and in-person data collection is not trivial (Sauter, Draschkow, & Mack, 2020), the main aim was to validate the morpheme interference paradigm using online testing. If participants found it more difficult to reject pseudomorphmic nonwords than control nonwords, as manifested by longer RTs and lower accuracy, then this would provide evidence that the morpheme interference paradigm can be reliably used in remote testing. We further explored the modulation effects of reading ability and suffix familiarity.

4.2.1 Method

4.2.1.1 Participants

Seventy-two adults were recruited online via Prolific (http://www.prolific.co). Adult participants who were older than 40 (N = 14) or whose overall accuracy on the experimental lexical decision task was below 75% (N = 1) were excluded. All the remaining 57 participants performed above chance levels on the reading ability assessment. The final sample consisted of participants aged between 18 and 40 years old (M = 29.09, SD = 6.95) who were native speakers of British English and had normal to correct-to-normal vision. One participant included in the final sample reported having a reading disorder of some kind.

Participants gave their consent online after reading the information about this research and were reimbursed at the rate of £10/hour for their time via Prolific. The study was approved by the Ethics Committee at the University of Oxford.

4.2.1.2 Materials and procedure

Reading ability. This was measured through an online reading assessment adapted from the Rapid Online Assessment of Reading Ability (ROAR) developed by Yeatman and colleagues (2021). This assessment was validated for individuals aged from 6 to 42 years, and the test performance was highly correlated with performance on other reading ability measures, such as Woodcock-Johnson Letter Word Identification (WJ-Word-ID) test (see Yeatman et al., 2021). The current study used stimuli taken from list A of ROAR (see Appendix 4A), consisting of 84 items (42 words and 42 pseudowords).

Lexical-decision task. The stimuli comprised a total set of 384 items (see Appendix 4B), which included two sets of nonwords (96 pseudomorphemic and 96 control) and two sets of word fillers (96 morphologically complex and 96 monomorphemic). The filler trials were used for the purpose of constructing the lexical decision task and were not included in data analysis. Each pseudomorphemic nonword was created by combining an existing English stem with an English suffix (e.g., 'birdly'= bird + ly). Stems and suffixes were selected from the children's reading material corpus Oxford Children's Corpus (OCC; Banerji, et al., 2013)[13]. Suffix familiarity was indexed using the frequency of suffixes (per million words) derived from the OCC, as described in Dawson et al. (2023). High frequency of occurrence in the corpus is taken as a proxy for high familiarity. See Table 4.1 for the token frequency counts of the suffixes. Each of the 12 suffixes was assigned to eight stems, making a total of 48 stems.

[13] According to Nation (2017), reading behaviour should be associate more closely with text-corpora statistics than with spoken-corpora statistics, and thus extracting frequency data from corpora that reflect children's written language exposure presents a promising approach to investigating the influence of reading experience on visual word recognition.

Table 4.1: Frequency (per million words) of Each Suffix in OCC

Suffix	Frequency per million words	Suffix	Frequency per million words
-ly	9532.01	-ish	490.15
-ful	1117.72	-ian	373.32
-ous	974.73	-ist	324.07
-ness	804.63	-ine	101.20
-ise	191.61	-acy	54.75
-oid	9.74	-ile	15.9

All the stems were three to six letters long and all ended in a consonant to avoid the need for spelling change when combined with a suffix. The 12 sets of eight stems were similar in number of letters, concreteness (Brysbaert et al., 2014), age of acquisition (Kuperman, et al., 2012), semantic diversity (Hoffman et al., 2013), orthographic neighbourhood size (Coltheart, 1981), and frequency based on the SUBTLEX-UK corpora (Van Heuven et al., 2014).

Each control nonword (e.g., 'birdla') was formed by changing one letter of the suffix of each pseudomorphemic nonword (e.g., 'birdly'), such that these nonwords contained non-morphological endings. Changes were made in the central position or the end of suffixes to keep morpheme boundaries consistent across conditions as morpheme boundary plays a role in morpho-orthographic segmentation (e.g., Diependaele et al., 2013). Therefore, pseudomorphemic and control nonwords were matched in orthographic similarity, length in letters, orthographic neighbourhood size, and mean log bigram frequency (see Table 4.2).

Table 4.2: Means (SD) for Linguistic Characteristics of Nonword Across Conditions

Lexical Characteristics	Pseudomorphemic	Control
Length	7.08 (.91)	7.08 (.91)
Orthographic N	.06 (.24)	.03 (.17)
Mean log bigram frequency	2.44 (.32)	2.33 (.32)

Note. Calculated using CELEX in N-Watch (Davis, 2005)

Procedure. The experiment was conducted online remotely via Gorilla Experiment Builder (https://gorilla.sc/), an online platform with visual display duration accuracy suitable for the purpose of lexical decision experiments (Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2020). Participants accessed the experiment through a link provided on Prolific, allowing them to complete the task on their personal computers.

For the lexical decision task, they were instructed to quickly and accurately decide whether each in a series of letter strings, presented one at a time on-screen, was a real English word or not. They were to press “J” for words and “K” for nonwords. Each trial began with a 1000-ms fixation cross, followed by the target displayed in the centre of the screen in lowercase Open-sans font in black on a white background. The target remained on the screen until a participant responded or 5000ms passed. Trials where no response was made within the 5000ms were recorded as 'timeout' and automatically progressed to the next. Prior to the experimental trials, participants undertook 12 practice trials, with feedback provided on accuracy and speed. No feedback was given in the experimental trials.

The ROAR task started immediately after the lexical-decision task. The procedure mirrored the previous task but employed a time-limited lexical decision paradigm where each item disappeared after 350ms. To maintain consistency, participants continued to press “J” for words and “K” for nonwords. There were six practice trials and feedback on accuracy (i.e., a red cross for a wrong answer and a green tick for a right answer) was provided on the screen for both practice and experimental trials, aligning with Yeatman et al. (2021). The test paradigm and the display duration were also consistent with Yeatman et al. (2021), apart from the removal of the child-oriented animations and sound effects.

In both tasks, each stimulus was shown only once, and their presentation was fully randomised within each task. Response latencies (RTs) and accuracy for each item were recorded and saved

online via Gorilla. Regular short breaks were provided and participants could skip the break if they wished to. The whole experiment took about 20 minutes.

4.2.2 Results

Data analyses were performed using RStudio (Rstudio Team, 2020). Only RTs and Accuracy for nonword stimuli in the lexical decision task were analysed using '*lmerTest*' package (Kuznetsova, Brockhoff, & Christensen, 2017) in R (R Core Team, 2021). Data wrangling was conducted using the '*tidyverse'* package (Wickham et al., 2019). RTs were log-transformed to correct for distribution skews and were analysed using the linear mixed-effects (LME) model. Accuracy (log odds) was analysed using the generalized linear mixed-effects (GLME; binomial family) model (Baayen, Davidson, & Bates, 2008). The significance of the fixed effects was determined using the ANOVA (type III) function in the '*car'* package (Fox & Weisberg, 2019). Significant interaction effects were displayed by figures produced using '*ggplot2'* (Wickham, Chang, & Wickham, 2016). The reading ability scores were derived from the number of correct answers in ROAR (Yeatman et al., 2021)[14], with the full score being 84.

The models were fitted with three fixed-effect factors: condition (pseudomorphemic vs. control), reading ability, and suffix familiarity. Suffix familiarities (i.e., frequency per million in the OCC; Banerji et al., 2013) were transformed to a log10 scale to reduce the range (original range = 9.74 for 'oid' – 9532.01 for 'ly'; transformed range = 0.99 – 3.98). For the purpose of statistical analyses, non-suffix endings of control nonwords were assigned the same suffix familiarity values as their corresponding pseudomorphemic nonwords. The two-level categorical variable, condition, was recoded using contrast coding (-0.5 for control, +0.5 for pseudomorphemic) and the continuous variables, suffix familiarity (log10 transformed) and reading ability, were mean-centred. Following Barr et al.'s (2013) 'keep it maximal' approach, we first fitted the maximal random effect structure.

[14] RTs for ROAR were also collected. However, Yeatman et al. (2021) found that response accuracy is a much stronger predictor of reading ability than RT which might be only informative in high-performing participants. Therefore, only accuracy (i.e., number of correct answers) was used in our analysis.

In the event of convergence failures or singular fits, we subsequently simplified it following the guidance outlined in Brauer and Curtin (2018). The significance and variance explained by each fixed effect and interaction, as determined by Type III ANOVA in the '*car'* package, are reported in tables in each section, whereas the detailed model summary, including estimates of each fixed effect from the '*lmerTest*::summary()' function, is provided in the appendix. Raw data and the R code used for the analyses are available on the Open Science Framework (OSF; https://osf.io/mx9wk).

4.2.2.1 Accuracy

Accuracy was generally quite high for participants (see Table 4.3). The main effects and interactions of the primary variables are shown in Table 4.4. A detailed model summary including estimates for fixed effects is at Appendix 4C. Results showed main effects of Condition, Reading ability, and Suffix familiarity, suggesting that participants responded less accurately to the pseudomorphemic nonwords compared to the control nonwords (β = -1.61, z = -6.83, p < .001) and their response accuracy increased with reading ability (β = 0.30, z = 2.83, p = .005) but decreased with the increase of suffix familiarity (β = -0.49 , z = -4.11 , p < .001). The interaction between Condition × Suffix familiarity was significant (β = -.08, z = -4.55, p < .001). As Figure 4.1 shows, the magnitude of the morpheme interference effect increased with suffix familiarity.

Table 4.3: Means (SDs) of RTs (untransformed), Accuracy, and Raw ROAR scores.

Measures / **Condition**	Accuracy (rate)	RTs (ms)	ROAR
Pseudomorphemic	.90 (.30)	804.48 (430.48)	78.75 (3.86)
Control	.98 (.14)	730.43 (361.81)	

Note. RTs for correct responses were calculated after data trimming. ROAR scores were the same across conditions as condition was a within-subjects factor.

Table 4.4: Output for Accuracy Model

Fixed effects	χ^2	*p value*
(Intercept)	659.96	< .001***
Condition	46.62	< .001***
Reading ability	8.00	.005**
Suffix familiarity	16.86	< .001***
Condition× Reading ability	.40	.53
Condition× Suffix familiarity	20.66	< .001***
Reading ability × Suffix familiarity	.0002	.99
Condition× Reading ability × Suffix familiarity	.53	.47

Note. Output of Model_ Accuracy ← glmer (Correct ~ Condition × Reading ability × Suffix familiarity + (1 | Participant) + (1 | Item), family = binomial).
* = p <.05, **= p <.01, *** = p <.001.

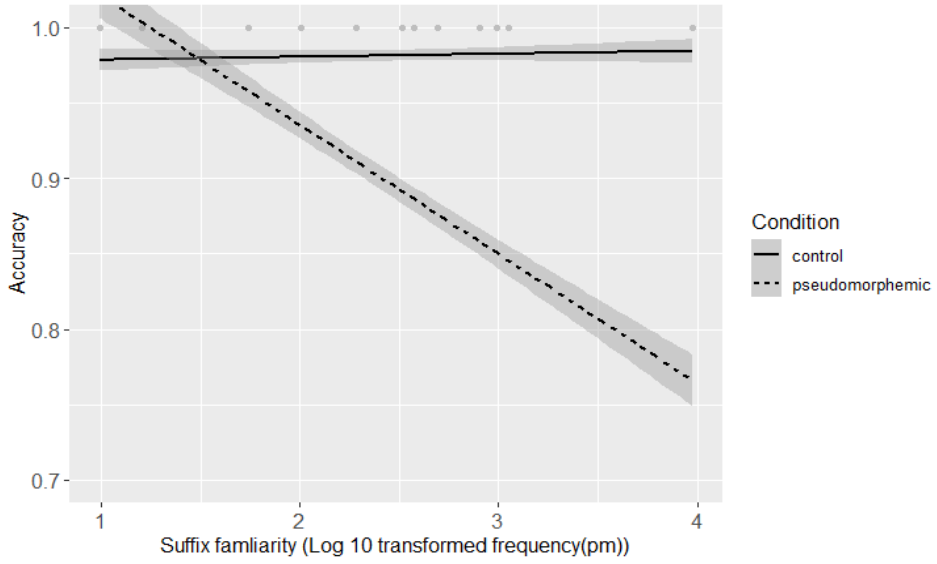


Figure 4.1: Two-way interactions between Condition × Suffix familiarity (i.e., log 10 transformed frequency of suffix in the OCC) for analysis of accuracy

4.2.2.2 RTs

RTs for incorrect trials was excluded (647 trials, 5.9%). RTs below 200ms were considered as premature button pressing and RTs exceeding 5000ms were recorded as timeout; both were therefore excluded[15]. Furthermore, outliers were identified following the procedure outlined by

[15] There were overlaps across incorrect, premature, and timeout responses. No more trials were further removed due to premature button pressing or timeout.

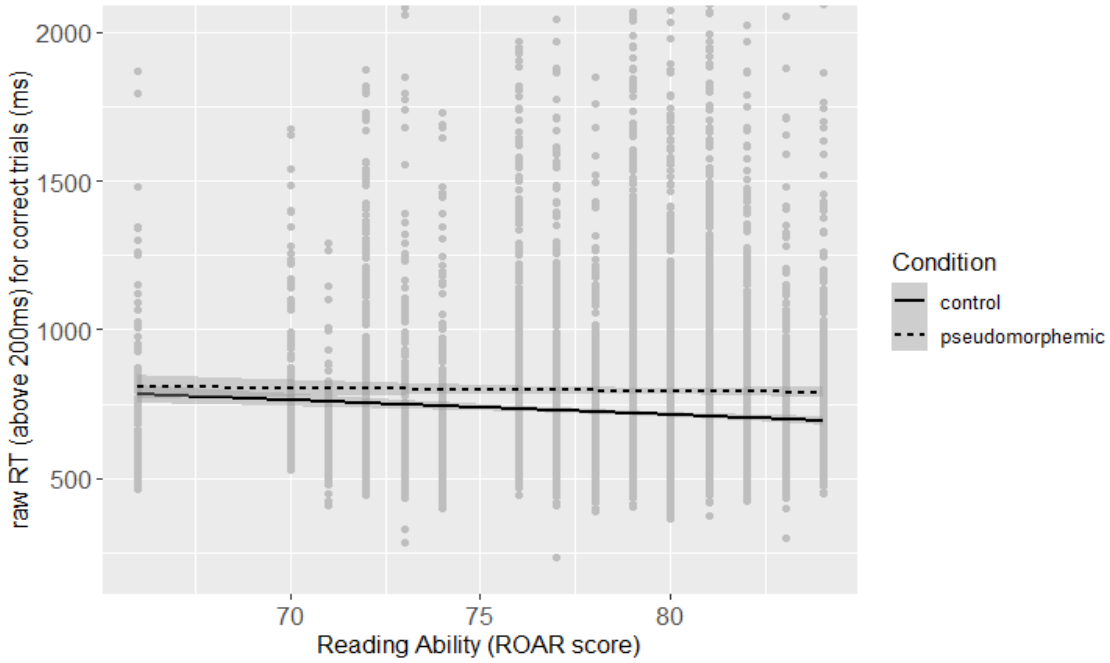
Baayen and Milin (2010): A base model with only random effects were first fitted for log-transformed RT and data points with absolute standardized residuals exceeding 2.5 SD were considered as outliers and were trimmed (329 trials, 3.0%). See Table 4.3 for the descriptive statistics of RTs after trimming.

The RT model output is presented in Table 4.5; detailed model output with estimates for each fixed effect is shown in Appendix 4D. There were significant main effects of Condition and Suffix familiarity, showing that participants responded significantly slower to pseudomorphemic nonwords compared to the control nonwords (Δ = 74.05ms, β = 0.095, t = 8.01, p < .001) and it took longer to reject nonwords containing relatively more familiar suffixes (β = 0.030, t = 4.93, p < .001). Significant two-way interactions were found for both Condition × Reading ability (β = 0.016, t = 2.57, p = .011) and Condition × Suffix familiarity (β = 0.058, t = 4.80, p < .001), as represented in Figure 4.2. An unexpected Reading ability × Suffix familiarity interaction was identified, β = .007, t = 2.10, p = .040. The pattern (see Appendix 4E for figures) showed that irrespective of Condition, as the familiarity of suffixes or the matched non-suffix endings in the control condition increased, better readers took longer to reject these nonwords, even though their RTs were consistently faster than less skilled readers.

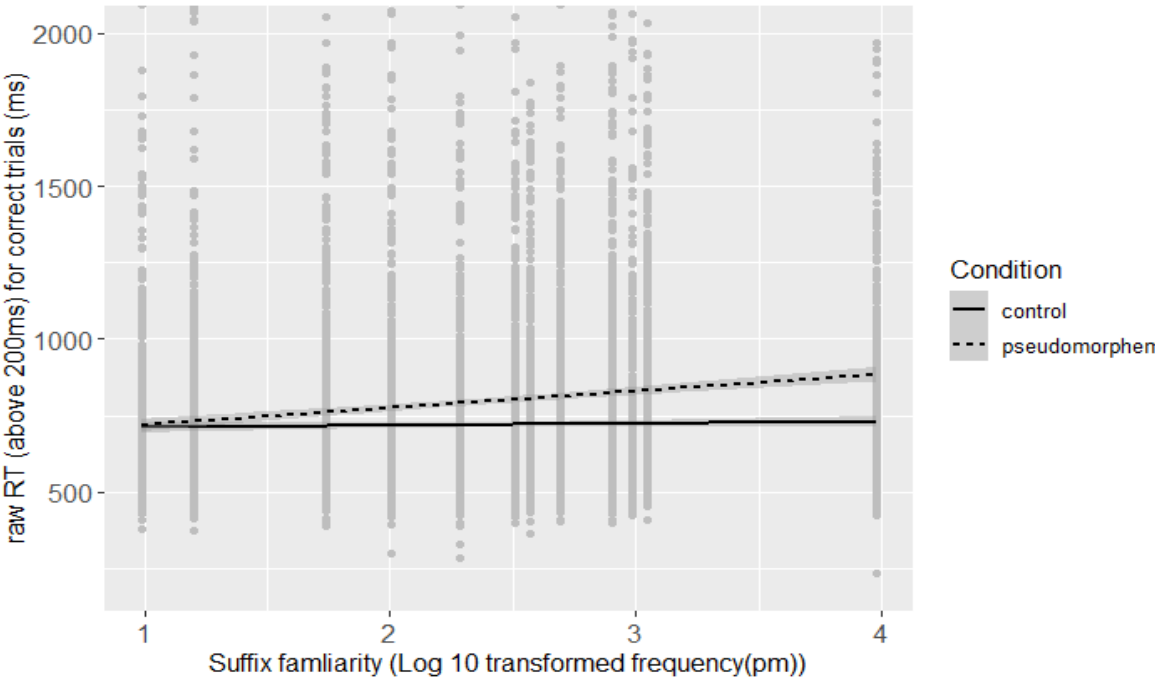
Table 4.5: Output for RT Model

Fixed effects	χ^2	*p value*
(Intercept)	104425.12	< .001***
Condition	64.14	< .001***
Reading ability	.045	.83
Suffix familiarity	24.32	< .001***
Condition× Reading ability	6.61	.010*
Condition× Suffix familiarity	23.04	< .001***
Reading ability × Suffix familiarity	4.42	.036*
Condition× Reading ability × Suffix familiarity	.51	.48

Note. Output of Model_ RT ← lmer (RT_transformed ~ Condition × Reading ability × Suffix familiarity + (1+ Condition × Suffix familiarity| Participant) + (1+ Reading ability| Item)).
* = p <.05, **= p <.01, *** = p <.001.



(a)



(b)

Figure 4.2: Two-way interactions for RT analysis between: (a) Condition × Reading ability (i.e., ROAR score) and (b) Condition × Suffix familiarity (i.e., log 10 transformed frequency of suffix in the OCC).

4.2.3 Experiment 2 discussion

Experiment 2 used the lexical decision paradigm to investigate morphological processing in skilled adult readers through remote online testing. Participants were found to be slower and less accurate when rejecting pseudomorphemic nonwords with an apparent morphological structure compared to the control nonwords without such a structure. The findings favoured the concept of pure morpho-orthographic decomposition during visual word recognition. This successful replication underscores the robustness of the morpheme interference effect and extends its validity to an online testing format.

Furthermore, the morpheme interference effect was larger for nonwords containing higher familiarity suffixes and among participants with stronger reading ability. This finding confirmed our hypothesis that reading experience influences morphological processing, aligning with existing research that suggests language experience and reading efficiency contribute to sensitivity to morphological structures (Nagy et al., 2014) and familiarity with morphemic lexical units influences the processing of morphologically complex stimuli (Burani & Thornton, 2003; Reichle & Perfetti, 2003).

The unexpected interaction observed between suffix familiarity and reading ability on RTs indicated that regardless of nonword types, as the familiarity of suffixes (or the corresponding non-suffix endings) increased, participants with a higher reading score took longer to reject the nonwords. In contrast, this RT pattern was not observed in average or less proficient readers. This result was not originally hypothesised as it did not pertain to the morpheme interference effect. The implications of this interaction are unclear, particularly because non-suffix endings should not be interpreted the same way as real suffixes, despite being assigned the same suffix familiarity values for statistical analysis. However, since we matched the pseudomorphemic and control nonwords for mean bigram frequencies, the non-suffix endings corresponding to highly familiar suffixes might be frequent letter clusters that skilled readers might have recognised early in the processing (see

Grainger & Ziegler, 2011; Crepaldi et al., 2020, for discussion). It is arguable that more proficient readers might be more adept at identifying the sub-lexical units in both conditions, potentially leading to longer RTs associated with more familiar suffixes or even non-suffix endings, and thus their decisions were delayed by the presence not only of real suffixes, but also of frequent letter clusters. This is in line with our prediction that reading ability contributed to the automaticity of language processing.

However, this interpretation is tentative, and further research, especially involving developing readers to cover a broader range of reading experiences, is required to validate it. Furthermore, given that the adult-like morpheme interference effect develops over time, possibly extending into late adolescence (Dawson et al., 2018), how reading experience influences this developmental trajectory remains unclear. Having demonstrated the morpheme interference effect in adults via online testing and confirmed the modulation effects of reading ability and suffix familiarity, we aimed to investigate how these factors might influence the morpheme interference effect in developing readers. These investigations could advance our understanding of how reading experience modulates morphological processing throughout its development.

4.3 Experiment 3

Experiment 3 aimed to investigate how reader and item characteristics influenced the morpheme interference effect in adolescent developing readers. Year 7 students (aged 11-12 years) were chosen as our youngest age group, one year younger than the youngest adolescents in Dawson et al. (2018). The aim was to examine the influence of reading ability and suffix familiarity on morpheme interference effect in participants with a wider range of reading experience. To capture a broad variation in responses from these developing readers, we adjusted our exclusion criteria – rather than excluding participants with below 75% overall accuracy for the lexical decision task (as was done with the adult participants in Experiment 1), we only excluded those whose overall accuracy for either the lexical decision task or the ROAR test fell below the chance level of 50%.

4.3.1 Method

4.3.1.1 Participants

The sample size[16] was determined using a power simulation conducted with the Monte Carlo method. The simulation was performed using the '*powerSim'* function in the '*SIMR*' package in R (Green & MacLeod, 2016). Older adolescents aged 16-17 years (N = 199) were recruited from a Sixth Form College in a city in the UK. Young-to-mid adolescents aged 11-15 years (N = 465) were recruited from three UK-based secondary schools. Participants who were non-native speakers of British English (N = 74) and those who scored below the chance level for either the lexical decision task (N = 4) or for the ROAR test (N = 5) were excluded. One more participant aged 18 years was further excluded. The final sample consisted of 77 children aged 11 years, 183 younger adolescents aged 12-13 years, 132 mid adolescents aged 14-15 years, and 188 older adolescents aged 16-17 years. Participants were instructed to provide their anonymised demographic information online. It should be noted that age-in-years information was obtained from all participants whereas age-in-months data were only collected for a subset of the sample aged 11-15 years.

Older adolescents recruited from the Sixth Form College were entered into a prize draw to win a £30 Amazon voucher and each participating secondary school received a gift voucher worth about £20. The study was approved by the Ethics Committee at the University of Oxford.

4.3.1.2 Material and procedure

The experimental design, tasks, and stimuli were identical to those used in Experiment 2. Participants were tested in groups/classes in their respective schools and monitored by a trained

[16] Based on the results obtained from reanalysing Dawson et al.'s (2018) raw data of adolescents, a fixed effect of β = 0.01 was used for the Condition * Reading ability* Suffix familiarity interaction for the current study. According to the results of 100 simulations, we found that a sample size of at least 195 was required to achieve 80% power for detecting the three-way interaction in a mixed linear model with Condition, Reading ability, and Suffix frequency as fixed effects.

researcher. To maintain an inclusive environment, all the students who chose to participate could complete the tasks via a link to the online experiment, although only data from native British English speakers were included in the analysis. Participants were offered regular short breaks and they could withdraw from the study at any point by closing their browser window if they wished to. The experiment lasted about 25 minutes for adolescents.

4.3.2 Results

Data analyses were conducted in a similar way as for the adult sample. To make our results more comparable to previous studies (e.g., Dawson et al., 2018), we divided our participants into four age groups. Age group was then treated as an ordinal variable with four levels in ascending order: 11 years, 12-13 years, 14-15 years, and 16-17 years. Given the ordered nature of the age groups and the meaningful distances between adjacent groups[17], we used orthogonal polynomial coding to better capture the trends across these groups (Schad, Vasishth, Hohenstein, & Kliegl, 2020), employing the '*contr.poly*' function in R. The ordinal age group variable was included as an additional fixed-effect predictor in the model. Post-hoc comparisons were performed using the '*emmeans*' package (Lenth et al., 2019). Raw data and the R code used for the analyses can be found on the Open Science Framework (OSF; https://osf.io/69ubr).

4.3.2.1 Accuracy

Descriptive statistics by age group are shown in Table 4.6. The model output is presented in Table 4.7, with detailed estimates of fixed effects being presented in Appendix 4G. The analysis of accuracy[18] showed a significant morpheme interference effect, as response accuracy was generally

[17] We acknowledge that the age gap between the 11-year-old group and the 12-13-year-old group is slightly smaller than that between other adjacent groups. Despite this, as our focus is on the general trend of age rather than precise transitions, age group was coded as an ordinal variable.

[18] The GVIF statistics (Sinclair, 2019) of the model were calculated using the '*car'* package (Fox & Weisberg, 2019). All GVIFs for significant terms were below 3, indicating that the multicollinearity is minor, and the correlation between the variables does not hinder the interpretation of their effects.

Model comparison using the LRT method showed showed that the current model leads to a significantly *improved* fit over the simpler model without the Age group variable.

lower for pseudomorphemic nonwords than for the control nonwords (β = -1.03, z = -9.21, p< .001). Additionally, significant main effects were observed for Reading ability, Suffix familiarity, and Age group, indicating that overall, accuracy increased with age (β = 0.42, z = 4.46, p < .001) and reading ability (β = 0.73, z = 16.83, p <. 001) but decreased with greater suffix familiarity (β = -0.36, z = -6.35, p < .001). Condition significantly interacted with all three variables: Age group (β = -0.27, z = -3.90, p < .001), Reading ability (β = -0.12, z = -3.32, p < .001), and Suffix familiarity (β = -0.68, z = -6.22, p < .001), respectively. There was an unexpected significant interaction between Age group × Suffix familiarity, β = -0.094, z = -2.66, p = .008 (see Appendix 4H for the figure). Besides, the two three-way interactions showed that the relationship between Condition × Suffix familiarity was modulated by both Reading ability (β = -0.13, z = -4.09, p <.001; see Figure 4.3) and Age group (β = -0.17, z = -2.99, p = .003; see Figure 4.4).

Table 4.6: Means (SDs) of Accuracy (rate), RTs (untransformed), and Raw ROAR scores.

Measure	Accuracy (rate)		RTs (ms)		ROAR
Age group	*Pseudo*	*Control*	*Pseudo*	*Control*	
11 years	.79 (.41)	.88 (.33)	1100.89 (700.22)	1048.52 (658.92)	71.1 (9.77)
12-13 years	.81 (.39)	.91 (.29)	969.89 (619.78)	933.76 (614.95)	72.0 (8.75)
14-15 years	.80 (.40)	.92 (.27)	1017.44 (677.56)	943.60 (612.6)	72.9 (8.98)
16-17 years	.88 (.33)	.97 (.18)	872.96 (460.10)	817.89 (433.06)	76.2 (5.73)

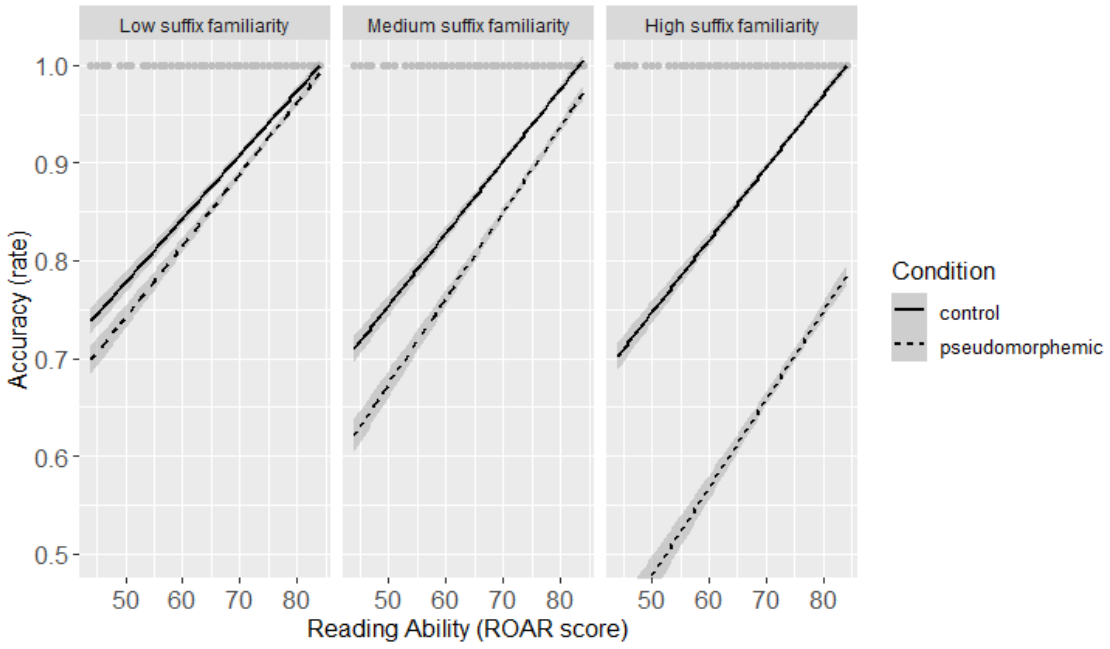
Note. RTs for correct responses were calculated after data trimming.

Table 4.7: Output for Accuracy Model

Fixed effects	χ^2	*p value*
(Intercept)	1569.13	<.001***
Condition	84.86	<.001***
Age group	34.34	<.001***
Reading ability	283.17	<.001***
Suffix familiarity	40.25	<.001***

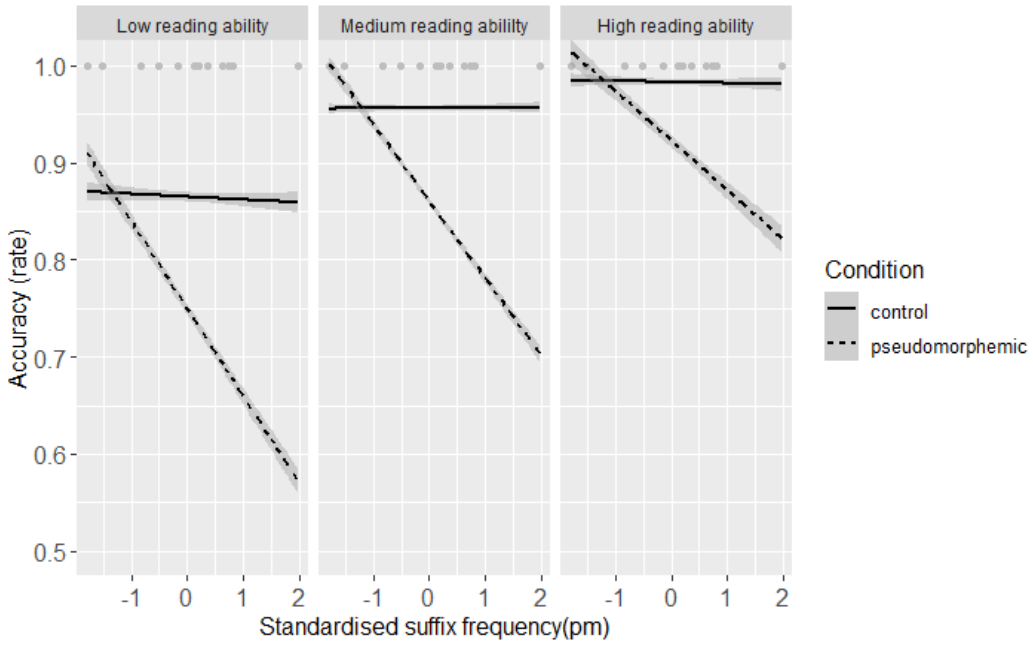
Condition × Age group	16.86	<.001***
Condition × Reading ability	11.04	<.001***
Age group × Reading ability	3.33	.34
Condition × Suffix familiarity	38.63	<.001***
Age group× Suffix familiarity	14.59	.002**
Reading ability × Suffix familiarity	2.72	.10
Condition × Age group × Reading ability	3.15	.37
Condition × Age group × Suffix familiarity	21.13	<.001***
Condition × Reading ability × Suffix familiarity	16.74	<.001***
Age group × Reading ability × Suffix familiarity	1.41	.70
Condition × Age group × Reading ability × Suffix familiarity	2.14	.54

Note. Output of Model_ Accuracy ← glmer (Correct ~ Condition × Age group × Reading ability × Suffix familiarity + (1 + Condition + Suffix familiarity | Participant) + (1 + Reading ability | Item), family = binomial); * = p <.05, **= p <.01, *** = p <.001.



