Second Language Processing of Derivational and Inflectional Morphology in English

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Yoonsang Song, M.A.

Washington, DC
May, 2015

SECOND LANGUAGE PROCESSING
OF DERIVATIONAL AND INFLECTIONAL MORPHOLOGY IN ENGLISH

Yoonsang Song, M.A.

Dissertation Advisor: Donna Lardiere, Ph.D.

ABSTRACT

This dissertation investigates how later second language (L2) learners process derivational and inflectional morphology to explore whether later L2 learners can develop lexical or syntactic representations qualitatively similar to native speakers'. As the current literature does not provide a conclusive answer to this question, two experiments were conducted to address this gap.

Experiment 1 investigates how later learners process English derived words. Native English speakers and native Korean-speaking advanced learners of English were compared in their performance on an overt-priming lexical decision task. The results show that not only the native speakers but also the L2 learners were able to morphologically decompose derived words in English. Both groups' data showed a robust morphological priming effect with semantically transparent morphologically related prime-target pairs (e.g., *politeness-polite*). Furthermore, both groups' priming effects were similarly affected by major determinants of the morphological decomposability of derived words (whole-word frequency, base frequency, and morphological productivity).

Experiment 2 examines how later learners of English process plural inflection in English. In a self-paced reading task, native English speakers and native Korean-speaking advanced learners of English were tested on their (in)sensitivity to plural errors in two different structures: a simple DP/QP structure (e.g., *those long Latin word*(s)*) and a partitive structure (*many of her*

book(s)*). The results indicate that despite the striking differences in plural marking between English and Korean, the learners were sensitive to plural errors in both structures, and that their sensitivity to the errors was affected by the structural distance of the feature-checking dependency related to plural inflection. These findings suggest that (1) L2 learners can acquire target-like L2 inflection knowledge, even if such inflection knowledge is not present in their L1, contra the representational deficit account (e.g., Hawkins & Hattori, 2006), and that (2) they compute hierarchically structured representations during real-time language comprehension, contra the shallow structure hypothesis (Clahsen & Felser, 2006).

Taken together, the results of the two experiments suggest that, although there are many differences between L1 and L2 learners' knowledge of the target language, advanced L2 learners can develop target-like mental representations of certain aspects of lexical and syntactic knowledge.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Donna Lardiere, for her guidance and support throughout all phases of my dissertation research and graduate study at Georgetown. My gratitude to her simply goes beyond words. My heartfelt gratitude also goes to the other members of my committee, Nan Jiang and Elissa Newport, for their interest in and support for this dissertation. In addition, I would like to extend my sincere gratitude to Lourdes Ortega for giving me invaluable advice in the process of getting published an earlier article on a study of my dissertation research (Song, 2015), and to William O'Grady for introducing me to linguistics.

I am also grateful to Héctor Campos for patiently discussing the syntactic analysis of partitive constructions, and to Amir Zeldes for spending his time and effort discussing various issues in morphological productivity. Furthermore, I am indebted to the reviewers of the Song (2015) article for their helpful comments, and to the participants of my dissertation research for their earnest participation.

Finally and most importantly, I would like to thank my wife, Youngah. Her support, encouragement, and love have been my greatest source of energy not only in writing my dissertation but also in my life.

TABLE OF CONTENTS

Chapter 1 Introduction 1

1.1 L2 Processing of Derivational Morphology 2

1.2 L2 Processing of Inflectional Morphology 4

1.3 Organization of the Dissertation 6

Chapter 2 L2 Processing of Derived Words 8

2.1 Previous Research on L2 Processing of Derived Words Using the Masked Priming Paradigm 8

2.2 The Central Representation of Derived Words in L2 & Overt Priming Paradigms 15

2.3 Summary 17

Chapter 3 Factors Influencing Processing of Derived Words 18

3.1 Whole-Word Frequency 18

3.2 Base Frequency (and Relative Frequency) 24

3.3 Productivity of Morphological Constituents 28

3.3.1 Productivity of Base: Family size 29

3.3.2 Productivity of affixes 33

3.4 Summary 47

Chapter 4 Experiment 1: Lexical Priming Experiment 49

4.1 Method 52

4.1.1 Participants 52

4.1.2 Materials 55

4.1.3 Procedure 64

4.2 Results 65

4.2.1 Native English Speakers 68

4.2.2 L2 Learners 71

4.3 Discussion 75

4.3.1 L1 and L2 Representation of Derived Words in English 76

4.3.2 Differences between L1 and L2 Processing of Derived Words in English 79

4.4 Limitations and Future Research 81

4.5 Conclusion 83

Chapter 5 Difficulties in L2 Inflection and L2 Processing of Plural Inflection in English 85

5.1 Representational Deficit Account vs. Performance Deficit Account 85

5.2 The Shallow Structure Hypothesis 90

5.3 L2 English Plural Inflection 92

5.3.1 Conflicting Results of Previous Studies 92

5.3.2 Different Types of Plural Inflection in English 93

Chapter 6 Experiment 2: Self-Paced Reading Experiment 98

6.1 Method 107

6.1.1 Participants 107

6.1.2 Materials 109

6.1.3 Procedure 113

6.2 Results 114

6.2.1 Comprehension Question Accuracy 114

6.2.2 Self-paced Reading 114

6.3 Discussion 126

6.3.1 L2 Representation of Plural Inflection in English 127

6.3.2 The SSH Revisited 129

6.3.3 Differences between L1 and L2 Processing 130

6.4 Conclusion 132

Chapter 7 Conclusion 136

7.1 Lexical and Syntactic Representations in L2 136

7.2 Differences between L1 and L2 Processing 138

7.3 Future Avenues of Study 139

Appendices 140

Appendix A: Critical Prime-Target Pairs and their Characteristics 141

Appendix B: Filler Prime-Target Pairs 146

Appendix C: RT Results of the Intermediate Korean ESL Learners 149

Appendix D: Test materials 150

Bibliography 154

Chapter 1 Introduction

This dissertation investigates one of the most controversial issues in the field of second language acquisition, namely, whether learners of a second language (L2) can develop mental representations (or grammatical knowledge) similar to those of native speakers of the language, or if instead the mental representations that learners develop are fundamentally different from those of native speakers. The scope of the study's investigation of this question is limited to derivational and inflectional morphology. The importance of pursuing this issue becomes apparent when we consider that the correct understanding of L2 acquisition can lead us to more appropriate theoretical and pedagogical approaches to L2 research. It is therefore no surprise that a large number of studies have delved into the issue. Unfortunately, however, our picture of the nature of L2 acquisition remains unclear. On the one hand, a group of researchers has argued that the development of second languages fundamentally differs from the development of native languages, especially for late learners who begin to learn their second languages after a hypothesized critical period (Lenneberg, 1967). This idea is encountered in various hypotheses in the current literature including the Fundamental Difference Hypothesis (Bley-Vroman, 1989, 2009), the Failed Functional Features Hypothesis (Hawkins & Chan, 1997), the Interpretability Hypothesis (Tsimpli & Mastropavlou, 2007), the Shallow Structure Hypothesis (Clahsen & Felser, 2006b), and the Morphological Congruency Hypothesis (Jiang, Novokshanova, Masuda, & Wang, 2011). On the other hand, other researchers have claimed that late L2 learners can eventually acquire various aspects of the target language and have provided evidence against these fundamental difference hypotheses (e.g., Dekydtspotter, Schwartz, & Sprouse, 2006; Hopp, 2010; Keating, 2009; Omaki & Schulz, 2011).

This dissertation provides another piece of supporting evidence for the latter position. To be specific, I provide evidence that advanced L2 learners are able to develop target-like knowledge of at least some aspects of lexical and syntactic representations (e.g., morphological structure or hierarchical sentence structure), demonstrating that L1 and L2 processing of derivational and inflectional morphology can be similar in some core aspects, although they may differ from each other in other ways.

1.1 L2 Processing of Derivational Morphology

This dissertation first focuses on L2 processing of derivational morphology (Chapters 2–4). Derivational morphology, which constitutes one of the three major subdomains of morphology along with inflectional morphology and compounding (e.g., Bauer, 1983), refers to the morphological process of creating a new lexeme from an already-existing lexeme. In English, this process often occurs by adding particular classes of affixes. For instance, when the suffix *re-* is added to a verb (e.g., CREATE[1]), it creates another verb (e.g., RECREATE). Although L2 morphology has always been one of the central issues of L2 research, unfortunately, derivational morphology has been overlooked by most of the second language research community, with more attention paid to inflectional morphology (e.g., Franceschina, 2005; Hawkins & Liszka, 2003; Hopp, 2010; Jiang et al., 2011; Keating, 2009; Lardiere, 1998, 2000, 2007; Prévost & White, 2000, among many others). Only a very small number of studies have investigated L2 knowledge of derivational morphology (e.g., Friedline, 2011; Hayashi & Murphy, 2010; Lardiere, 2006; Schmitt & Zimmerman, 2002) and L2 processing of derivational morphology (Clahsen &

[1] Lexemes are usually written in (small) capital letters to distinguish them from actual word-forms,

Neubauer, 2010; De Grauwe, Lemhöfer, Willems, & Schriefers, 2014; Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Kirkici & Clahsen, 2013; Rehak & Juffs, 2011; Silva & Clahsen, 2008).

Against this background, the present dissertation investigates how native English speakers and native Korean-speaking late learners[2] of English process derived words in English (e.g., *polite-ness* or *humor-ous*) to better understand how these words are represented in the L2, using an overt priming paradigm (see Chapter 2). Two specific research questions are addressed in the dissertation. The first question is whether late advanced L2 learners of English are able to decompose English derived words into their morphological constituents at the central lexical representation level (see Chapter 2). Previous studies with various experimental methodologies provide ample evidence that across languages, at least certain types of derived words are decomposed in L1 lexical processing (e.g., Bölte, Schulz, & Dobel, 2010; Duñabeitia, Perea, & Carreiras, 2007; Lehtonen, Monahan, & Poeppel, 2011; Longtin, Segui, & Halle, 2003; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Orfanidou, Davis, & Marslen-Wilson, 2011; Vannest & Boland, 1999). However, L2 research on this question is scarce, and the results are, moreover, inconsistent (Clahsen & Neubauer, 2010; Diependaele et al., 2011; Silva & Clahsen, 2008).

The second question is whether L2 processing of English derived words is affected by the same factors that L1 studies have found to be major determinants of the morphological decomposability of given derived words—namely, whole-word frequency, base frequency, and morphological productivity (e.g., Boudelaa & Marslen-Wilson, 2011; Ford, Davis, & Marslen-Wilson, 2010; Hay, 2001; Meunier & Segui, 1999). The goal of this research question is to

---

[2] In this dissertation, late learners of a second language are defined as learners who began to learn their second language at around 13 years old in classroom contexts.

compare L1 and L2 processing of English derived words in detail in order to examine whether any substantial qualitative similarity between L1 and L2 processing of derivational morphology exists. To the best of my knowledge, no previous study has explored the extent to which such lexical or distributional properties influence L2 processing of derived words.

1.2 L2 Processing of Inflectional Morphology

In Chapter 5, the dissertation shifts focus to L2 processing of inflectional morphology. The major function of inflectional morphology is providing grammatically appropriate forms of a single lexeme. For instance, via inflection, the lexeme CREATE, an abstract vocabulary unit with a specific sound /kɹieɪt/ and a specific meaning 'to bring something into existence', is realized by one of its word forms *create*, *creates*, *created*, or *creating* in actual speech or writing. Here, it should be noted that the appropriate word form is determined by the syntactic context in which the lexeme occurs. For instance, the word form *created* appears in a past tense or perfect aspect context, whereas the word form *creating* appears in the progressive aspect context. For this reason, inflection is regarded as relevant to syntax (see Anderson, 1982 for discussion).

L2 acquisition of inflectional morphology has received much attention, and it has been consistently reported in the current literature that inflectional morphology is extremely difficult for L2 learners to master. In many cases, highly advanced learners who otherwise are very proficient in the target language (TL) nonetheless display variability in their use of inflectional morphemes (e.g., Haznedar & Schwartz, 1997; Lardiere, 2007; Long, 2003; Prévost & White, 2000). For example, the Chinese participant in Lardiere's (2007) longitudinal case study revealed very low accuracy in the use of the English regular past-tense suffix (about 5%), although she spoke English very fluently and had been living in the United States for over 20 years. Late L2 learners' variability in target inflectional morphology seems not to be limited to production but also extends to comprehension. For instance, McCarthy (2008) reports that her advanced English-speaking learners of Spanish exhibited variability in Spanish number and gender inflection in both a production task and a comprehension task. In addition, some studies have observed that in online reading comprehension tasks, advanced learners were insensitive to inflectional errors in some contexts (Jiang, 2011; Keating, 2009).

Traditionally, the variability in L2 comprehension and production of inflection has been explained by two contrasting theoretical approaches. Some researchers attribute it to representational deficits in L2 grammar due to incomplete acquisition of target representations (e.g., Franceschina, 2001; Hawkins & Hattori, 2006; Hawkins & Liszka, 2003; Jiang et al., 2011; Sabourin & Stowe, 2008). I call this explanation the "representational deficit account" (RDA), following Hawkins and colleagues (see, e.g., Snape, Leung, & Sharwood Smith, 2009). In contrast, other researchers maintain that the variability of L2 inflection reflects a performance deficit, rather than a representational deficit (e.g., Haznedar & Schwartz, 1997; Keating, 2009; Lardiere, 1998; Prévost & White, 2000; Sagarra & Herschensohn, 2010; Wen, Miyao, Takeda, Chu, & Schwartz, 2010; White, Valenzuela, Kozlowska-MacGregor, & Leung, 2004). On this view, L2 learners' variable performance with target inflections is due to processing limitations, such as relatively lower L2 working memory span (e.g., Service, Simola, Metsänheimo, & Maury, 2002). According to this explanation, higher working memory demands in the L2 make it harder for learners to hold in working memory information about grammatical features (e.g., person, gender, number, etc.) until those features are checked for inflection, as compared to native speakers. I call this approach the "performance deficit account" (PDA). The difficulty of acquiring L2 inflection has recently been accounted for by another theoretical approach as well, namely, Clahsen and Felser's (2006b) Shallow Structure Hypothesis (SSH) (e.g., Keating, 2009, 2010). According to this hypothesis, L2 learners construct syntactically less detailed representations, and thus they are capable of checking grammatical features only in local domains. On this view, L2 learners are expected to be incapable of native-like processing of inflection that involves nonlocal checking of grammatical features.

The present dissertation tests the RDA and the SSH by investigating whether late L2 learners whose L1 differs considerably in plural inflection from English can attain target-like knowledge of the English plural inflection. Specifically, using an online reading comprehension task, I examine whether native Korean-speaking advanced learners of English as a second language (ESL) are sensitive to missing plural inflections even when successful nonlocal checking of grammatical features is required for such sensitivity. In order to further test the SSH, I also explore whether they are sensitive to the structural distance of a feature-checking dependency. Regarding the PDA, the background literature as well as the current findings relevant to that account will be discussed where appropriate, although the current study does not provide direct evidence for or against the PDA because it did not employ measures of processing capacity, such as working memory.

1.3 Organization of the Dissertation

The dissertation is organized as follows. Chapters 2–4 focus on L2 processing of derivational morphology. First, previous L2 studies on the processing of derived words are reviewed in Chapter 2, and then previous L1 studies are reviewed in Chapter 3, with a focus on the major determinants of the morphological decomposability of derived words. Chapter 3 also discusses relevant morphological processing models where appropriate. Chapter 4 presents a morphological priming experiment designed to investigate highly advanced late L2 learners' processing of derived words in English. The chapter first explains the experimental design and then reports the results and discusses them in detail. Chapter 5 shifts the scope of the dissertation's investigation to L2 processing of inflectional morphology. In this chapter, the relevant theoretical approaches to L2 inflection, namely the RDA, the PDA, and the SSH, are thoroughly discussed, and previous studies on L2 processing of English plural inflection are reviewed. Chapter 6 presents a self-paced reading experiment designed to test the RDA and the SSH. As in Chapter 4, the experimental design is discussed first, and then the results are reported and discussed in detail. Chapter 7 summarizes the findings and concludes the dissertation.

Chapter 2 L2 Processing of Derived Words

2.1 Previous Research on L2 Processing of Derived Words Using the Masked Priming Paradigm

Although studies on L2 processing of derivational morphology are scarce as discussed in Section 1.1, they have consistently pursued the same issue—namely, whether there are fundamental differences between L1 and L2 in processing derived words. To be specific, assuming that L1 processing of derived words involves morphological parsing, the L2 research explores whether L2 processing involves morphological parsing as well. As briefly mentioned in Section 1.1 and to be discussed in more detail in Chapter 3, the current literature provides ample evidence that native speakers process at least some derived words via morphological decomposition. In contrast, the L2 studies on this issue are limited in number, and their results are apparently inconsistent, although most of them have employed the same experimental methodology of *masked* morphological priming paradigms.

In morphological priming paradigms, for each trial, participants decide whether or not a letter string—which is called the target—is an existing word in a particular language; typically, the target is preceded by either a word that is morphologically related to the target (related prime, e.g., *baker-bake*) or a word that is not related in any way to the target (unrelated prime, e.g., *flood-bake*). Participants' reaction times for such lexical decisions are measured, and the lexical decision times in the related and unrelated conditions are compared. The major interest is whether the lexical decision time in the related condition is significantly less than that in the unrelated condition. Such a reduced lexical decision time is called a morphological priming effect.