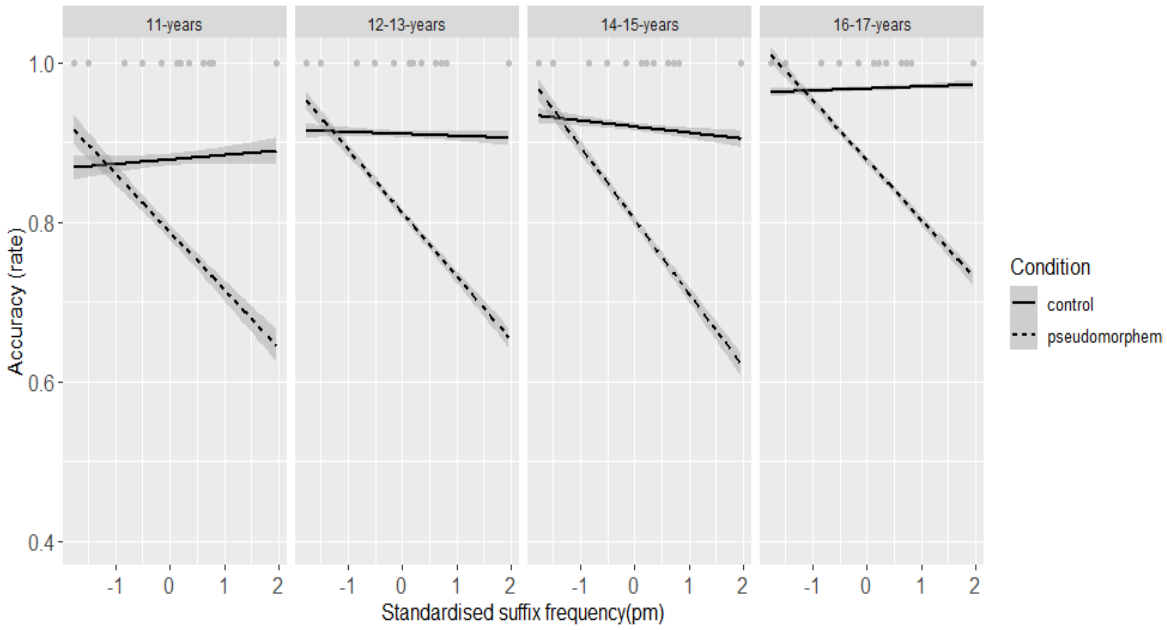
(a) [19]

[19] While the model indicates an increasing difference in accuracy between the pseudomorphological and control conditions with increasing reading ability (and suffix familiarity), this trend is not visually represented in the figure due to a ceiling effect in the control codition at 100% accuracy.

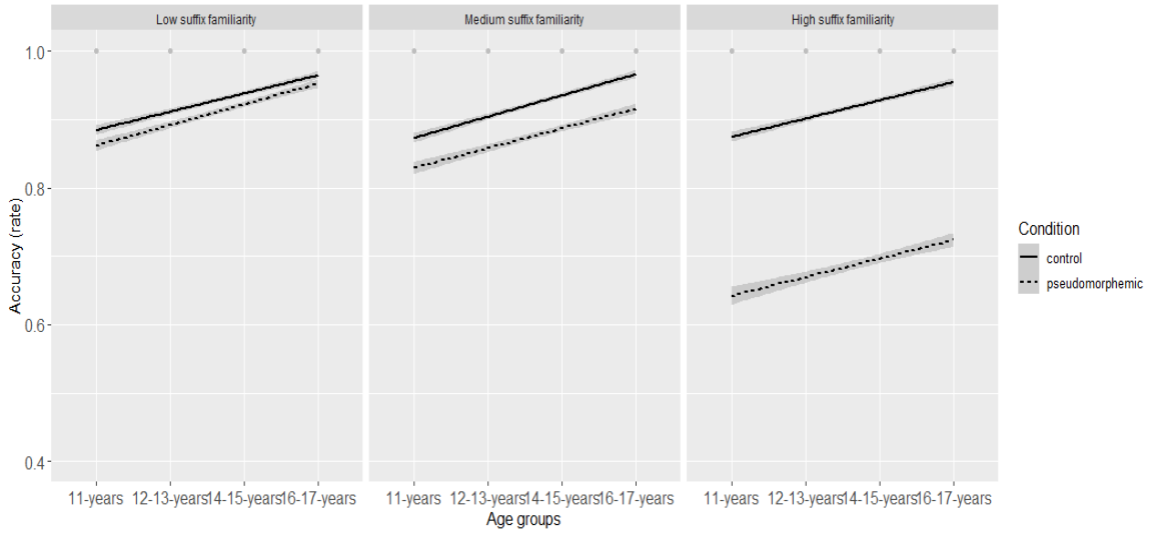


(b)

Figure 4.3: Accuracy: three-way interaction of Condition × Suffix familiarity × Reading ability: (a) grouped by Reading ability; (b) grouped by Suffix familiarity (See Appendix 4F for the interaction plot grouped by individual suffixes).



(a)



(b)

Figure 4.4: Accuracy: three-way interaction of Condition × Suffix familiarity × Age group: (a) grouped by Age groups; (b) grouped by Suffix familiarity

The three-way interactions were further examined by categorising Suffix familiarity into low, medium, and high familiarity levels based on ascending frequency rankings (four suffixes per level) and grouping participants into three reading ability groups based on standardised ROAR scores, i.e., poor (−1 SD), average (± 1 SD), and good (+1 SD) readers[20]. It is important to note that those reader groups are named for convenience and do not necessarily imply that the participant is a poor, average or good reader compared to their peers, as reading ability was not standardised within each age group. We refitted the models using the categorical variables for Suffix familiarity and Reading ability, respectively, and examined the modulation effects of the two factors using the '*emtrends'* function in '*emmeans'* (Lenth, 2018). The results of the posthoc trend analyses are shown in Table 4.8, which indicates the influence of one covariate (either continuous Suffix familiarity or Reading ability) on the morpheme interference effect (i.e., difference between performance in pseudomorphemic condition compared to the control condition) across different levels of the other

[20] Both the categorical (ordinal) Suffix familiarity and Reading ability variables were also coded as ordered variables using orthogonal polynomial coding.

categorical factor (either Reader group or Suffix familiarity level). Negative estimates represent a larger decrease in Accuracy for responses to pseudomorphemic nonwords compared to control nonwords as the covariate (i.e., suffix familiarity or reading ability, respectively) increased, indicating a stronger morpheme interference effect associated with the increase of the covariate. As shown in Figure 4.3 (b) and Figure 4.4 (a), in all reader and age groups, the morpheme interference effect became more salient with the increase of suffix familiarity. However, when examining different levels of suffix familiarity separately (as shown in Figure 4.3 (a)), only when the suffix was highly familiar did the morpheme interference effect became stronger as reading ability increased.

Table 4.8: Effects of covariates on the morpheme interference effect for different reader and age groups and different suffix familiarity levels (*Contrast = Pseudomorphemic− Control*)

Covariate	Group	Estimate	SE	$z.ratio$	p
	Poor reader	- .43	.11	- 3.82	<.001***
	Average reader	- .71	.10	- 6.79	<.001***
	Good reader	- .63	.16	- 3.85	<.001***
Suffix familiarity					
	11 years	-.64	.12	-5.14	<.001***
	12-13 years	-.59	.12	-5.08	<.001***
	14-15 years	-.62	.12	-5.18	<.001***
	16-17 years	-.89	.12	-7.44	<.001***
	Low familiarity	.00044	.057	.008	.99
Reading ability	Medium familiarity	-.073	.055	- 1.32	.19
	High familiarity	- .28	.052	-5.54	<.001***

Note. * = p <.05, **= p <.01, *** = p <.001.

Further posthoc analysis were conducted on the significant two-way interactions using '*emmeans*' function. A significant morpheme interference effect on Accuracy was found for participants from all Age groups (11 years: $estimate$ = -0.85, SE = 0.13, z = -6.40; 12-13 years: $estimate$ = - 0.96, SE = 0.12, z = - 8.04; 14-15 years: $estimate$ = -1.13, SE = 0.12, z = - 9.18, 16-17 years: $estimate$ = -1.19, SE = 0.12, z = - 9.66, all ps< .001). Particularly, older adolescents

aged 14-15 showed a stronger morpheme interference effect on accuracy compared to 11- and 12- to 13-year olds[21]. Additionally, when averaged over the levels of Age group, the morpheme interference effect was also found in all readers (poor reader: $estimate$ = -0.75, SE = 0.12, z = - 6.29; average reader: $estimate$ = - 1.08, SE = 0.11, z = - 10.05; good reader: $estimate$ = -1.24, SE = 0.17, z = - 7.29, all ps< .001) and for suffixes of all levels of familiarity (low familiarity: $estimate$ = -0.37, SE = 0.18, z = - 2.05, p = .040; medium familiarity: $estimate$ = -0.70, SE =0.18, z = - 3.95; high familiarity: $estimate$ = -2.02, SE = 0.18, z = - 11.50, both ps< .001).

4.3.2.2 RTs

RTs for incorrect trials (13665 trials, 12.2%) or trials that fell outside the 200-5000ms range (0.5%, 608 trials) were excluded. Outliers were identified in the same way as in Experiment 1, and were further trimmed (3207 trials, 2.9%). Descriptive statistics are shown in Table 4.6.

Table 4.9 shows the output from the RT model[22] (See Appendix 4I with detailed estimates of each fixed effect). Significant main effects were found for all four fixed-effect variables. Younger participants responded more slowly than older participants in general, β = -0.14, t = -7.19, p < .001. RTs generally decreased in participants with higher reading ability (β = - 0.018, t = -1.97, p < .001) but increased with suffix familiarity (β = 0.028, t = 5.29, p < .001). Notably, the main effect of Condition showed that participants took longer to reject pseudomorphemic nonwords compared to the control nonwords in general (Δ = 52.98ms, β = 0.076, t = 7.21, p < .001), indicating an interference effect on RTs for nonwords with (pseudo)morphological. This morpheme interference effect was further modulated by both Reading ability and Suffix familiarity as indicated by the

[21] For comparison with Dawson et al. (2018) on this result, the "Age group" variable was recoded using the repeated contrasts function '*(contr.sdif(4))*', and were expressed as fractions via the '*fractions()'* function in R (Schad et al., 2020) . This recoding allowed the "Condition * Age" interaction to represent the differences in log-odds of accuracy between pseudomorphemic and control nonwords across adjacent age groups. The corresponding analyses are available from the authors upon request.

[22] All GVIFs (Sinclair, 2019) were below 2, indicating that there was minor collinearity which should not hinder the interpretation of the model. Notably, model comparison showed that the current model leads to a significantly *improved* fit over the simpler model without the Age group variable.

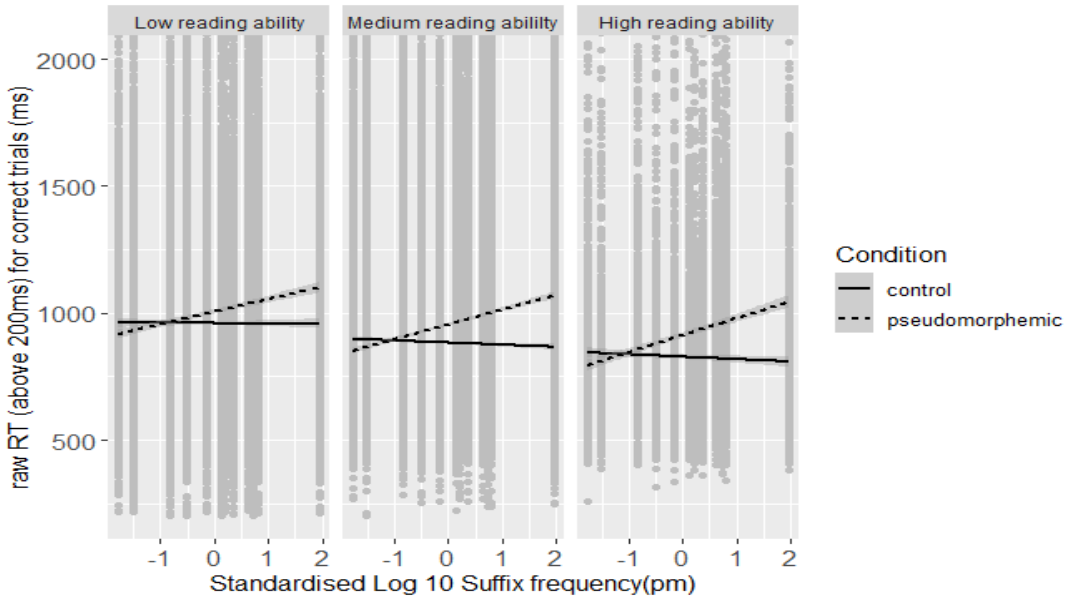
significant interactions between Condition × Reading ability (β = 0.010, t = 3.23, p =.001) and Condition × Suffix familiarity (β = 0.065, t = 6.20, p < .001), as well as the three-way interaction effect of Condition × Reading ability × Suffix familiarity (β = 0.008, t = 2.71, p = .007) (see Figure 4.5)[23]. The interactions suggested that the morpheme interference effect was larger for nonwords containing high familiarity suffixes (β = 0.07, t = 6.20, p < .001) and for participants with higher reading ability (β = 0.01, t = 3.23, p = .001). Furthermore, the effect of reading ability on the morpheme interference effect became more pronounced when responding to nonwords containing higher familiarity suffixes (β = 0.008, t = 2.71, p = .007).

Table 4.9: Output for RT Model

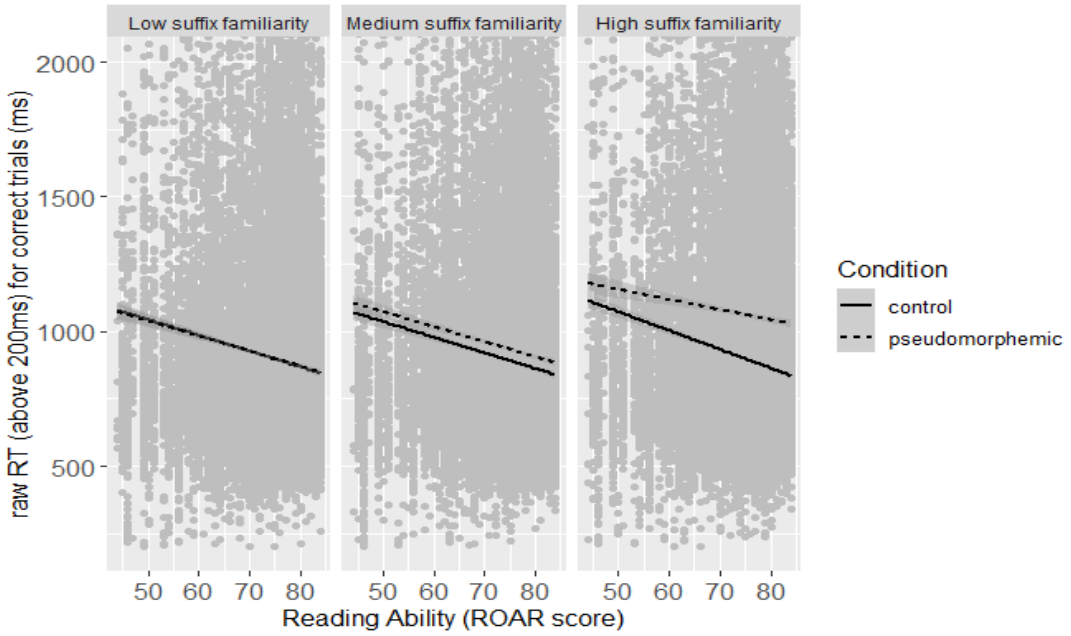
Fixed effects	χ^2	*p value*
(Intercept)	409469.90	<.001***
Condition	52.03	<.001***
Age group	64.15	<.001***
Reading ability	3.89	.05 *
Suffix familiarity	28.00	<.001***
Condition × Age group	4.41	.22
Condition × Reading ability	10.46	.001**
Age group × Reading ability	.64	.89
Condition × Suffix familiarity	38.39	<.001***
Age group × Suffix familiarity	3.07	.38
Reading ability × Suffix familiarity	.89	.35
Condition × Age group × Reading ability	6.62	.085
Condition × Age group × Suffix familiarity	4.35	.23
Condition × Reading ability × Suffix familiarity	7.33	.007*
Age group × Reading ability × Suffix familiarity	2.04	.56
Condition × Age group × Reading ability × Suffix familiarity	4.08	.25

Note. Output of Model_ RT ← lmer (RT_transformed ~ Condition × Age group × Reading ability × Suffix familiarity + (1+ Suffix familiarity| Participant) + (1+ Reading ability| Item)).
* = p <.05, **= p <.01, *** = p <.001.

[23] To confirm that the overall pattern was not particularly dominated by responses to one or more individual suffixes, we also draw an interaction plot for each individual suffix, see Appendix 4F



(a)



(b)

Figure 4.5: RTs: three-way interaction of Condition × Reading ability × Suffix familiarity: (a) grouped by Reading ability: low, medium, and high reading ability; (b) grouped by Suffix familiarity: low, medium, and high suffix familiarity based on the ascending rank of suffix frequency (four suffixes per group; See Appendix 4F for the interaction plot grouped by each of the 12 individual suffixes).

Posthoc trend analysis of the three-way interactions was conducted in the same manner as the Accuracy analysis. The models were refitted using the categorised Suffix Familiarity and Reading Ability variables. Positive estimates indicate a larger increase in RTs with a covariate in the pseudomorphemic than in the control condition, representing a greater morpheme interference effect with an increase in the covariate. The results (Table 4.10) corroborated the trends observed in Figure 4.5: the morpheme interference effect on RTs increased with greater suffix familiarity across all reader groups. However, significant effects of reading ability on the morpheme interference effect were only observed for highly familiar suffixes. No significant influence of reading ability on the morpheme interference effect were for nonwords containing low- to medium-familiarity suffixes.

Table 4.10: Effects of covariates on the morpheme interference effect for different reader groups and different suffix familiarity levels (*Contrast = Pseudomorphemic− Control*)

Covariate	Group	Estimate	SE	$t.ratio$	p
Suffix familiarity	Poor reader	.049	.012	4.11	<.001***
	Average reader	.067	.011	6.26	<.001***
	Good reader	.079	.013	6.37	<.001***
Reading ability	Low familiarity	.0023	.0051	.44	.66
	Medium familiarity	.0012	.0052	.24	.81
	High familiarity	.027	.0054	5.00	<.001***

Note. * = p <.05, **= p <.01, *** = p <.001.

Further posthoc analyses on the significant two-way interactions showed that averaged across the levels of Age groups, the morpheme interference effect was evident in all reader groups (poor readers: $estimate$ = 0.065, SE = 0.012, z = 5.38; average readers: $estimate$ = 0.073, SE = 0.011, z = 6.83; good readers: $estimate$ = 0.091, SE = 0.013, z = 7.33, all ps< .001). However, morpheme interference was only observed for nonwords containing medium-to-high familiarity suffixes (medium familiarity: $estimate$ = 0.050 , SE = 0.017, z = 2.98 , p = .003; high familiarity: $estimate$ = 0.17, SE = 0.017, z = 10.04, p <.001). No significant morpheme interference was

observed for nonwords containing low-familiarity suffixes ($estimate$ = 0.0084, SE = 0.017, z = 0.50, p = .62).

4.3.2.3 Exploratory analysis using age as a continuous predictor

We further explored the role of age by including Age in months as a continuous predictor in the model. Only data obtained from a subset of the participants were included since we did not collect age in months from all participants. After excluding ineligible participants based on the criteria stated above, the final sample consists of 230 participants aged 11-15 years ($N_{11\ years}$ = 39; $N_{12\ years}$ = 37; $N_{13\ years}$ = 70; $N_{14\ years}$ = 66; $N_{15\ years}$ =18).

For accuracy, the pattern of results was similar to the accuracy analysis reported in the main analysis (see Appendix 4J for detailed model output), except that the main effect of Age (in months) was not significant anymore (note that Age *group* was used in previous analysis), z = 1.38, p =.17 (ns). Moreover, aside from the Age × Suffix familiarity interaction, another unexpected interaction between Reading ability × Suffix familiarity reached significance, β = - 0.052, z = - 2.21, p = .027 (see Appendix 4K for the figures).

For RTs, only the main effects of Condition (Δ = 42ms, β = 0.061, t = 6.25, p < .001) and Suffix familiarity (β = 0.023, t = 4.65, p < .001) reached significance. The two-way interaction between Condition × Suffix familiarity (β = 0.053, t = 5.44, p < .001) was still significant. Reading ability × Age in month interaction became marginally significant (β = 0.025, t = 1.88, p = .060). No other two-way interactions reached significance.

The analysis also revealed a three-way interaction of Condition × Reading ability × Age in month (β = 0.009, t = 2.26, p = .024) that was not observed in the main analysis. As Figure 4.6 shows, the morpheme interference effect in RTs was less pronounced in older poor readers (possibly indicating genuine reading difficulties) compared to younger students with low reading ability (see Figure 4.7 for reading ability across age). However, model comparison (LRT) indicated

that the inclusion of the Age in months variable in the model did *not* significantly improve the model fit compared to a simpler model without this variable, $\chi^2(8) = 14.02$, $p = .081$ (ns) (see Appendix 4L for detailed model output).

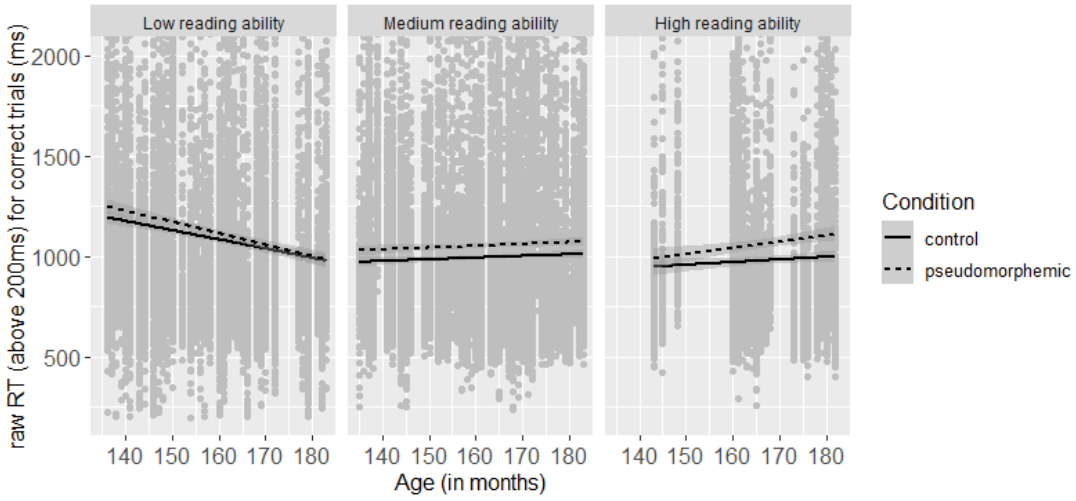


Figure 4.6: Three-way interaction of Condition × Reading ability × Age in month

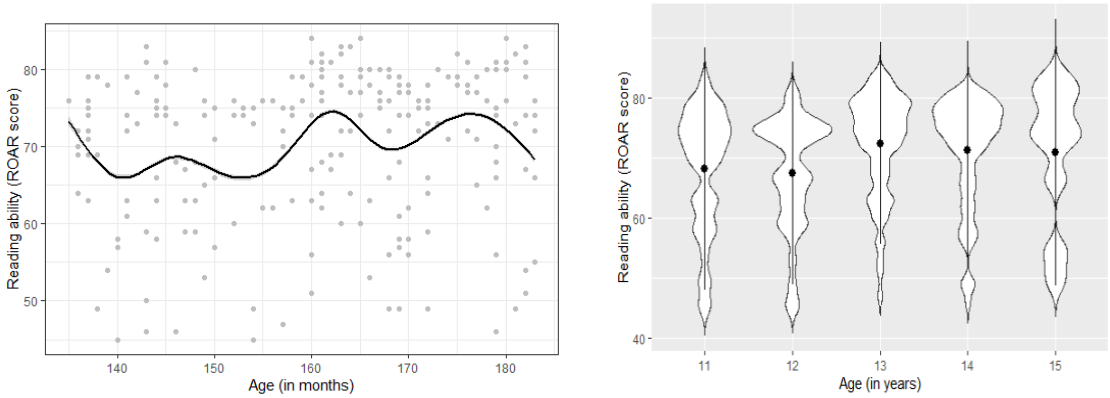


Figure 4.7: Reading ability across Age Ranges

4.3.3 Experiment 3 discussion

This experiment revealed a morpheme interference effect in adolescents aged 11-17 years. Similar to their adult counterparts, these developing readers exhibited longer RTs and reduced response accuracy for pseudomorphemic nonwords compared to control nonwords. In both RTs and

accuracy, the morpheme interference effect was modulated by factors characterising reading experience, specifically reading ability and suffix familiarity.

The results underscored the intertwined influence of suffix familiarity and individual reading ability on the morpheme interference effect, evident in both RTs and accuracy (i.e., the three-way interactions). While an increase in suffix familiarity generally intensified the morpheme interference effect across all reading abilities, this effect became notably larger with increasing reading ability only for nonwords containing highly familiar suffixes. In contrast, for nonwords with low to medium familiarity suffixes, reading ability did not significantly modulate the morpheme interference effect.

Posthoc analyses separately examining two dimensions of reading experience—reading ability and suffix familiarity (i.e., two-way interactions with condition)—reconfirmed that when averaged across suffix familiarity, participants of varying reading abilities took longer and made more errors when rejecting pseudomorphemic nonwords compared to control nonwords. However, when averaged across reading ability, while suffix presence influenced response accuracy regardless of familiarity, it did not significantly interfere with RTs when suffixes were unfamiliar. This underscores that compared to explicit suffix knowledge, having sufficient experience with suffixes might be more crucial for the development of the ability to implicitly activate morphological knowledge during reading.

Unlike reading ability, age group did not significantly influence the morpheme interference effect on RTs. This could be because RTs that reflect the automaticity and rapidness of morphological processing are more directly linked with reading ability than age. On the other hand, for accuracy, the magnitude of the morpheme interference effect varied across age groups, with older adolescents (14-15 years) exhibiting a significantly larger interference effect compared to participants aged 12-13 and below.

In addition, the interaction between age group and suffix familiarity on accuracy, as shown in Appendix 4H, was unexpected as it did not involve morphological status of the nonwords. For nonwords containing both high familiarity suffixes and corresponding control non-suffix endings, the response accuracy difference converged among the 11-, 12-13, and 14-15-year age groups— the accuracy of the 12-15 year-olds dropped more drastically and became similar to that of 11-year-olds. This might be attributed to random variation in our data, given the overall trend of the suffix familiarity effect on accuracy was essentially consistent across all age groups. It might also be possible that because adolescents, like adults, recognised the letter clusters after stem activation, and their reliance on explicit morphological knowledge, as mentioned earlier, may have interfered with their response accuracy. The suffix recognition of 16- to 17-year-olds was more fine-tuned and more adult-like, whereas younger adolescents seemed more prone to overgeneralise familiar letter clusters. This might suggest that the adolescent period is pivotal in developing precise and accurate orthographic representations of suffixes. However, this explanation is speculative and not fully substantiated by the current study, which necessitates further research.

Similarly, findings from the exploratory analysis should be interpreted with caution, primarily due to potential biases stemming from imbalanced sample sizes across age groups. This imbalance gave rise to a questionable non-linear association between reading ability and age in months. Despite this, the accuracy results largely mirrored the main analysis, with a new interaction between reading ability and suffix familiarity emerging. Compared to the unexpected finding in the main accuracy analysis, this interaction was more inexplicable, and likely caused by statistical noise that might be further amplified by the imbalanced sample sizes. The RT results demonstrated a robust influence of suffix familiarity on the morpheme interference effect. However, contrary to our expectations, reading ability ceased to significantly modulate this effect and its overall influence on RTs became non-significant. This might be due to RTs being particularly sensitive to the unexpected non-linear relationship between reading ability and age. This was hinted at by the trend

observed in the interaction between reading ability and age in months as well as the newly observed three-way interaction of age, reading ability, and condition. This interaction might indicate that older poor readers (possibly indicating genuine reading difficulties) exhibited a less pronounced morpheme interference effect in RTs compared to low-reading-ability counterparts. However, the potential sampling biases and the fact that including age in the model did not significantly improve its fit cast doubt on the reliability of these results.

4.4 General discussion

The present study investigated the effects of reading ability and suffix familiarity on the morpheme interference effect. The key aspect of the present study was the incorporation of two factors that characterise reading experience at both the individual and the item level. Crucially, we examined adolescent participants across a wide age range of adolescent years, bridging the gap between the findings obtained from children and adults. This allowed us to capture a broad spectrum of reading experience and delve into the development of morphological processing during this pivotal stage of reading development— a phase often neglected in previous research (see Ricketts et al., 2020).

For the individual-level reading ability, we used an adapted version of the Online Assessment of Reading Ability (ROAR; Yeatman et al., 2021), a self-administered measure that has been validated for remote online testing. The total accuracy on the ROAR has been shown to be a reliable measure of reading ability in both adult and developing readers (Yeatman et al., 2021). Suffix familiarities were taken from suffix frequency counts from the OCC (Banerji et al., 2013), a large-scale, developmental corpus of children's reading materials. This approach differed from prior research that typically sourced suffix frequency from adult-oriented corpora (e.g., the CELEX database, as used in Beyersmann et al., 2012). Extracting lexical statistics from corpora that capture children's exposure to written language offers an insightful approach to studying how cumulative

experiences with orthographic patterns influence reading, especially considering morphological regularities are a key feature of the English writing system (Dawson et al., 2023; Nation, 2017).

The results replicate the morpheme interference effect, as manifested by slower RTs and lower accuracy for pseudmorphemic nonwords compared to the control, in both adults and developing readers. The findings favour the concept of morpho-orthographic decomposition during visual word recognition, which contradicts supralexical accounts which propose that morphological processing depends on morpho-semantic decomposition, occurring only after a word has been retrieved from the mental lexicon (e.g., Giraudo & Grainger, 2001, Marslen-Wilson et al., 1994). While some might argue that phonological strategies play a role and that pseudomorphemic nonwords might be phonologically more accessible compared to control nonwords (despite our attempts to match them), this seems unlikely for several reasons. Firstly, the nature of the task did not require participants to phonetically articulate the stimuli, eliminating the necessity for phonological activation. Moreover, existing research indicates that the processing of morphologically complex stimuli is primarily based on orthographic representations rather than phonology. For instance, masked priming studies show comparable priming for complex words, regardless of phonological variations (e.g., healer–HEAL vs. health–HEAL; Tang & Witzel, 2020), suggesting that phonological information might not have a significant influence during the early stages of morphological processing. Therefore, our findings effectively replicated the morpheme interference effect observed in previous in-person studies (e.g., Crepaldi et al., 2010; Dawson et al., 2018).

Moreover, the findings indicate that the morpheme interference effect was influenced by both reading ability and suffix familiarity in both adults and adolescents. As anticipated, individuals with superior reading abilities showed a stronger morpheme interference effect, and suffixes with higher familiarity showed greater interference on response times and accuracy. As stronger readers are equipped with more high-quality representations of sub-lexical units like suffixes (Perfetti, 2007), they should have enhanced sensitivity to morphological structures. Therefore, these individuals

could activate implicit morphological knowledge more rapidly and automatically during reading. Besides, as suggested by the "Matthew effect" of reading, more proficient readers tend to engage more with reading. This frequent engagement allows them to encounter suffixes repeatedly, further consolidating the orthographic patterns of these units—a point particularly relevant for developing readers. In parallel, due to the regular appearance of familiar suffixes in reading material, as readers encounter these suffixes more often, their visual word recognition system becomes finely attuned to those morphological structures (Grainger & Beyersmann, 2017). As a result, these suffixes are identified and processed more easily and rapidly during visual word recognition. This aligns with Nagy et al. (2014) who suggested that frequent exposure to suffixes not only strengthens explicit suffix knowledge but also consolidates implicit suffix knowledge, thus facilitating the automaticity of morphological decomposition.

Besides the two primary factors of interest, the morpheme interference effect on accuracy was also modulated by age in the adolescent cohort (Experiment 3). All adolescent participants showed reduced accuracy when rejecting pseudomorphemic nonwords compared to control nonwords, and notably, the 14-15 age group exhibited a significantly larger magnitude of such effect than their 12-13-year-old and 11-year old counterparts. This hinted that a key transition in morphological processing might occur during the mid-to-late adolescent years. Drawing from the edge-aligned embedded stem activation account (Grainger & Beyersmann, 2017), participants are likely to identify stems early in the reading process. They might then use apparent morphological structures to infer word meanings when encountering unknown items after stem identification, as suggested in previous research on developing readers (e.g., Dawson et al., 2018; Hasenäcker, Solaja et al., 2021). Employing this morphological strategy, participants should rely heavily on explicit morphological knowledge, leading to more misjudgements of pseudomorphemic nonwords. This explicit knowledge about suffixes is often acquired over years of reading instruction. As such, age group, as a proxy for years of formal reading instruction, uniquely contributes to the variance in response

accuracy, alongside reading ability. The pattern of accuracy aligns with the findings in Dawson et al. (2018), with the transition emerging somewhat earlier in our study. Dawson et al. (2018) identified a more pronounced morpheme interference effect on accuracy among 16- to 17-year-olds compared to 12-13 year olds. This discrepancy could be because Dawson et al. (2018) only included two adolescent groups (12-13 years and 16-17 years) and did not include participants aged 14-15 years in their sample as well as different factors considered between the two studies (e.g., Dawson et al., 2018 did not consider reading ability and suffix familiarity as in Experiment 3).

By contrast, age did not modulate the morpheme interference effect on RTs. Regardless of the age group, responses to pseudomorphemic nonwords were consistently slower than to control nonwords. This contrasts with the findings of Dawson et al. (2018), who did not observe this effect among 12- to 13-year-olds. A potential reason could stem from our consideration of interindividual differences in reading ability, a factor Dawson et al. (2018) did not account for in their models. When considering developing readers, the variability in reading-related skills becomes increasingly important, potentially leading to diverse results across studies, as highlighted by Hasenäcker et al. (2020). Although both reading ability and age were taken into account, it was predominantly reading ability that significantly modulated the morpheme interference effect on RTs. These findings might suggest that reading ability is more closely associated with RTs, implying it was a stronger predictor than age in assessing the rapidness and automaticity of morphological processing. However, this explanation warrants caution as suffix familiarity was also a key consideration in our study. A full understanding of how these factors interplayed and collectively influenced the morpheme interference effect remains essential.

Indeed, beyond their individual influences, both individual- and item-level factors collaboratively influenced morphological decomposition in the developmental sample in Experiment 3. While stronger reading ability enhanced the morpheme interference effect, this modulation was conditioned by suffix familiarity. Specifically, the interference effect became more

pronounced with increasing reading ability only for nonwords containing highly familiar suffixes. This indicates that for developing readers, morphological decomposition is notably influenced by reading ability, but only when they can regularly encounter certain suffixes in various written contexts (e.g., Rastle et al., 2004). Even for proficient developing readers, consistent exposure to suffixes is essential in refining their mental representations of those sublexical units. This underscores the importance of accounting for both facets of reading experience when investigating the complexities of morphological decomposition.