In masked priming paradigms (Forster & Davis, 1984), the prime is preceded (i.e., masked) by a mask of marks (e.g., #######), which typically stays on the screen for 500 ms. Then, the prime is presented for such a short time (often less than 60 ms) that it cannot be consciously perceived. The target then immediately appears, replacing the prime, and participants make a lexical decision on the target. Crucially, in the current literature, the masked priming paradigm is assumed to tap into the *surface* level of lexical representation, and thus, the results of previous L2 studies do little to help us understand the central level of lexical representation, which includes syntactic and semantic representations (see, e.g., Havas, Rodríguez-Fornells, & Clahsen, 2012; Marslen-Wilson, 2007; Schreuder & Baayen, 1995, for discussion of lexical level distinctions; Havas et al.'s and Schreuder & Baayen's ideas are visualized in Figure 1). This assumption of the masked priming paradigm is based on a consistent finding of L1 studies (e.g., Lehtonen, Monahan, & Poeppel, 2011; Longtin & Meunier, 2005; McCormick, Rastle, & Davis, 2008, among many others; cf. Feldman, O'Connor, & del Prado Martín, 2009): Robust priming effects are observed not only with prime-target pairs in genuine (i.e., semantically strongly related) morphological relationship (e.g., *baker-bake*) but also with those in *pseudo* (or apparent) morphological relationship (e.g., *corner-corn*), but not with only orthographically related prime-target pairs (e.g., *brothel-broth*). Note that the prime in the pseudo morphological relationship can be parsed into a free standing base and an affix (e.g., *corner* → *corn* + *-er*), while the prime in the orthographic relationship (e.g., *brothel*) cannot be morphologically parsed, although both types of prime-target pairs are equally semantically unrelated. These observed effects have been interpreted to suggest that at an early phase of lexical processing, any word whose *surface* form can be segmented into a legitimate stem and a legitimate affix in the language (i.e., words that are morpho-orthographically segmentable) is obligatorily decomposed into the potential morphemic constituents regardless of the semantic relation between the constituents and the whole word.

(a) (b)

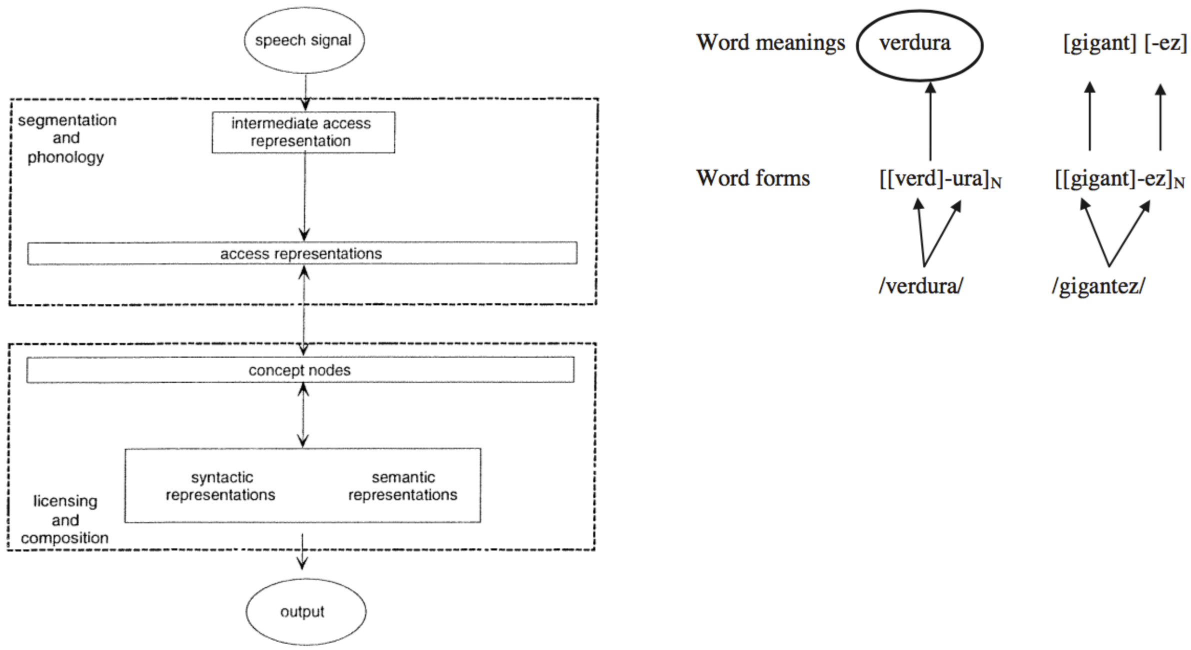


Figure 1. Distinction of surface and central levels of lexical representation: (a) the general outline of Schreuder and Baayen's (1995) morphological processing model and (b) lexical representations of morphologically complex words proposed by Havas et al. (2012)

Silva and Clahsen (2008) applied such a masked priming paradigm to the investigation of L2 processing of derived words. One of their specific interests was how native Chinese-speaking and native German-speaking advanced adult learners of English process derivatives of deadjectival suffixes *-ness* and *-ity*, in comparison with native English speakers (their Experiments 3 and 4). The design and results of the study are summarized in Table 1.

Table 1. Design and results of Silva & Clahsen's (2008) Experiments 3 & 4

	Priming Conditions	Prime	Target	Lexical Decision Reaction Times (ms)*		
				L1	Chinese L2	German L2
Experiment 3 (-*ness*)	Identity	*bold*	*bold*	454	642	548
	Test	*boldness*	*bold*	460	745	617
	Unrelated	*rough*	*bold*	504	842	669
Experiment 4 (-*ity*)	Identity	*fertile*	*fertile*	511	696	588
	Test	*fertility*	*fertile*	496	768	619
	Unrelated	*strange*	*fertile*	533	883	702

As Table 1 shows, three priming conditions were compared in terms of lexical decision latency: (1) identity, (2) test, and (3) unrelated. In these conditions, the same target (i.e., bases of -*ness* or -*ity* derivatives) was presented, but different types of primes preceded it. In the identity condition, the prime was identical to the target, while in the unrelated condition, the prime was not related to the target. The prime in the test condition was the -*ness* or -*ity* derivative of the target.

Native speakers and L2 learners exhibited somewhat different priming effects. First, native speakers exhibited a *full* priming effect for both -*ness* and -*ity* derivatives. Their reaction times (RTs) to the target in the lexical decision were significantly reduced to the same degree in the test condition and in the identity condition, compared to their RTs in the unrelated condition. On the other hand, only a *partial* priming effect was found with L2 learners. That is, although the

RTs to the target were significantly reduced in the test condition, compared to the RTs in the unrelated condition, they were not reduced as much as the RTs in the identity condition. Based on these results, Silva and Clahsen (2008) concluded that, although L2 learners have morphologically-structured representations for derived words, some or all of the processes involved in derivational morphological processing (e.g., extracting the base and the affix from the morphologically-structured representation of a derived word or retrieving them from entries in the lexicon) function less efficiently in L2 learners than in native speakers.

Furthermore, Clahsen and Neubauer (2010) failed to observe even a partial priming effect with advanced learners of German. In a masked priming experiment (Experiment 2), native German speakers and native Polish-speaking advanced adult learners of German were compared in terms of how they process the German deverbal suffix *-ung*, which is considered fully productive and phonologically transparent. The experimental design was similar to that of Silva and Clahsen's (2008) experiments, and is summarized along with the results in Table 2.

Table 2. Design and results of Clahsen & Neubauer's (2010) Experiment 2

Priming Conditions	Prime	Target	Lexical Decision RTs (ms)*	
			L1	L2
Identity	*nutzen* 'to use'	*nutzen*	660 (139)	676 (146)
Test	*Nutzung* 'use'	*nutzen*	661 (120)	725 (172)
Unrelated	*Mischung* 'mixture'	*nutzen*	715 (160)	721 (162)

*Standard deviation in parentheses

As illustrated in Table 2, in all three priming conditions, the targets were always the infinitival forms of the given verbs—which end mostly with *-en*, but in some cases with *-ern*, *-eln*, *-n*—and the test primes were the -*ung* derivatives of the verbs. The results showed that for native German speakers, the target verbs were fully primed by their -*ung* derivatives, whereas for the L2 learners, the target verbs were not primed at all by their -*ung* derivatives. With L2 learners, priming occurred only when the target and prime were the same (i.e., only in the identity condition). Clahsen and Neubauer (2010) interpreted their results within the framework of Ullman's (2004, 2005) declarative/procedural model, according to which late L2 learners tend to depend more on the declarative memory system (e.g., lexical storage) and less on the procedural memory system (e.g., rule application) for maturational reasons: Adult L2 learners rely heavily on direct look-up rather than morphological parsing for derived word processing, since they mainly rely on the declarative memory system for processing of derived words.

On the other hand, some studies provide evidence that native speakers and L2 learners rely on the same strategies in the processing of derived words. For instance, Diependaele et al. (2011) showed that native Spanish-speaking and native Dutch-speaking learners of English behaved in similar ways to native English speakers when they processed derived words, in a masked priming experiment. There were three experimental conditions: (1) transparent, (2) opaque, and (3) form. In the transparent condition, the critical prime was a genuine derivative of the target (e.g., *viewer-view*). In contrast, the critical prime in the opaque condition was morphologically unrelated to the target, but could apparently be parsed into the target and a suffix (e.g., *corner-corn*). The critical prime in the form condition was related to the target only in orthography (e.g., *freeze-free*). In all these conditions, each target was also preceded by a control prime (in addition to the critical prime), which was unrelated to the target. Both native speakers and L2 learners exhibited the largest priming effects in the transparent condition, moderate priming effects in the opaque condition, and no priming effect in the form condition.

The authors claimed that these results suggest that similar principles underlie L1 and L2 processing of derived words.

In another study by Kirkici and Clahsen (2013), L2 Turkish speakers performed in a native-like way when processing derivatives of the deadjectival suffix -*lIk*, which is a productive and transparent Turkish derivational affix. In terms of its function, this suffix corresponds to the suffix -*ness* in English. In a masked priming experiment (Experiment 1), native speakers of Turkish and adult L2 learners of Turkish with diverse L1 backgrounds were compared with regard to how they processed -*lIk* derivatives.[3] The design and results of the experiment are given in Table 3.

Table 3. Design and results of Kirkici and Clahsen's (2013) Experiment 1

	Prime	Target	Lexical Decision RTs (ms)*	
			L1	L2
Related Condition	*delilik* 'madness'	*deli* 'mad'	592 (82)	673 (132)
Unrelated Condition	*çağdaş* 'contemporary'		627 (117)	722 (173)
Priming Effect			35	49

*Standard deviation in parentheses

As illustrated in Table 3, the experiment had two priming conditions, related and unrelated. In the related conditions, the adjective target was preceded by the derivative of the suffix -*lIk*, while in the unrelated conditions, the target was preceded by a word unrelated morphologically,

[3] Although I discuss only the results of -*lIk* derivative processing, this experiment also investigated processing of words inflected by a regular inflectional Turkish suffix.

orthographically, or semantically to the target. A priming effect was observed with both the L1 and L2 groups, and the effect for the L2 group was numerically larger than the effect for the L1 group.

2.2 The Central Representation of Derived Words in L2 & Overt Priming Paradigms

Unlike surface lexical representation, central lexical representation, which includes syntactic and semantic representations, is assumed to be tapped into by employing *overt* (or unmasked immediate) priming paradigms, in which primes are completely visible (fully-visible paradigm) or audible (auditory-visual cross-modal paradigm), such that participants can clearly recognize and fully process them (see Marslen-Wilson, 2007, for discussion). This assumption is based on the finding, reported by several studies, that in overt priming paradigms, only prime-target pairs in genuine (i.e., semantically closely related) morphological relation, such as *sadness*-*sad*, produce priming effects; unlike in the masked priming paradigm, prime-target pairs in pseudo morphological relation, such as *corner-corn*, do not exhibit priming effects (e.g., Lavric, Rastle, & Clapp, 2011; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Rastle, Davis, Marslen-Wilson, & Tyler, 2000, among others). These studies' results indicate that the semantic transparency of derived words is an important factor in overt priming. Given that semantic information is considered a property of central lexical representation (e.g., Schreuder & Baayen, 1995), the results further suggest that overt priming is mainly induced by properties of the central lexical representation of derived words.

At this point, it should be emphasized that the overt priming effect from morphologically related prime-target pairs is *morphological* in nature, and cannot be simply reduced to the semantic and orthographic overlap between the prime and the target. Previous studies have demonstrated that prime-target pairs related only orthographically, such as *electrode*-*elect*, yield no priming effect or even an inhibitory effect (e.g., Longtin, Segui, & Halle, 2003; Rastle et al., 2000), and those related only semantically, such as *cello*-*violin*, yield a much weaker priming effect than morphologically related pairs (e.g., Feldman & Soltano, 1999; Rastle et al., 2000). Furthermore, Feldman and Soltano (1999) demonstrated that the morphological priming effect is not a combination of semantic and orthographic effects. In addition, the current literature provides neurological evidence that the priming effects reported in behavioral studies (i.e., reduced RTs for morphologically related pairs) cannot be attributed only to semantic and orthographic overlap (e.g., Beyersmann, Iakimova, Ziegler, & Colé, 2014; Lavric et al., 2011). For example, Beyersmann et al.'s (2014) ERP data show significant morphological but not semantic priming at 100–250 ms (i.e., significantly strong positive potentials [P200] only for morphologically related pairs). And other studies have shown that only morphologically related pairs reduced the amplitude of the N400 (a negative-going waveform that peaks around 400 ms after stimulus onset), which is interpreted as a morphological priming effect (e.g., Morris, Frank, Grainger, & Holcomb, 2007). Therefore, researchers have generally interpreted such a priming effect as evidence for repeated access to the lexical representation shared by the prime and the target (e.g., the root morpheme {sad}), and therefore as evidence for morphological decomposition, following Marslen-Wilson et al. (1994).

Considering the assumptions of the overt priming paradigms (and the masked priming paradigm), the central lexical representation of derived words in L2 has not yet been explored. As mentioned in the previous section, previous studies on derived words in L2 have uniformly used a masked priming paradigm. To the best of the author's knowledge, no study has employed overt priming paradigms to explore the central representation of derived words in L2.

2.3 Summary

In sum, only a limited number of studies have explored the question of how L2 speakers process derived words, and their results are apparently inconsistent. Furthermore, as these studies uniformly employed masked priming paradigms, which have been assumed to tap into only the surface representation level (i.e., orthography) of lexical items, the domain of the previous studies on L2 processing of derived words is limited to the surface (orthographic) level of the representation of derived words, and the central representation level of derived words in L2 has been left unexplored. In addition, all the previous L2 studies investigated the issue of whether or not L2 learners are able to process derived words via morphological parsing, using as experimental stimuli derivatives of certain affixes (e.g., -*ness* or -*ity*) matched in root and whole-word frequency (e.g., Silva & Clahsen, 2008) or derivatives controlled according to a number of lexical or distributional properties (such as root family size and affix type count; e.g., Diependaele et al., 2011). None of these studies, however, has explored the question of whether and how each of the lexical or distributional properties that have been claimed to be important determinants of morphological decomposability in L1 processing influences L2 processing of derived words, and how they interact with one another. In Chapter 3, I will discuss these lexical and distributional properties in detail.

Chapter 3 Factors Influencing Processing of Derived Words

The current literature suggests that various properties of derived words may influence their decomposability. Among those properties, the *whole-word frequency* and the *base frequency* of derived words and the *productivity of morphological constituents* of derived words stand out in the literature. This chapter discusses these properties in detail.

3.1 Whole-Word Frequency

Whole-word frequency (also known as surface or word frequency) of a derived word, which refers to the number of occurrences of the derived word as a specific whole word (Burani & Caramazza, 1987; Colé, Beauvillain, & Segui, 1989), seems to stand out as one of the most important factors in determining how derived words are processed. First, in most morphological processing models, this property of morphologically complex words (including derived words) is assumed to be a major determinant of how the words are processed and/or how they are represented in the mental lexicon (e.g., Baayen, 1992, 1993; Burani & Laudanna, 1992; Bybee, 1985; Caramazza, Miceli, Silveri, & Laudanna, 1985; Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995). Roughly, these models assume that morphologically complex words of high frequency are effectively processed via whole-word representations, whereas those of low frequency are effectively processed via the representations of their morphological constituents.

One such model is the augmented addressed morphology (AAM) model. Initially, this model postulated that lexical access occurs through morphological decomposition only with novel derived words (i.e., derived words of effectively zero frequency) or nonwords (Caramazza, Laudanna, & Romani, 1988; Caramazza et al., 1985). In this model, two processing routes exist for derived words: a *direct route*, which “addresses” whole-word access representations, and a *morphological parsing route*, which makes use of access representations of morphological constituents (i.e., roots and affixes). The direct route is employed for previously encountered derived words, whereas the parsing route is employed only for novel derived words as a backup option. A later version of this model distinguishes previously encountered forms and novel forms less strictly, but it still assigns an important role to whole-word frequency: Derived words of *medium-low* frequency are processed via the morphological parsing route when their constituent morphemes have a higher frequency than the derivatives themselves (Burani & Laudanna, 1992). Furthermore, in the AAM model, whole-word frequency is a major determinant of the lexical status of derived words. High frequency derivatives have their own lexical representations, irrespective of their other properties such as phonological and semantic transparency (Burani & Caramazza, 1987).

Another model is Frauenfelder and Schreuder’s (1992) morphological race (MR) model, which assumes that whole-word frequency is the most important determinant of how a derived word is processed. In this model, morphologically complex word forms can be processed either via whole-word access representations (i.e., direct route) or via access representations of morphological constituents, as in the AAM model. These two lexical access processes run in parallel, and the faster one wins the race. For access to morphologically complex words of high whole-word frequency, the direct route generally wins the race regardless of a word’s morphological properties such as the productivity of the derivational affix that it contains. This is because frequently encountered words have high levels of resting activation, and such high resting activation levels give the direct access route a head start. Moreover, these high frequency words are assumed to have their own semantic representations as well (see Schreuder and Baayen, 1995). In other words, the meanings of the words are independently stored, rather than computed from the meaning of the root and the meaning of its affixes.

The literature also provides some empirical evidence that whole-word frequency is an important determinant of how derivatives are processed or represented (e.g., Burani & Caramazza, 1987; Holmes & O'Regan, 1992; Leminen, Leminen, Kujala, & Shtyrov, 2013; Meunier & Segui, 1999; Raveh, 2002). For example, the results of Meunier and Segui's (1999) cross-modal priming experiments with native French speakers seem to suggest that only derived words of low frequency in French are recognized through morphological decomposition, while derived words of high frequency in French are recognized through whole-word representation. In their first experiment (Experiment 1), they compared three priming conditions with regard to lexical decision latency: (1) stem priming, (2) high-frequency (HF) derivative priming, and (3) low-frequency (LF) derivative priming, as shown in Table 4.