Notably, in adults, both reading ability and suffix familiarity individually influenced the morpheme interference effect, but they did not interact as they did in younger readers. We observed a larger morpheme interference effect in stronger readers and for nonwords with more familiar suffixes. These findings suggest that, even with their extensive reading experience, adult morphological processing is still influenced by individual reading proficiency (see Andrews & Lo, 2013; Beyersmann et al., 2015) and how often they encounter specific suffixes. This highlights the consistent influence of reading ability in morphological decomposition and underscores the ongoing significance of accumulated suffix exposure, even in skilled adult readers. Despite the evident link between suffix familiarity and reading ability, the absence of the three-way interaction in adults suggests that one factor did not modulate the impact of the other. It is possible that the extended reading experience in adults provides even the less proficient readers with ample exposure to common suffixes. As a result, their familiarity with these suffixes may not vary as much by reading proficiency, unlike in developing readers. Such extensive exposure could potentially lead to shifts in language processing for adults, enabling them to establish a more stabilised morphological decomposition. This stability might reduce the variation in morpheme interference tied to the interplay between reading proficiency and suffix familiarity.

I will not discuss in depth the results from the exploratory analysis, primarily due to potential sampling bias warranting interpretative caution. These findings hint at a potential interaction

between reading ability and age, an interplay not evident in our main analysis. Individual reading skill variance may arise from both the developmental stage of reading, indicated by age, and relative reading proficiency within a peer group (cf. Hasenäcker et al., 2020). Standardising reading scores within each age group might offer insights into this peer-relative reading ability. We refrained from conducting this analysis for two reasons: Firstly, our primary objective was to capture a broad spectrum of reading abilities, focusing on the overarching effect of reading ability on the morpheme interference effect, rather than differentiating between the influences of years of schooling and peer-related reading proficiency. Secondly, the reading ability measure, ROAR (Yeatman et al., 2020), was not tailored for assessing word reading in individuals across specific age ranges, potentially undermining its reliability in assessing peer-relative reading ability. Future studies aiming to distinguish the influences of age and peer-related reading ability could consider tools designed to evaluate reading proficiency within distinct age cohorts, which would allow accurate assessment.

A limitation of our study is the lack of position-specific frequency matching between the non-suffix endings in control nonwords and the suffixes in pseudomorphemic nonwords. Even with bigram frequencies matched across conditions, potential frequency discrepancies between the non-suffix endings and the suffixes might exist. If the suffixes naturally occurred more frequently than the non-suffix endings in word-final positions, it could introduce ambiguity in conclusively determining if the observed differences arose due to morphological structures or simply letter co-occurrence frequencies. Furthermore, while stem activation is position-independent, affix recognition is position-dependent, as indicated by studies like Crepaldi et al. (2010). Consequently, if these non-suffix endings were less commonly found in word-final positions compared to the suffixes, this could lead to confounds between position and morphological structure. Future studies could address this limitation by computing position-specific letter cluster frequencies when selecting non-suffix endings for control nonwords. This could be achieved by extracting all words ending in specific letter clusters and then aggregating the frequencies of those words.

In addition, this study relied solely on suffix token frequencies to infer familiarity. It is widely accepted that frequent exposure or accumulated experience with morphemes increases their familiarity (e.g., Carlisle, & Katz, 2006). While there are the strong correlations between suffix type and token frequencies, type frequency reflects the diversity of word forms a suffix can generate, which could be seen as a measure of productivity, whereas token frequency measures the cumulative occurrence of a suffix across all words (e.g., Bybee, 1995). Since our focus was not on the word form diversity of a given suffix, we found token frequency — which reflects the overall prevalence of a suffix — to be the most appropriate measure of suffix familiarity. For future research focusing on the influence of stem diversity in morphological processing, a manipulation crossing type and token frequencies across various suffixes might be beneficial.

Another important consideration is that the suffixes used in this study differ not only in familiarity but also in other aspects such as productivity, semantic consistency, and transparency. For instance, as mentioned previously, the productivity of the suffixes varies, with some like -ish being more widely combined to different stems to create new words. Additionally, these suffixes differ in their semantic transparency, which can affect the interpretability of the pseudowords they created (e.g., Burani et al., 1999). For example, suffixes like -ful and -ly are highly semantically transparent and consistent across different words. Furthermore, some suffixes are relatively 'invariant', meaning they do not change the stem spelling when added (e.g., -ly), while others, like -acy, often require alterations to the stem endings. However, in this study, all pseudowords were created by adding the target suffixes to an unaltered stem. Future research could explore how these factors influence morphological processing or attempt to control for them. However, this is challenging because English has a relatively limited number of affixes, making it difficult to fully match these aspects across different suffixes.

In conclusion, the findings of Chapter 4 add evidence to a growing body of studies investigating the modulation and the development of morphological processing. Reading ability and suffix

familiarity are critical factors modulating morphological processing, especially among developing adolescent readers. Adolescents have already developed an automatic morphological decomposition system, albeit not fully adult-like. Their morphological processing was associated with individual differences in reading ability and the frequency of occurrence of suffixes in reading material. This is consistent with reading experience shaping the development of skilled word recognition.

Chapter 5: Influence of suffix exposure during reading on morphological processing

5.1 Introduction

Skilled readers automatically and rapidly decompose complex words into their morphological components, a process that is termed morpho-orthographic decomposition (Taft, 2004). Evidence in support of this comes from the morpheme interference effect (e.g., Crepaldi et al., 2010; Dawson et al., 2018). Adolescence has been shown to be a crucial phase in progressing toward adult-like morphological decomposition, yet this phase has been underrepresented in much of the prior research (cf. review in Chapter 2; Experiment 3 in Chapter 4; Dawson et al., 2021a, 2018; Ricketts et al., 2020). Results in Chapter 4 have shown that in adolescents, nonwords with an apparent morphological structure, such as 'earist', are harder to reject as words than their matched controls without such structures, like 'earilt'. These observations highlight the pivotal role of reading experience—and specifically, suffix familiarity—in morphological decomposition. However, the underlying mechanisms that tie suffix exposure to its implicit activation during reading, and the possible causal link between suffix familiarity and the advancement of morphological processing remain to be underexplored. This chapter seeks to investigate whether exposure to suffixes during reading may increase their familiarity, such that they are activated in subsequent visual word recognition tasks. Additionally, it examines factors that might enhance suffix acquisition through silent self-paced reading. The aim of this study was to establish whether there is a causal link between suffix familiarity and proficient morphological processing.

Print exposure is associated with reading ability (e.g., Chateau & Jared, 2000). The idea that exposure to affixes through reading and then positively impacts morphological processing is supported by the item-level factor of suffix frequency. Suffix frequency acts as a marker of exposure and a proxy for familiarity with morphological patterns. Specifically, research has shown that words with high-frequency suffixes are read and recognised more efficiently in reading aloud

and lexical decision (e.g., Angelelli et al., 2014, 2017; Burani et al., 2018; Lázaro et al., 2016; Traficante et al., 2011). In addition, the extent of morphological priming, as measured by RTs, increases with suffix frequency (Beyersmann et al., 2012).

While the frequency effect reflects cumulative experiences with written language (Nation, 2017), studies have indicated that even short-term, repeated exposures in lab settings could lead to the learning of novel or unfamiliar lexical units. For instance, studies utilising the self-teaching paradigm have shown that learners can acquire the orthographic patterns of novel words, whether presented in isolation or embedded in meaningful contexts via either reading aloud (e.g., Nation et al., 2007) or by silent reading (e.g., de Jong & Share, 2007). The strength of the learning effect improves with the number of exposures (e.g., Bowey & Muller, 2005), but is also evident after just a single-trial encounter (e.g., Nation et al., 2007). Additionally, even incidental encounters with specific words influences subsequent word recognition (Protopapas et al., 2017).

Although it has been suggested that stems are the more salient processing units (as the semantic function of suffixes is less transparent, Grainger & Beyersmann, 2017), research demonstrates that throughout the learning process, readers are sensitive to morphological structures and affix units (Dawson et al., 2021b; Pacton et al., 2013). For instance, Dawson et al. (2021b) investigated how suffixes (e.g., ‘ist’) in novel words with a nonword stem (e.g., ‘clantist’) influence stem learning by establishing two scenarios: one where the definitions of the words were congruent with the typical semantic and syntactic roles of their suffixes, and one where they were not. After two training sessions, adolescents aged 12-13 and 16-19 showed significantly improved semantic recall for the novel stem words in the congruent condition compared to those in the incongruent condition. This result highlights that developing readers are sensitive to suffixes as meaningful sublexical units during reading. In addition to word learning, Tamminen et al. (2015) specifically investigated the acquisition of novel suffixes. They found that adults, when repeatedly exposed to novel suffixes embedded within nonwords in sentence contexts, were able to generalise the semantic function of

these suffixes to new, untrained stems. However, to my knowledge, there have not been studies investigating whether intensive exposure to low-frequency real suffixes during reading facilitates the sensitivity to these suffix representations, thereby making them more readily activated during morphological processing.

To address this gap in the literature, Experiment 4 in this chapter investigates the effects of intensive exposure to low-frequency real suffixes on their processing. Reflecting on findings from Chapter 4, where the morpheme interference effect (i.e., the pseudomorphemic words, comprising a stem + suffix, e.g., 'birdly', were more difficult to reject compared to its control nonword with the same stem and a nonsuffix like 'birdla') was found to be absent for low-frequency suffixes—presumably due to the limited familiarity with these sublexical patterns among children. The morpheme interference paradigm could serve as a sensitive measure for assessing suffix familiarity and the quality of their representations in the mental lexicon. In addition, unlike artificial suffixes, using real suffixes that are embedded in real words should help avoid 'weird looking' items that stand out in sentences, and might therefore attract explicit attention (cf. Protopapas et al., 2017). My approach aimed to more closely mirror a natural reading context and therefore potentially enhance the generalisability of the findings to real-world morphological acquisition processes.

Using low-frequency real suffixes means that participants were likely to have previously encountered these suffixes and therefore have some (albeit weak) knowledge of their orthographic forms. I hypothesised that more intensive exposure would bolster this familiarity. This increased familiarity, in turn, was expected to facilitate rapid and automatic activation of the suffixes during reading, which could elicit or amplify a morpheme interference effect associated with these suffixes. However, the frequency of exposures should not be considered merely in terms of the number of times they appeared. The linguistic context in which these suffixes were experienced might influence their acquisition. As suggested by Nation (2017), the frequency effect in word learning and processing is not merely about simple repetition of a word in isolation or repeated in

identical contexts. Semantic diversity of the contexts matters beyond raw frequency, with words experienced in more diverse contexts influencing reading and lexical decision (Nation, 2017). It could be reasonably hypothesised that this applies to suffixes as well, although suffixes, unlike stem words, do not typically appear on their own. Relevant though might be morphological family size (De Jong et al., 2000) and suffix productivity (Lázaro et al., 2015) and both of these factors influence lexical decision performance when processing morphologically complex words. Therefore, an additional objective of Experiment 4 in this Chapter 5 was to explore whether the diversity of stems to which a suffix is attached influences the acquisition and processing of these suffixes.

My experiment used a morpheme interference paradigm at both pre-test and post-test. In between in the exposure phase participants encountered multiple exposures to target suffixes in sentence contexts. Of interest was whether this experience was associated with differences in morphological processing, as estimated by the morpheme interference effect at post-test. Considering that age in months and reading ability were identified as influential factors on the morpheme interference effect in Chapter 4, these variables were included as continuous factors in the analysis. Furthermore, Chapter 4 revealed that the suffix '-ise', with a token frequency of 191.61 per million in the Oxford Children's Corpus (OCC), exhibited minimal evidence of the morpheme interference effect. Therefore, this chapter focused on suffixes with lower frequencies than '-ise'. Given their low frequency, these suffixes were anticipated to be less likely to show a morpheme interference effect at pre-test. To test the hypothesis regarding stem-diversity, suffixes were presented as either only attached to two stems (i.e., low stem diversity) or eight different stems (i.e., high stem-diversity) during the exposure phase, but with overall frequency controlled across conditions. The hypotheses were:

1. The morpheme interference effect (i.e., the difference in the difficulty of rejecting pseudomorphemic vs. control nonwords) would be more evident at the post-test compared

to the pre-test in the exposure conditions, but not in the non-exposure control condition, i.e., a three-way interaction of Nonword type (pseudomorphemic vs. control) × Test (pre vs. post) × Exposure condition (mean of high stem-diversity and low stem-diversity vs. non-exposure);

2. The exposure effect as described in hypothesis 1 would be more pronounced in the high stem-diversity exposure condition compared to the low stem-diversity condition, i.e., a three-way interaction of Nonword type (pseudomorphemic vs. control) × Test (pre vs. post) × Exposure condition (high stem-diversity vs. low stem-diversity).

5.2 Method

This study was pre-registered on the Open Science Framework (OSF; https://osf.io/vmpcd). The raw data, R code for these analyses, and experimental materials are available at the same location for transparency and replication purposes.

A mixed factorial design was adopted. There were two within-participant factors, Nonword type in the lexical decision task (pseudomorphemic vs. control) and Test phase (pre- test vs. post- test). Exposure condition was manipulated between-participants (high stem-diversity vs. low stem-diversity vs. non-exposure). All participants completed a pre- and post-exposure lexical decision task, and were randomly assigned to one of the three exposure conditions. To ensure uniformity and avoid potential confounds between suffix exposure effects and year group, participants were equally and randomly assigned to the three exposure conditions within their respective year groups.

5.2.1 Sample size rationale

The sample size was determined through a power simulation using the Monte Carlo method, as implemented in the powerSim function from the 'SIMR' package in R (Green & MacLeod, 2016). The simulated RT data were informed by the results of Experiment 3 with adolescents in Chapter 4 on the effects of reading ability and suffix familiarity on the morpheme interference effect. We

established an RT range of 200 - 3000ms, with average RTs of 900ms, and a standard deviation of 600ms. An effect size of β = .03 was employed for the three-way interaction of Nonword type (pseudomorphemic vs. nonmorphemic) ×Test (pre vs. post) ×Exposure condition (high diversity vs. low diversity vs. non-exposure). Previous results reported in Experiment 3 revealed an effect size of β = 0.065 for the interaction of Nonword type and Suffix familiarity. Therefore, if the suffix exposure manipulation in the current study increased the familiarity with target suffixes, thereby eliciting a comparable suffix familiarity effect, a similar interaction effect between Nonword type and Test could be anticipated. However, considering that in Experiment 4 in the current chapter there was another variable, i.e., Exposure condition and that suffix exposure was via a single short session only (compared with cumulative reading experience), we anticipated a much smaller effect size for the three-way interaction. According to the results of 300 simulations, a minimum of 150 participants would be needed to achieve 80% power for detecting the three-way interaction. To ensure adequate power, we aimed to recruit at least 180 participants, accounting for potential exclusions and ensuring that 150 eligible participants would be included in the analysis.

5.2.2 Participants

A total of 220 students, ranging from Year 7 to Year 10 (ages 11-15), were initially recruited from two secondary schools in different cities within the UK. Twenty-two participants were excluded because they did not finish the reading ability measure (ROAR) at the end of the experiment. An additional 36 participants were removed from the final sample because they scored below the chance level on either the word or nonword stimuli (or both), in the pre-exposure test, post-exposure test (or both).

It is essential to note a deviation from our pre-registered protocol. Originally, we had planned to exclude participants who scored below chance level on the overall lexical decision task (including both words and nonwords) in the pre- or post-tests. However, during data cleaning, we noticed an unexpected pattern: Some participants displayed near-perfect accuracy for either words or nonwords

but showed nearly zero accuracy for the alternate stimuli type. This trend was more pronounced during the post-test. We suspect that this pattern might indicate uncooperative actions, possibly due to fatigue, boredom, or a lack of engagement. This might have led participants to press a single response button repeatedly without responding appropriately to the stimuli. For example, they might consistently press the 'J' or the 'F' key regardless of the stimuli presented. This type of response behaviour may be more likely in contexts such as online testing, pre- and post-test designs, group administration, and when working with adolescent participants, where motivation and attention can fluctuate. All these factors were present in the in the current experiment, making it more likely that this response pattern would occur.

Given these observations, we amended our initial exclusion criteria. We decided to evaluate accuracy separately for word and nonword stimuli. Only participants who demonstrated above-chance accuracy for both types of stimuli were included. We believe that this modification helped to preserve the integrity of our data by minimising the impact of potential anomalous response patterns.

The final sample consisted of 147 adolescents (N_{Year7}= 19, N_{Year8} = 15, N_{Year9} = 96, N_{Year10} = 17). Among these, eight were non-native speakers of British English, and 10 reported having a diagnosed reading disorder. There was one participant who fell into both categories, being a non-native speaker and reporting a reading disorder. All participants completed the ROAR (Yeatman et al., 2021) as in Chapter 4 which served as a background reading ability measure (see Appendix 4A for the stimulus list). The study was approved by the Central University Research Ethics Committee (CUREC).

5.2.3 Materials

Suffix exposure phase. Four low-frequency suffixes were chosen based on their frequencies derived from the Oxford Children's Corpus (OCC; Banerji et al., 2013), as described by Dawson et al. (2023). All the four suffixes had token frequencies below 150 per million words (but above 30

per million words) in the OCC (frequency per million: '-ism': 139.15, '-ee': 93.32, '-ling': 50.68, '-let': 32.87). These suffixes were selected as low-frequency suffixes because as shown in Experiment 3 (Chapter 4), the suffix '-ise' with an OCC token frequency of 191.61 per million demonstrated little to no evidence of the morpheme interference effect. In addition, only invariant suffixes were chosen as they do not induce changes in the stem.

A total of 32 exposure words, each containing one of these four suffixes (eight words per suffix), were selected (see Table 5.1). All the words are semantically transparent and have a free-standing stem. We aimed to preserve the stem, except in a few cases where the final 'e' of the stem was omitted or overlapped with the suffix (e.g., 'fledgling', 'refugee'). Although the embedded suffixes had low frequencies, the 32 exposure words are generally known across the population of adults, as indicated by their prevalence values being well above zero[24] (Brysbaert et al., 2019). Exposure words were assigned to the high- or low diversity condition and their prevalence values were summarised in Table 5.1.

Table 5.1: Means (SDs) of Word Prevalence Values for Exposure Words containing Exposed Suffixes, across High- and Low-diversity Exposure Conditions.

Suffix	**Embedding words**	**Mean prevalence (SD)**	
		High diversity	***Low diversity***
- ism	**tourism**, **vandalism**, magnetism, symbolism, socialism, realism, liberalism, skepticism	2.33 (.26)	2.51 (.10)
-ee	**attendee**, **refugee**, devotee, referee, trainee, trustee, employee, committee	2.21 (.19)	2.2 (.14)
-let	**leaflet**, **piglet**, coverlet, booklet, droplet, ringlet, bracelet, wavelet	1.72 (.83)	2.18 (.36)
-ling	**nestling**, **seedling**, sapling, fledgling, duckling, dumpling, changeling, hatchling	1.85 (.31)	1.87 (.24)

Note. The two words used in the low diversity conditions are **highlighted in bold** within each word set. Prevalence values for each word are presented in Appendix 5A.

[24] According to Brysbaert et al. (2019), word prevalence has negative values for words known by less than 50% of their participants, and those words are considered not known well enough. A word known by 97.5% of the participants would have a prevalence of + 1.96 whereas a word known by 2.5% of the participants would have a word prevalence of –1.96.

Several stories were crafted to contain the exposure words. Each story consisted of six story fragments, amounting to 160-170 words, with each fragment comprising 1-2 sentences of 25-30 words (see Table 5.2 for sample text). Three versions of each story were created for the three exposure conditions:

a. *High stem-diversity exposure condition*: Each story incorporated all eight exposure words containing a specific target suffix (e.g., for suffix '-let', 'leaflet', 'piglet', 'coverlet', 'booklet', 'droplet', 'ringlet', 'bracelet', and 'wavelet').

b. *Low stem-diversity exposure condition*: Each story used only two exposure words containing the target suffix. Each word was repeated four times throughout the story (e.g., 'leaflet', 'piglet', each repeated four times). The mean prevalence value of these two words was comparable to that of the eight-word set in the high-diversity condition, as summarised in Table 5.1.

c. *Non-exposure condition*: Target suffixes were entirely absent from the stories.

For both the high stem-diversity and low-stem diversity exposure conditions, two unique stories were created for each target suffix, yielding a total of eight stories for the four suffixes. The same set of exposure words was incorporated across both stories corresponding to each target suffix. Each suffix only appeared in the two stories assigned to it, without appearing in any other stories. This way, participants assigned to these two exposure conditions encountered each of the four suffixes a total of 16 times. To diminish the potential primacy and recency effects, no exposure words were placed at the beginning or end of any story fragment.

To boost participant engagement, a comprehension question unrelated to the exposure words was presented after each story fragment. The same questions were used across all three conditions for the equivalent fragments. All stories and questions are reported in Appendix 5B and Table 5.2 presents sample fragments from the stories.

Efforts were made to keep the stories across conditions as similar as possible by preserving similar story content, sentence structures, and non-exposure words. We also ensured that the same words would not appear more than once in the same fragment in the low diversity condition.

Table 5.2: Examples of the First Three Fragments from Stories Featuring '-let' Suffix Exposure Words and Corresponding Comprehension Questions across Three Conditions

Condition	Story Fragments	Comprehension Questions
High stem-diversity	1) Sophia loved to visit her aunt in the countryside where she enjoyed cosying up under a soft **coverlet** and admiring the stunning views outside the window. 2) In the forest behind the house, she had a secret garden - a hidden forest that couldn't be found on any travel **leaflet**, her own personal retreat. A. On her adventures to the forest, Sophia always brought along her favourite **booklet** of forest animals hoping to catch a glimpse of the creatures she had read about.	Sophia loved to visit___? A. Her grandmother B. Her auntie C. Her uncle What was behind the house? A. A forest B. A park C. A garden What did Sophia read about? A. Mountains B. Oceans C. Animals
Low stem-diversity	1) Sophia loved to visit her aunt in the countryside, a place advertised in the local travel ***leaflet*** as a must-visit. She enjoyed cosying up and admiring the stunning views. 2) In the forest behind their house, she had a secret garden - a hidden forest that couldn't be found on any travel ***leaflet***, her own personal retreat. 3) On her adventures to the forest, Sophia always took her pet ***piglet*** and her favourite ***leaflet*** of forest animals, hoping to find other creatures she had read about.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
Non-exposure	1) Sophia loved to visit her aunt in the countryside where she enjoyed cosying up under a soft blanket and admiring the stunning views outside the window. 2) In the forest behind their house, she had a secret garden - a hidden forest that couldn't be found on any travel guide, her own personal retreat.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

3)	On her adventures to the forest, Sophia always took her pet dog and her favourite picture book about forest animals, hoping to find other creatures she had read about.

Note. Exposure words were not highlighted in bold in the experiment.

Pre- and Post-tests. Parallel lexical decision tasks (List A and List B) were used in pre- and post-test to assess the implicit activation of exposed suffix representations.

Pseudomorphemic nonwords were created by combining word stems with one of the four target suffixes to which participants were exposed during the suffix exposure phase (e.g., king + ling = kingling). Control nonwords were formed by modifying one letter in the suffix of each pseudomorphemic nonword, resulting in a stem with a non-suffix ending (e.g., king + lang = kinglang). In each list, each suffix was combined with nine different stems, totalling 36 pseudomorphemic nonwords and 36 corresponding controls. Thus, the experimental stimuli consisted of 16 sets of nine nonwords across both List A and List B.

The nonwords across sets were matched on various linguistic characteristics (see Table 5.3), including length in letters, orthographic neighbourhood size N (Coltheart, 1981), and mean log bigram frequency (MLBF), obtained using N-Watch (Davis, 2005). The stems used were common English words, ranging from three to six letters in length, and each ended with a consonant to avoid the need for spelling alternations when combined with a suffix. The sets of stems were matched on log10 frequency in the OCC (Banerji et al., 2013; Dawson et al., 2023), concreteness (Brysbaert et al., 2014), age of acquisition (AoA; Kuperman et al., 2012), semantic diversity (Hoffman et al., 2013), orthographic neighbourhood size (Coltheart, 1981), MLBF, and lexical decision RT in the British Lexicon Project (Keuleers et al., 2012). The 16 sets of stimuli were divided into two lists, List A and List B, each containing 72 nonwords (36 pseudomorphemic and 36 controls). List A and List B were assigned as either the pre-test or the post-test, with the order of administration being counterbalanced across participants. The complete stimulus list is reported in Appendix 5C.

To ensure morpheme boundaries maintained consistent across nonword types, the non-suffix endings of the control nonwords (i.e., "-ea", "-irm", "lat", "-lang") were formed by changing one letter in the central position or at the end of each target suffix (i.e., "-ee", "-ism", "-let", "-ling"). Besides orthographic overlap, the non-suffix endings and their suffix counterparts also had similar bigram frequencies.

Additionally, two sets of 74 word fillers (74 for List A and 74 for List B) were included for the purpose of making up the lexical decision task. Responses to words were not analysed.

Table 5.3: Means (SDs) for Lexical Characteristics of Nonwords and Their Stems for Pseudomorphemic and Control Nonwords

Lexical Characteristics	**List A**		**List B**	
	Pseudo	*Control*	*Pseudo*	*Control*
Length	7.06 (1.07)	7.06 (1.07)	7.11 (1.14)	7.11 (1.14)
Orthographic N	.17 (.38)	.08 (.28)	.16 (.38)	.11 (.32)
MLBF	2.50 (.27)	2.31 (.27)	2.45 (.30)	2.24 (.36)
Stem log10 frequency in OCC	1.72 (.30)		1.79 (.29)	
Stem length	4.06 (.75)		4.11 (.78)	
Stem concreteness	4.53 (.59)		4.73 (.28)	
Stem AoA	5.06 (1.59)		5.18 (1.20)	
Stem semantic diversity	1.60 (.15)		1.55 (.16)	
Stem N	9.08 (5.41)		8.08 (4.84)	
Stem MLBF	2.92 (.41)		2.93 (.41)	
Stem RTs (in BLP)	526 (22)		531 (29)	

Note. Length, Orthographic N, and MLBF were calculated using N-Watch (Davis, 2005)

5.2.4 Procedure

Participants took part in the study at their schools under the supervision of a trained researcher or their class teacher. Data were collected online using the Gorilla Experiment Builder (https://gorilla.sc/). Instructions for each task were provided on-screen.

The experiment began with the pre-test lexical decision task, where participants saw a series of letter strings one at a time. They were instructed to decide whether each string was a real English word as quickly and accurately as possible, pressing 'J' for words and 'K' for nonwords. Each trial began with a 500-ms fixation cross, followed by the target displayed at the centre of the screen for a maximum of 3000ms or until the participant responded. Trials with no responses within the 3000ms limit were recorded as 'timeout' and immediately proceeded to the next trial. Each stimulus was presented only once, and the order of presentation was fully randomized. The test consisted of three blocks of 48 trials, with short breaks offered after each block.

Following the pre-test, participants were randomly assigned to one of the three suffix exposure conditions. In each condition, they read designated stories and answered corresponding multiple-choice comprehension questions as instructed. A fragment from the stories was displayed on the screen each time. Participants read at their own pace and progressed by pressing the space bar once they finished reading. The story fragment automatically moved on after a maximum duration of 18 seconds in the absence of a key pressing. This duration was chosen based on silent reading rate and comprehension in previous studies (e.g., Hayden et al., 2019; Daniel, 2015). A comprehension question with three options appeared after each story fragment disappeared, with no time limit for answering. Participants chose one answer from three options via a mouse click, and once they answered, the subsequent fragment was presented. No feedback was provided for these questions. The fragments within each story were presented in a fixed order, whereas the order of the eight stories was randomised.

Upon completion of the stories during the exposure phase, participants completed a digit span task before the post-test. This task, not involving lexical components, served as an effective distractor without interfering with the experimental tasks. They saw a sequence of digits ranging from 3-8 digits on the screen. Each number was displayed for 400ms. After each sequence, participants were prompted to enter the numbers in the correct order in an answer box. Sequences

began with three digits and increased by one, up to eight, with each length presented in two consecutive trials in a fixed, ascending order.

The post-test mirrored the pre-test format, but with a different set of stimuli. Half of the participants completed List A for the pre-test and List B for the post-test, with the other half doing the opposite. The ROAR (Yeatman et al. 2021), the same measure used in Chapter 4, was administered immediately after the post-test.

Participants were offered regular short breaks and could withdraw from the study at any time by closing their browser window. The experiment lasted approximately 25 minutes.

5.3 Results

Data analysis was conducted using the 'lmerTest' package (Kuznetsova et al., 2017) in RStudio (RStudio Team, 2020). Data wrangling was facilitated using the 'tidyverse' package (Wickham et al., 2019). The significance of the fixed effects was determined using the ANOVA (type III) function in the car package (Fox & Weisberg, 2019). Figures were generated using 'ggplot2' (Wickham et al., 2016). The significance and variance explained by each fixed effect and interaction, as determined by Type III ANOVA, are reported in tables within the corresponding sections, whereas the detailed model summary, including estimates of fixed effect significance from the 'lmerTest::summary()' function, is provided in the appendix.

We utilised linear mixed-effects models (LME) for RT data and generalized linear mixed-effects models (GLME; binomial family) for accuracy data (binary). Both models incorporated random effects for item and participant. The two categorical variables, 'Nonword type (pseudomorphemic vs. control)' and 'Test (pre vs. post)', were recoded using deviation contrast coding with a coding scheme of 0.5 and -0.5. Additionally, Helmert contrasts (Schad et al., 2020) were employed to code the levels of the 'Exposure Condition (non-exposure vs. high vs. low)', allowing comparisons between the high and low stem-diversity exposure conditions (contrast [1]:

low = -1, high = 1, non-exposure = 0), as well as between the non-exposure condition and the average of high and low stem-diversity exposure conditions (contrast [2]: low = -1, high = -1, non-exposure = 2). The continuous control variables (e.g., age in months, ROAR scores) were standardised by mean-centring and scaling.

5.3.1 Accuracy

Accuracy across the three exposure conditions did not differ significantly at the pre-test (*p*s > .30). Descriptive data as a function of exposure condition and test phase are summarised in Table 5.4. Accuracy in rejecting nonwords containing each suffix collapsed across pre- and post-test phases is shown in Figure 5.1 which showed varied but clear morpheme interference (i.e., lower accuracy in responding to pseudomorphemic nonwords) for each suffix.

Table 5.4: Means (SDs) of Accuracy (Proportion Correct) for Pseudomorphemic and Control nonwords.

Test / **Exposure condition**	Pre-test		Post-test	
	Pseudo	*Control*	*Pseudo*	*Control*
High diversity	0.81 (0.39)	0.91 (0.29)	0.82 (0.38)	0.89 (0.32)
Low diversity	0.82 (0.39)	0.92 (0.27)	0.81 (0.39)	0.88 (0.32)
Non- exposure	0.81 (0.39)	0.91 (0.28)	0.87 (0.34)	0.91 (0.29)

The main effects and interactions of the primary variables in model:

$Accuracy \neg\ glmer\ (Correct_{\log-odds} \sim Nonword\ type \times Test \times Exposure\ condition +$

$Reading\ ability +\ Age\ in\ months\ + (1 + Nonword\ type + Test \mid Participant)\ +$

$(1 \mid Item)$ are presented in Table 5.5. A detailed model summary is at Appendix 5D.

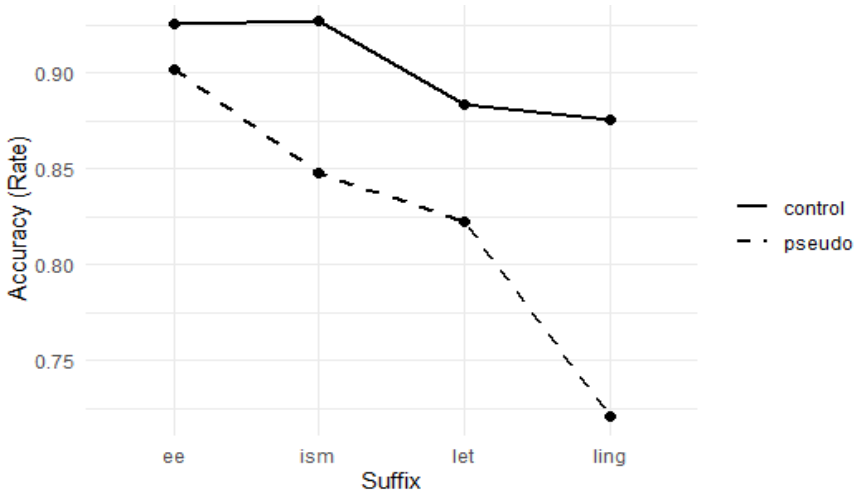


Figure 5.1: Proportion Correct for Pseudomorphemic and Control Conditions Grouped by Each Suffix (Pre- and Post-tests Combined)

There was a significant main effect of Nonword type (β = -0.82, z = -7.23, p < .001), with pseudomorphemic nonwords being harder to reject than control nonwords, indicating a morpheme interference effect. Pairwise comparisons with Bonferroni corrections showed that the morpheme interference effect was significant in both pre- and post- tests for each exposure condition (see Appendix 5E for detailed pairwise comparisons). Additionally, a significant interaction was observed between Nonword type × Test (β = 0.42, z = 4.42, p < .001). As illustrated in Figure 5.2, this indicates that the difference in accuracy across two nonword types diminished from the pre- to the post-test. In addition, although the overall interaction between Exposure condition and Test showed only a marginal contribution to the variability of the model ($c^2 = 5.57$, p = .062, see Table 5.5), the specific interaction involving Test and the second contrast of the Exposure condition was significant (β = 0.13, z = 2.24, p = .025, detailed in Appendix 5D). This contrast compared the

'non-exposure' condition to the average of the two exposure conditions, indicating that accuracy increased more from the pre to post-test in the non-exposure condition compared to the average effect in the two exposure conditions (Figure 5.3). Nevertheless, since Nonword type was not involved in this interaction, this result could not be directly attributed to changes in the morpheme interference effect from the pre- to the post-test. Despite this, to provide a comprehensive view of the data, Figure 5.4 illustrates accuracy across Tests and Exposure conditions for each type of nonword.