Table 4. Design, sample stimuli, and results of Meunier & Segui's (1999) Experiment 1

Priming Condition	Prime	Target	Lexical Decision Reaction Times (ms)*
Identical priming	*travail*	*Travail*	488 (54)
HF derivative priming	*travailleur*	*Travail*	516 (68)
LF derivative priming	*travailliste*	*Travail*	492 (62)

*Standard deviation in parentheses

In these conditions, the same target (i.e., a stem) was presented, but different types of primes preceded it. In the identical priming condition, the prime was identical to the target (i.e., the prime was the stem as well); in the HF derivative priming condition, the prime was a high-frequency suffixed derivative of the target; lastly, in the LF derivative priming condition, the prime was a low-frequency suffixed derivative of the target. As in typical cross-modal priming, all primes were presented acoustically, while all targets were visually presented on the screen; participants listened to the prime first, and then made a lexical decision on the target, which was presented immediately after the priming procedure.

The results show that only derived words of low frequency yielded a *full priming* effect, which is regarded in the literature as indicative of repeated access to a single lexical representation shared by the prime and the target (e.g., Clahsen & Neubauer, 2010; Sonnenstuhl, Eisenbeiss, & Clahsen, 1999; Stanners, Neiser, Hernon, & Hall, 1979). That is, in the LF derivative priming condition, the participants identified the target as quickly as they did in the identical priming condition, but they identified it in the HF derivative priming condition much more slowly than in the other two conditions (see Table 4). The interpretation of Meunier and Segui was that only derivatives of low frequency (but not derivatives of high frequency) are accessed via morphological parsing ([*travail-* + *-eur*]), whereby the stem (*travail-*) is isolated and directly primes the target (i.e., the same stem, *travail-*).

In addition, the electroencephalography (EEG) data of Leminen et al.'s (2013) recent study shows that whole-word frequency influences the lexical status of derived words. They recorded mismatch negativity (MMN) brain responses of native Finnish speakers, while the participants were listening to different types of morphologically complex words and pseudo-words in Finnish including high- and low-frequency derived words and pseudo-derivatives. One of Leminen et al.'s interests was whether derived words in Finnish develop stronger memory traces for their whole-word representations, as they are used more frequently. In the current literature, MMN response is regarded as an index of the strength of memory trace for a lexical representation, based on the finding that high-frequency monomorphemic words produce stronger MMN activity than low-frequency monomorphemic words, and pseudo-words produce the weakest activity (see Leminen et al., 2013 for a brief review of MMN).

Hence, Leminen et al. argued that if high-frequency derived words produce stronger MMN activity than low-frequency derived words, this could suggest that high-frequency derivatives have stronger memory traces for their whole-word representations than low-frequency derivatives. The EEG data showed that real derived words elicited stronger MMN responses than pseudo-derivatives, and that high-frequency derived words elicited stronger MMN responses than low-frequency derived words. If, as the results suggest, the high-frequency derived words develop stronger memory traces for their whole-word representations, they are also more likely to be subjected to holistic processing than low-frequency derived words.

However, it should be noted that in many morphological processing models, including the AAM model and the MR model, whole-word frequency is not the only determinant of how a derived word is processed or represented. These models assume that properties of morphological constituents play significant roles as well. For instance, in the AAM model, not every derived word of low frequency is represented in decomposed or combinatorial form (Burani & Caramazza, 1987). The model presumes that low-frequency derived words are represented in this way only in the case that they are phonologically transparent and their derivational affix is productive. In addition, the MR model assumes that the outcome of the race between the two lexical processing routes (i.e., the direct route vs. the morphological parsing route) for medium to low frequency derivatives is a function of the parsability (i.e., the semantic and phonological transparency) of the word and the cumulative frequency of its root and affix(es).[4]

In addition, previous empirical research on the processing of derived words has provided evidence that calculating the probability of morphological parsing for a derived word involves complex balancing between properties of morphological constituents and whole-word frequency, rather than whole-word frequency alone (Burani & Thornton, 2003; Hay, 2001; Kuperman, Bertram, & Baayen, 2010; Winther Balling & Baayen, 2008). For example, Burani and Thornton (2003) demonstrated that not all derived words of low frequency may be accessed via morphological decomposition, and that low frequency derivatives are accessed in this way only when the root has a high frequency. In a lexical decision task in Italian (Experiment 3), they measured reaction times (RTs) and accuracy on four types of low-frequency derived words—matched for phonological/orthographical transparency, semantic transparency, morphological family size, and rated familiarity—with factorial variation of cumulative frequency of root and suffix, and compared them with RTs and accuracy on low-frequency non-derived words, as illustrated in Table 5.

---

[4] Frauenfelder and Schreuder (1992) included these two frequency values in the MR model as factors that reflect the resting activation levels of the root and affixes. They further mentioned, however, that the frequency values are not exact measures of the resting activation levels, because the activation levels increase only when a successful parse occurs (i.e., when the parsing route wins). The cumulative frequencies of the root and affixes include the root and affix frequencies of high-frequency derived words, which the MR model assumes to be accessed via their whole-word access representation without parsing.

Table 5. Design and results of Burani & Thorton's (2003) Experiment 3

	HH	HL	LH	LL	ND
Mean RTs* (ms)	603	605	641	645	640
Error rates (%)	5.4	8.6	14.1	17.1	13.6

*The authors did not provide the standard deviation. HH: (Low-frequency derived words with) high-frequency root and high-frequency suffixes; HL: high-frequency root and low-frequency suffix; LH: low-frequency root and high-frequency suffix; LL: low-frequency root and low-frequency suffix; ND: non-derived words.

The results show that low-frequency derived words that contain a high-frequency root (i.e., HH and HL words), regardless of suffix frequency, resulted in faster and more accurate access than non-derived words of comparable frequency. On the other hand, low-frequency derived words that do not contain a high-frequency root (i.e., LH and LL words) did not exhibit such facilitative effects. Burani and Thornton (2003), interpreting their results in the framework of the Meta Model (Schreuder & Baayen, 1995), proposed that low-frequency derivatives with a high-frequency root are accessed via morphological parsing, while those without a high-frequency root are accessed via whole-word processing. Based on this interpretation, they suggested that the access to a derived word of low frequency is considerably influenced by distributional properties of its morphemic components as well as by its whole-word frequency.

3.2 Base Frequency (and Relative Frequency)

As discussed in the previous section, it is widely assumed in the current literature that whole-word frequency is a major determinant of the morphological decomposability of derived words. That is, derived words of high frequency tend to be subject to whole-word processing, whereas derived words of low frequency tend to be processed via morphological parsing. Some researchers, however, argue that the base frequency of derived words also plays an important role in determining the morphological decomposability of the words. According to the researchers, it is the relative frequency of a given derived word and the base, rather than the absolute whole-word frequency of the derivative, that influences the decomposability of the word (e.g., Hay, 2001; Hay & Baayen, 2002). In other words, even a high-frequency derived word may well be highly decomposable, if its base has a high frequency. At the same time, processing of a low-frequency derived word may not involve morphological parsing, if its base is of relatively much lower frequency.

Relative frequency has been most extensively discussed by Hay (2001, 2003). She first argues that current models of morphological processing actually predict a relation between relative frequency and decomposability rather than a relation between absolute (whole-word) frequency and decomposability, although absolute frequency is evoked as a primary factor in the models. As discussed earlier, most current models assume two processing routes, namely, the direct route and the parsing route, and these models involve a competition between the routes at least to some extent.[5] These models presume that the speed of the direct and parsing routes, which determines the winner of the competition, is affected by the frequency of the base and the whole word, respectively. Hay claimed that therefore, in these models, the frequency of a derived

---

[5] The extent varies from model to model. For example, some models such as Frauenfelder and Schreuder's (1992) MR model assume that the direct route and the parsing route are independent and thus do not interact with each other, whereas others such as Schreuder and Baayen's (1995) meta model assume that the two routes may interactively converge on the desired semantic representation.

word relative to the base arises as a much more important factor in the access to the derived word than its absolute surface frequency.[6]

In addition, Hay (2001) argued that the previously observed effects of high whole-word frequency are actually artifacts of relative frequency, and she demonstrated that absolute whole-word frequency and relative frequency are not independent of each other. She investigated absolute and relative frequencies of English derived words in the CELEX corpus, and found that these two values are significantly positively correlated. That is, it is more likely that the whole-word frequency of a derived word is higher than the frequency of its base when the whole-word frequency is relatively high than when it is relatively low. This tendency was especially clear among suffixed derivatives, as shown in Table 6 below.

[6] Recall that a version of the AAM model explicitly states that derived words are accessed through their morphemic components when "the frequency of the stem is much higher than the frequency of the surface form" (Chialant & Caramazza, 1995, p. 63), as mentioned in Section 2.1.

Table 6. Number and percentage of suffixed derivatives with derivatives more frequent than base, broken down by frequency range of derivative (Hay, 2001, p. 1053)

Whole-word frequency (log-transformed)	Derived > Base	Total forms	Percentage
log(WWF) < 2	13	914	1%
2 < log(WWF) < 4	53	633	8%
4 < log(WWF) < 6	39	355	11%
6 < log(WWF) < 8	43	115	38%
8 < log(WWF)	9	11	82%

WWF: Whole-word frequency

Table 6 shows that as the whole-word frequency increases, the proportion of derived words whose whole-word frequency is higher than their base frequency increases.

Hay (2001) also conducted an experiment to investigate relative frequency effects on decomposability or parsability of derived words. In the experiment, participants were presented with a list of pairs of derived words in English. One of each pair had higher frequency relative to its base, whereas the other had lower frequency relative to its base. The derivatives in a pair shared an affix (e.g., *uncouth* vs. *unkind*) and were matched for their whole-word frequency. For each pair, participants were asked to mark which of the pair they thought more "complex." The results showed that derived words that were less frequent relative to their bases were significantly consistently rated as more complex than derived words that were more frequent relative to their bases. About 65% of the responses of participants indicated that derived words of higher frequency than their bases were considered more complex than their counterparts. Hay

(2001) interpreted this result as strong evidence that the decomposability of derived words relies on the relative frequency of the whole word and the base.

The results of a study by Colé, Segui, and Taft's (1997) lend support to Hay's interpretation. In one of their experiments (Experiment 2), they compared French derived words whose whole-word frequency is higher than their base frequency[7] and French derived words whose whole-word frequency is less than their base frequency in a lexical decision task and a word naming task. The reaction time data showed that derived words of high base frequency were accessed faster than those of low base frequency only when whole-word frequency was relatively lower compared to base frequency. In addition, the results of another lexical decision task (Experiment 3) demonstrated that derived words of high whole-word frequency was accessed faster than those of low whole-word frequency only when whole-word frequency was higher than base frequency. Based on these results, Colé et al. (1997) argued that there is competition between access via the base morpheme and access via the whole word, and that the winner of the competition is determined by the difference between the base frequency and whole-word frequency.

3.3 Productivity of Morphological Constituents

The productivity of morphological constituents seems to be one of the most influential factors in the processing of morphologically complex words. A large number of studies

[7] Unlike Hay (2001), Colé et al. (1997) used cumulative frequency, which includes the frequency of all its derivationally related words and excludes the frequency of the base itself, rather than the frequency of the base as a free-standing word.

demonstrate that productive morphemic units are more readily exploited than unproductive ones in complex word processing than unproductive ones (Bertram, Laine, & Karvinen, 1999; Bertram, Schreuder, & Baayen, 2000; Boudelaa & Marslen-Wilson, 2011; Ford et al., 2010; Forster & Azuma, 2000; Krott & Nicoladis, 2005; Kuperman, Bertram, & Baayen, 2008). The results of these studies suggest that productive morphemes are easily parsed out of the whole word form, and consequently they facilitate lexical access via morphological decomposition. Hence, in this section, I introduce several measures of productivity of morphological constituents and discuss their effects on derived word processing.

3.3.1 Productivity of Base: Family size

In the current literature, morphological family size (or simply, family size) appears to be the most prominent index of productivity of bases (or roots). This concept, which was introduced by Schreuder and Baayen (1997), refers to a type count of the number of words formed with a particular base (such as *open*) through either derivation (as in *openness*) or compounding (as in *open-bar*). Thus, family size quantifies the productivity of a particular base by measuring how frequently the base has been involved in word formation (i.e., lexeme creation).

The family size effects on lexical processing were first observed with monomorphemic words in visual lexical decision tasks. Schreuder and Baayen (1997) found that Dutch nouns with a large family size (e.g., *ketel* 'kettle')—that is, those incorporated as a stem in a large number of derivatives or compounds—elicited shorter latencies and higher frequency ratings than nouns with a small family size (e.g., *regio* 'region'). These results were replicated in English (Baayen, Lieber, & Schreuder, 1997). Furthermore, Bertram, Baayen, and Schreuder (2000) showed that family size effects on lexical decisions with monomorphemic words extend to derived words:

Dutch derived words with a large family size (e.g., *zwemmer* 'swimmer') were responded to faster and received higher subjective frequency ratings than derived words with a small family size (e.g., *peller* 'peeler'). These effects of family size on derived words suggest that derivatives containing a base with a large family size have a higher activation level in the lexicon and thus, are recognized faster than derivatives containing a base with a small family size.

An important aspect of family size is that the family size effect is driven by *semantically transparent* morphologically related words. Schreuder and Baayen (1997) and Bertram, Baayen et al. (2000) both found a higher or clearer correlation between the reaction times in lexical decision tasks and family size when semantically opaque words were excluded in counting family sizes than when they were included (e.g., *gemeente* 'municipality' was excluded from the initial family size count for the derived word *gemeenheid* 'meanness'). In addition, Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, and Baayen (2004) showed that it is the subset of family members that were directly derived from a complex Finnish word itself and thus semantically more closely related to the word, rather than the complete morphological family of the word, that gives rise to the family size effect (i.e., shorter response latencies with words with larger family sizes). Another important aspect is that the family size effect is modulated by the *morphological context* in which a base appears (e.g., Bertram, Baayen et al., 2000; de Jong, Schreuder, & Baayen, 2003). For instance, Bertram, Baayen et al. (2000) first failed to find a correlation between response latencies and the adjectival family sizes (i.e., the type count of adjectival family members) of derivatives containing the deadjectival suffix -*heid*, but they found such a correlation when they excluded adjectival family members that are not compatible with the suffix -*heid* such as color compounds (e.g., *zeegroen* 'sea-green') and intensified adjectives (e.g., *steenkoud* 'stone-cold', meaning 'very cold'). Bertram, Baayen, et al.'s (2000) interpretation was that the adjectival family members of a *-heid* derivative that are incompatible with the selection restrictions of *-heid* are not activated by the access of the derivative. (See de Jong, Schreuder, and Baayen, 2000, for discussion of a similar morphological aspect of the family size effect for a verbal inflectional suffix.)

Crucially, studies on derived word processing show that the family size of the base is a major determinant of how a derived word is processed (Boudelaa & Marslen-Wilson, 2011; Forster & Azuma, 2000; Pastizzo & Feldman, 2004). For example, Boudelaa and Marslen-Wilson (2011) demonstrated that the lexical access process of a given derived word in Arabic is determined by the family size of its root. In Arabic, as in other Semitic languages such as Hebrew, a word is constituted by at least two abstract bound morphemes: The *root* and the *word pattern* (see Boudelaa & Marslen-Wilson, 2011, for a detailed account). A root consists exclusively of consonants, and it conveys a general meaning (e.g., {ʃrb} 'related to drinking'); on the other hand, a word pattern consists primarily of vowels (but can include subsets of consonants), and functions as a phonological structure that carries some syntactic and semantic information (e.g., {mafʕalun} 'place where to do the action related to the meaning of the root', a singular and masculine noun).[8] A specific word is derived by infixation of a word pattern into a root (e.g., {ʃrb}+{mafʕalun} → [maʃrabun] 'place where to drink', a singular and masculine noun).

---

[8] The consonants 'f' and 'ʕ' just indicate where the first and second consonant of the root are to be inserted, respectively. In other words, they are used as place holders for the consonants of a root.

Using masked and cross-modal priming, Boudelaa and Marslen-Wilson (2011) investigated how family sizes of roots and word patterns[9] affect lexical processing of derived words in Arabic. Specifically, they examined whether word pattern priming—which is an indication of morphological decomposition (e.g., Boudelaa & Marslen-Wilson, 2005; Deutsch, Frost, & Forster, 1998; Frost, Forster, & Deutsch, 1997)—occurs depending on the family sizes of roots or the family sizes of word patterns, co-varying these two factors in a 2 × 2 factorial design (see Table 7). The data showed that word pattern priming occurs depending entirely on the family sizes of roots: In both masked and cross-modal priming, word pattern priming was observed only when the word prime contained a root with a large family size, irrespective of the family size of the word pattern, as shown in Table 7.

Table 7**.** Design and results of Boudelaa & Marslen-Wilson's (2011) priming experiments

		PWP & PR	UWP & PR	PWP & UR	UWP & UR
Word Pattern Priming effects (ms)	Experiment 1: masked-priming	31*	33*	-7	3
	Experiment 2: cross-modal priming	33*	35*	11	9

*Statistically significant; Productive word pattern (PWP); productive root (PR); unproductive word pattern (UWP); unproductive root (UR).

[9] The term "family size" had not been used for morphemic units other than roots (or stems), such as affixes or word patterns, although for affixes the same concept had been termed "type (count or frequency)" and used as a measure of their productivity (e.g., Baayen, 1992). Boudelaa and Marslen-Wilson (2011) extended the concept of family size to word patterns.

Further data analysis using linear mixed-effect techniques showed that the larger the family size of the prime root, the larger the priming effects that occurred. These results suggest the importance of root productivity in driving morphological decomposition of Arabic derivatives.

In addition, Forster and Azuma (2000) observed comparable effects of root family size for masked priming between English word pairs sharing a bound stem (e.g., *submit-permit*). The magnitude of bound stem priming was modulated by the family size of the bound root: Word pairs sharing a bound root with a larger family size (such as *submit-permit*) produced larger priming effects than pairs sharing a bound root with a smaller family size (such as *survive-revive*). Taken together with the results of Boudelaa and Marslen-Wilson's (2011), this result suggests a potential cross-linguistic parallel in the role of base productivity in online lexical processing: Productive roots enhance morphological decomposition.