Table 5.5: Accuracy Model Output for the Main Effects

Fixed effects	c^2	*p value*
(Intercept)	760.38	< **.001**
Nonword type	52.28	< **.001**
Test	0.012	.91
Exposure Condition	0.51	.78
Nonword type × Test	19.52	< **.001**
Nonword type × Exposure condition	3.16	.21
Test× Exposure Condition	5.57	.062
Nonword type × Exposure Condition× Test	0.48	.79

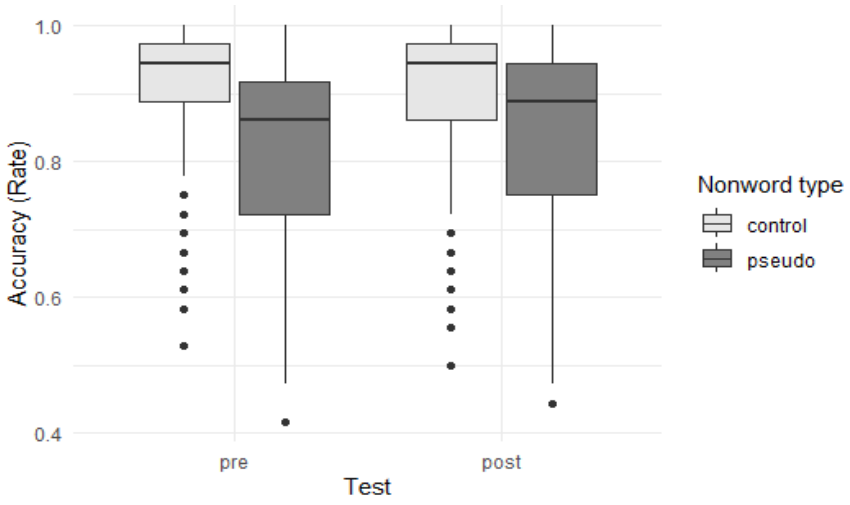


Figure 5.2: Proportion Correct for Pseudomorphemic and Control Nonwords in Pre- and Post-test (the lines represent the medians)

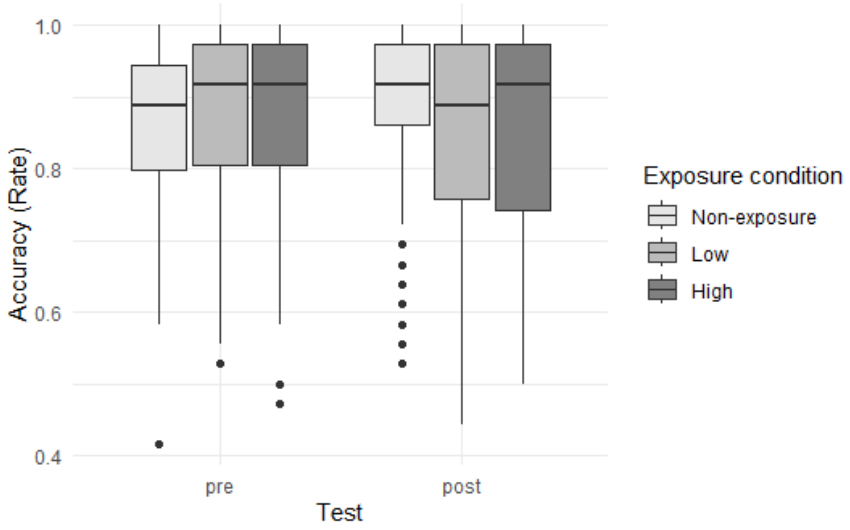


Figure 5.3: Proportion Correct in Different Exposure Conditions in Pre- and Post-test

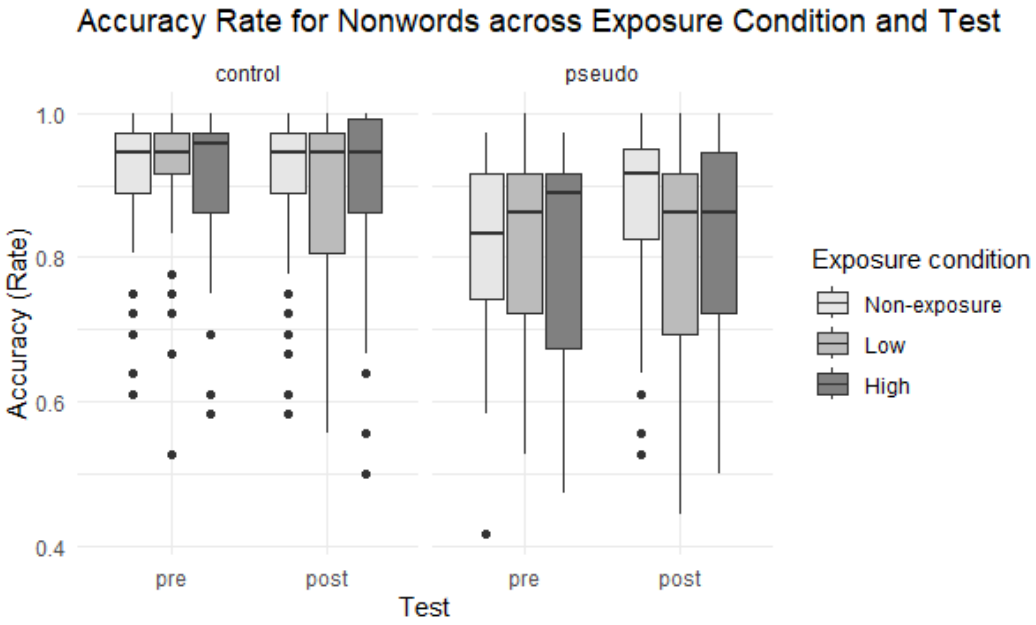


Figure 5.4: Pre- and Post- test Performance across Exposure Conditions for Different Nonword Types

5.3.2 RTs

Following pre-registered criteria, RTs for incorrect or timed-out trials (> 3000ms) trials were excluded (2898 trials, 13.69%; 1438 [13.59%] in pre-test and 1460 [13.79%] in post-test) along with RTs below200ms (218 trials, 1.19% of the correct trials). Outliers were identified following

the procedure outlined by Baayen and Milin (2010): a base model with only random effects was fitted for log-transformed RT and data points with absolute standardised residuals exceeding 2.5 SD were considered as outliers and were trimmed (in total 658 trials, 3.11%). This was carried out separately for pre- and post-tests to capture potential variability at both time points, as we are interested in changes in RT distributions from pre-test to post-test. Figure 5.5 shows performance for each suffix, allowing for an inspection of whether the overall pattern was particularly dominated by responses to one individual suffix. RTs across the three exposure conditions did not differ significantly at the pre-test (*p*s > .50). Descriptive data are shown in Table 5.6.

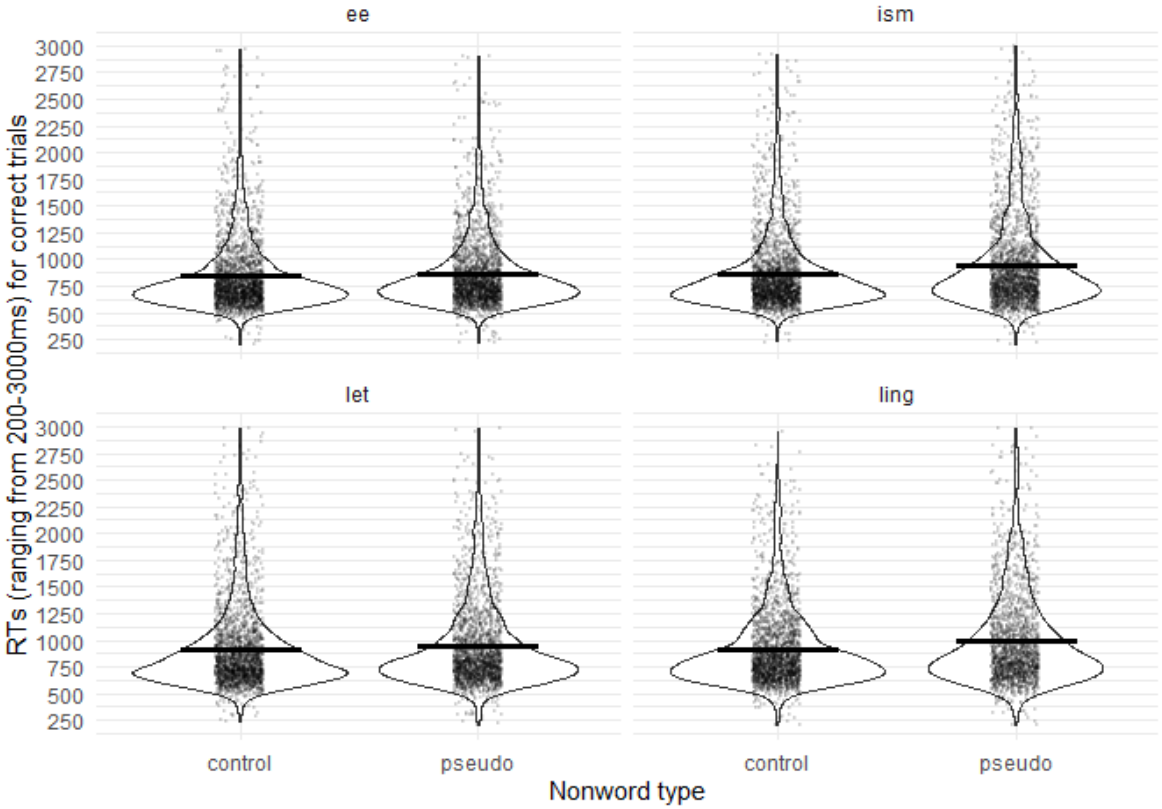


Figure 5.5: RTs (pre- and post-tests combined) for Pseudomorphemic and Control Nonwords Grouped by Suffix (the lines represent the means)

Table 5.6: Means (SDs) of Raw RTs (ms) for Pseudomorphemic and Control Nonwords.

Test / Exposure condition	Pre-test		Post-test	
	Pseudo	*Control*	*Pseudo*	*Control*
High diversity	963.03 (402.63)	920.69 (370.54)	857.38 (360.43)	814.47 (324.98)
Low diversity	1002.35 (410.55)	954.66 (404.80)	843.31 (338.91)	799.15 (300.47)
Control	1029.94 (454.78)	949.18 (413.98)	885.56 (365.90)	839.73 (355.28)

Note. RTs for correct responses were calculated after data trimming as described above.

The main effects and interactions of the primary variables in the final model: $RT \leftarrow lmer(RT\ log-transformed\ \sim Nonword\ type\ \times Test \times Exposure\ condition\ + Reading\ ability + Age\ in\ month\ +\ (1 + Test + Nonword\ type\ |\ Participant) +\ (1 + Test\ |\ Item))$ are summarised in Table 5.7 (See Appendix 5F for detailed model summary).

Table 5.7: RT Model Output for the Main Effects

Fixed effects	χ^2	*p value*
(Intercept)	268941.95	< **.001**
Nonword type	31.97	< **.001**
Test	132.91	**<.001**
Exposure Condition	1.47	.48
Nonword type× Test	1.85	.17
Nonword type× Exposure Condition	2.63	.27
Test× Exposure Condition	3.90	.14
Nonword type× Exposure Condition× Test	1.29	.52

The main effects of Nonword type (β = 0.054, t = 5.65, p < .001) and Test (β = -0.15, t = -11.53, p < .001) were significant. These findings suggested that the pseudomorphemic nonwords were rejected more slowly than the control nonwords (ΔRT = 49.40ms), indicating an overall morpheme interference effect. Participants responded faster overall in the post-test than in the pre-

test (ΔRT = 129.89ms), suggesting a potential practice effect (see Appendix 5G for detailed pairwise comparisons). No other effects were significant.

5.4 Discussion

Experiment 4 aimed to explore the impact of intensive exposure to low-frequency suffixes on morphological processing by adolescents. The morpheme interference paradigm was used for both the pre-exposure and post-exposure tests, aiming to build a causal link between suffix familiarity and the morpheme interference effect. My first hypothesis was that enhanced exposure to suffixes would trigger or increase the morpheme interference effect because increased familiarity with these suffixes would make it harder to reject pseudomorphemic nonwords that contained them. My second hypothesis was that the diversity of stems to which these suffixes were attached could further facilitate suffix acquisition during the exposure phase (i.e., higher stem diversity leading to more robust learning). However, the lack of a statistically significant three-way interaction involving Nonword type, Test, and Exposure condition indicated that the data did not fully support either hypothesis.

The morpheme interference effect on both accuracy and RTs was unexpectedly prominent at pre-test, given that the suffixes had lower frequencies than those which did not exhibit significant morpheme interference in Chapter 4. This suggests that additional factors beyond frequency might have played a role. Previous studies point to factors like the productivity (Lázaro et al., 2015) and diagnosticity (Ulicheva et al., 2020) of suffixes, along with imageability (Lázaro et al., 2023) and semantic interpretability (Burani et al., 1999) of whole pseudowords, as potential influencers on performance. For instance, according to Ulicheva et al. (2020), the four suffixes used in the current study, namely ‘-ee’, ‘-let’, ‘-ism’, and ‘-ling’, are high in noun diagnosticity and specificity, meaning that they can efficiently predict lexical categories and tend to be the preferred options for a lexical category for a certain sound pattern. Moreover, these suffixes often represent stable semantic functions. For example, ‘-ee’ typically denotes a person in a certain state or a recipient of an action,

making them ‘non-homonymous’, as described by studies like Bertram et al (2000). Bertram et al (2000) suggested that words with productive and non-homonymous suffixes tended to facilitate morpheme-based processing due to reduced processing ambiguity. Furthermore, the four suffixes in the current study could easily attach to various stems to form novel words that are interpretable and imageable— for instance, the novel word ‘textee’ has been used on social media as it can be easily understood as someone receiving a text. This is also true for nonwords like ‘rocklet’ used in the current study. These characteristics highlighted above might trigger more pronounced morphological processing, resulting in significant morpheme interference, despite the relative infrequency of the suffixes. The unexpected result of the morpheme interference effect observed with the low-frequency suffixes in the current study invites further investigation to clarify the influence of other properties of suffixes, beyond just frequency, on modulating the morpheme interference effect.

In addition, interpreting the two two-way interactions is challenging as they did not involve the third factor that was crucial to demonstrate differential change in response to the two types of nonwords (i.e., pseudomorphemic vs. control) between pre- and post- exposure. The results indicated that mean accuracy in the two exposure conditions increased less from the pre-test to post-test compared to the non-exposure condition where participants were not exposed to the target suffixes. This interaction might hint at an exposure effect: the increase in accuracy could be attributed to a repeated-measures practice effect (to be discussed later in this section); however, this relatively smaller increase in the exposure conditions might be due to the fact that exposure to the target suffixes could have made these suffixes more interfering (i.e., leading to less accurate responses to pseudomorphemic nonwords containing them), thereby offsetting the increase in accuracy associated with practice. This might partially be attributed to the interplay between increased familiarity due to exposure and the practice effect which could have led to a smaller increase in accuracy in the exposure conditions. However, this interpretation is speculative and

potentially overly ambitious. Nonword type was not a factor in this interaction, and thus caution is advised in drawing conclusions from it. One reason for the lack of significant three-way interaction could be that one single-session training was insufficient to cause meaningful changes to the sensitivity to the target suffixes—a point further discussed below. It is also possible that the experiment design lacked sufficient power to detect an effect (that might also be too small to be of practical significance). The actual effect size of the three-way interaction might be much smaller than the one (β = 0.03) used in the per-experiment power simulation for determining sample size. Additionally, a larger number of items in each condition (i.e., more data points) might be necessary to better account for variability within the data caused by other unexpected underlying factors mentioned above. In Chapter 4 the lexical decision task included 12 suffixes attached to eight stems each, but in Chapter 5, only four suffixes were used due to practical issues such as challenges in finding eligible suffixes (which will be discussed later), and shortening the time required to complete each task to ensure the experiment would not be overly lengthy.

Another interaction showing that the difference in accuracy across two nonword types was reduced from pre- to post-test may appear counter-intuitive because an increase in morpheme interference was expected after the exposure phase. However, as the exposure condition was not involved in this interaction —indicating the inclusion of data from participants not exposed to the suffixes—this makes the interpretation less straightforward. Inspection of the change from pre- to post-tests in mean accuracy for each type of nonword across exposure conditions (Table 5.4) shows that the narrowed difference between accuracy for pseudomorphemic vs. control nonwords appears to be largely driven by a significant increase in response accuracy to the pseudomorphemic nonwords in the post-test. This increase might stem from repeated assessment effects, with participants becoming more adept at making judgements about the pseudomorphemic nonwords, independent of the influence of exposure. This may be because the lexical decision tasks, having consistent settings for both pre- and post-tests (despite utilising different counterbalanced stimulus

lists), could have led to lexical decision strategies during the experiment. Children might have identified the recurring patterns of the 'stem + suffix', which aided their responses in the post-test. Thus, instead of demonstrating the influence of suffix familiarity due to exposure (if any), their response might be more influenced by explicit processes such as task-specific strategies arising from the close implementation of the pre- and post-tests. They might have strategically recognised those patterns as linked with nonwords in making lexical judgements rather than being interfered with by the familiar suffixes. This interpretation is supported by the timepoint-related effects observed solely in accuracy, not in RTs, as the former might be more related indicative of explicit processes, whereas the latter reflects more implicit automatic processing (cf. Dawson et al., 2018).

These results question the assumption that a single intensive exposure session could adequately increase suffix familiarity in the same way that real-world suffix experience might. Such an effect, if it exists, might require exposure to suffixes over time and multiple sessions (cf. Bowey & Muller, 2005; Nation et al., 2007). This also challenges the effectiveness of implementing pre- and post-tests in the same session. One solution might be to include real words containing those suffixes (the word fillers used in this Experiment 4 did not contain the same suffixes that comprise the experimental nonwords) to prevent associating these suffixes with nonwords, though this might lead to unnecessary exposure to suffixes embedded in real words during the pre-test. Another relevant point is the 'temporal spacing effect', which suggests that items appearing once in each block across multiple blocks are learned better than the same items learned in a 'massed' manner (i.e., read for the same number of times but on consecutive trials) (Wegener et al., 2023). In this Chapter 5 experiment, each suffix was presented in a massed format within each story (with each story containing the same suffix embedded in eight words) during the exposure phase, potentially preventing children from effectively gaining sensitivity to these suffixes. In addition, the changes between pre- and post- tests appeared less significant for control nonwords, making it harder to detect any changes in the morpheme interference effect. Why might this be the case? One reason

could be the already high accuracy (above 90%) in responding to control nonwords in the pre-test compared to their pseudomorphemic counterparts, indicating a ceiling effect. Additionally, unlike real suffixed patterns to which children might already have weak orthographic representations, potentially leading them to recognise them during the lexical decision task and then link them with a 'nonword' decision over trials, and thus lead the suffixed nonwords to be more likely be influenced repeated-measures practice effect, novel orthographic control patterns (e.g., '-lang' for -ling) were harder to recognise. This might make the control nonwords less susceptible to possible practice effects. Such differences in the potential impact of repeated measures could introduce further unexpected variations to the data.

Therefore, the unexpected significant morpheme interference effect at pre-test, potentially attributed to the choice of suffixes, might limit the potential for a notable increase in morpheme interference post-exposure. Additionally, the variability in data stemming potentially from repeated-measures effects suggests that any potential change in morpheme interference might be too subtle to outweigh the other underlying influences. Again, this interpretation remains tentative and warrants further empirical investigation. Reflecting on the selected suffixes for the current study, besides the limitations discussed above, the similarities in semantic function between '-ling' and '-let', as well as the fact that the suffix -let is also a word, might have introduced additional confounds. For example, Crepaldi et al. (2010) avoided selecting suffixes that were homographic with existing words (e.g., -able). Additionally, the choice of 'invariant' suffixes to keep the stems unaltered was based on the assumption that intact stems would better support the parsing and learning of these suffixes, as highlighted in my systematic review (Chapter 2, regarding morphological structure integrity). However, finding suffixes that are both 'invariant' and embedded in enough prevalent words for exposure proved challenging, thus limiting the suffix options available. This challenge, coupled with their multifaceted properties affecting morphological processing in varied ways, suggests that using real suffixes might not have been the optimal approach for exploring the

exposure effects. Despite the initial discussions suggesting that results obtained from real suffixes are more generalisable to natural reading and language acquisition processes than entirely novel and artificial items, artificial language paradigms may indeed offer a potential solution to isolate the frequency (exposure) effects from those attributable to other properties.

In addition to the linguistic factors mentioned above, another potential limitation affecting the results is the quality of the data. We had to exclude several participants due to potential uncooperative behaviour and below-chance test scores. This issue is not uncommon in studies involving young participants. Al-Salom and Miller (2019) found that online research can lead to less valid data, especially in certain populations, such as undergraduate students. Their study showed that over 10% of undergraduate participants—similar in age to those in my study—provided data that was likely invalid, particularly in responses to computer-based surveys. Several factors can affect the quality and validity of data, especially when testing adolescents in groups. Demographic factors such as age can increase the likelihood of careless responses, leading to higher rates of invalid data; issues like noncompliance (i.e., not following instructions) and boredom are particularly common among student volunteers. As a result, research involving these participants may be biased or limited without proper validity testing to account for potential noncompliance (see Al-Salom & Miller, 2019; DeRight & Jorgensen, 2015; for a review). Participants might exhibit performance that does not accurately reflect their true abilities (Green & Merten, 2013). This could be the case even for those who did not score below chance in the lexical decision task in the current study.

To address these issues, researchers have proposed strategies such as attention or consistency checks (e.g., Abbey & Meloy, 2017; Aust et al., 2013) and instructional manipulation checks (e.g., Oppenheimer et al., 2009) to ensure full participant engagement. Additionally, fostering a sense of being ‘individually monitored’ could encourage more serious participation, which is inspired by Burnham and Hare (2007), who found that the perception of being watched—such as by an image

of a robot—can positively influence behaviour, such as increasing contributions to the public good. If I were to redesign my experimental tasks, it would be beneficial to incorporate various validity checks or create the perception that participants are being closely and individually monitored, either by the researcher or by an icon on the screen during the task. These measures, together with careful control and design of the stimuli taking into consideration the psycholinguistic factors mentioned above, might help mitigate the issues identified in the current study.

In conclusion, the current data did not support the hypothesis that a single session of intensified suffix exposure would amplify the morpheme interference effects of these suffixes. It is possible that beyond suffix familiarity, various factors such as suffix homonymity, diagnosticity, and pseudoword interpretability could have shaped morphological processing. Therefore, more rigorous control of the stimuli and targeted analysis are crucial to disentangle the impact of each specific aspect from the rest. In addition, extending the interval between pre- and post-tests might help provide clearer insights into the effects of training or exposure. It is also possible that the intensive exposures may differ greatly from the rich and prolonged suffix experience accumulated through years of reading in a natural context. Such 'suffix booster' exposure sessions might fail to allow readers the chance to consolidate and cultivate mental representations of these suffixes. Yet, due to existing confounds and limitations, it is possible that the suffix exposure could have marginally increased suffix familiarity, yet any resulting changes in morpheme interference owing to increased suffix familiarity following exposure were too subtle, and thus overshadowed by other effects. As such, no definitive conclusions can be drawn from the current study. In spite of this, on the positive side, the unexpected outcomes, including the null results, present numerous avenues for future investigation which I will return to in the General Discussion.

Chapter 6. General discussion

6.1 Summary of findings

Reading is essential both for individuals and for humanity as a whole. As a key component of language, morphology plays a critical role in reading. Morphologically complex words, which constitute the majority in alphabetic language systems, are prevalent in texts encountered by developing readers (e.g., Dawson et al., 2023; Grainger & Ziegler, 2011). This thesis reviews and investigates how morphological knowledge is implicitly and automatically activated during word reading, and how morphological processing is acquired and develops over the course of adolescence. A developmental perspective was adopted by examining participants of different ages using cross-sectional designs, and individual differences were also investigated.

Chapter 2 presented a systematic review on morphological processing in developing readers from ages 6 to 18 across Indo-European languages. This review synthesised three major themes concerning the modulation of morphological processing: first, at the item level—specially, the influence of morphological structure, morpheme type, and their distributional properties; second, at the individual level—namely, the impacts of years of reading instruction, reading ability and deficits; and third, the variations across different paradigms and tasks. Factors extracted at both the item and individual levels are either directly linked or indirectly relevant to reading experience, i.e., how frequently morphemes are encountered in a language and how a reader engages with them, which in turn shapes their morphological processing. Regarding task factors, the specific aspects or stages of processing measured by various tasks may vary. For instance, the ‘reading aloud’ task relies on phonological ability, whereas lexical decision tasks do not. The masked priming paradigm, where primes are usually subliminally presented, often taps the initial stage of morphological processing whereas the performance of un-primed lexical decision tasks may reveal both automatic and rapid morphological processing (often reflected in RTs) and later stages of judgement that may involve some explicit morphological strategy (which typically affects accuracy).

Experiment 1 (Chapter 3) asked whether the morphological and pseudomorphological priming effects observed in masked priming studies with adults in lab (e.g., Beyersmann et al., 2012) replicate when data collection is remote via the internet-based Gorilla platform online. Although the morphological priming effect was replicated, the pseudomorphological priming was not. A posthoc power simulation indicated that a substantially larger number of participants would be needed to detect the effect (N= 560). The lack of replication of the pseudomorphological priming effect might stem from the noise and variability in online data collection, which might arise from several sources, including hardware and software differences, delays in stimulus presentation, and a less stringent control over participant attention in remote settings. Pseudomorphological priming usually has smaller effect sizes compared to morphological priming (e.g., Beyersmann et al., 2012) and may be particularly sensitive to timing inaccuracies in the display of primes. For instance, extending the display duration risks conscious perception of pseudomorphological primes by participants, which can potentially undermine any pseudomorphological priming effect due to the semantic irrelevance between the prime to the target (e.g., 'corner-CORN'). However, this issue is less likely to affect morphological priming, which might even benefit from conscious perception of the primes. The results of Experiment 1 highlight the challenges of conducting experiments that require precise timing online. The lack of significant pseudomorphological priming and the potential need for a large sample size suggest using this paradigm through remote testing may not be suitable for exploring my further research questions, such as those investigating factors that influence morphological processing, particularly among developing readers. Nevertheless, the successful replication of morphological priming suggests that Gorilla remains an effective option for conducting experiments that do not require highly precise timing or those that typically generate larger effect sizes. Alternative paradigms like the un-primed lexical decision task might offer more robust results in remote testing environments.

Chapter 4 explored how individual-level and item-level proxies of reading experience—specifically reading ability and suffix familiarity, influenced the morpheme interference effect in both adults (Experiment 2) and developing readers aged 11-17 years (Experiment 3). Reading ability was assessed using the ROAR (Yeatman, 2021), and familiarity with suffixes was captured by their frequency in a corpus of children's reading materials (OCC; Banerji, et al., 2013). Both adults and adolescents showed a morpheme interference effect: they found it more challenging to reject pseudomorphemic nonwords like 'birdly' compared to closely matched control nonwords like 'birdla'. This replicates the morpheme interference effect previously observed in in-person studies (e.g., in adults: Crepaldi et al., 2010; in developing readers: Dawson et al., 2018) and in doing so, it provides evidence for form-based morpho-orthographic decomposition during visual word recognition. Importantly, the results show that the morpheme interference is more pronounced (larger difference in responding to pseudomorphemic nonwords compared to responding to the control nonwords) when the nonwords contain more familiar suffixes, and among participants with higher reading abilities. Both of these observations underscore that reading experience facilitates morpho-orthographic processing. Especially, in adolescents, the effect of reading ability (or something correlated with reading ability) on the morpheme interference effect was further modulated by suffix familiarity (i.e., significant three-way interactions between nonword type, suffix familiarity, and reading ability). The influence of reading ability on the morpheme interference effect became more pronounced in nonwords containing higher familiarity suffixes. This will be discussed in more detail in Section 6.1.2.

Chapter 5 (Experiment 4) investigated whether inducing familiarity by increasing encounters with low-frequency suffixes in one exposure session affects morphological decomposition in developing readers aged 11-15 years (Year 7- Year10). This study used the morpheme interference paradigm to assess whether intensified exposure to four suffixes, embedded within different words in stories, would increase the morpheme interference effect from pre- to post-test. These suffixes

were selected because they were less frequent than those which did not yield morpheme interference in Experiment 3. The aim of Experiment 4 was therefore to establish a causal relationship between suffix familiarity and morpheme interference. Stem diversity (i.e., the number of different words containing the target suffixes) was also manipulated to explore whether a more diverse linguistic context would enhance suffix learning. Despite the hypotheses that the exposure session would enhance familiarity with the suffixes and consequently make them more 'interfering', with stem diversity aiding learning, the data did not support these expectations—there was no significant three-way interaction involving Nonword type, Test (pre vs. post), and Exposure conditions (non-exposure control vs. low stem-diversity vs. high stem-diversity). Unexpectedly, there was already a morpheme interference effect at pre-test, and this might have restricted the potential for observable changes, induced by training. Additionally, it was also possible that the amount or nature of the exposure was insufficient to enhance sensitivity to these suffixes, implying that factors beyond mere exposure affect suffix learning—a point that will be discussed later. The results underscore the complexity of acquiring sublexical units like suffixes and suggest that other stimulus characteristics and external factors should also be considered in investigating suffix acquisition.

6.1.1 Reading ability and development of morphological processing

As reviewed in Chapter 2, morphological processing takes time to develop, and individual differences in reading ability modulate morphological processing. An increase in reading ability could be associated with higher lexical quality and more stable orthographic representations of morphemes (Perfetti, 2007). In turn, this could enhance sensitivity to morphological structures during reading, and allow for developing readers to rapidly and automatically activate sublexical units via morpho-orthographic processing, similar to skilled readers. For example, Dawson et al. (2021) found that larger pseudo-morphological priming based on surface morphological structure ('corner-CORN') compared to non-morphological form priming ('mother – MOTH') became more

pronounced as a function of increased reading ability. Hasenäcker et al. (2017) reported that children with stronger vocabulary knowledge benefitted from morphology in both compounds and derived words over simple ones, compared to their peers with weaker vocabulary who found reading complex words challenging. The findings in Experiments 2 and 3, as summarised above, support this premise, finding that even in skilled adult readers, morpheme interference on RTs was larger for those individuals with higher reading ability. This suggests greater automatic morpho-orthographic processing is associated with increases in reading ability. This is consistent with the lexical legacy hypothesis (Nation, 2017), which suggests that even in skilled adults, the ease of processing a word is a product of accumulated lexical experience.

Importantly, when analysing the RT data in Experiment 3 with developing readers, although both reading ability and age group were positively associated with stronger morpheme interference, when both factors were included in the same model, only reading ability predicted interference – age was no longer a significant factor. This suggests that reading ability influences sensitivity to morphological structures beyond age or grade, and potentially, it accounts for at least part of the effects of age. Similar findings were reported by Traficante et al. (2018) who analysed children's ability to read suffixed words and found that incorporating reading proficiency (calculated using factor analysis on word and nonword reading speed) improved model fit, and rendered the influence of grade insignificant. Therefore, although reading ability generally increases with age and grade, it could be a stronger predictor of morphological processing, presumably due to large individual differences in reading ability within each year group or age band (e.g., Hasenäcker et al., 2020). Although in this thesis reading ability was mainly treated as an individual-level indicator or 'product' of reading experience, it could also be seen as a contributor to reading experience in a broader sense. For instance, akin to the 'Matthew effect' of reading—which suggests that children with strong early reading abilities tend to engage more frequently in reading—strong readers with a greater sensitivity to morphological regularities are better equipped to activate and utilise these

regularities during reading (see review in Chapter 2). This proficiency allows them to further refine the orthographic patterns of morphological units in their mental lexicon.

Turning to the accuracy data, age group also significantly influenced the magnitude of the morpheme interference effect, alongside reading ability. This could stem from developing readers using morphological strategies to infer the meanings of unknown items, leading to more misjudgements of pseudomorphemic nonwords (e.g., Dawson et al., 2018; Hasenäcker, Solaja et al., 2021). Such explicit knowledge about suffixes is typically acquired through years of schooling (correlated with age) as well as reading. Moreover, the finding that reading ability, rather than age, influences RTs suggests that reading proficiency is more directly linked to the speed and automaticity of activating morphological knowledge during processing. Traficante et al. (2018) also noted that while reading ability reflects reading-related skills and implicit learning, age or grade is more closely associated with explicit learning. Similarly, other research, such as Hasenäcker et al. (2020), found that age/grade and reading ability influenced morphological processing in different ways (see Chapter 2).

Since reading ability does not uniformly increase with years of schooling across all individuals, age is not an appropriate proxy for reading expertise. It is important to consider both factors when investigating morphological processing (e.g., Hasenäcker et al., 2020). This may also partially explain the difference between my finding that adolescents from 11- 17 years year showed significant morpheme interference effects on both RTs and accuracy (Experiment 3) and the findings of Dawson et al. (2018), who did not observe this effect among their younger 12-13-year-old adolescents. As Dawson et al. (2018) did not account for reading ability in their models, it is possible that proficient or even average readers in their younger age group might already have demonstrated some morpho-orthographic processing, albeit not fully adult-like.

In sum, higher reading ability is generally associated with fine-tuned orthographic representations which facilitate morpho-orthographic processing that requires such representations

of morphemes. Therefore, reading ability could be a stronger predictor than age in the development of automatic and rapid morphological processing. However, this might not be the complete picture as the item-level reading experience (suffix familiarity) was also involved in the interaction of reading ability and morpheme interference effect, as demonstrated in Experiment 3. This underscores the need to consider both factors collectively. Additionally, computational simulations by Reichle and Perfetti (2003) showed that both availability and familiarity influence base word activation, with the former aiding in the ease of retrieval, serving as a proxy lexical quality, and both familiarity and availability being influenced by token frequencies. It could be argued that similar mechanisms might apply to sublexical units like suffixes.

6.1.2 Revisiting the frequency effect

In Experiments 2, 3, and 4, presented across Chapters 3 and 4, suffix familiarity was proxied by their token frequency in children's reading material, as estimated by the Oxford Children's Corpus (Banerji, et al., 2013). Token frequency reflects the overall prevalence of a suffix, indicating how frequently a suffix is encountered in children's reading materials. This seems suitable for assessing suffix familiarity as expertise in reading suffixes could develop from cumulative exposure to them in diverse texts (Dawson et al., 2021; Nation, 2017). Results from Experiments 2 and 3 showed that higher familiarity suffixes were more interfering. This is likely due to the ease of activating their orthographic representations during visual word recognition, compared to the less familiar suffixes. The trend indicated that for lower-frequency suffixes, there was no difference in responses (both RTs and accuracy) to pseudomorphemic nonwords compared to those in the control condition, suggesting that these suffixes were processed similarly to non-morphemic letter strings (i.e., no morpheme interference effect). Experiment 3 with developing readers revealed a morpheme interference effect on RT only for nonwords that contained medium-to-high familiarity suffixes (i.e., a two-way interaction of Nonword type and Suffix familiarity). Although increased suffix familiarity led to larger morpheme interference effect across all reading ability levels, the effect

became more pronounced with higher reading ability only in the case of familiar suffixes (i.e., a three-way interaction of Nonword type, reading ability, and Suffix familiarity). These findings highlight that regardless of their level of reading proficiency, people need sufficient exposure to specific suffixes in print. This exposure helps them become familiar with the suffixes, acquire their orthographic patterns and establish refined representations in the mental lexicon. This is further supported by the notion that there are item-level variations in lexical quality due to the frequency of encounters with these words— although the 'average' lexical quality in the mental lexicon of an individual improves with their general reading experience, within an individual the quality of specific items varies: frequently encountered items are of high quality, whereas less frequent items possess lower quality (Castles et al., 2018).

The hypothesis that morpho-orthographic processing may be partially driven by increasing suffix familiarity needs to be empirical validation. To explore this relationship, controlling for word reading ability, Experiment 4 temporarily increased the 'frequency' (i.e., the opportunity of encounters) of four low-frequency suffixes by having participants read stories containing words with these suffixes. However, there was no change in the morphemic interference effect as a function of exposure. Although seeing a word more frequently is expected to adjust its recognition threshold, making it more easily processed (e.g., Brysbaert et al., 2018), the absence of a training effect prompts a reconsideration of what frequency captures beyond mere times of exposure. Despite the limitations of this study, which will be detailed in the limitations section, some key questions regarding the frequency of exposure need to be considered.