3.3.2 Productivity of affixes

3.3.2.1 Quantitative measures of affix productivity[10]

In the current literature, the productivity of a particular affix seems to be generally defined as the extent to which the language users are likely to use the affix to derive new words in a systematic way (e.g., Aronoff & Anshen, 1998; Plag, 2006; for a detailed discussion about defining affix productivity, see Plag, 1999, Chapter 2). This definition implies two important characteristics of affix productivity. First, as the "extent" part indicates, affix productivity is a *continuum* rather than a dichotomous notion. In other words, affixes can be categorized either as

[10] I would like to thank you Amir Zeldes for patiently discussing various issues in this section.

completely unproductive or productive, but among productive affixes, some may be relatively more productive than others. At the ends of the continuum, we have completely unproductive affixes, such as the nominal suffix *-th* (as in *warmth*), and highly productive affixes, such as the nominal suffix *-ness* (as in *kindness*). At the same time, between these two ends, we find affixes such as the nominal suffix *-ity* (as in *vanity*), which are productive, but relatively less productive than the suffix *-ness*. Second, the "likely" part of the definition suggests that affix productivity is considered as a probability, and thus that it can in principle be *quantified*. This approach to affix productivity can be regarded as a reflection of a definition proposed by Bolinger (1948): "the statistically determinable readiness with which an element enters into new combinations" (p. 18).

The question, is then, how we can quantify the productivity of affixes that differs along a continuum. The literature provides a number of quantitative measures of affix productivity, most of which rely on computerized corpora. Here I focus on three of the measures that appear to have been most widely and frequently adopted in the literature, namely, type count (V), hapax-conditioned degree of productivity ($\mathscr{P}^*$), and category-conditioned degree of productivity ($\mathscr{P}$). These three measures are particularly relevant to the proposed study in that their effects on lexical processing have been experimentally and computationally attested in at least a few previous studies (e.g., Baayen, Wurm, & Aycock, 2007; Hay & Baayen, 2002; Kuperman et al., 2010; Plag & Baayen, 2009), and in that they quantify different aspects of morphological productivity, as will be discussed below. For an extensive and detailed review on various quantitative measures of affix productivity, readers are referred to Zeldes (2012, Chapter 3).

First, the type count V of a given derivational affix estimates an important aspect of affix productivity namely, the "realized productivity" (Baayen, 2009, p. 901, or "extent of use" Baayen, 1992, p. 181), which refers to the extent that a given affix has produced neologisms.

Ideally, the realized productivity of an affix can be measured by the total count of attested word types (i.e., lexemes) containing that affix in a given language community at a given point in time. However, practically, it is calculated by counting the number of word types derived by the affix in a given corpus. Such a word type count for an affix is termed as the "type count", "type frequency", or "vocabulary size" of the affix, and is mathematically formalized as in (1) below (Baayen, 2009, p. 902).

(1) $V(C, N)$

This mathematical expression is read as the type count for an affix *C* in a given corpus of the size *N* (i.e., word tokens).

The measure *V* should be considered in determining the productivity of an affix, given that there is a general consensus in the current literature that type count is an important determinant of morphological productivity (e.g., Baayen & Lieber, 1991; Boudelaa & Marslen-Wilson, 2011; Bybee, 1995; Bybee & Newman, 1995; Hay & Baayen, 2002; Pierrehumbert, 2001). At the same time, many researchers do not rely solely on type count as a measure of affix productivity due to its major problem: It reflects the *past* productivity rather than the current productivity of a given affix. In other words, although it suggests how productive the affix was at some period in the past, it does not necessarily indicate that the *current* likelihood that language users will use the affix to coin new words, as Plag (1999) notes.

On the other hand, the hapax-conditioned degree of productivity $\mathscr{P}^*$ estimates the productivity of an affix with regard to the rate at which the affix is "expanding" its lexical inventory (i.e., vocabulary size) by deriving new words (Baayen, 2009, p. 902). Since this measure gauges an aspect of *current* productivity (i.e., the current expanding rates of lexical inventories of affixes), it is considered to complement the type count V, which assesses past (i.e., realized) productivity. Theoretically, the expanding productivity of a given affix can be estimated by the extent to which the affix contributes by coining neologisms to the overall vocabulary growth rate of a language community. However, for practical reasons, it is estimated by the extent to which the affix contributes to the overall vocabulary growth rate of a given corpus by adding so-called *hapax legomena* to the corpus (Baayen, 2009). Hapax legomena (hapaxes for short) refers to words that occur only once in a given corpus, and their counts are used to indirectly estimate the count of neologisms. Note that hapax legomena are not necessarily neologisms. In small corpora, neologisms are likely to constitute only a small proportion of hapax legomena. As the size of the corpus increases, however, the proportion of neologisms among hapaxes increases, and neologisms predominantly appear among hapaxes rather than among words that occur two times, three times, and so on. That is, the number of hapax legomena is highly correlated with the number of neologisms (see Baayen & Renouf, 1996 and Plag, 2003, Chapter 3, for detailed discussion of this issue).

The relative contribution of an affix to the overall vocabulary growth in a corpus, namely, the hapax-conditioned degree of productivity $\mathscr{P}^*$, is calculated by the conditional probability that when the corpus expands by an additional word token, this additional token is derived by the affix, given that the word token constitutes a new word type in the corpus (Baayen, 1993). The calculation of this conditional probability involves apparently complex mathematical operations, but the outcome of the operations is rather simple. The $\mathscr{P}^*$ value for a particular affix eventually equates to the ratio of the number of hapax legomena derived by the affix to the number of total hapax legomena within the corpus. Readers who are interested in the details of this mathematical derivation are referred to Baayen, 1993.

The hapax conditioned degree of productivity $\mathscr{P}^*$ can be mathematically formalized as in (2) (Baayen, 2009, p. 902).

(2) $$\mathscr{P}^* = \frac{V(1, C, N)}{V(1, N)}$$

The numerator $V(1, C, N)$ denotes the type count of words derived by an affix C that appear only once in a corpus of N tokens, and the denominator $V(1, N)$ denotes the total type count of words occurring only once in the corpus. According to Baayen (1993), the $\mathscr{P}^*$-based rankings of affix productivity generally correspond to our linguistic intuitions about degrees of affix productivity. Also, notice that ranking affixes based on the measure $\mathscr{P}^*$ is in essence ranking them on the basis of their *number of hapaxes*, because the denominator $V(1, N)$ is constant for any affix, given that $\mathscr{P}^*$ for every affix is calculated using an identical corpus. For Dutch and English data, see Baayen, 1993; for Italian data, see Gaeta and Ricca, 2006.

Lastly, the category-conditioned degree of productivity $\mathscr{P}$ gauges the "potential productivity" of a given affix (Baayen, 2009, p. 902). Baayen explains this aspect of affix productivity by analogy with the potential productivity of a company. If a company has gained a large market share, but has few potential buyers left to whom it might sell its product, the company has become less productive in the sense that its growth rate has become much lower compared to when it started out. Likewise, if an affix has already derived many words (i.e., if it has already used up most of its potential bases), it can be considered relatively unproductive in that it cannot produce many new members. The measure $\mathscr{P}$ estimates this aspect of affix productivity. Note that the measure P is complementary to the measures *V* and $\mathscr{P}$*, which assess past and current productivity, because it assesses an aspect of *future* productivity (i.e., how saturated the market for an affix is, and thus how likely it is to derive novel words in the future).

The category-conditioned degree of productivity $\mathscr{P}$ is also estimated using hapax legomena. The mathematical definition of the measure $\mathscr{P}$ is given in (3) (Baayen, 2009, p. 902).

(3) $$\mathscr{P} = \frac{V(1, C, N)}{N(C)}$$

Note that $\mathscr{P}$ and $\mathscr{P}$* are calculated with the same numerator $V(1, C, N)$, the type count (i.e., number) of hapaxes that are derived by the affix *C* in a corpus with N tokens. The denominator $N(C)$ refers to the total token number of words that are derived by the affix *C* in the corpus. The ratio in (3) can be easily understood with the "vocabulary growth curves" exemplified in Figure 2.