First, higher frequency of exposure is likely to be associated with experience in different forms across various contexts, encompassing both linguistic environment (e.g., semantic diversity) and contextual diversity. Variations in both factors influence lexical quality and word reading (e.g., Adelman et al., 2006; Hsiao & Nation, 2018; Plummer et al., 2014). In Experiment 4, linguistic diversity was manipulated by presenting a suffix in either two or eight words. Each suffix was also

exposed in two different stories, providing a degree of contextual diversity (see Appendix 5B). However, this level of diversity might not have been sufficient to trigger a change in suffix learning.

Second, learning through frequent, multiple encounters with lexical items typically occurs over time rather than consecutively. Unlike in Experiment 4, each suffix appeared eight times within a single story in a 'massed' exposure manner in a single session, which did not yield significant training effects. Critical factors that influence the acquisition of sensitivities to trained items include learning through multiple sessions (Bowey & Muller, 2005), intervals between learning and testing (e.g., Tamminen et al., 2015), and the temporal spacing (e.g., learning an item multiple times in one block or across several blocks, as in Wegener et al., 2022). Tamminen et al. (2015) observed training effects in explicit judgments about novel items immediately after training, whereas implicit activation and the generalisation of the newly learned items required a period of time for memory consolidation. This might partially explain why, in Experiment 4, a slightly larger trend of changes was observed in accuracy (albeit still insignificant) but not in RTs.

In summary, frequency as indexed by token frequency in a corpus reflects encounters with lexical items throughout reading experience. This is likely to reflect on learning in a way that is slow and cumulative, rather than 'all or none', as measured by short-term training experiments. The frequency of an item in a corpus integrates the co-occurrence of its diverse linguistic features (e.g., semantic diversity, productivity) shaped by experience (Nation, 2017). Therefore, frequency quantifies and qualifies extensive and varied interactions with lexical items, through which the mental representations of these items are refined over time. This highlights the primary challenge and major difference between the prolonged effects of reading experience and temporary, short-term boosts in frequency. The latter cannot fully capture the complexity of the 'lexical legacy' that cumulative frequency cultivates (e.g., Castles et al., 2018; Nation, 2017). Despite these challenges, integrating carefully controlled training studies with corpus analysis could offer valuable insights into the frequency effects on morphological processing. Additionally, although not directly

examined in my experiments, type frequency is another important index worth considering alongside token frequency. While some might assume that type and token frequency are statistically indistinguishable, the distinction between these two types of frequency has been a key aspect of linguistics (cf. Berg, 2014). Furthermore, the computational model outlined by Reichle and Perfetti (2003), though primarily focused on stems rather than suffixes, provides useful insights into understanding experience with sublexical units like affixes. Their findings suggest that repeated exposure to stems across multiple inflected forms increases familiarity (as measured by token frequency), while type frequency (the diversity of derived forms) is more important for the availability and retrieval of phonological and semantic information. Therefore, although token frequency was selected for my study as a more relevant index of familiarity, investigating both frequency indices, individually and in combination, might offer additional insights.

6.2 Implications

The systematic review presented in Chapter 2 is the first to synthesise findings on morphological processing of written language in developing readers across Indo-European languages. Despite the diversity in research designs and heterogeneity of findings, this review clearly demonstrates that morphological processing is associated with the development of skilled reading. It indicates that adult-like morpho-orthographic decomposition, which relies on refined representations of sublexical units like suffixes, develops through experience with these units. This underscores the importance of considering both individual-level and item-level factors and employing varied analytical perspectives to fully understand morphological processing. A detailed discussion of theoretical significance is provided in the review in Chapter 2. My review also highlighted adolescence as an under-studied period in reading research, supporting the argument made by Ricketts et al. (2020). Moreover, only studies from 10 Indo-European languages met the inclusion criteria. In addition to potential limitations in the literature search, which will be discussed later, this limited scope might also result from research conducted on other Indo-European

languages not being reported in English. While this does not necessarily indicate a complete absence of studies, it does suggest that research on morphological processing in developing readers is relatively scarce in languages such as Swedish, Russian, and Greek.

While not the central focus of this thesis, an important point that emerged during the review is that skilled readers can effectively activate both the whole word and its constituent morphemes during reading, utilising this information flexibly depending on the task demand. On the other hand, 'poor readers', despite general processing limitations, also benefit from morphological structure, albeit in a slightly different manner. For example, skilled readers were found to benefit from morphological structure when naming low-frequency derived words, but not when whole-word frequency was high enough to allow direct activation of the whole word. In contrast, their peers with poorer reading abilities consistently processed derived words faster than simple ones (Marcolini et al., 2011). This pattern indicates that skilled readers take the whole-word pathway where it is more efficient, especially in transparent orthographies, yet they also remain flexible in benefitting from morphological structure with unfamiliar words. On the other hand, children with limited reading abilities rely more on morpheme-based strategies, as whole-word processing is more challenging for them. They benefit from morphological structure, as breaking down complex words into smaller morphological units facilitates processing. This echoes the triangle model of reading (Harm & Seidenberg, 2004): for those children with relatively weak association between phonology and orthography (i.e., O-P; the phonological pathway), they might rely on the semantic pathway to aid their reading. In this pathway, orthography is directly mapped onto semantics, which then activates word pronunciation (i.e., O-S-P). This insight is relevant both for researchers and particularly educators: explicit instruction on morphology in teaching could be crucial not only for its demonstrated role in reading development but also as a tool to aid children with poor reading skills. These children could compensate for their phonological processing difficulties to some extent, and effectively process long derived words, approaching that of their more skilled peers.

Experiments 3 and 4 involving English-speaking adolescents provided clear evidence that from early adolescence (11 years) onwards, students can implicitly activate morphological knowledge while reading words and nonwords. In addition to increased reading proficiency relative to younger children, adolescents have typically been exposed to a significantly higher volume of morphologically complex words. These complex words make up a large proportion of the vocabulary in their reading materials, compared to those targeted at younger children as shown in a study analysing reading materials for 5- to 14-year-olds (Dawson et al., 2023). The findings demonstrate that morphological processing skills continue to develop throughout mid-to-late adolescence and vary depending on both individuals' reading ability and how frequently suffixes are encountered. During upper secondary education, morphological processing ability becomes increasingly important as curriculum content in higher grades demands greater comprehension of academic vocabulary, which is often morphologically complex. Again, the findings contribute to the ongoing investigation into the mechanisms underlying the development of morphological processing across a wide range of adolescent years (11-17 years). While this age group has begun to receive more attention (e.g., Dawson et al., 2018, 2021), it remains relatively neglected in the field.

Turning to the theoretical accounts explaining morphological processing, the results of Experiments 3 and 4 provide strong evidence to contradict the supralexical account of morphological processing (e.g., Giraudo & Grainger, 2001) which claims that morphological decomposition occurs after the whole word representation is activated. My experiments used nonword stimuli and as these lack whole-word representations in the mental lexicon, they cannot be activated before the activation of their constituent morphemes. Although it was not an aim of the thesis to distinguish different theoretical accounts, the results could be explained by the localist sub-lexical account (e.g., Rastle & Davis, 2008) which suggests that morphologically complex items are automatically decomposed into morphemes based on the apparent morphological structure, prior to lexical access. The results also do not contradict the dual-route model which posits that both the

whole word and its constituent morphemes can be activated during processing (e.g., Baayen & Schreuder, 2000). Although Experiments 3 and 4 did not adopt word stimuli that could be activated as whole-word representations, some findings reviewed in Chapter 2 provide evidence supporting the occurrence of both decomposition and whole-word access (e.g., Marcolini et al., 2011, as mentioned earlier in this section). In addition, while this thesis primarily attributes the automatic processing of complex words to high-quality morpho-orthographic representations, the association of morphological processing development with increasing familiarity and higher reading ability can also be explained by distributed models. These models propose that morphological sensitivity emerges as the language processing system becomes more finely tuned to statistical regularities encountered throughout language experience, without requiring explicit morpheme representations (see Stevens & Plaut, 2022, for a review). Despite varying perspectives on whether morphemes are explicitly represented as meaningful sublexical units, there is a consensus among these theories that, morphological processing is underpinned by 'learned' sensitivity to regularities, which is cultivated through extensive reading experience.

The online data collection methods used in this thesis demonstrate the viability of conducting some psycholinguistic experiments via remote online testing. Despite challenges such as noisier data and less precise timing, detailed in Chapter 3 (Experiment 1) and summarised in Section 6.1, my experience of collecting data online has demonstrated that online testing is not only a practical choice during the pandemic but also a viable option for future research (cf. Rodd, 2024). While larger sample sizes are often required to mitigate variability in data collected remotely, the online method allowed for recruiting and testing large numbers of participants more easily. Additionally, online recruitment supports a more demographically diverse participant pool than traditional lab-based studies, broadening the demographic scope and facilitating research in varied settings. This could enhance the generalisability of findings (Majid, 2023). For instance, many psychology experiments often rely on a homogeneous group of undergraduate students as their adult

participants but Experiments 1 and 2 in the thesis collected data from participants aged 18-40 years. Furthermore, online platforms like Gorilla provide secure online storage, offering robust version control for experiments and securely storing all experiment data online. This helps enhance data protection and data access by eliminating the need for local storage. It also facilitates sharing experiment tasks and data securely with researchers in the same project, thereby enhancing open science and research collaboration. However, remote testing requires carefully designed experiments with clear instructions and additional practice trials to ensure participants can easily understand and complete the tasks on their own. Detailed instructions should include specific actions, such as maximising the browser window, to standardise the testing environment across participants. Overall, while online testing introduced certain challenges to the experiments in the thesis, it benefitted them in terms of accessibility, diversity, and efficiency. The methodologies and outcomes of this thesis could offer some references for future research practices in related fields.

6.3 Limitations and future directions

For the review in Chapter 2, there were two deviations from the pre-registered procedure (see Chapter 2 for details). Initially, it was planned for both reviewers to screen 25% of all retrieved studies separately, but due to the substantial volume of literature extracted from the original search, it was not feasible to double-screen 25% of all. Therefore, the primary reviewer conducted the first round of screening alone. Additionally, the inter-rater agreement was 70%, below the originally planned 80%. However, all disagreements were resolved through discussion between the two reviewers, and a third expert in the field was consulted for further guidance. Moreover, conducting a meta-analysis was not feasible due to the heterogeneity of paradigms and the small sample sizes within each of those paradigms. Overall, these limitations illustrate the challenges in quantifying and defining experimental effects, such as morphological processing, via systematic reviews and meta-analyses, compared to the approach taken to synthesise intervention studies. Despite these challenges, the review identifies several promising future research directions. For instance, as

previously discussed, exposure to lexical items develops through frequent encounters, making print exposure a promising individual-level factor to be investigated. Additionally, expanding research to include more cross-linguistic and longitudinal studies, particularly with adolescent participants, would be beneficial.

In addition to the limitations discussed in previous chapters, a key constraint of this thesis was the exclusive use of nonword stimuli to test morphological processing. Given the potential differences in processing words and nonwords (e.g., Laxon et al., 1992), this approach may limit insights into factors influencing the morphological processing of real words, such as semantics and whole-word frequency. Furthermore, items were presented in isolation in the experimental tasks measuring morphological processing, whereas real-world reading typically occurs within contexts, such as meaningful sentences. Future research should explore morphological processing by embedding target words within sentential contexts. Such studies often require real-time tracking of reading, potentially employing techniques like eye-tracking (e.g., Traficante et al., 2018). A promising behavioural paradigm is the 'flankers' task, which has only recently been used to study morphological processing. The task involves making lexical decisions about a word in the middle flanked by two others (e.g., 'corn corner corn'; Grainger et al., 2021). This approach moves closer to 'mimicking' the reading of words in a multiword context.

In addition, the sample size for each age group in Experiment 3 was unbalanced, potentially introducing sampling biases. This limitation may have restricted the ability of Experiment 3 to demonstrate a clear age trend by analysing age as a continuous variable in the exploratory analysis. This underscores the need for large and evenly distributed sample sizes across age groups to ensure data integrity. Moreover, logistical challenges during the COVID-19 pandemic have limited the feasibility of conducting research with children. Future studies should aim to include a broader age range of participants to examine factors modulating morphological processing across various stages of development.

6.4 Conclusion

Taken together, the systematic review and the findings of the experiments shed light on factors modulating morphological processing in visual word recognition among children (the review) and adolescents. It is evident that sensitivity to morphological regularities is cultivated through extensive reading experience, characterised by both individual reading ability and item-level experience with morphological regularities in the reading input. These aspects are essential for the development of refined, high-quality morpheme representations in the mental lexicon, which can be rapidly, effortlessly, and automatically activated during processing. This high quality of sublexical representations facilitates the transition from novice to expert (e.g., Castles et al., 2018). While some key questions remain unanswered, e.g., the direct causal link between suffix familiarity and the development of morpho-orthographic processing, the training study remains a promising approach to enhance our understanding of the mechanism driving the development of morphological processing. By taking a developmental approach, the thesis adds evidence to a growing body of studies investigating how morphological processing is shaped by accumulated reading experience.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. Journal of Operations Management, 53, 63-70.

Adelman, J. S., Brown, G. D. A. & Quesada, J. F. Contextual diversity, not word frequency, determines word-naming and lexical decision times. Psychol. Sci. 17, 814–823 (2006).

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., ... & Davis, C. J. (2014). A behavioral database for masked form priming. Behavior research methods, 46, 1052-1067.

Al-Salom, P., & Miller, C. J. (2019). The problem with online data collection: Predicting invalid responding in undergraduate samples. Current Psychology, 38, 1258-1264.

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*. https://doi.org/10.3389/fpsyg.2012.00232

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. https://doi.org/10.1016/j.jml.2012.12.001

*__Angelelli, P., Marinelli, C. V., & Burani, C. (2014). The effect of morphology on spelling and reading accuracy: a study on Italian children. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01373__

*__Angelelli, P., Marinelli, C. V., De Salvatore, M., & Burani, C. (2017). Morpheme-based Reading and Spelling in Italian Children with Developmental Dyslexia and Dysorthography. *Dyslexia*, *23*(4), 387–405. https://doi.org/10.1002/dys.1554__

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods*, *53*(4), 1407-1425.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Online participants in the wild: Realistic precision & accuracy of platforms, web-browsers, and devices. *PsyArxiv Preprints*.

Apel, K. (2014). A Comprehensive Definition of Morphological Awareness. *Topics in Language Disorders*, *34*(3), 197–209. https://doi.org/10.1097/tld.0000000000000019

Apel, K., & Henbest, V. S. (2020). Morphological Awareness Skills of Second- and Third-Grade Students With and Without Speech Sound Disorders. *Language, Speech, and Hearing Services in Schools*, *51*(3), 603–616. https://doi.org/10.1044/2019_lshss-19-00045

***Assink, E. M. H., Vooijs, C., & Knuijt, P. P. N. A. (2000). Prefixes as access units in visual word recognition: A comparison of Italian and Dutch data. *Reading and Writing, 12*(3/4), 149–168. https://doi.org/10.1023/a:1008179825696**

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. Behavior research methods, 45, 527-535.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International journal of psychological research, 3*(2), 12-28. https://doi.org/10.21500/20112084.807

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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.

Baguley, T., & Kaye, W. (2010). Review of: Understanding psychology as a science: An introduction to scientific and statistical inference, by z. dienes. *British Journal of Mathematical and Statistical Psychology*, 63 (3), 695–698.

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. https://doi.org/10.3758/bf03193014

Banerji, N., Gupta, V., Kilgarriff, A., & Tugwell, D. (2013). Oxford Children's Corpus: A corpus of children's writing, reading, and education. *Corpus Linguistics 2013*, 315-317.

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. https://doi.org/10.1016/j.jml.2012.11.001

Bauer, L. (2001). *Morphological productivity* (Vol. 95). Cambridge University Press.

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2011). The development of cortical sensitivity to visual word forms. *Journal of cognitive neuroscience*, *23*(9), 2387-2399.

Berg, T. (2014). On the relationship between type and token frequency. Journal of Quantitative Linguistics, 21(3), 199-222.

Berko, J. (1958). The Child's Learning of English Morphology. *WORD*, *14*(2-3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

***Berthiaume, R., & Daigle, D. (2014a). Are Dyslexic Children Sensitive to the Morphological Structure of Words When They Read? The Case of Dyslexic Readers of French. *Dyslexia*, *20*(3), 241–260. https://doi.org/10.1002/dys.1476**

***Berthiaume, R., & Daigle, D. (2014b). Morphological Processing and Learning to Read: The Case of Deaf Children. *Deafness & Education International*, *16*(4), 185–203. https://doi.org/10.1179/1557069x14y.0000000036**

Bertram, R., Baayen, R. Harald., & Schreuder, R. (2000). Effects of Family Size for Complex Words. *Journal of Memory and Language*, *42*(3), 390–405. https://doi.org/10.1006/jmla.1999.2681

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *Journal of experimental psychology: Learning, memory, and cognition*, 26(2), 489.

Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review*, *22*(4), 1054–1061. https://doi.org/10.3758/s13423-014-0752-9

***Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *Quarterly Journal of Experimental Psychology*, *65*(7), 1306–1326. https://doi.org/10.1080/17470218.2012.656661**

Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded Stem Priming Effects in Prefixed and Suffixed Pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230. https://doi.org/10.1080/10888438.2016.1140769

Beyersmann, E., Grainger, J., & Castles, A. (2019). Embedded stems as a bootstrapping mechanism for morphological parsing during reading development. *Journal of

***Experimental Child Psychology*, *182*, 196–210. https://doi.org/10.1016/j.jecp.2019.01.010**

***Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology*, *139*, 115–126. https://doi.org/10.1016/j.jecp.2015.06.001**

***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*, 105140. https://doi.org/10.1016/j.jecp.2021.105140**

Beyersmann, E., Ziegler, J. C., & Grainger, J. (2015). Differences in the processing of prefixes and suffixes revealed by a letter-search task. Scientific Studies of Reading, 19(5), 360-373.

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

Blevins, J. P. (2006). English inflection and derivation. *The handbook of English linguistics*, 507-536.

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology*, *52*(5), 704-716. doi:10.1037/dev0000111

***Bowey, J. A. (1996). Grammatical Priming of Visual Word Recognition in Fourth grade Children. *The Quarterly Journal of Experimental Psychology*, *49*(4), 1005–1023. https://doi.org/10.1080/713755675**

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of experimental child psychology*, 92(3), 203-219.

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), 389–411. https://doi.org/10.1037/met0000159

***Breadmore, H. L., & Carroll, J. M. (2016). Effects of Orthographic, Morphological and**

Semantic Overlap on Short-Term Memory for Words in Typical and Atypical Development. *Scientific Studies of Reading*, *20*(6), 471–489. https://doi.org/10.1080/10888438.2016.1246554

Breadmore, H. L., & Carroll, J. M. (2016). Morphological spelling in spite of phonological deficits: Evidence from children with dyslexia and otitis media. *Applied Psycholinguistics*, *37*(6), 1439-1460.

*__Breadmore, H. L., & Carroll, J. M. (2017). Sublexical and syntactic processing during reading: evidence from eye movements of typically developing and dyslexic readers. *Journal of Cognitive Psychology*, *30*(2), 177–197. https://doi.org/10.1080/20445911.2017.1414222__

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. Current directions in psychological science, 27(1), 45-50.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2018). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Bulut, T. (2022). Neural correlates of morphological processing: An activation likelihood estimation meta-analysis. *Cortex*. https://doi.org/10.1016/j.cortex.2022.02.010

Burani, C., & Thornton, A. M. (2003). Morphological Structure in Language Processing, 157–208. https://doi.org/10.1515/9783110910186.157

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., Dovetto, F. M., Thornton, A. M., & Alessandro Laudanna. (1997). Accessing and naming suffixed pseudo-words. Accessing and naming suffixed pseudo-words. In *Yearbook of morphology 1996*, Geert E. Booij and Jaap van Marle (eds.), 55– 72. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3718-0_5

*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. https://doi.org/10.1006/brln.2001.2548

*Burani, C., Marcolini, S., De Luca, M., & Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition*, *108*(1), 243–262. https://doi.org/10.1016/j.cognition.2007.12.010

*Burani, C., Marcolini, S., Traficante, D., & Zoccolotti, P. (2018). Reading Derived Words by Italian Children With and Without Dyslexia: The Effect of Root Length. *Frontiers in Psychology*, *9*(9). https://doi.org/10.3389/fpsyg.2018.00647

Burani, C., Thornton, A. M., & Caramazza, A. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. Morphological Structure in Language Processing. *Visible language*, 18(4), 157–208. https://doi.org/10.1515/9783110910186.157

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425-455.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332. https://doi.org/10.1016/0010-0277(88)90017-0

Carlisle, J. F (1995). Morphological awareness and early reading achievement. In Feldman, L. B. (Eds.), Morphological Aspects of Language Processing (pp. 189-209). Psychology Press.

*Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, *12*(3/4), 169–190. https://doi.org/10.1023/a:1008131926604

*Carlisle, J. F., & C. Addison Stone. (2003). The Effects of Morphological Structure on Children's Reading of Derived Words in English. *Springer EBooks*, 27–52. https://doi.org/10.1007/978-1-4757-3720-2_2

Carlisle, J. F., & Feldman, L. B. (1995). Morphological awareness and early reading achievement. *Morphological aspects of language processing*, pp.*189-209*.

Carlisle, J. F., & Fleming, J. (2003). Lexical Processing of Morphologically Complex Words in the Elementary Years. *Scientific Studies of Reading*, *7*(3), 239–253. https://doi.org/10.1207/s1532799xssr0703_3

*Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing*, *19*(7), 669–693. https://doi.org/10.1007/s11145-

005-5766-2

Carlisle, J. F., & Stone, C. A. (2003). The effects of morphological structure on children's reading of derived words in English. In Reading complex words: Cross-language studies (pp. 27-52). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-3720-2_2

***Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, *40*(4), 428–449. https://doi.org/10.1598/rrq.40.4.3**

***Carlisle, J. F., Stone, C. A., & Katz, L. A. (2001). The effects of phonological transparency on reading derived words. *Annals of Dyslexia*, *51*(1), 249–274. https://doi.org/10.1007/s11881-001-0013-2**

***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. https://doi.org/10.1348/026151008x389575**

***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*. https://doi.org/10.3389/fpsyg.2015.00452**

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. Psychological science in the public interest, 19(1), 5-51. https://doi.org/10.1177/1529100618772271

***Cauchi, C., Beyersmann, E., Lété, B., & Grainger, J. (2022). A developmental perspective on morphological processing in the flankers task. *Journal of Experimental Child Psychology*, *221*, 105448. https://doi.org/10.1016/j.jecp.2022.105448**

Clahsen , H., & Fleischhauer, E. (2013). Morphological priming in child German. *Journal of Child Language*, *41*(6), 1305–1333. https://doi.org/10.1017/s0305000913000494

***COLÉ, P., BOUTON, S., LEUWERS, C., CASALIS, S., & SPRENGER-CHAROLLES, L. (2011). Stem and derivational-suffix processing during reading by French second and third graders. *Applied Psycholinguistics*, *33*(1), 97–120. https://doi.org/10.1017/s0142716411000282**

COLE, P., MAREC-BRETON, N., ROYER, C., & GOMBERT, J. E. (2003). Morphologie des mots et apprentissage de la lecture. *Morphologie Des Mots et Apprentissage de La Lecture*,

41(213), 57–76. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=14727600

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Condry, S. M. (1979). A developmental study of processes of word derivation in elementary school children and their relation to reading (Doctoral dissertation, Cornell University).

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition*, *38*(3), 312–321. https://doi.org/10.3758/mc.38.3.312

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). "Fell" primes "fall", but does "bell" prime "ball"? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, *63*(1), 83–99. https://doi.org/10.1016/j.jml.2010.03.002

***D'Alessio , M. J., Wilson, M. A., & Jaichenco, V. (2019). Morphological De-com-pos-it-ion Helps Recognize Low-er Frequency Words in Typically Developing Spanish-Speaking Children. *Journal of Psycholinguistic Research*, *48*(6), 1407–1428. https://doi.org/10.1007/s10936-019-09665-8**

***D'ALESSIO, M. J., JAICHENCO, V., & WILSON, M. A. (2018). The role of morphology in word naming in Spanish-speaking children. *Applied Psycholinguistics*, *39*(5), 1065–1093. https://doi.org/10.1017/s0142716418000127**

Daniel, M. (2015). *aimswebPlus (Technical manual)*. Bloomington, MN: Pearson. Retrieved from https://www.marshfieldschools.org/cms/lib/WI01919828/Centricity/Domain/82/Plus%20Technical%20Manual.pdf

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. https://doi.org/10.3758/bf03206399

Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2023). Effects of target age and genre on morphological complexity in children's reading material. *Scientific Studies of Reading*, *27*(6), 529-556. https://doi.org/10.1080/10888438.2023.2206574

***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. https://doi.org/10.1037/xlm0000485**

***Dawson, N., Rastle, K., & Ricketts, J. (2021a). Finding the man amongst many: A**

developmental perspective on mechanisms of morphological decomposition. *Cognition, 211*, 104605. https://doi.org/10.1016/j.cognition.2021.104605

Dawson, N., Rastle, K., & Ricketts, J. (2021b). Bridging form and meaning: support from derivational suffixes in word learning. *Journal of Research in Reading*, *44*(1), 27–50. https://doi.org/10.1111/1467-9817.12338

De Jong, N. H., Schreuder, R., & Harald Baayen, R. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*(4-5), 329–365. https://doi.org/10.1080/016909600501196 25

De Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific studies of reading*, *11*(1), 55-71.

DeRight, J., & Jorgensen, R. S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. The Clinical N*f*europsychologist, 29(1), 101-117.

De Rosa, M., & Crepaldi, D. (2022). Letter chunk frequency does not explain morphological masked priming. *Psychonomic bulletin & review, 29(*2*)*, 589-599. https://doi.org/10.3758/s13423-021-02010-y

***de Zeeuw (2013). Chapter 4: Processing derivational morphology in first and second language reading acquisition: Evidence from the identification of prefixed and pseudoprefixed words in Morphological processing in first and second language reading acquisition [Doctoral Dissertation]. Retrieved from https://repository.ubn.ru.nl/bitstream/handle/2066/107675/107675.pdf**

***de Zeeuw, M., Schreuder, R., & Verhoeven, L. (2013). Processing of Regular and Irregular Past-Tense Verb Forms in First and Second Language Reading Acquisition. *Language Learning, 63*(4), 740–765. https://doi.org/10.1111/lang.12023**

***de Zeeuw, M., Schreuder, R., & Verhoeven, L. T. W. (2014). Lexical processing of nominal compounds in first- and second-language learners across primary grades. *Writing Systems Research, 7*(2), 133–156. https://doi.org/10.1080/17586801.2014.926806**

***Deacon, S. H., & Francis, K. A. (2017). How Children Become Sensitive to the Morphological Structure of the Words That They Read. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01469**

***Deacon, S. H., Campbell, E., Tamminga, M., & Kirby, J. (2010). Seeing the harm in harmed**

and harmful: Morphological processing by children in Grades 4, 6, and 8. *Applied Psycholinguistics, 31*(4), 759–775. https://doi.org/10.1017/s0142716410000238

Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The Relation Between Morphological Awareness and Reading Comprehension: Evidence From Mediation and Longitudinal Models. *Scientific Studies of Reading, 18*(6), 432–451. https://doi.org/10.1080/10888438.2014.926907

Deacon, S. H., Parrila, R., & Kirby, J. R. (2008). *A review of the evidence on morphological processing in dyslexics and poor readers: a strength or weakness?* (G. Reid, A. J. Fawcett, F. Manis, & L. S. Siegel, Eds.). The SAGE handbook of dyslexia (pp. 212–237). SAGE Publications.

***Deacon, S. H., Tong, X., & Francis, K. (2015). The relationship of morphological analysis and morphological decoding to reading comprehension. *Journal of Research in Reading, 40*(1), 1–16. https://doi.org/10.1111/1467-9817.12056**

***Deacon, S. H., Whalen, R., & Kirby, J. R. (2011). Do children see the danger in dangerous? Grade 4, 6, and 8 children's reading of morphologically complex words. *Applied Psycholinguistics, 32*(3), 467–481. https://doi.org/10.1017/s0142716411000166**

***Deilen, S., Hansen-Schirra, S., & Nagels, A. (2022). The Effect of Segmentation Signs in Compounds on Reading Behavior: an Eye-Tracking Study. *Lebende Sprachen, 67*(1), 45–99. https://doi.org/10.1515/les-2022-0003**

Dienes, Z. (2008). Understanding psychology as a science: An introduction to scientific and statistical inference. Palgrave Macmillan.

Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational Morphology and Skilled Reading. *The Cambridge Handbook of Psycholinguistics*, 311–332. https://doi.org/10.1017/cbo9781139029377.016

Diependaele, K., Morris, J., Serota, R. M., Bertrand, D., & Grainger, J. (2013). Breaking boundaries: Letter transpositions and morphological processing. *Language and Cognitive Processes, 28*(7), 988–1003. https://doi.org/10.1080/01690965.2012.719082

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

Domínguez Martínez, A., Cuetos Vega, F., & Segui, J. (2000). Morphological processing in word recognition: A review with particular reference to Spanish data. *Psicológica*, *21*. https://digibuo.uniovi.es/dspace/handle/10651/33897

***Drewnowski, A. (1981). Missing -ing in reading: Developmental changes in reading units. *Journal of Experimental Child Psychology*, *31*(1), 154–168. https://doi.org/10.1016/0022-0965(81)90009-6**

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental science*, *13*(2), 331-338.

***Duncan, L. G., Gray, E., Quémart, P., & Casalis, S. (2011). Do good and poor readers make use of morphemic structure in English word recognition? *Journal of Portuguese Linguistics*, *10*(1), 143. https://doi.org/10.5334/jpl.104**

***Duranovic, M., Gangl, M., Finke, S., Smajlagic, S., & Landerl, K. (2020). Development of morphological priming effects in reading aloud in the biscriptal Bosnian orthography. *Reading and Writing*. https://doi.org/10.1007/s11145-020-10036-2**

***Egan, J., & Pring, L. (2004). The Processing of Inflectional Morphology: A Comparison of Children with and without Dyslexia. *Reading and Writing*, *17*(6), 567–591. https://doi.org/10.1023/b:read.0000044433.30864.23**

***Elbro, C., & Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Annals of Dyslexia*, 46(1), 209–240. https://doi.org/10.1007/bf02648177**

***Feldman, L. B., Rueckl, J. G., Diliberto, K. A., Matthew John Pastizzo, & Vellutino, F. R. (2002). Morphological analysis by child readers as revealed by the fragment completion task. *Psychonomic Bulletin & Review*, *9*(3), 529–535. https://doi.org/10.3758/bf03196309**

***Fleischhauer (2013). Morphological processing in children: an experimental study of German past participles- Chapter 4. [Doctoral Dissertation]. Retrieved from https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/6807/file/fleischhauer_diss.pdf**

***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*, *44*(1), 90–109.**

https://doi.org/10.1111/1467-9817.12340

Fortson IV, B. W. (2011). *Indo-European language and culture: An introduction*. John Wiley & Sons.

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: *Sage*. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox, J., & Weisberg, S. (2019). Using car functions in other functions. *CRAN R*.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. Perspectives on psychological science, 9(6), 641-651.

Gil-López, C., Perea, M., Moret-Tatay, C., & Carreiras, M. (2011). Can masked priming effects be obtained with handwritten words?. Attention, Perception, & Psychophysics, 73, 1643-1649.

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review*, *8*(1), 127–131. https://doi.org/10.3758/bf03196148

Giraudo, H., & Voga, M. (2014). Measuring morphology : the tip of the iceberg? A retrospective on 10 years of morphological processing. *Carnets de Grammaire*, *22*, 136. https://hal.science/hal-01919128/

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. https://doi.org/10.1037/0096-3445.136.2.323

Goodwin, A. P., Petscher, Y., Carlisle, J. F., & Mitchell, A. M. (2017). Exploring the Dimensionality of Morphological Knowledge for Adolescent Readers. *Journal of Research in Reading*, *40*(1), 91–117. https://doi.org/10.1111/1467-9817.12064

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

Grainger, J., & Beyersmann, E. (2017, January 1). *Chapter Nine - Edge-Aligned Embedded Word Activation Initiates Morpho-orthographic Segmentation* (B. H. Ross, Ed.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/abs/pii/S0079742117300208

Grainger, J., & Ziegler, J. C. (2011). A Dual-Route Approach to Orthographic Processing. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00054

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210x.12504

Hamawand, Z. (2011). Morphology in English: Word formation in cognitive grammar. Bloomsbury Publishing.

***Hasenäcker, J., & Schroeder, S. (2016). Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, *38*(3), 733–753. https://doi.org/10.1017/s0142716416000412**

***Hasenäcker, J., & Schroeder, S. (2019). Compound reading in German: Effects of constituent frequency and whole-word frequency in children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(5), 920–933. https://doi.org/10.1037/xlm0000623**

***Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2016). Masked Morphological Priming in German-Speaking Adults and Children: Evidence from Response Time Distributions. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00929**

***Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2020). Morphological Priming in Children: Disentangling the Effects of School-grade and Reading Skill. *Scientific Studies of Reading*, 1–16. https://doi.org/10.1080/10888438.2020.1729768**

***Hasenäcker, J., Ktori, M., & Crepaldi, D. (2021). Morpheme Position Coding in Reading Development as Explored With a Letter Search Task. *Journal of Cognition*, *4*(1), 16. https://doi.org/10.5334/joc.153**

***Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1093–1108. https://doi.org/10.1037/xlm0000353**

Hasenäcker, J., Solaja, O., & Davide Crepaldi. (2019). Food in the corner and money in the cashews: Semantic activation of embedded stems in the presence or absence of a morphological structure. *Psychonomic Bulletin & Review*, *27*(1), 155–161. https://doi.org/10.3758/s13423-019-01664-z

*Hasenäcker, J., Solaja, O., & Davide Crepaldi. (2021). Does morphological structure modulate access to embedded word meaning in child readers? *Memory & Cognition, 49*(7), 1334–1347. https://doi.org/10.3758/s13421-021-01164-3

Hayden, E., Hiebert, E. H., & Trainin, G. (2019). Patterns of Silent Reading Rate and Comprehension as a Function of Developmental Status, Genre, and Text Position. *Reading Psychology*, *40*(8), 731–767. https://doi.org/10.1080/02702711.2019.1673602

*Hess, S., Mousikou, P., & Schroeder, S. (2021). Morphological processing in developmental handwriting production: evidence from kinematics. *Reading and Writing, 35*(4), 899–917. https://doi.org/10.1007/s11145-021-10204-y

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19-24.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2012). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. Journal of Memory and Language, 103, 114-126.

Jarvella, R.J. & Wennstedt, O. (1993). Recognition of partial regularity in words and sentences. *Scandinavian Journal of Psychology*, *34*(1), 76–85. http://doi:https://doi.org/10.1111/j.1467-9450.1993.tb01102.x.

Jeffreys, H. (1939). The theory of probability. Oxford University Press, Oxford

*Jöbstl, V., Kargl, R., Prattes, A. E., Beyersmann, E., & Landerl, K. (2021). Effects of a Morpheme-based Spelling Intervention Challenging Previous Results. *Lnternational Electronic Journal of Elementary Education, 13*(5), 651–671. https://doi.org/10.26822/iejee.2021.219

Joseph, H. S., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, *42*(2), 207-222.

*Kearns, D. M. (2015). How elementary-age children read polysyllabic polymorphemic words. *Journal of Educational Psychology*, 107(2), 364–390. https://doi.org/10.1037/a0037518

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2011). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

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. https://doi.org/10.3758/s13428-011-0118-4

Kruk, R. S., & Bergman, K. (2013). The reciprocal relations between morphological processes and reading. *Journal of Experimental Child Psychology*, *114*(1), 10–34. https://doi.org/10.1016/j.jecp.2012.09.014

Kuo, L., & Anderson, R. C. (2006). Morphological Awareness and Learning to Read: A Cross-Language Perspective. *Educational Psychologist*, *41*(3), 161–180. https://doi.org/10.1207/s15326985ep4103_3

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

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. https://doi.org/10.18637/jss.v082.i13

*Law, J. M., & Ghesquière, P. (2021). Morphological Processing in Children with Developmental Dyslexia: A Visual Masked Priming Study. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.450

Laws, J., & Ryder, C. (2014). Getting the measure of derivational morphology in adult speech a corpus analysis using MorphoQuantics. *Language Studies Working Papers*, *6*, 3–17. https://centaur.reading.ac.uk/38479/

*Laxon, V., Rickard, M., & Coltheart, V. (1992). Children read affixed words and non-words. *British Journal of Psychology*, *83*(3), 407–423. https://doi.org/10.1111/j.2044-8295.1992.tb02450.x

*Lázaro, M., Acha, J., Rosa, Seila García, & Sainz, J. (2016). Exploring the derivative suffix

frequency effect in Spanish speaking children. *Reading and Writing*, *30*(1), 163–185. https://doi.org/10.1007/s11145-016-9668-2

***Lázaro, M., Camacho, L., & Burani, C. (2013). Morphological Processing in Reading Disabled and Skilled Spanish Children. *Dyslexia*, *19*(3), 178–188. https://doi.org/10.1002/dys.1458**

***Lázaro, M., García, L., & Burani, C. (2015). How orthographic transparency affects morphological processing in young readers with and without reading disability. *Scandinavian Journal of Psychology*, *56*(5), 498–507. https://doi.org/10.1111/sjop.12213**

Lázaro, M., García-Gutiérrez, A., García, L., & Hinojosa, J. A. (2023). Pupillary responses to pseudowords with different morphological and imageability features. *Journal of Psychophysiology*.

***Lázaro, M., Illera, V., Acha, J., Escalonilla, A., García, S., & Sainz, J. S. (2018). Morphological effects in word identification: tracking the developmental trajectory of derivational suffixes in Spanish. *Reading and Writing*, *31*(7), 1669–1684. https://doi.org/10.1007/s11145-018-9858-1**

Lázaro, M., Sainz, J., & Illera, V. (2015). The Role of Derivative Suffix Productivity in the Visual Word Recognition of Complex Words. *Psicologica: International Journal of Methodology and Experimental Psychology*, *36*(2), 165–184. https://eric.ed.gov/?id=EJ1068799

Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex*, *116*, 4–44. https://doi.org/10.1016/j.cortex.2018.08.016

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package 'emmeans'.

***Leong, C. K. (1989). The effects of morphological structure on reading proficiency?A developmental study. *Reading and Writing*, *1*(4), 357–379. https://doi.org/10.1007/bf00386267**

***Leong, C. K. (2000). Rapid Processing of Base and Derived Forms of Words and Grades 4, 5, and 6 Children's Spelling. *Reading and Writing: An Interdisciplinary Journal*, *12*, 277–302. https://eric.ed.gov/?id=EJ606259**

***Leong, C. K., & Parkinson, M. E. (1992). Sensitivity to orthotactic rules in visual word recognition by below average readers. *Reading and Writing*, *4*(1), 1–17. https://doi.org/10.1007/bf01027069**

*Leong, C. K., & Parkinson, M. E. (1995). Processing of English morphological structure by poor readers. In C. K. Leong & R. M. Joshi (Eds.), Developmental and acquired dyslexia: *Neuropsychological and neurolinguistic perspectives* (pp. 237–261). Dordrecht: Springer Netherlands. http://ndl.ethernet.edu.et/bitstream/123456789/7216/1/809.pdf#page=233

*Lukatela, G., Carello, C., & Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes*, *2*(1), 1–17. https://doi.org/10.1080/01690968708406349

*Madziak, R. (2021). Morphological segmenting skills in reading-impaired children: An ERP study. [Unpublished master's thesis], Retrieved from https://mspace.lib.umanitoba.ca/server/api/core/bitstreams/15c27a11-ade6-48d3-9e2f-ee5bb7458001/content

Majid, A. (2023). Establishing psychological universals. *Nature Reviews Psychology*, *2*(4), 199-200.

Mak, M. H., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. Journal of Memory and Language, 118, 104203.

*Mann, V., & Singson, M. (2003). Linking Morphological Knowledge to English Decoding Ability: Large Effects of Little Suffixes. *Reading Complex Words*, 1–25. https://doi.org/10.1007/978-1-4757-3720-2_1

*MARCOLINI, S., TRAFICANTE, D., ZOCCOLOTTI, P., & BURANI, C. (2011). Word frequency modulates morpheme-based reading in poor and skilled Italian readers. *Applied Psycholinguistics*, *32*(3), 513–532. https://doi.org/10.1017/s0142716411000191

Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 823–836. https://doi.org/10.1098/rstb.2007.2091

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

Massol, S., Grainger, J., Dufau, S., & Holcomb, P. (2010). Masked priming from orthographic neighbors: an ERP investigation. Journal of Experimental Psychology: Human Perception and Performance, 36(1), 162.

*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. https://www.jstor.org/stable/25655464

McKinnon, R., Allen, M., & Osterhout, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *NeuroReport, 14*(6), 883. https://journals.lww.com/neuroreport/Fulltext/2003/05060/Morphological_decomposition_involving.22.aspx

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological bulletin*, *138*(2), 322.

Montgomery, J. W., & Leonard, L. B. (1998). Real-Time Inflectional Processing by Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1432–1443. https://doi.org/10.1044/jslhr.4106.1432

Moret-Tatay, C., & Perea, M. (2011). Is the go/no-go lexical decision task preferable to the yes/no task with developing readers? *Journal of Experimental Child Psychology, 110*(1), 125–132. https://doi.org/10.1016/j.jecp.2011.04.005

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. https://doi.org/10.1080/01690965.2010.495482

*Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., Grainger, J., & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science, 23*(6). https://doi.org/10.1111/desc.12952

Muncer, S. J., Knight, D., & Adams, J. W. (2013a). Bigram Frequency, Number of Syllables and Morphemes and Their Effects on Lexical Decision and Word Naming. *Journal of Psycholinguistic Research, 43*(3), 241–254. https://doi.org/10.1007/s10936-013-9252-8

Muncer, S. J., Knight, D. C., & Adams, J. W. (2013b). Lexical decision and the number of morphemes and affixes. Scandinavian Journal of Psychology, 54(5), 349–352. https://doi.org/10.1111/sjop.12064

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological Knowledge and Literacy Acquisition. *Journal of Learning Disabilities*, *47*(1), 3–12. https://doi.org/10.1177/0022219413509967

Nation, K. (2009). Form–meaning links in the development of visual word recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3665-3674.

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*(1). https://doi.org/10.1038/s41539-017-0004-7

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of experimental child psychology*, *96*(1), 71-84.

***Oliveira, B. S. F. de, & Justi, F. R. dos R. (2017). Morphological priming development in Brazilian Portuguese-speaking children. *Psicologia: Reflexão E Crítica*, *30*(1). https://doi.org/10.1186/s41155-017-0058-8**

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of experimental social psychology, 45(4), 867-872.

Pacton, S., Foulin, J. N., Casalis, S., & Treiman, R. (2013). Children benefit from morphological relatedness when they learn to spell new words. *Frontiers in psychology*, *4*, 696.

Pennington, B. F., & Bishop, D. V. M. (2009). Relations Among Speech, Language, and Reading Disorders. *Annual Review of Psychology*, *60*(1), 283–306. https://doi.org/10.1146/annurev.psych.60.110707.163548

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. Scientific Studies of Reading, 11(4), 357–383. https://doi.org/10.1080/10888430701530730

***Pescuma, V. N., Ktori, M., Beyersmann, E., Sowman, P. F., Castles, A., & Crepaldi, D. (2022). Automatic morpheme identification across development: Magnetoencephalography (MEG) evidence from fast periodic visual stimulation. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.932952**

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. https://doi.org/10.1080/01690960050119661

Plummer, P., Perea, M. & Rayner, K. The influence of contextual diversity on eye movements in reading. J. Exp. Psychol. -Learn. 40, 275–283 (2014).

Protopapas, A., Mitsi, A., Koustoumbardis, M., Tsitsopoulou, S. M., Leventi, M., & Seitz, A. R. (2017). Incidental orthographic learning during a color detection task. *Cognition, 166*, 251-271.

***Quémart, P., & Casalis, S. (2014). Effects of Phonological and Orthographic Shifts on Children's Processing of Written Morphology: A Time-Course Study. *Scientific Studies of Reading*, *18*(5), 363–382. https://doi.org/10.1080/10888438.2014.912218**

***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. https://doi.org/10.1017/s014271641300026x**

***Quémart, P., & Lambert, E. (2019). The influence of the morphological structure of words on the dynamics of handwriting in adults and fourth and sixth grade children. *Reading and Writing*, *32*(1), 175–195. https://doi.org/10.1007/s11145-017-9762-0**

***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. https://doi.org/10.1016/j.jecp.2011.02.008**

***Quémart, P., Casalis, S., & Duncan, L. G. (2012). Exploring the Role of Bases and Suffixes When Reading Familiar and Unfamiliar Words: Evidence From French Young Readers. *Scientific Studies of Reading*, *16*(5), 424–442. https://doi.org/10.1080/10888438.2011.584333**

Quémart, P., Gonnerman, L. M., Downing, J., & Deacon, S. H. (2017). The development of morphological representations in young readers: a cross-modal priming study. *Developmental Science*, *21*(4), e12607. https://doi.org/10.1111/desc.12607

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for

***Rabin, J., & Deacon, H. (2008). The representation of morphologically complex words in the developing lexicon*. *Journal of Child Language*, *35*(2), 453–465. https://doi.org/10.1017/s0305000907008525**

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. https://doi.org/10.1016/j.cortex.2018.02.008

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. https://doi.org/10.1016/j.cortex.2018.02.008

Rastle, K. (2022). Word Recognition III. In: M.J. Snowling, C. Hulme and K. Nation, eds., *The Science of Reading: The Handbook, Second Edition*. pp.102–119. https://doi:https://doi.org/10.1002/9781119705116.ch5

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7-8), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/bf03196742

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in Word Identification: A Word-Experience Model That Accounts for Morpheme Frequency Effects. Scientific Studies of Reading, 7(3), 219–237. https://doi.org/10.1207/s1532799xssr0703_2

Richard D. Morey & Jeffrey N. Rouder (2022). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.4. https://CRAN.R-project.org/package=BayesFactor

Ricketts, J., Lervåg, A., Dawson, N., Taylor, L. A., & Hulme, C. (2020). Reading and oral vocabulary development in early adolescence. *Scientific Studies of Reading*, *24*(5), 380-396.

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language*, *134*, 104472.

***Rosenberg, L., & Kruk, R. S. (2021). Coarse or fine? Grain size and morpho-orthographic segmentation in struggling readers. *Annals of Dyslexia*. https://doi.org/10.1007/s11881-021-00240-2**

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Rueckl, J. (2010). Connectionism and the role of morphology in visual word recognition. *The Mental Lexicon*, *5*(3), 371–400. https://doi.org/10.1075/ml.5.3.07rue

***Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing*, *21*(5), 481–504. https://doi.org/10.1007/s11145-007-9074-x**

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, Hosting and Recruiting: A Brief Introduction to Running Behavioral Experiments Online. *Brain Sciences*, *10*(4), 251. https://doi.org/10.3390/brainsci10040251

Savinova, E., & Malyutina, S. (2021). Evidence for dual-route morphological processing across the lifespan: data from Russian noun plurals. *Language, Cognition and Neuroscience*, *36*(6), 730–745. https://doi.org/10.1080/23273798.2021.1879182

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schmalz, X., Biurrun Manresa, J., & Zhang, L. (2021). What is a Bayes factor?. *Psychological methods*.

Segbers, J., & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*, *34*(3), 297-320.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current directions in psychological science*, *14*(5), 238-242.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151-218.

Silvey, C., Dienes, Z., & Wonnacott, E. (2021, December 3). Bayes factors for logistic (mixed effect) models. https://doi.org/10.31234/osf.io/m4hju

Sinclair, Michael. (2006). Multiple Linear Regression. Retrieved from: https://rpubs.com/CPEL/multlinearregression

Smith, P. T. (1995). Are morphemes really necessary. In Feldman, L. B. (Eds.), Morphological Aspects of Language Processing (pp. 365-382). Psychology Press. Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior research methods*, *51*, 1042-1058.

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review*, *29*(5), 1673–1702. https://doi.org/10.3758/s13423-022-02086-0

*Suárez-Coalla, P., & Cuetos, F. (2013). The Role of Morphology in Reading in Spanish-Speaking Children with Dyslexia. *The Spanish Journal of Psychology, 16*. https://doi.org/10.1017/sjp.2013.58

*Suárez-Coalla, P., Martínez-García, C., & Cuetos, F. (2017). Morpheme-Based Reading and Writing in Spanish Children with Dyslexia. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01952

Taft, M. (1979). Recognition of affixed words and the word frequency effect. Memory & Cognition, 7(4), 263–272. https://doi.org/10.3758/bf03197599

Taft, M. (2004). Morphological Decomposition and the Reverse Base Frequency Effect. *The Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765. https://doi.org/10.1080/02724980343000477

Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. The Mental Lexicon, 1(2), 183–199. https://doi.org/10.1075/ml.1.2.02taf

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/s0022-5371(75)80051-x

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39. https://doi.org/10.1016/j.cogpsych.2015.03.003

Tang, R., & Witzel, N. (2020). The role of phonology in processing morphologically complex words. Psychonomic Bulletin & Review, 27(2), 379–384. https://doi.org/10.3758/s13423-019-01683-w

To, N. L., Tighe, E. L., & Binder, K. S. (2014). Investigating morphological awareness and the processing of transparent and opaque words in adults with low literacy skills and in skilled readers. *Journal of Research in Reading*, *39*(2), 171–188. https://doi.org/10.1111/1467-9817.12036

Topor, M., Pickering, J. S., Barbosa Mendes, A., Bishop, D. V. M., Büttner, F. C., Elsherif, M. M., Evans, T. R., Henderson, E. L., Kalandadze, T., Nitschke, F. T., Staaks, J., Van den Akker, O., Yeung, S. K., Zaneva, M., Lam, A., Madan, C. R., Moreau, D., O'Mahony, A., Parker, A. J., & Riegelman, A. (2020). *An integrative framework for planning and conducting Non-*

Intervention, Reproducible, and Open Systematic Reviews (NIRO-SR). https://doi.org/10.31222/osf.io/8gu5z

***Traficante, D., Marcolini, S., Luci, A., Zoccolotti, P., & Burani, C. (2011). How do roots and suffixes influence reading of pseudowords: A study of young Italian readers with and without dyslexia. *Language and Cognitive Processes*, *26*(4-6), 777–793. https://doi.org/10.1080/01690965.2010.496553**

***Traficante, D., Marelli, M., & Luzzatti, C. (2018). Effects of Reading Proficiency and of Base and Whole-Word Frequency on Reading Noun- and Verb-Derived Words: An Eye-Tracking Study in Italian Primary School Children. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02335**

***Traficante, D., Marelli, M., Luzzatti, C., & Burani, C. (2014). Influence of verb and noun bases on reading aloud derived nouns: evidence from children with good and poor reading skills. *Reading and Writing*, *27*(7), 1303–1326. https://doi.org/10.1007/s11145-013-9488-6**

Treiman, R., & Kessler, B. (2005). *Writing Systems 1 Writing Systems and Spelling Development*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9e4f44dae8c9a6733c9257947fd99cf92569e60e

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, *195*, 103810. https://doi.org/10.1016/j.cognition.2018.09.013

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

***van Hoogmoed, A. H., Knoors, H., Schreuder, R., & Verhoeven, L. (2013). Complex word reading in Dutch deaf children and adults. *Research in Developmental Disabilities*, *34*(3), 1083–1089. https://doi.org/10.1016/j.ridd.2012.12.010**

***VAN HOOGMOED, A. H., VERHOEVEN, L., SCHREUDER, R., & KNOORS, H. (2011). Morphological sensitivity in deaf readers of Dutch. *Applied Psycholinguistics*, *32*(3), 619–634. https://doi.org/10.1017/s0142716411000245**

Vannest, J., Newport, E. L., Newman, A. J., & Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. Brain Research, 1373, 144–159. https://doi.org/10.1016/j.brainres.2010.12.022

Verhoeven, L., & Perfetti, C. A. (2011). Morphological processing in reading acquisition: A cross-linguistic perspective. *Applied Psycholinguistics, 32*(3), 457–466. https://doi.org/10.1017/s0142716411000154

***Verhoeven, L., & Schreuder, R. (2011). Morpheme frequency effects in Dutch complex word reading: A developmental perspective. *Applied Psycholinguistics, 32*(3), 483–498. https://doi.org/10.1017/s0142716411000178**

***Verhoeven, L., Schreuder, R., & Baayen, H. (2003). Units of Analysis in Reading Dutch Bisyllabic Pseudowords. *Scientific Studies of Reading, 7*(3), 255–271. https://doi.org/10.1207/s1532799xssr0703_4**

***Verhoeven, L., Schreuder, R., & Haarman, V. (2006). Prefix identification in the reading of Dutch bisyllabic words. *Reading and Writing, 19*(7), 651–668. https://doi.org/10.1007/s11145-005-1912-0**

Wegener, S., Wang, H. C., Beyersmann, E., Reichle, E. D., Nation, K., & Castles, A. (2023). The effect of spacing versus massing on orthographic learning. *Reading Research Quarterly*, *58*(3), 361-372.

***Whylie, S. (2022). *Morphological development : from the beginner to the proficient reader*. [Doctoral Dissertation, Coventry University]. Retrieved from https://pure.coventry.ac.uk/ws/portalfiles/portal/54013416/Whylie_2022.pdf**

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Takahashi, K. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version, 2(1), 1-189. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=af53fd2f5b9e81b6edec0c13e1b3babd34bda399

Wiley, J. (2020, May 31). Linear Mixed Models (LMMs) - Model Comparisons, retrieved from https://joshuawiley.com/MonashHonoursStatistics/LMM_Comparison.html

Yablonski, M., Menashe, B., & Ben-Shachar, M. (2021). A general role for ventral white matter pathways in morphological processing: Going beyond reading. *Neuroimage*, *226*, 117577.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 597–613. https://doi.org/10.1037/xlm0000064

Yeatman, J. D., Tang, K. A., Donnelly, P. M., Yablonski, M., Ramamurthy, M., Karipidis, I. I., Caffarra, S., Takada, M. E., Kanopka, K., Ben-Shachar, M., & Domingue, B. W. (2021). Rapid online assessment of reading ability. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-85907-x

Yeatman, J. D., Tang, K. A., Donnelly, P. M., Yablonski, M., Ramamurthy, M., Karipidis, I. I., Caffarra, S., Takada, M. E., Kanopka, K., Ben-Shachar, M., & Domingue, B. W. (2021). Rapid online assessment of reading ability. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-85907-x

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current opinion in neurobiology*, *17*(2), 251-257.

Appendix for Chapter 2

Appendix 2A: Reading aloud

Authors (Year)	Paradigm notes	Participant type	Age group	Sample size	Language	Factors	Outcomes
Angelelli et al. (2014)		TD	A,B	15; 15	Italian	Task; Grade; Morphology	Accuracy
Angelelli et al. (2017)		TD; DYS	A	16; 16	Italian	Group; Task; Morphology	Accuracy
Bowey (1996)	Reading aloud with grammatical priming	TD	B	32; 40	English	Word Class; Inflection; Constraint	RTs; Accuracy
Burani et al. (2008)		CA; RA; DYS	A,B	17; 34; 17	Italian	Group; Pseudoword type; Word type	RTs; Accuracy
Burani et al. (2002, Experiment 1)		TD	A,B	90	Italian	Grade; Morphology	RTs; Accuracy
Burani et al. (2018)		DYS; TD; CA	B,C	20; 40	Italian	Word length; Word frequency; Root length; Root frequency; Root family size; Suffix frequency	RTs; Accuracy
Carlisle (2000)		TD	A,B	34; 25	English	Grade; Reading test scores; Word type	Word Reading Test scores
Carlisle & Katz (2006)		TD	A,B	79; 73	English	Grade; Reading ability; Derived word frequency; Family size; Average family frequency; Morphemic Composition; Exposure to Word Family	Reading complex words score
Carlisle & Stone (2003) (*Book Chapter-Chapter 2*)		TD (Study 1); poor reader & CA & RA (Study 2)	A,B	Study 1:34; 26; Study 2: 15; 40; 40	English	Study 1: Grade; Morphological awareness; Word type; Word length; Base frequency Study 2: Group; Morphemic structure; base frequency; Number of syllables	Study 1: Word Reading Test scores Study 2: RT; Accuracy
Carlisle & Stone (2005)		TD	Study 1:A, B; Study 2: B, C	Study 33; 39; Study 15; 18	English	Study 1: Grade; Morphemic structure; Surface frequency; Base frequency; number of syllables; Word reading ability Study 2: Grade group; Derived word type	RTs; Accuracy

Carlisle et al. (2001)		Poor readers; Average readers;	B,C,D	18; 33	English	Reading ability; Phonological transparency; Orthographic patterns; Task; RT to nonwords	RTs; Accuracy
Cole et al. (2012)	Exp 1. Pseudoword reading; Exp 2. Word reading	TD	A, B	24; 24	French	School Grade; Morphological condition	RTs; Accuracy
D'Alessio et al. (2018)	Longitudinal; 6th graders could not be retested at Time 2	TD	A,B,C	Time 1: 30; 32; 35; Time 2: 16; 14	Spanish	Grade; Morphology; Word frequency	RTs; Accuracy
Deacon et al. (2017)		TD	A, B	53; 46	English	Grade; Base Frequency	Accuracy
Deacon & Francis (2017)		TD	A, B	43; 32	English	Family size; Base frequency; Family frequency; Grade	Accuracy
Deacon et al. (2011)		TD	A,B,C	30; 31; 24	English	Grade; Base frequency; Surface frequency; Base transparency	RTs; Accuracy
Egan & Pring (2004)	Words with visual disruption at morpheme or phoneme boundary	CA, RA, DYS	A, B	48	English	Type of Disruption; Word type; Group	RTs; Accuracy
Elbro & Arnbak (1996) (Exp 1)		Dyslexic; RA	B, D	26; 26	Danish	Reader groups; Word types	RTs; Accuracy
Kearns (2015)		TD	A	202	English	Item-Specific Measures (e.g., Affix Reading; Grapheme-Phoneme Correspondence Reading; Phonogram Reading) Measures of Child Abilities (e.g., Affix Reading; Grapheme-Phoneme Correspondence Knowledge; Morphological Awareness) Word Characteristic Measures (e.g., Bigram Frequency; Word Frequency; Root Frequency)	Accuracy
Laxon et al. (1992)		TD	A, B	60	English	Age; Reading ability; Word type; Word endings; Phonological property of the suffix -ed; Lexicality;	Accuracy

						Lexicality of the stem; Frequency	
Mann & Singson (2003) *(Book chapter- Chapter 1)*		TD	A, B	49; 38; 38; 28	English	Grade; Phonological complexity; Base frequency; Derived frequency; Morphological awareness; Phonological awareness; Vocabulary skill	Accuracy
Marcolini et al. (2011)		Poor readers; CA	B	21; 42	Italian	Group; Morphological type; Whole word frequency	RTs; Accuracy
Mousikou et al. (2020)		TD	A	30; 32; 32; 32 .	English; French; German; Italian	Language; (presence of) Stem; (presence of) Suffix; Reading Ability	RTs; Accuracy
Quemart & Lambert (Task 1, 2019)		TD	B	25; 35	French	Grade; Word construction; Task	RTs
Saiegh-Haddad & Geva (2008)		TD	A,B	14; 9; 9; 11	English	Phonological awareness; Morphological awareness; Oral language proficiency	Word reading fluency
Suarez-Coalla & Cuetos (2013)		Spanish-speaking children with dyslexia; CA	A,B	16; 16	Spanish	Lexicality; Morphological complexity; Reading group	RTs; Accuracy
Suarez-Coalla et al. (2017)		Spanish-speaking children with dyslexia; CA	A,B	24; 24	Spanish	Reading group; Morphological condition; Lexicality	RTs; Reading durations; Reading critical segment durations; Accuracy
Traficante et al. (2011)	Reading aloud pseudowords	Children with dyslexia; CA	B	14; 44	Italian	<u>For the mixed-effect linear model:</u> Group; Root; Suffix <u>For the regression analysis:</u> Group; Bigram frequency of the left part of the string; Bigram frequency of the right part of the string; Frequency of the root; Frequency of the suffix	RTs; Accuracy
Traficante et al. (2018)	Sentence reading with eye tracking	TD	A	20; 22; 20	Italian	Grade; Word length; Word frequency; Root length; Base frequency; Base category; Proficiency in reading	Accuracy; Eye-movement

Traficante et al. (2014)		TD	B	36; 18	Italian	Reader group; Grammatical class of the base word; Word frequency; Word length in letters; Suffix length in letters; Base frequency	RTs; Accuracy
Verhoeven et al. (2003)	Reading aloud pseudowords	TD	A,B	33	Dutch	Pseudoword type; Stress pattern	The relative numbers of /ɔ/ and /ɛ/ realisations; Stress assignment
Verhoeven et al. (Exp. 1, 2006)		TD	A,B	33; 46	Dutch	Word type; Grade	Accuracy ; Pronunciation of the vowel E in the first syllable BE for pseudowords;

Note.
TD= typically developing, DYS = dyslexic children, RA= Reading ability matched control, CA= Chronological age matched control;
Age group: A = (6-9 years]; B = (9-12 years]; C = (12-15 years]; D = (15-17 years]

Appendix 2B: Lexical decision

Authors (Year)	Paradigm notes	Participant type	Age group	Sample size	Language	Factors	Outcomes
Assink et al. (2000, Exp. 1)		TD	B	32; 53	Dutch	Prefixedness; Corruptness; Frequency of prefix	RTs; Accuracy
Burani et al. (2002) (Exp. 2)		TD	A, B	90	Italian	Grade; Morphology;	RTs; Accuracy
Carlisle & Stone (2005) (Study 2)		TD	B, C	15; 18	English	Grade; Derived word type	RTs; Accuracy
Carlisle et al. (2001)	Direct comparison or reading aloud and lexical decision	Poor readers; Average readers	B, C, D	18; 33	English	Reading ability; Phonological transparency; Orthgraphic patterns; Task	RTs; Accuracy
Casalis et al. (2015)		TD	A, B	40; 32; 32	English; French	Root presence; Suffix presence; Language group; Grade /Age	RTs; Accuracy
D'Alessio et al. (2019)		TD	A, B, C	30; 30; 30	Spanish	Grade; Morphology; Frequency	RTs; Accuracy
Dawson et al. (2018)		TD	A, C, D	50; 37; 36	English	Morphological condition; Age group;	RTs; Accuracy
de Zeeuw (2013) (*Thesis chapter: Chapter 4*)		TD	A, B	24; 26	Dutch	Prefix status	RTs; Accuracy
de Zeeuw et al. (2013)		TD	A, B	24; 26	Dutch	Regularity; Grade	RTs; Accuracy
de Zeeuw et al. (2015)		TD	A, B	35; 24; 40	Dutch	Frequency of the whole compound; Frequency of the first constituent; Frequency of the second constituent; Grade level	RTs; Accuracy

Duncan et al. (2011)		Good readers; Poor readers	A, B	20; 10	English	Presence of embedded word; Presence of suffix	RTs; Accuracy
Fleischhauer & Elisabeth (2013) (*Thesis chapter: Chapter 8*)		TD	A, B	20; 21	German	Reader group; Participle type; Surface frequency; Working memory	RTs; Accuracy
Hasenacker & Schroeder (2017)		TD	A, B	51; 20	German	Morphological status; Disruption position; Age group	RTs; Accuracy
Hasenäcker & Schroeder (2019)		TD	Exp 1: A Exp 2: A, B	Exp 1: 22; Exp 2: 105	German	1st Constituent Frequency; 2nd Constituent Frequency; Whole-word Frequency; Vocabulary Size	RTs; Accuracy
Hasenacker et al. (2017)		TD	A, B	146; 62; 70; 61	German	Age group; Morphological type	RTs; Accuracy
Lázaro et al. (2016)	With go/no-go lexical decision	TD	A, B	30; 30; 30	Spanish	Grade ; Suffix frequency Vocabulary and reading accuracy scores; Mean decision times	RTs; Accuracy
Lazaro et al. (2013)		Children with reading disorders; CA; RA.	A	60	Spanish	Reader group; Base frequency ;	RTs; Accuracy
Lazaro et al. (2015, Experiment 2)	with go/no-go lexical decision	Children with reading deficits; CA; RA	A	60	Spanish	Morpho-orthographic alteration of the base; Reader group; Surface frequency; Neighbourhood density; Letter length; Base frequency; Suffix productivity;	RTs; Accuracy

Leong (1989)		TD	B	103; 88 ; 107	English	Grade; Reading level; Lexicality; Morphemic parsing condition	RTs
Leong & Parkinson (1992)		Exp 1**: poor readers;** Exp 2: less skilled readers; CA	Exp 1: A, B Exp 2: B, C	Exp 1: 75; Exp 2: 88	English	Grade; Reading/spelling level; Lexicality : Morphology parsing condition; Affixation condition	RTs
Lukatela et al. (1987)		TD	C	36	Serbo-Croatian	Regularity : Case; Noun Gender	RTs
Quémart et al. (2012)		TD	A, B	30; 30	French	Grade; Presence of Base; Presence of Suffix	RTs; Accuracy
van Hoogmoed et al. (2013)		Deaf children; Hearing CA	B	21; 42	Dutch	Word category: Hearing status	RTs; Accuracy
van Hoogmoed et al. (2011, Experiment 1)		Deaf children; Hearing children CA & RA	A, B	25; 43; 50	Dutch	Reader group; Word category	RTs; Accuracy
Verhoeven & Schreuder (2011, Experiment 2)		DYS; CA; RA	A, B	37; 43; 33	Dutch	Reader group; Plural form frequency; Single root form frequency	RTs; Accuracy
Verhoeven et al. (2006) (Exp. 2)		TD	A, B	33; 46	Dutch	Age group; Word type	RTs; Accuracy

Note.
TD= typically developing, DYS = dyslexic children, RA= Reading ability matched control, CA= Chronological age matched control;
Age group: A = (6-9 years]; B = (9-12 years]; C = (12-15 years]; D = (15-17 years]

Appendix 2C: Masked priming

Authors (Year)	*Paradigm notes*	Participant type	Age group	Sample size	Language	Factors	Outcomes
Beyersmann et al. (2012, Exp. 2)	50ms prime	TD	A, B	42; 50	English	Prime type; Item type; Age group	RTs; Accuracy
Beyersmann et al. (2015)	50ms prime	TD	A, B	191	French	Prime type; Ending frequency; Reading proficiency; Grade; Stem frequency; Trial order	RT; Accuracy
Beyersmann et al. (2019)	50ms prime	TD	A, B, C, D	33; 46; 35	English	Prime type; Relatedness; Grade; Word reading proficiency	RTs; Accuracy
Beyersmann et al. (2021)	50ms prime	TD	A, B	252	French; German	Grade; Reading ability; Affix type; Prime type; Language; Reading ability:	RTs; Accuracy * **Different priming type: different** contrasts.
Casalis et al. (2009)	75ms vs. 250ms prime	TD	A, B	26; 27	French	Prime condition; Prime Duration	RTs
Dawson et al. (2021)	50ms prime	TD	B, C, D	48; 57; 48; 51	English	Prime type; Item type; Age; Word reading ability	RTs; Accuracy
de Oliveira & Justi (2017)	Go/ no go; 60ms vs. 250ms prime	TD	A, B	141	Brazilian Portuguese	SOA; Prime; Grade; Morphological awareness	RTs
Duranovic et al. (2020)	With read aloud response; 50ms vs. 250ms prime	TD	A, B, C	40; 40; 40	Bosnian	Grade; Prime duration; Morphology; Levenshtein distance	RTs
Fleischhauer et al. (2021)	67ms prime	TD	A, B	48; 44; 51; 75	German	Prime type; Test condition; Age group	RTs; Accuracy
Hasenacker et al. (2016)	50ms prime	TD	A, B	40	German	Prime type; RT quartiles	RTs;
Hasenacker et al. (2020)	50ms prime	TD	A	98	German	Prime type; Grade; Reading skills	RTs; Accuracy

Law & Ghesquière (2021)	60ms prime	Children with dyslexia(DR); CA	B	46; 20	Dutch	Morphology; Priming; Group; Second-grade phonological awareness	RTs;
Lazaro et al. (2018)	Go/ no-go lexical decision; 50ms prime	TD	A, B	96 (33; 31; 25)	Spanish	Grade; Word type; Prime type	RTs; Accuracy
Madziak & Rebecca (2021) (*Master's Dissertation*)	with ERP recording; 67ms prime	Reading impaired; CA; RA	A, B	10; 16; 12	English	Prime type; Suffix type; Target word frequency	RTs (behavioural); N400 amplitude (ERP); N250 amplitude (ERP).
Quemart & Casalis (2014)	60ms prime in Exp.1; 250ms prime in Exp. 2; 800ms prime in Exp. 3	TD	Exp 1: A, B Exp 2: A, B	**Exp 1:** 20; 20; 20 **Exp 2**: 20; 22; 22.	French	**Exp 1**: Morphology; Prime; Grade; **Exp 2**: **same** variables as in Exp 1.	RTs;
Quemart & Casalis (2015)	60ms prime	Exp 1: DYS; Reading level (RL) control children; CA; Exp 2: Same as in Exp. 1	Exp 1: A, B Exp 2: B, C	16; 16; 16	French	Group; Morphology; Prime	RTs
Quémart et al. (2011)	60ms prime in Exp. 1; 250ms prime in Exp. 2; 800ms prime in Exp. 3	TD	A, B, C	21; 19; 20	French	Grade; Morphology; Prime	RTs

Note.

TD= typically developing, DYS = dyslexic children, RA= Reading ability matched control, CA= Chronological age matched control;

Age group: A = (6-9 years]; B = (9-12 years]; C = (12-15 years]; D = (15-17 years]

Appendix 2D: Other paradigms

Authors (Year)	Paradigm notes	Participant type	Age group	Sample size	Language	Factors	Outcomes
Berthiaume & Daigle (2014a)	Plausibility/probability judgement task	DYS; RA; CA	B	26; 30; 26	French	Age; Reading ability	Percentage for choosing the legal pseudowords
Berthiaume & Daigle (2014b)	Plausibility/probability judgement task	Deaf children; RA; CA	A, B	21; 22; 24	French	Age; Reading ability	Percentage for choosing the legal pseudowords
Breadmore & Carroll (2016)	Probe detection task	TD Children with a history of Otitis Media RA & CA DYS	A, B	Exp 1: 135 Exp 2: 29; 29; 29 Exp 3: 27; 27; 27	English	Relationship between the probe and the target lure; Reader groups	d' : a measure of sensitivity based on signal detection theory
Breadmore & Carroll (2018)	Silent reading (sentence); Eye-tracking	TD DYS; CA; RA	A, B	Exp 1: 38;37 Exp 2: 19; 19; 19	English	Reading ability group; Verb error	First fixation; Gaze duration; Total duration
Cauchi et al. (2022)	Flankers task	TD	A,B	Cycle 2: 44; Cycle 3: 42	French	Target type; Relation type; Relatedness	RTs; Accuracy
Deacon et al. (2010)	Fragment completion task	TD	B, C	32; 32; 24	English	Priming condition; Grade	Accuracy (as **number** of fragments completed with the **target word**)
Deilen et al. (2022)	Exp 1. Word picture matching (eye-tracking); Exp 2. Sentence reading (eye-tracking)	Hearing impaired	B,C, D	9; 9	German	Segmentation ; Number of morphemes;	Number of fixations; First fixation duration; Total reading time; Rate of regressions.
Drewnowski (1981)	Letter search	TD	A,B	235	English	Number of Syllables; Word Ending: -ing vs. -ent; Passage Type; Grade Level	Accuracy
Feldman et al. (2002)	Fragment completion task	TD	B	88	English	Relationship between prime and target; Mrphological transparency	Completion rate (complete the fragment as the target)

Hasenäcker et al. (2021)	Letter search	TD	Exp 1: A, B Exp 2: A, B	Exp 1: 38; 33; Exp 2: 26; 31	Italian	Affix status; Suffix position; Grade	RTs; Accuracy
Hasenäcker et al. (2021)	Semantic categorisation	TD	A, B	82	Italian	Congruency; Suffix status	RTs; Accuracy
Hess et al. (2021)	Copying task	TD	A, B	134	German	Visual Disruption; Letter Position	Writing Onset Duration; Letter Duration.
Leong & Chen (2000)	Sentence completion (verbally)	TD	B	79; 67; 80	English	Base form shifts; Spelling ability; Grade	RTs
Leong & Parkinson (1995) (*Book chapter-Chapter 15*)	Repetition priming	Below-average & above-average readers	B	18; 23; 23; 24	English	Grade; Reading level; Priming condition	RTs
McCutchen et al. (2009)	Continuous lexical decision priming	TD	B,C	163	English	Grade; Reading skill group; Prime type; Phonological stability	RTs; Accuracy
Pescuma et al. (2022)	Oddball with MEG	TD	B	17	English	Types of pseudowords	Signal-to-Noise Ratio in MEG
Quemart & Lambert (2019, Task 2)	Copying	TD	B	25; 35	French	Grade; Word; Task	Naming latency ; Writing latency Letter writing duration; Inter-letter interval; Letter writing duration
Rabin & Deacon (2008)	Fragment completion task	TD	A,B	100	English	Prime condition; Grade	Completion rates
S, Whylie (2021) (Thesis Chapter: Chapter 4 & Chpater 5)	Sentence reading (eye-tracking)	TD	A,B	25	English	Chapter 4: Word surface frequency Base frequency; Chapter 5: Preview condition/ word type	Total time; Go-Past duration; Gaze duration; First Fixation duration

Note. TD= typically developing, DYS = dyslexic children, RA= Reading ability matched control, CA= Chronological age matched control;

Age group: A = (6-9 years]; B = (9-12 years]; C = (12-15 years]; D = (15-17 years]

Appendix for Chapter 3

Appendix 3A. Experimental Stimuli

1. Morphological Condition:

Target	Related Prime	Unrelated Prime
WALK	walked	smelly
FILL	filled	lovely
TOAST	toaster	grocery
GOLD	golden	frosty
CRY	crying	posted
BAD	badly	liked
DRY	drying	weaker
OPEN	opened	boards
SHY	shyly	mower
FLY	flying	softer
PLAY	playing	tighter
MIX	mixer	doing
BUY	buying	louder
FIX	fixing	boiler
TEACH	teacher	robbery
ACT	acting	nearer
MOOD	moody	waved
MAIN	mainly	fuller
FARM	farmer	stormy
LUCK	lucky	named
BOX	boxer	messy
HARD	harder	filthy
TRY	trying	soften
EAT	eating	locker
SLY	slyly	fixed
LAY	layer	milky
BUSH	bushy	moved
CREAM	creamy	darker
SLOW	slowly	leader
DEEP	deeply	banker
AIM	aimed	rainy
SAD	sadly	loved
SEW	sewed	windy

Target	Related Prime	Unrelated Prime
DIRT	dirty	stars

2. Pseudomorphological Condition:

Target	Related Prime	Unrelated Prime
MISS	mission	longest
SLIM	slimy	eater
EAST	easter	likely
LAD	lady	eggs
SHOULD	shoulder	fighting
CORN	corner	sticky
OFF	offer	dolly
SHOW	shower	fallen
SCAR	scary	older
MAST	master	grassy
FOR	forest	prayer
POST	poster	bricks
DRAW	drawer	postal
MOTH	mother	greedy
FLOW	flower	saving
PART	party	tower
LIST	listen	sleepy
MAN	many	used
MET	metal	sandy
ARM	army	cats
NAUGHT	naughty	painter
BELL	belly	eaten
FAST	fasten	nearly
FAIR	fairy	beans
BOTH	bother	widely
NUMB	number	fluffy
MILL	million	clearly
EVER	every	lower
BUS	busy	aged
FACT	factory	cheaper
SAND	sandal	wooden
COUNT	country	filling
HUNG	hungry	warmer
FIN	finish	caller

3. Form(nonsuffixed) Condition:

Target	Related Prime	Unrelated Prime
ADD	address	speaker
FREE	freeze	tender
SING	single	curled
AGAIN	against	tidying
THIN	think	early
TEA	tease	salty
WIND	window	fruity
HOW	howl	ants
CAR	carrot	sooner
BEE	beer	maps
TWIN	twinkle	lighter
SIGH	sight	curly
HOT	hotel	risky
WIN	wink	legs
CART	carton	player
ARE	area	cars
CHIN	china	jelly
TOO	tooth	bumpy
BEG	begin	snowy
SKI	skirt	dusty
SPIN	spinach	magical
MEN	menu	bags
CROW	crown	going
TURN	turnip	slower
YELL	yellow	hunter
STAR	starve	camped
SHOVE	shovel	lately
COME	comet	cried
WON	wonder	gloomy
PAST	pasta	rocky
DRAG	dragon	poetry
PILL	pillow	lesser
CAME	camel	bossy
LESS	lesson	richer

Appendix 3B: Contrast Comparisons for RT (200 – 2000ms) Analysis

Contrast	Estimate	SE	*z* ratio	*p* value
Related Form - Unrelated Form	-0.014877	0.00613	-2.428	0.1464
Related Form - Related Pseudomorphological	-0.000832	0.01288	-0.065	1.0
Related Form - Related Morphological	0.038383	0.01285	2.988	**0.0335**
Unrelated Form - Unrelated Pseudomorphological	-0.014673	0.01371	-1.07	0.8934
Unrelated Form - Unrelated Morphological	0.020468	0.01368	1.496	0.6666
Related Pseudomorphological - Unrelated Pseudomorphological	-0.028717	0.00622	-4.619	**0.0001**
Related Pseudomorphological - Related Morphological	0.039214	0.01287	3.046	**0.0281**
Related Morphological - Unrelated Morphological	-0.032791	0.00608	-5.392	**<.0001**

Appendix for Chapter 4

Appendix 4A. Items of reading ability assessment

Words	Nonwords
play	uoy
we	fror
miss	om
black	ujst
become	inni
sun	wared
shine	leld
liberty	arsher
however	blans
back	ggnoi
without	bledge
pickle	cuppom
resolve	ancrusinishness
elements	crit
brought	ef
arm	ip
will	cibsus
best	slellen
custom	tfaer
answer	trood
testify	losh
them	loms
waves	nied
library	sint
from	beeltisul
guess	telechinct
describe	sudsage
inside	bulbet
purchase	fattify
faster	hus
mystery	groon
dryer	strosh
invent	coth
reverse	nop
once	trull
whose	hust
fatter	streat
telescoped	bads
sit	tis
listen	stenil
lunch	nulch
night	ginth

Adapted from List A in Yeatman, et al. (2021). Rapid online assessment of reading ability. *Scientific reports*, *11*(1), 6396.

Appendix 4B. Lists of experimental stimuli for the lexical decision task

Pseudomorphemic nonwords		Control nonwords	
ballness	maidoid	ballnels	maidoud
bandish	mapile	bandith	mapige
beachise	meatful	beachime	meatfil
beanine	messness	beanire	messnels
beastine	milkous	beastire	milkoes
beltish	moonly	beltith	moonla
birdly	mouthian	birdla	mouthien
boatly	mudoid	boatla	mudoud
bowlish	nutful	bowlith	nutfil
boxile	painacy	boxige	painagy
branchacy	petful	branchagy	petfil
brickise	poolian	brickime	poolien
bushacy	railness	bushagy	railnels
capian	rainile	capien	rainige
caracy	rockness	caragy	rocknels
cellous	roofous	celloes	roofoes
chairful	saltous	chairfil	saltoes
cheekile	sandacy	cheekige	sandagy
chestoid	scentian	chestoud	scentien
chinile	seatly	chinige	seatla
cliffine	seedoid	cliffire	seedoud
cloakness	sheetist	cloaknels	sheetilt
clockful	shipness	clockfil	shipnels
coachile	skinish	coachige	skinith
craftness	skyist	craftnels	skyilt
crewian	smellous	crewien	smelloes
cupise	songine	cupime	songire
deckoid	sonine	deckoud	sonire
deeracy	soulise	deeragy	soulime
deskful	sportist	deskfil	sportilt
duckian	starful	duckien	starfil
earian	stormise	earien	stormime
eggine	strawly	eggire	strawla
fanise	streamful	fanime	streamfil
feastly	sunoid	feastla	sunoud
flagacy	tailise	flagagy	tailime
floodoid	tankist	floodoud	tankilt
fogine	threadish	fogire	threadith
fuelacy	toolile	fuelagy	toolige
gasile	toyist	gasige	toyilt
glassist	waistish	glassilt	waistith
grassine	windian	grassire	windien
guardous	wingise	guardoes	wingime
hallist	woodness	hallilt	woodnels
hatous	wristist	hatoes	wristilt
jawly		jawla	
jobish		jobith	

kingoid	kingoud
laply	lapla
legish	legith
lipous	lipoes

Appendix 4C. Model summary for accuracy analysis of Experiment 2

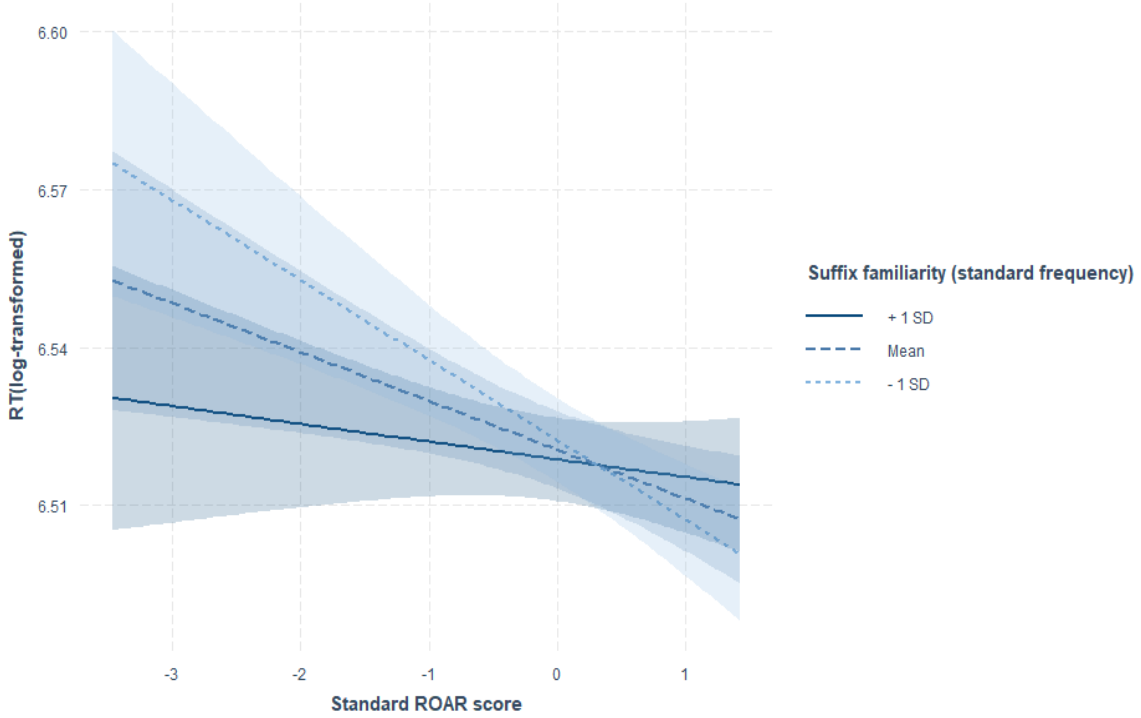
Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	4.0661095	0.1582785	25.690	< .001 ***
Condition	-1.6123871	0.2361397	-6.828	<0.001***
Suffix familiarity	-0.4901288	0.1193822	-4.106	<.001***
Reading ability	0.2955802	0.1044807	2.829	0.0047 **
Condition: Suffix familiarity	-1.0843692	0.2385494	-4.546	<.001 ***
Condition:Reading ability	-0.0683578	0.1077453	-0.634	0.526
Suffix familiarity: Reading ability	-0.0007111	0.0541660	-0.013	0.990
Condition: Suffix familiarity: Reading ability	-0.0792643	0.1085147	-0.730	0.465

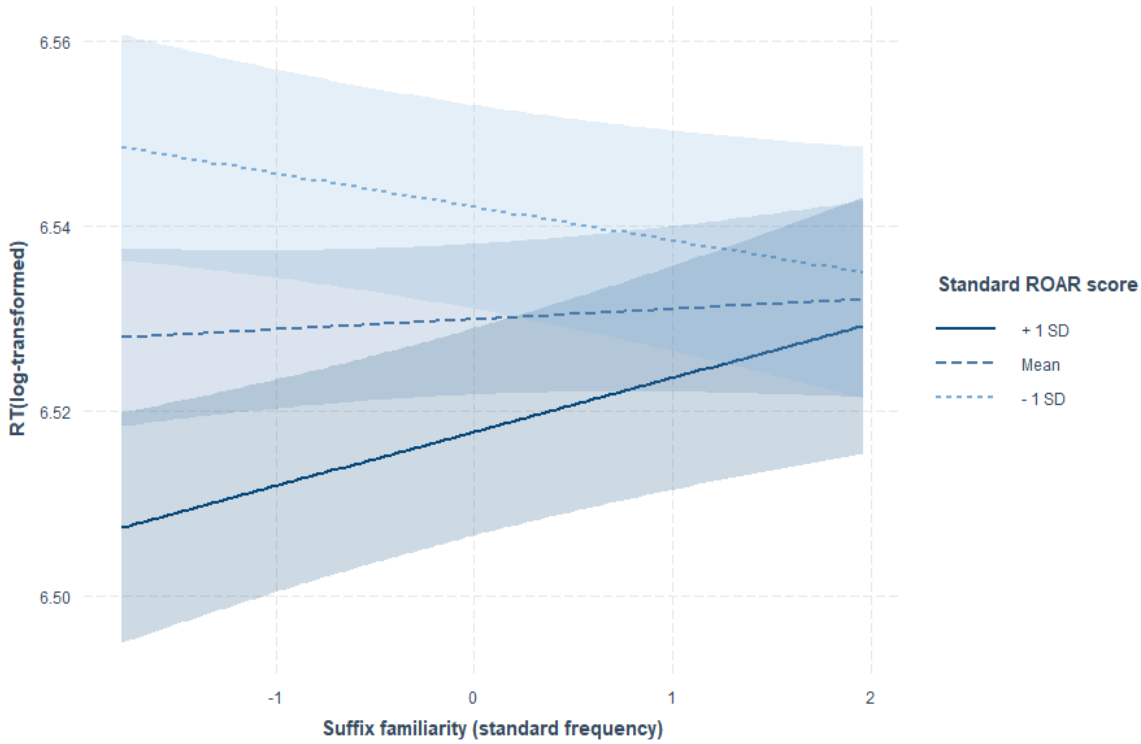
Appendix 4D. Model summary for RT Analysis of Experiment 2

Fixed effects	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.577408	0.020354	323.149	< .001 ***
Condition	0.094567	0.011808	8.009	<0.001 ***
Reading ability	-0.004180	0.019740	-0.212	0.8331
Suffix familiarity	0.029774	0.006038	4.931	<.001 ***
Condition:Reading ability	0.016365	0.006367	2.570	0.0109 *
Condition: Suffix familiarity	0.057485	0.011977	4.800	<.001 ***
Reading ability:Suffix familiarity	0.007191	0.003421	2.102	0.039 *
Condition: Reading ability: Suffix familiarity	0.004748	0.006667	0.712	0.4787

Appendix 4E. (Experiment 2)

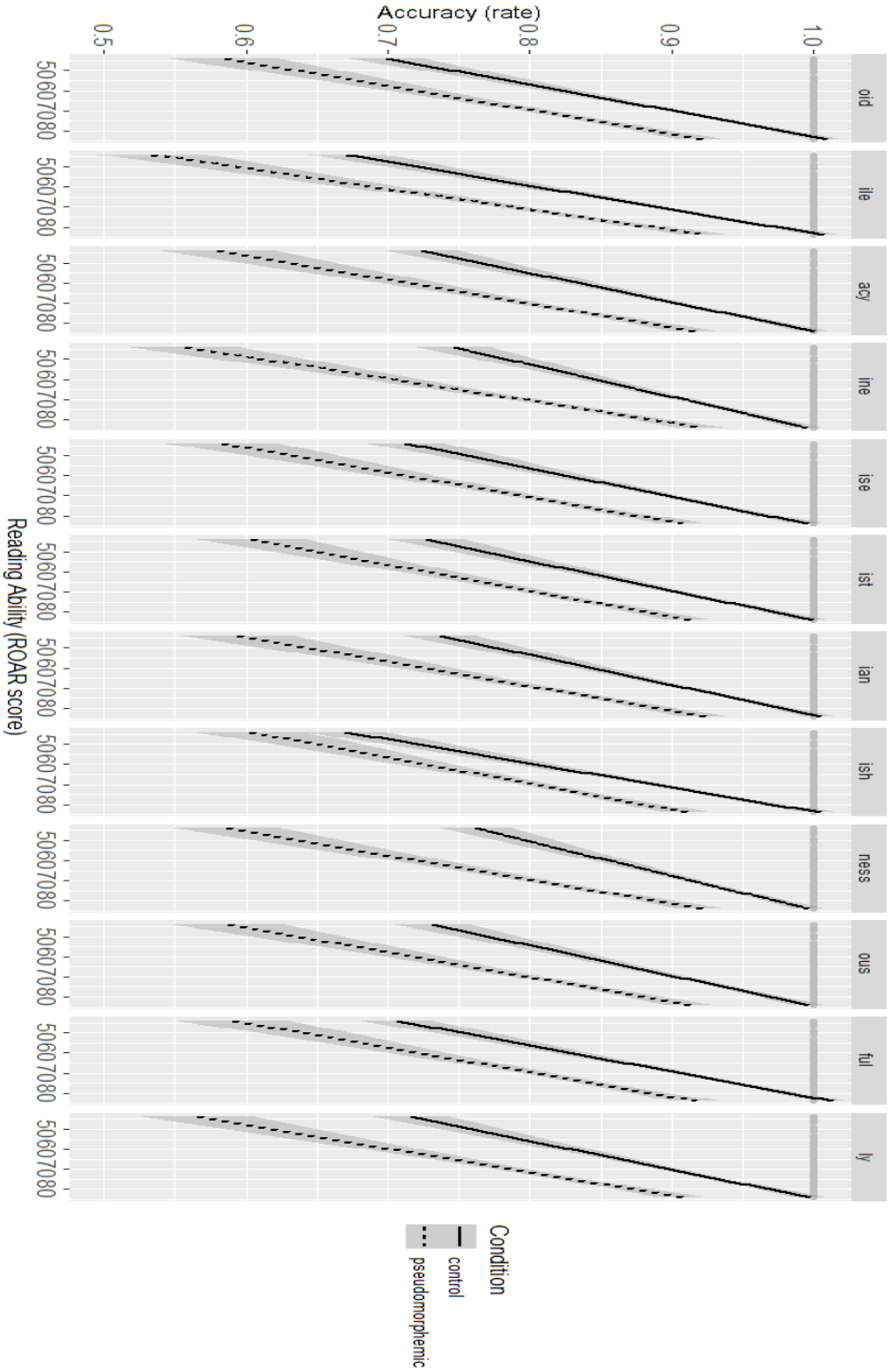
Plots showing raw Log-transformed RTs (smoothed conditional means and data points) by Reading ability (standard ROAR score) and Suffix familiarity (standard frequency per million in OCC)

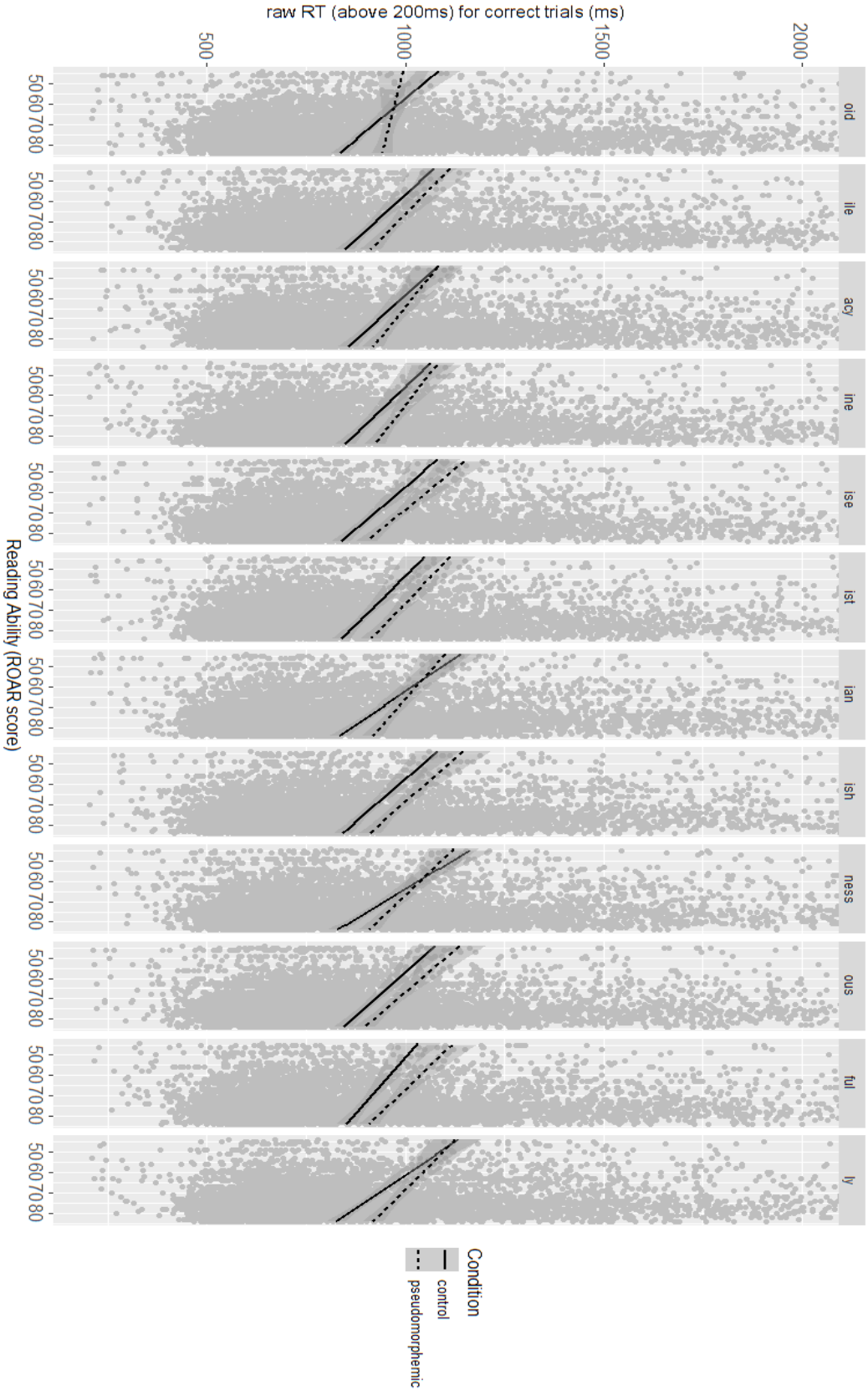




Appendix 4F. (Experiment 3)

Plots showing the interaction effect of Condition × Reading ability × Suffix familiarity (grouped by each individual suffix in ascending order of familiarity, e.g., -oid is the least familiar whereas -ly is the most familiar) for Accuracy and RT analyses.

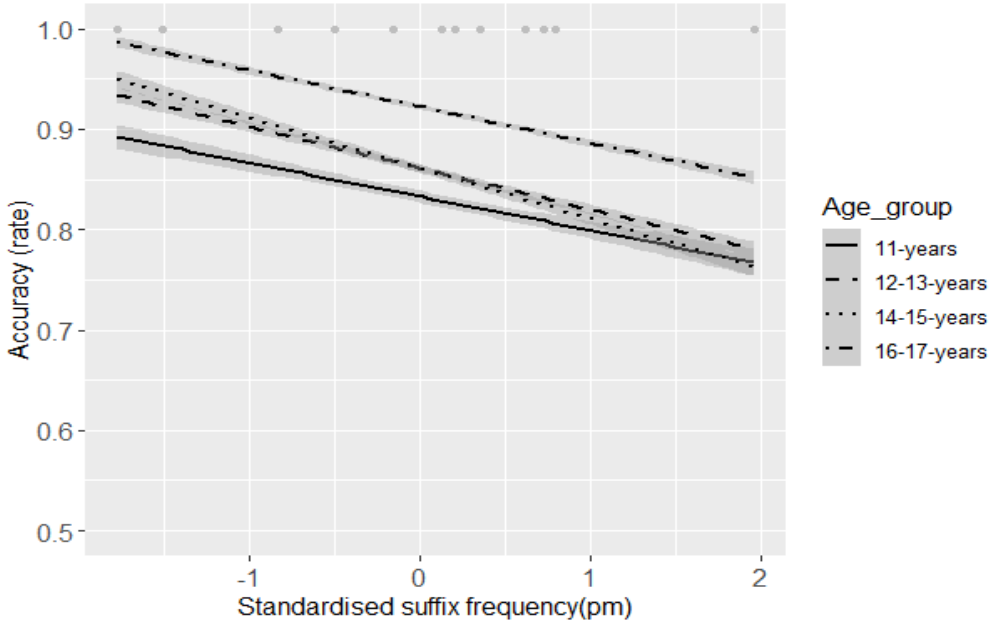



raw RT (above 200ms) for correct trials (ms)
500
1000
1500
2000
Reading Ability (ROAR score)
50 60 70 80
oid
ile
acy
ine
ise
ist
ian
ish
ness
ous
ful
ly
Condition
control
pseudomorphemic

Appendix 4G. Accuracy model output for Experiment 3

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.725416	0.068802	39.612	< .001 ***
Condition	-1.030778	0.111895	-9.212	< .001 ***
Age group.	0.416976	0.093404	4.464	< .001 ***
Suffix familiarity	-0.356427	0.056179	-6.345	< .001 ***
Reading ability	0.730785	0.043428	16.828	< .001 ***
Condition: Age group	-0.265387	0.068081	-3.898	< .001 ***
Condition: Suffix familiarity_	-0.682467	0.109810	-6.215	< .001 ***
Age group: Suffix familiarity	-0.093645	0.035262	-2.656	.007914 **
Condition: Reading ability	-0.118210	0.035576	-3.323	< .001 ***
Age group: Reading ability	0.151369	0.093987	1.611	.107285
Reading ability: Suffix familiarity	-0.030043	0.018227	-1.648	.099298
Condition: Age group: Suffix familiarity	-0.174684	0.058369	-2.993	.002765 **
Condition: Age group: Reading ability	0.098385	0.067105	1.466	.142609
Condition: Reading ability: Suffix familiarity	-0.127703	0.031217	-4.091	< .001 ***
Age group: Reading ability: Suffix familiarity_	-0.004132	0.034615	-0.119	0.904977
Condition: Age group: Reading ability: Suffix familiarity_	-0.073137	0.056975	-1.284	0.199258

Appendix 4H. (Experiment 3) Two-way interaction between Age group × Suffix familiarity on Accuracy



Appendix 4I. RT Model output for Experiment 3

Fixed effects	Estimate	Std. Error	t value	p value
(Intercept)	6.7814164	0.0105976	639.898	< .001 ***
Condition	0.0757263	0.0104981	7.213	< .001 ***
Age group	0.1441930	0.0200653	-7.186	.001***
Reading ability_	-0.0181227	0.0091837	-1.973	.048 *
Suffix familiarity	0.0278541	0.0052634	5.292	<.001 ***
Condition: Age group	0.0046532	0.0050722	0.917	0.36
Condition: Reading ability	0.0101282	0.0031319	3.234	.001 **
Age group: Reading ability	0.0096149	0.0202289	0.475	0.63
Condition: Suffix familiarity_	0.0650489	0.0104980	6.196	<.001 ***
Age group: Suffix familiarity	0.0012259	0.0026716	0.459	.646513
Reading ability: Suffix familiarity	0.0015191	0.0016108	0.943	.346826
Condition: Age group: Reading ability	0.0134364	0.0052271	2.571	.01 *
Condition: Age group: Suffix familiarity	0.0004092	0.0050729	0.081	.94
Condition: Reading ability: Suffix familiarity	0.0084692	0.0031278	2.708	.007**
Age group: Reading ability: Suffix familiarity_	0.0032567	0.0027456	1.186	.24
Condition: Age group: Reading ability: Suffix familiarity_	0.0066503	0.0052227	1.273	.20

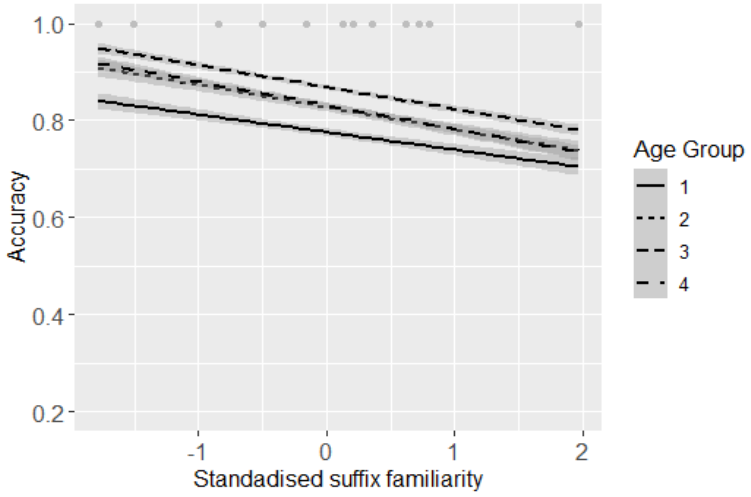
Appendix 4J. (Exploratory analysis) Accuracy model output of the exploratory analysis

		Correct	
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	9.55	8.12 – 11.23	<**0.001**
Condition1	0.37	0.30 – 0.46	<**0.001**
Suffix familiarity	0.71	0.63 – 0.79	<**0.001**
ROAR score	2.10	1.84 – 2.39	<**0.001**
Age in months	1.09	0.96 – 1.24	0.168
Condition1 × Suffix familiarity	0.56	0.46 – 0.69	<**0.001**
Condition1 × ROAR score	0.81	0.74 – 0.89	<**0.001**
Suffix familiarity × ROAR score	0.95	0.91 – 0.99	**0.027**
Condition1 × Age in months	0.89	0.83 – 0.96	**0.002**
Suffix familiarity × Age in months	0.94	0.91 – 0.98	**0.003**
ROAR score × Age in months	1.00	0.89 – 1.13	0.959
(Condition1 × Suffix familiarity) × ROAR score	0.88	0.81 – 0.95	**0.001**
(Condition1 × Suffix familiarity) × Age in months	0.98	0.92 – 1.05	0.613
(Condition1 × ROAR score) × Age in months	1.03	0.96 – 1.10	0.419
(Suffix familiarity × ROAR score) × Age in months	1.00	0.96 – 1.03	0.845
(Condition1 × Suffix familiarity × ROAR score) × Age in months	0.98	0.93 – 1.04	0.497

***Note*.** Output of Model_ Accuracy ← glmer (Correct ~ Condition × Age in months × Reading ability × Suffix familiarity + (1 + Suffix familiarity | Participant) + (1 + Reading ability | Item), family = binomial);
All the continuous variables: Age in months, ROAR score, and Suffix familiarity were standardised in this model.

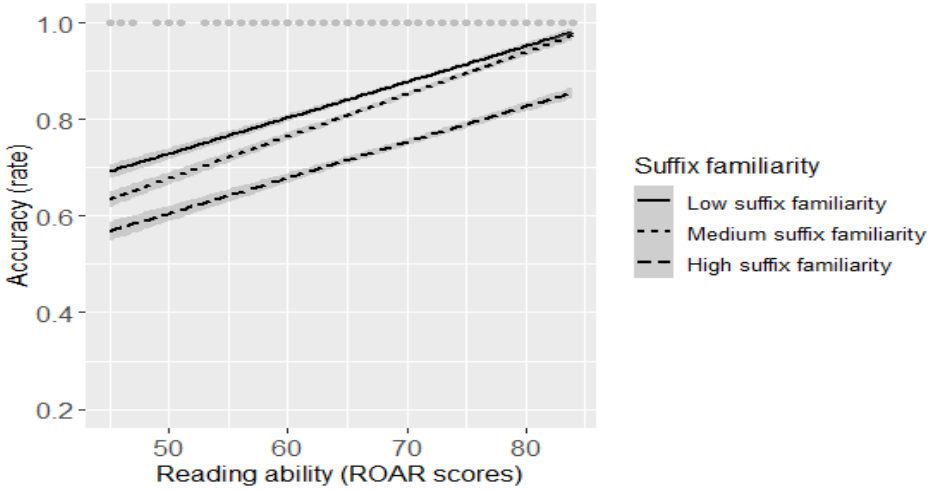
Appendix 4K. (Exploratory analysis)

Two-way interaction of Age × Suffix familiarity (Figure (a)) Reading ability × Suffix familiarity (Figure (b) & (c)) on Accuracy

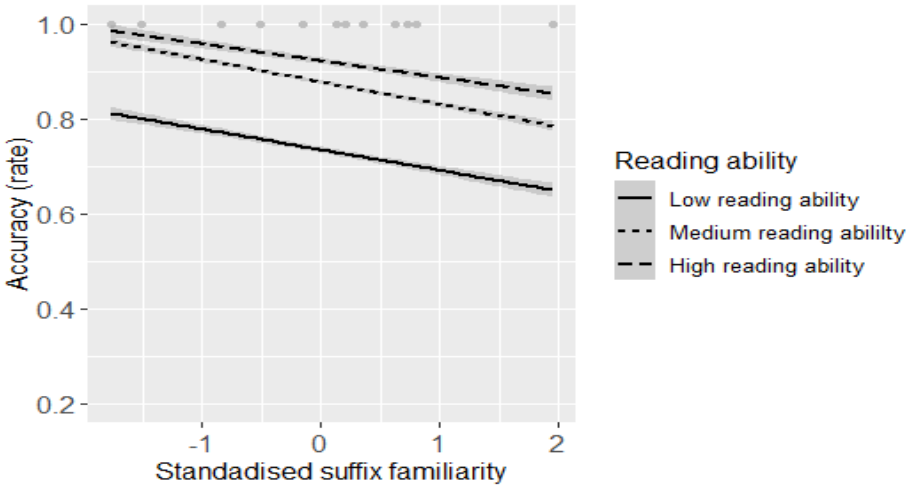


Note. Age groups are organized in ascending order based on age in months.

(a)



(b)


1.0
0.8
0.6
0.4
0.2
Accuracy (rate)
-1
0
1
2
Standadised suffix familiarity
Reading ability
Low reading ability
Medium reading abililty
High reading ability

(c)

Appendix 4L. (Exploratory analysis)

RT model output of the exploratory analysis

Predictors	***Estimates***	***CI***	***p***
(Intercept)	6.87	6.84 – 6.89	**<0.001**
Condition1	0.06	0.04 – 0.08	**<0.001**
Age in months	-0.02	-0.05 – 0.01	0.124
ROAR score	-0.00	-0.03 – 0.02	0.729
Suffix familiarity	0.02	0.01 – 0.03	**<0.001**
Condition1 × Age in months	0.00	-0.00 – 0.01	0.448
Condition1 × ROAR score	0.01	-0.00 – 0.02	0.156
Age in months × ROAR score	0.02	-0.00 – 0.05	0.060
Condition1 × Suffix familiarity	0.05	0.03 – 0.07	**<0.001**
Age in months × Suffix familiarity	0.00	-0.00 – 0.01	0.740
ROAR score × Suffix familiarity	-0.00	-0.01 – 0.00	0.757
(Condition1 × Age in months) × ROAR score	0.01	0.00 – 0.02	**0.024**
(Condition1 × Age in months) × Suffix familiarity	-0.00	-0.01 – 0.01	0.532
(Condition1 × ROAR score) × Suffix familiarity	0.00	-0.01 – 0.01	0.386
(Age in months × ROAR score) × Suffix familiarity	0.00	-0.00 – 0.01	0.215
(Condition1 × Age in months × ROAR score) × Suffix familiarity	0.00	-0.01 – 0.01	0.489

Note. Model_ RT ← lmer (RT_log_transformed ~ Condition × Reading ability × Suffix familiarity × Age in months + (1+ Suffix familiarity| Participant) + (1+ Reading ability| Item));
All the continuous variables: Age in months, ROAR score, and Suffix familiarity were standardised in this model.

Appendix for Chapter 5

Appendix 5A: Prevalence value of each word containing the target suffixes used in the stories during suffix exposure phase.

Target suffix	Words	Prevalence value
-let	coverlet	0.723
	leaflet	1.925
	booklet	2.148
	droplet	2.149
	ringlet	1.512
	bracelet	2.576
	wavelet	0.274
	piglet	2.432
-ee	devotee	1.811
	referee	2.319
	trainee	2.151
	attendee	2.137
	refugee	2.331
	trustee	2.339
	employee	2.425
	committee	2.196
-ling	sapling	1.771
	fledgling	1.556
	nestling	1.698
	duckling	2.181
	dumpling	2.321
	changeling	1.41
	seedling	2.034
	hatchling	1.86
-ism	tourism	2.576
	magnetism	2.444
	symbolism	2.246
	vandalism	2.435
	socialism	2.576
	realism	2.436
	liberalism	2.134
	skepticism	1.796

Note. The prevalence values were obtained from Brysbaert et al. 2019.

Appendix 5B: Stories Used across Three Exposure Conditions

Condition	Story Fragments	Comprehension Questions
High stem-diversity	***Story 1. Sophia (-let)*** 1) Sophia loved to visit her aunt in the countryside where she enjoyed cosying up under a soft **coverlet** and admiring the stunning views outside the window. 2) In the forest behind the house, she had a secret garden - a hidden forest that couldn't be found on any travel ***leaflet***, her own personal retreat. 3) On her adventures to the forest, Sophia always brought along her favourite **booklet** of forest animals hoping to catch a glimpse of the creatures she had read about. 4) Sophia felt a refreshing chill as a ***droplet*** landed on her nose while walking in the forest. She knew she had to find shelter before a sudden downpour. 5) As she ran faster, a bouncy ***ringlet*** of hair swung from her head like a ***wavelet*** on the ocean's surface. Sophia spotted a bush and quickly took cover. 6) Raindrops on the twigs shimmered like a bead ***bracelet***. There, she discovered a new friend - a little wild boar ***piglet*** gazing up at her with inquisitive eyes.	Sophia loved to visit___? A. Her grandmother B. Her auntie C. Her uncle What was behind the house? A. A forest B. A park C. A garden What did Sophia read about? A. Mountains B. Oceans C. Animals Why did Sophia need to find a shelter? A. She was tired. B. It was going to rain. C. She felt cold. What did Sophia find? A. A hut B. A bush C. A cave Was Sophia there alone? A. No B. Yes C. Did not mention
Low stem-diversity	1) Sophia loved to visit her aunt in the countryside, a place advertised in the local travel ***leaflet*** as a must-visit. She enjoyed cosying up and admiring the stunning views. 2) In the forest behind their house, she had a secret garden - a hidden forest that couldn't be found on any travel ***leaflet***, her own personal retreat. 3) On her adventures to the forest, Sophia always took her pet ***piglet*** and her favourite ***leaflet*** of forest animals, hoping to find other creatures she had read about. 4) Sophia felt a raindrop land on her nose while she was walking with her ***piglet*** in the forest. She knew she had to find shelter before a sudden downpour.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	5) As they ran faster, her ***piglet*** squealed happily, sounding like a high-pitched violin that filled the forest with delightful melody. Sophia spotted a bush and took cover. 6) There, they discovered a new friend - a wild boar ***piglet*** that Sophia had read about in her wildlife ***leaflet***, gazing up at her with inquisitive eyes.	
Non-exposure	1) Sophia loved to visit her aunt in the countryside where she enjoyed cosying up under a soft blanket and admiring the stunning views outside the window. 2) In the forest behind their house, she had a secret garden - a hidden forest that couldn't be found on any travel guide, her own personal retreat. 3) On her adventures to the forest, Sophia always took her pet dog and her favourite picture book about forest animals, hoping to find other creatures she had read about. 4) Sophia felt a raindrop land on her nose while she was walking with her dog in the forest. She knew she had to find shelter before a sudden downpour. 5) As they ran faster, Sophia's hair swung wildly, and her dog barked in excitement. Sophia quickly spotted a nearby bush and quickly took cover. 6) There, they discovered a new friend - a wild boar baby that Sophia had read about in her wild animal book, gazing up at them with inquisitive eyes.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
	Story 2. A Company (-let)	
High stem-diversity	1) WISE, a start-up company, was developing a smart ***bracelet*** that can track various health metrics like heart rate and blood pressure and provide real-time feedback to the wearer.	What was the name of the company? A. SMART B. RISE C. WISE
	2) Before testing on humans, the team conduct trials on their experimental ***piglet***, Wilbur, who had a ***ringlet*** of fur around its neck and a similar skin texture to humans.	What similarities did Wilbur share with humans? A. His weight B. His skin texture C. Did not mention
	3) When strapping the product onto Wilbur's front leg, the team noticed a small ***droplet*** of sweat on his snout, indicating that he was a bit stressed.	Where on Wilbur did the team placed their product? A. His front leg B. His neck C. His ear

	4) To comfort him, they placed a soft ***coverlet*** over him, and Wilbur immediately nestled into it and visibly relaxed, allowing the team to collect his physiological data. 5) Utilising ***wavelet*** techniques, they discovered a small glitch in the device, jotted down a note in their experiment ***booklet***, and set to work resolving the issue. 6) After data analysis, the team made some adjustments to the product information ***leaflet*** based on their findings, and they decided to call this new version Wilbur 2.0.	Did Wilbur get relaxed? A. Yes B. No C. Hard to say What did the team discover? A. A new function B. A glitch C. A new technique What would be the name of the new version? A. WISE 2.0 B. Wilbur 1.0 C. Wilbur 2.0
Low stem-diversity	1) WISE, a start-up company, was revising the product information ***leaflet*** for a smartwatch they developed that can track various health metrics and provide real-time feedback. 2) Before testing on humans, the team conduct trials on their experimental ***piglet***, Wilbur, whose picture was printed on the ***leaflet*** and had a skin texture similar to humans. 3) When strapping the product onto the front leg of the ***piglet***, the team noticed a small bead of sweat forming on his snout, indicating that he was stressed. 4) To comfort him, they placed a soft blanket over the ***piglet***, and Wilbur immediately nestled into it and visibly relaxed, allowing the team to collect his physiological data. 5) Thanks to the ***piglet***, the team discovered a small glitch in the device. They jotted down a note in their draft ***leaflet*** and set to work resolving the issue. 6) After data analysis, the team made some adjustments to the product information ***leaflet*** based on their findings, and they decided to call this new version Wilbur 2.0.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
Non-exposure	1) WISE, a start-up company, was developing a smartwatch that can track various health metrics like heart rate and blood pressure and provide real-time feedback to the wearer. 2) Before testing on humans, the team conduct trials on their experimental pig,	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	Wilbur, who had a patch of fur around its neck and a similar skin texture to humans. 3) When strapping the product onto Wilbur's front leg, the team observed a small bead of sweat forming on his snout, indicating that he was a bit stressed. 4) To comfort him, they placed a soft blanket over him, and Wilbur immediately nestled into it and visibly relaxed, allowing the team to collect his physiological data. 5) Thanks to Wilbur, the team discovered a small glitch in the device, jotted down a note in their lab notebook, and set to work resolving the issue. 6) After data analysis, the team made some adjustments to the device based on their findings, and they decided to call this new version of product Wilbur 2.0.	
High stem-diversity	***Story 3. Mia (-ee)*** 1) Mia was always a ***devotee*** of human rights, so she knew she had to get involved when she learned that a local charity was seeking new members. 2) She reached out to the charity and asked her former manager to be her **referee**. The manager strongly believed that she was the right person for the job. 3) Mia was thrilled to be accepted. As a new member, she had to work as a ***trainee*** first, under the supervision of an official ***employee*** of the charity. 4) At her first meeting, Mia was a bit nervous. Luckily, another ***attendee*** greeted her and introduced her to the charity and the work they had accomplished so far. 5) Her name was Nora, a former ***refugee*** who had been helped by the charity to resettle in the area and had become a ***committee*** member of the charity herself. 6) Nora's story touched Mia deeply. So Mia asked Nora to be her supervisor and ***trustee***, to manage a fundraising event for the homeless she was going to organise.	Who was looking for new members? A. A school B. A company C. A charity Who did Mia ask help? A. Her friend B. Her former manager C. Her family Did Mia get accepted by the charity? A. No B. Yes C. Did not mention How did Mia feel at her first meeting? A. Nervous B. Exciting C. Happy What was the name of the lady who talked to Mia? A. Nova B. Nora C. Nina Who was going to organise the fundraising event? A. Nora B. The charity C. Mia
Low stem-diversity	1) Mia was always committed to protecting human rights, so she signed up to be an	Same question for each story fragment as used for the

	attendee when she heard that a charity was organising a ***refugee*** recognition forum 2) . She reached out to the ***refugee*** charity and asked a for reference letter from her manager. The manager strongly believed that she would be a valuable ***attendee*** at the event. 3) Mia was thrilled to be invited. As a forum ***attendee***, she talked with the homeless representatives, under the supervision of a staff member of the charity. 4) At her first meeting, Mia was a bit nervous. Luckily, another ***attendee*** greeted her and introduced her to the charity and the work they had accomplished so far. 5) Her name was Nora, a former ***refugee*** who had been helped by the charity to resettle in the area and had become a member of the charity herself. 6) Nora's story touched Mia deeply. So Mia invited Nora to be her collaborator and adviser to manage a ***refugee*** fundraising event she was going to organise.	corresponding fragment in the high diversity condition
Non-exposure	1) Mia was always committed to protecting human rights, so she knew she had to get involved when she learned that a local charity was seeking new members. 2) She reached out to the charity and asked for a reference letter from her manager. The manager strongly believed that she was the right person for the job. 3) Mia was thrilled to be accepted. As a new member, she talked with the homeless representatives, under the supervision of a staff member of the charity. 4) At her first meeting, Mia was a bit nervous. Luckily, someone greeted her and introduced her to the charity and the work they had accomplished so far. 5) Her name was Nora, a former homeless person who had been helped by the charity to resettle in the area and had become a member of the charity herself. 6) Nora's story touched Mia deeply. So Mia invited Nora to be her collaborator to work together on a fundraising event she was going to organise.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

Story 4. David (-ee)

High stem-diversity	1) David, a ***trainee*** journalist working at a well-known local news agency, was eager to attend a charity football match to cover the event and report its impact.	What was the event that David was going to cover? A. A football match B. A baseball match C. A basketball game
	2) Ms. Smith, the owner of a local business, generously agreed to sponsor the event. She was also a charity ***trustee*** and would manage the whole fundraising process.	Who was the sponsor of the event? A. David B. Mr. Smith C. Ms. Smith
	3) Mr. Brown was another enthusiastic ***devotee*** of the charity. He was a renowned ***referee*** in the world of football who had kindly volunteered to officiate the match.	Was Mr. Brown famous? A. Yes B. No C. Did not mention
	4) David noticed a young man at the football field. His name was Ahmed. He was a ***refugee*** who had fled his home country with his family due to war.	What was the name of the young man? A. Hamid B. Amir C. Ahmed
	5) Despite all the challenges he faced, Ahmed managed to become an ***employee*** at a local company and joined the community team to pursue his passion for football.	What did Ahmed like to play? A. Basketball B. Football C. Did not mention
	6) Another ***attendee***, Ms. Jones, who worked on the ***committee*** of community services, told David that it was clear how sport brought people joy and a sense of community.	Who told David that sport brought people joy? A. Ms. James B. Ms. Jones C. Mr. James
Low stem-diversity	1) David, a young journalist working at a local news agency, was an eager ***attendee*** at a charity football match to cover this event and report its impact. 2) An important ***attendee***, Ms. Smith who owned a local business, generously agreed to sponsor the event. She also worked for a ***refugee*** charity and would manage the whole fundraising process. 3) Mr. Brown was another enthusiastic ***attendee*** at the event. He was a respected figure in the world of football who had kindly volunteered to officiate the match. 4) David noticed an ***attendee*** at the football match. His name was Ahmed. He was a ***refugee*** who had fled his home country with his family due to war. 5) Despite all challenges he faced as a ***refugee***, Ahmed managed to find a job at a local company and joined the community team to pursue his passion for football.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	6) Ms. Jones, who worked for the ***refugee*** services division, told David that it was clear how sport brought people joy and a sense of community.	
Non-exposure	1) David, a young journalist working at a well-known local news agency, was eager to attend a charity football match to cover the event and report its impact. 2) Ms. Smith, the owner of a local business, generously agreed to sponsor the event. She also worked for a charity and would manage the whole fundraising process. 3) Mr. Brown was another key participant at the event. He was a respected figure in the world of football who had kindly volunteered to officiate the match. 4) David noticed a young man at the football field. His name was Ahmed. He was a new immigrant who had fled his home country with his family due to war. 5) Despite all the challenges he faced, Ahmed managed to find a job at a local company and joined the community team to pursue his passion for football. 6) Ms. Jones, who worked for the community services department, told David that it was clear how sport brought people joy and a sense of community.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
	Story 5. Samantha (-ling)	
High stem-diversity	1) Samantha lived in the countryside. One day, she decided to plant a ***sapling*** in her backyard. She dug a hole and carefully placed the delicate ***seedling*** in the soil. 2) When watering her tiny plant, Samantha was greeted by a soft chirping. There was a ***fledgling*** bird spreading its wings and a ***hatchling*** whose eyes were still tightly shut. 3) Samantha felt amazed and blessed. She believed that this delicate ***nestling*** would soon grow stronger and start to learn how to fly and explore the world outside. 4) Over the bird's nest, Samantha saw a splash in the pond made by a tiny ***duckling*** trying to move forward in the water using its webbed feet.	Where did Samantha decided to plant the plant? A. In her garden B. In her house C. In her backyard What did Samantha hear? A. Ringing B. Chirping C. Laughter How did Samantha feel? A. Amazed and blessed B. Surprised C. Inspired What did Samantha see in the pond? A. A boat B. A reflection C. A splash

	5) It was so small but so round and fluffy. It bobbed up and down in the water like a cute little ***dumpling*** floating on the surface. 6) Samantha sat there and read a folk tale about a **changeling** raised by fairies. She didn't believe in magic, all things in nature were already magical to her!	How did Samantha feel? A. Happy B. Surprised C. Did not mention What did Samantha read? A. A folk tale B. A history book C. A picture book
Low stem-diversity	1) Samantha lived in the countryside. One day, she decided to plant something in her backyard. She dug a hole and carefully placed the delicate ***seedling*** in the soil. 2) When watering her ***seedling***, Samantha was greeted by a soft chirping. There was a young bird spreading its wings and a ***nestling*** whose eyes were still tightly shut. 3) Samantha felt amazed and blessed. She believed that this delicate ***nestling*** would soon grow stronger and start to learn how to fly and explore the world outside. 4) Not far from her ***seedling***, Samantha saw a splash in the pond made by a duck. There was also a ***nestling*** on the bank looking at its mom swimming in the water. 5) The ***nestling*** was still covered in soft downy feathers. It huddled closely with other baby ducks like a cute little marshmallow staying in a mug. 6) Samantha sat there and read a folk tale about a ***seedling*** grown by fairies. She didn't believe in magic, but all things in nature were already magical to her!	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
Non-exposure	1) Samantha lived in the countryside. One day, she decided to plant something in her backyard. She dug a hole and carefully placed the delicate tiny plant in the soil. 2) When watering her tiny plant, Samantha was greeted by a soft chirping. There was a bird spreading its wings and a baby bird whose eyes were still tightly shut. 3) Samantha felt amazed and blessed. She believed that this delicate baby bird would soon grow stronger and start to learn how to fly and explore the world outside. 4) Over the bird's nest, Samantha saw a splash in the pond made by a baby duck	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	trying to move forward in the water using its webbed feet. 5) It was so small but so round and fluffy. It bobbed up and down in the water like a cute little rice ball floating on the surface. 6) Samantha sat there and read a folk tale about a girl raised by fairies. She didn't believe in magic, all things in nature were already magical to her!	
	Story 6. Tim and Sarah (-ling)	
High stem-diversity	1) Tim and his younger sister Sarah enjoyed exploring the woods behind their house. One day, they discovered a small ***seedling*** sprouting from the ground. "Spring has arrived." Tim murmured. 2) As they headed to a nearby river to fetch water for the tiny plant, they noticed a ***duckling*** quacking and wandering around the water's edge. 3) Suddenly, Sarah's attention was seized by a peculiar, brown object gradually inching closer to them. "A **changeling**!" Sarah exclaimed. "I read about them in storybooks. They are little magical creatures!" 4) Tim approached and discovered that it was a featherless ***hatchling***. They wrapped it in a soft cloth, carefully shaping it into a small, snug ***dumpling*** to block the chilly air. 5) They decided to take the ***nestling*** home and nurse it carefully. The baby bird soon grew into a ***fledgling***, following Tim and Sarah wherever they went. 6) Years later, as they reflected on their childhood adventures, Tim and Sarah still remember how they placed their baby bird on their ***sapling***, trying to teach it how to fly.	Where did Tim and Sarah enjoy exploring? A. A mountain B. The woods C. The garden Where did they head to? A. The woods B. The house C. A river Where did they head to? A. The woods B. The house C. A river Who approached the object? A. Tim B. Sarah C. Did not mention Did they take the baby bird home? A. Yes B. No C. Did not mention What did they try to teach the bird? A. How to find food. B. How to sing. C. How to fly.
Low stem-diversity	1) Tim and his younger sister Sarah enjoyed exploring the woods behind their house. One day, they discovered a small ***seedling*** sprouted from the ground. "Spring has arrived." Tim murmured. 2) As they headed to a nearby river to fetch water for the ***seedling***, they noticed a ***nestling*** hiding under its mother's wings near the water's edge.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	3) Suddenly, Sarah's attention was seized by a peculiar, brown object inching closer to them. "A ***nestling*** bird!" Sarah exclaimed. "It must have fallen from a height. They cannot fly!" 4) Tim approached and confirmed that it was a featherless ***nestling***. They wrapped it in a soft cloth, carefully shaping it into a snug bundle to block the chilly air. 5) They decided to take the ***nestling*** home and nurse it carefully. Just like the ***seedling***, the baby bird soon grew bigger, following Tim and Sarah wherever they went. 6) Years later, reflecting on their childhood adventures, Tim and Sarah still laughed about how they placed their baby bird on their ***seedling***, hoping to teach it how to fly.	
Non-exposure	1) Tim and his younger sister Sarah enjoyed exploring the woods behind their house. One day, they discovered a small sprout from the ground. "Spring has arrived." Tim murmured. 2) As they headed to a nearby river to fetch water for the tiny plant, they noticed a baby duck quacking and wandering around the water's edge. 3) Suddenly, Sarah's attention was seized by a peculiar, brown object gradually inching closer to them. "A baby bird!" Sarah exclaimed. "It must have fallen from a height. It cannot fly!" 4) Tim approached and discovered that it was a featherless tiny creature. They wrapped it in a soft cloth, carefully shaping it into a small, snug bundle to block the chilly air. 5) They decided to take the baby bird home and nurse it carefully. The baby bird soon grew bigger, following Tim and Sarah wherever they went. 6) Years later, as they reflected on their childhood adventures, Tim and Sarah still remember how they placed their bird on a little plant, trying to teach it how to fly.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
	Story 7. Wigston (-ism)	
High stem-diversity	1) There is a remote town called Wigston. The town council decided to invest in a new ***tourism*** campaign to promote the	Who decided to invest the campaign? A. The town council B. The mayor

	town's stunning scenery and attract more visitors. 2) Drawn to the ***magnetism*** of the natural beauty as well as the value of ***liberalism*** of the town, a street artist Max arrived and began creating wall paintings. 3) As a talented artist, Max's work was rich in ***symbolism***, and it was initially met with ***skepticism***. But his vibrant art soon drew the attention of tourists and became popular. 4) Unfortunately, the increase in visitors led to an increase in ***vandalism*** in the meantime. One morning, Max was devastated to find his artwork was ruined. 5) To his surprise, the townspeople came together to help him restore the paintings, as they had a strong tradition of ***socialism*** and belief in the value of community. 6) Inspired by the generosity of the people, Max decided to incorporate elements of ***realism*** into his work, depicting the daily life of people in the town.	C. The townspeople What was Max's job? A. Designer B. Street artist C. Singer Was Max's work popular from the beginning? A. Yes B. No C. Did not mention What happened to Max's artwork? A. It's faded B. Nothing happened C. It was ruined Who helped Max? A. People of the town B. Another artist C. The town council Did Max leave the town? A. Yes B. No C. Did not mention
Low stem-diversity	1) There is a remote town called Wigston. The town council decided to invest in a new ***tourism*** campaign to promote the town's stunning scenery and attract more visitors. 2) Due to the natural beauty as well as the friendly people of the town, ***tourism*** began to boom in the town and started to impact residents' lives. 3) As a local street artist, Max's wall paintings were criticised as a form of ***vandalism*** initially. But his vibrant art soon drew the attention of tourists and became popular. 4) Unfortunately, the increase in ***tourism*** led to an increase in ***vandalism*** in the meantime. One morning, Max was devastated to find his artwork was ruined. 5) To his surprise, the townspeople came together to help him restore the paintings, as they had a strong aversion to ***vandalism*** and belief in the value of community. 6) Inspired by the ideas of sustainable ***tourism***, Max cleaned up the ***vandalism*** and put up signs around his artwork,	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	encouraging people to appreciate it without defacing it.	
Non-exposure	1) There is a remote town called Wigston. The town council decided to invest in a new campaign to promote the town's stunning scenery and attract more visitors. 2) Drawn to the natural beauty as well as the diverse culture of the town, a street artist Max arrived and began creating wall paintings. 3) As a talented artist, Max's paintings were rich in imagination, which were critisised as being too abstract initially. But his art soon drew the attention of tourists and became popular. 4) Unfortunately, the increase in visitors led to some negative impacts on the town. One morning, Max was devastated to find his artwork was ruined. 5) To his surprise, the townspeople came together to help him restore the paintings, as they had a strong tradition of charity and belief in the value of community. 6) Inspired by the generosity of the people, Max decided to incorporate daily life of people in the town into his work, transforming the everyday experiences into art.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
High stem-diversity	***Story 8. Emily (-ism)*** 1) As a travel enthusiast, Emily had been a supporter of preserving historical heritage. She felt that it was fascinating to appreciate the ***realism*** and ***symbolism*** of local landmarks. 2) One day, Emily went to a small town that was known for its ***socialism*** and hospitality. People there were cooperative and friendly, sharing resources with each other. 3) However, Emily noticed that some beautiful landmarks were defaced by some tourists because they falsely saw ***vandalism*** as a way of expressing their own beliefs and ideas. 4) Despite believing that the ***magnetism*** of the town lay in its landmarks and the kindness of its people, Emily found that the	What did Emily like to do? A. Travel B. Read C. Take pictures Where did Emily go? A. A city B. Her hometown C. A small town What happened to the landmarks? A. They were defaced. B. They were removed. C. They were repainted. What were the people in the town like? A. Traditional B. Kind

	misbehaviour of tourists was ruining her experience. 5) With a healthy dose of ***skepticism***, Emily believed that the act of damaging those landmarks was disrespectful of the values of ***liberalism*** and the town's history. 6) Emily started to worry that even though ***tourism*** brought economic growth, it could also be a threat to the local culture and people's way of life.	C. Creative How did Emily feel about the act of damaging the landmarks? A. Acceptable B. Disrespectful C. Inspiring How did Emily start to feel? A. Excited B. Curious C. Worried
Low stem-diversity	1) As a travel enthusiast, Emily had been a supporter of developing ***tourism*** in small towns. She believed that it was important to attract travellers to appreciate local landmarks. 2) One day, Emily went to a small town that was known for its ***tourism*** and hospitality. People there were cooperative and friendly, sharing resources with each other. 3) However, Emily noticed that some beautiful landmarks were defaced because some tourists falsely saw ***vandalism*** as a way of expressing their own beliefs and ideas. 4) Despite believing that the success of ***tourism*** in the town was due to its landmarks and the kindness of its people, Emily found that ***vandalism*** was ruining her experience. 5) With a deep concern about ***vandalism***, Emily believed that the act of some tourists was disrespectful to the value of public properties and the town's history. 6) Emily started to worry that even though ***tourism*** brought economic growth, the issue of ***vandalism*** could threaten the local culture and people's way of life.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition
Non-exposure	1) As a travel enthusiast, Emily had been a supporter of preserving historical heritage. She felt that it was fascinating to appreciate the beauty and culture of local landmarks. 2) One day, Emily went to a small town that was known for its unique culture and hospitality. People there were cooperative and friendly, sharing resources with each other.	Same question for each story fragment as used for the corresponding fragment in the high diversity condition

	3) However, Emily noticed that some beautiful landmarks were defaced by tourists because some of them falsely saw defacing public property as a form of self-expression. 4) Despite believing that the charm of the town lay in its landmarks and the kindness of its people, Emily found that the misbehaviour of tourists was ruining her experience. 5) With a deep concern about the cultural heritage, Emily believed that the act of damaging those landmarks was harmful and disrespectful of the town's history. 6) Emily started to worry that even though visitors brought economic growth, their improper conduct could also be a threat to the local culture and people's way of life.

Note. The eight stories were presented in randomised order (story fragment order was fixed).

Appendix 5C: Stimuli Lists (A and B) for the Pre- and Post- Lexical Decision Task.

Suffix	**List A**		**List B**	
	Pseudomorphemic	*Control*	*Pseudomorphemic*	*Control*
-ism	milkism	milkirm	saltism	saltirm
	gasism	gasirm	deerism	deerirm
	starism	starirm	chestism	chestirm
	fanism	fanirm	toolism	toolirm
	brickism	brickirm	cupism	cupirm
	smellism	smellirm	tankism	tankirm
	boxism	boxirm	mapism	mapirm
	cellism	cellirm	cheekism	cheekirm
	fuelism	fuelirm	sonism	sonirm
-ee	toyee	toyea	poolee	poolea
	hallee	hallea	jawee	jawea
	waistee	waistea	deskee	deskea
	flagee	flagea	wristee	wristea
	ballee	ballea	railee	railea
	crewee	crewea	songee	songea
	strawee	strawea	guardee	guardea
	tailee	tailea	skyee	skyea
	soulee	soulea	shipee	shipea
-let	sheetlet	sheetlat	mouthlet	mouthlat
	laplet	laplat	earlet	earlat
	feastlet	feastlat	meatlet	meatlat
	kinglet	kinglat	decklet	decklat
	sandlet	sandlat	caplet	caplat
	bandlet	bandlat	clocklet	clocklat
	carlet	carlat	threadlet	threadlat
	foglet	foglat	boatlet	boatlat
	chairlet	chairlat	windlet	windlat
-ling	rockling	rocklang	cliffling	clifflang
	lipling	liplang	maidling	maidlang
	messling	messlang	stormling	stormlang
	beachling	beachlang	grassling	grasslang
	floodling	floodlang	coachling	coachlang
	beanling	beanlang	sunling	sunlang
	craftling	craftlang	moonling	moonlang
	glassling	glasslang	roofling	rooflang
	nutling	nutlang	beltling	beltlang

Appendix 5D. Detailed Summary of Accuracy Model

Fixed effects	β	SE	z value	p-value
Intercept	2.373	0.086	27.576	< 0.001***
Nonword type	-0.817	0.113	-7.231	< 0.001***
Test	0.009	0.086	0.108	0.914
Exposure_condition[1]	0.046	0.086	0.542	0.588
Exposure_condition[2]	0.022	0.050	0.438	0.661
Nonword type: Test	0.416	0.094	4.419	< 0.01**
Nonword type: Exposure_condition[1]	0.053	0.065	0.814	0.416
Nonword type: Exposure_condition[2]	0.059	0.038	1.548	0.122
Test: Exposure_condition[1]	0.066	0.103	0.646	0.518
Test: Exposure_condition[2]	0.134	0.060	2.243	0.025*
Nonword type: Test: Exposure_condition[1]	0.002	0.110	0.021	0.983
Nonword type: Test: Exposure_condition[2]	0.045	0.065	0.689	0.491

Note: The estimates presented in the table were obtained using the '*summary()'* function from the '*lmerTest'* package in R.

Appendix 5E. Pairwise Comparisons of Accuracy (Bonferroni Corrected)

Test Time	Exposure Condition	Contrast	Estimate	SE	z-Ratio	Adjusted p-Value
Post	Low	Pseudo - Control	-0.745	0.151	-4.932	< 0.0001
Pre	Low	Pseudo - Control	-1.113	0.157	-7.109	< 0.0001
Post	High	Pseudo - Control	-0.637	0.159	-4.000	0.0001
Pre	High	Pseudo - Control	-1.010	0.162	-6.242	< 0.0001
Post	Non-exposure	Pseudo - Control	-0.446	0.161	-2.775	0.0055
Pre	Non-exposure	Pseudo - Control	-0.951	0.156	-6.110	< 0.0001

Appendix 5F. Detailed Summary of RT Model

Fixed effects	β	SE	t value	p-value
Intercept	6.7444	0.0130	518.596	< 0.001***
Nonword type	0.0545	0.0096	5.654	< 0.001***
Test	-0.1452	0.0126	-11.529	< 0.001***
Exposure_condition [1]	-0.0009	0.0151	-0.062	0.9508
Exposure_condition [2]	0.0106	0.0088	1.211	0.2278
Nonword type: Test	-0.0126	0.0093	-1.359	0.1765
Nonword type: Exposure_condition[1]	-0.0003	0.0063	-0.049	0.9607
Condition1: Exposure_condition[2]	0.0059	0.0037	1.621	0.1074
Test: Exposure_condition[1]	0.0277	0.0154	1.802	0.0738 .
Test: Exposure_condition[2]	0.0066	0.0089	0.737	0.4625
Nonword type: Test: Exposure_condition[1]	0.0030	0.0113	0.261	0.7938
Nonword type: Test: Exposure_condition[2]	-0.0072	0.0065	-1.114	0.2653

Note: The estimates presented in the table were obtained using the '*summary()'* function from the '*lmerTest'* package in R.

Appendix 5G. Pairwise Comparisons of RTs (Bonferroni Corrected)

Test Time	Exposure Condition	Contrast	Estimate	SE	z-Ratio	Adjusted p-Value
Post	Low	Pseudo - Control	-0.745	0.151	-4.932	< .0001
Pre	Low	Pseudo - Control	-1.113	0.157	-7.109	< .0001
Post	High	Pseudo - Control	-0.637	0.159	-4.000	0.0001
Pre	High	Pseudo - Control	-1.010	0.162	-6.242	< .0001
Post	Non-exposure	Pseudo - Control	-0.446	0.161	-2.775	0.0055
Pre	Non-exposure	Pseudo - Control	-0.951	0.156	-6.110	< .0001