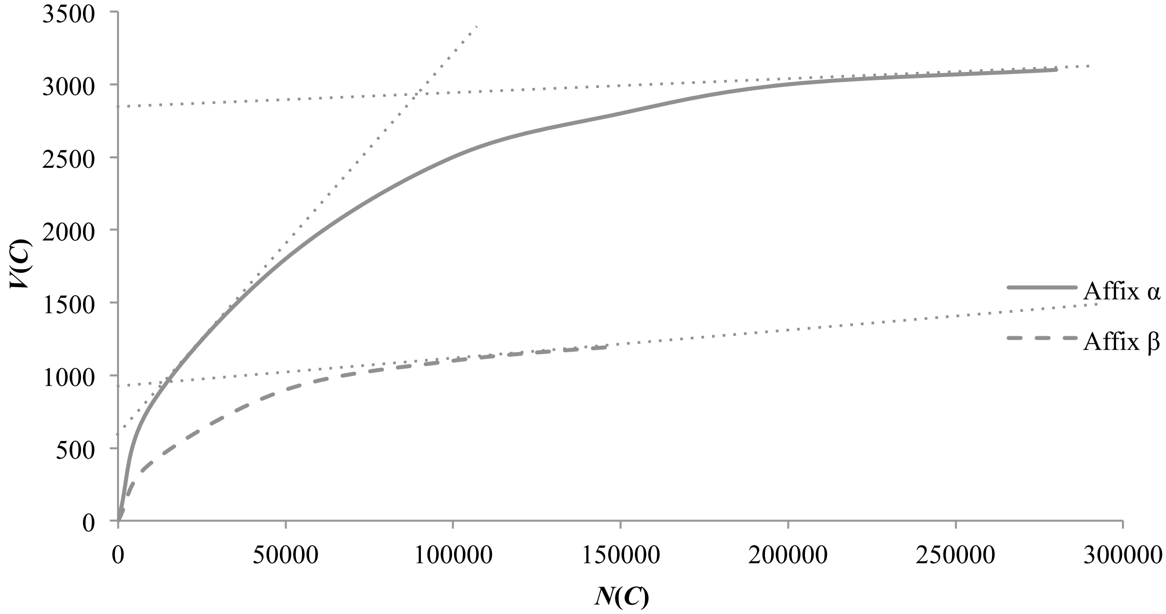


Figure 2. Vocabulary growth curves for two hypothetical affixes α and β

As Figure 2 demonstrates, the growth of the vocabulary of an affix *C* is charted by plotting the type count of words derived by the affix (i.e., *V*(*C*)) against the token number of these words (i.e., *N*(*C*)). This vocabulary growth curve shows to what extent the vocabulary size of the affix increases, as the token number of the affix increases (i.e., as the sample size increases). Hence, the slope of the tangent to the curve (i.e., derivative) at a certain point (*N*(*C*), *V*(*C*, *N*)) represents the growth rate of the vocabulary size of the affix at that point. Notice that the slope flattens out as the type count increases, indicating that the growth rate of the vocabulary size decreases as new types keep coming into the sample and the market for the affix is saturated. The ratio in (3) is equal to the slope of the tangent to the vocabulary growth curve at its endpoint, which represents the growth rate of the vocabulary size at the end (Baayen, 1992, 2009). Note that this ratio can also be interpreted as the probability that the curve will increase by one type at the endpoint. Therefore, the measure P can be understood as the likelihood that a new word type will be encountered (i.e., sampled) at this point of sampling given that this word is derived by the affix *C* (Baayen, 1992).

3.3.2.2 Affix productivity and morphological decomposition

In general, researchers seem to agree that affix productivity and morphological decomposition are highly related to each other, although their specific claims vary. First, a number of morphological models state that words derived by a productive affix are more likely to be morphologically decomposed than words derived by an unproductive affix (e.g., Baayen, 1992; Burani & Caramazza, 1987; Bybee, 1995; Frauenfelder & Schreuder, 1992; Laudanna & Burani, 1995). For instance, according to Frauenfelder and Schreuder's (1992) morphological racing (MR) model, morphologically complex words with a productive affix tend to be accessed by morphological parsing. This is because the model presumes that *whole-word frequency* and *transparency* are the most important two factors in determining the winner of the race between the parsing and direct routes. In this model, the parsing route has the highest chance of winning the race when the given word has low frequency and is semantically and phonologically transparent. Since the model further assumes that these two properties are exactly the properties of words derived by productive affixes,[11] words derived by productive affixes are better candidates for the morphological parsing route than words derived by unproductive affixes in the model.

Furthermore, a large number of previous studies provide empirical evidence that the morphological parsability of a derived word is affected by the productivity of the affix that it contains (e.g., Bertram et al., 1999; Bertram, Schreuder, et al., 2000; Ford et al., 2010; Havas, Rodríguez-Fornells, & Clahsen, 2012; Laine, 1996; Marslen-Wilson, Ford, Older, & Xiaolin, 1996; Vannest & Boland, 1999; Vannest, Polk, & Lewis, 2005). For instance, Bertram et al. (1999) demonstrate that affix productivity is an important determinant of the morphological parsability of derived words in Finnish. The results of their lexical decision tasks suggested that derivatives of an unproductive denominal suffix, *-lA,* and derivatives of a prototypical productive denominal suffix, *-stO*, were processed in different ways, although both of the suffixes were unambiguous (i.e., they do not have homonymic suffixes). Specifically, compared to monomorphemic words matched in various properties including surface frequency, the derivatives of the suffix *-lA* were processed equally fast, whereas the derivatives of the suffix *-stO* were processed significantly faster. Interpreting these results as an indication that the *-lA*

[11] In the current literature, there is a general consensus that the lexical inventory of a productive affix tends to consist of a large number of word types with low frequencies, such that their mean frequency is low, whereas that of an unproductive affix tends to consist of a small number of word types with high frequencies, such that their mean frequency is high (e.g., Aronoff & Anshen, 1998; Baayen, 1992; Baayen & Lieber, 1991; Bybee, 1985; Laudanna & Burani, 1995). Also, it has been generally assumed that productive affixes tend to be phonologically neutral and semantically compositional such that their derivatives are likely to be phonologically and semantically transparent (e.g., Aronoff, 1976; Baayen, 2009; Burani, 1993; Kiparsky, 1983; Plag, 2003; Vannest, Newport, Newman, & Bavelier, 2011).

derivatives are processed via the full-form representations like monomorphemic words, while the -*stO* derivatives are processed via their morphemic units, Bertram et al. (1999) claimed that high affix productivity enhances morphological parsing.

In addition, Vannest and Boland (1999) show that native English speakers' processing of words derived by relatively productive suffixes involves morphological decomposition, but their processing of words derived by relatively unproductive suffixes does not. In a lexical decision task, participants were presented with either words derived by a productive suffix (i.e., -*ness*) or words derived by relatively unproductive suffixes (i.e., -*ity* and -*ation*).[12] They were asked to decide whether the given suffixed words were real English words or not. The results show that when they encountered words with the productive suffix, the participants' decision times were affected not only by whole-word frequency, but also by cumulative base frequency (i.e., the summed frequency of all words containing the base). When they were presented with words with the relatively unproductive affixes, however, their reaction times were influenced only by the words' whole-word frequency. Vannest and Boland's (1999) interpretation was that -*ness* derivatives require morphological parsing, whereas -*ity* and -*ation* derivatives do not. Based on this interpretation, they argued that the productivity of an affix may be an important factor in determining whether the affix has a discrete representation in the mental lexicon.

---

[12] To be precise, Vannest and Boland (1999) distinguished the suffix -*ness* and the other suffixes -*ity* and -*ation* loosely based on Kiparsky's (1983) lexical phonology and morphology theory. According to this theory, derivational affixes can be assigned to two different "levels" referred to as Level I and Level II. Level I affixes are in general relatively unproductive and (phonologically and semantically) non-transparent. On the other hand, Level II affixes are by and large productive and transparent. For a concise review of Kiparsky's theory, see (Plag, 2003, Chapter 7).

Furthermore, Vannest et al. (2005) provide additional evidence for Vannest and Boland's (1999) claim, using functional magnetic resonance imaging (fMRI). In their experiment, they compared four types of words: monomorphemic words, derivatives of relatively unproductive suffixes (*-ity* and *-ation*), derivatives of relatively productive suffixes (*-ness*, *-less*, and *-able*), and words with inflectional suffixes (*-ed* and *-ing*).[13] According to the current literature, rule-governed combinatorial linguistic processes such as inflection increase activation in frontotemporal regions (including Broca's area) in the left hemisphere and the basal ganglia (Ullman, 2004; Ullman et al., 1997). The fMRI data showed that activation in Broca's area and the basal ganglia was increased not only when participants were reading inflected words but also when they were reading derivatives of productive affixes. However, this pattern was not observed when they were reading derivatives of unproductive affixes or monomorphemic words. These results seem to suggest that derivatives of relatively productive suffixes have combinatorial representations like inflected words, while derivatives of relatively unproductive suffixes have holistic representations like monomorphemic words. For further neurological evidence for representational and processing differences between productive and unproductive affixes, see Havas et al., 2012 and Vannest, Newport, Newman, and Bavelier, 2011.

The claim that productive affixes are represented as independent lexical units is further supported by Marslen-Wilson et al.'s (1996) finding that productive affixes are primed, while unproductive affixes are not. Using a cross-modal priming paradigm (auditory primes and visual targets), Marslen-Wilson et al. (1996) compared productive affix processing and unproductive affix processing (Experiment 1). The distinction between productive and unproductive affixes

---

[13] Vannest et al. (2005) also distinguished the two types of derivational affixes on the basis of the lexical phonology and morphology theory, as in Vannest and Boland's (1999) study.

was based on whether the given affixes were currently being used by English speakers to coin new words. The productive affixes included suffixes such as -*ation*, -*able*, and -*ness* and prefixes such as *re*-, *pre*-, and *dis*-; the unproductive affixes included suffixes such as -*ment*, -*al*, and -*ate* and prefixes such as *en*-, *in*-, and *mis*-. The design and results of the experiment are given in Table 8.

Table 8. Design and results of Marslen-Wilson et al.'s (1996) Experiment 1

Priming conditions	Test prime	Target	Priming effects (ms)
Productive suffix	*darkness*	*toughness*	27*
Unproductive suffix	*development*	*government*	12
Productive prefix	*rearrange*	*rethink*	31**
Unproductive prefix	*enslave*	*encircle*	23(*)
Stem	*absurdity*	*absurd*	30**

** $p < .01$, * $p < .05$, (*) $p < .10$

As Table 8 shows, the prime-target pairs consisted of derivative pairs sharing one of the productive or unproductive affixes plus derivative-stem pairs sharing an identical stem. Besides the related primes, each target was paired with an unrelated prime matched in frequency and number of syllables, which provided a baseline for measuring priming effects.

The results provided in Table 8 also show that significant priming effects were found with both productive suffixes and productive prefixes, but not with unproductive affixes. In addition, the priming for productive affixes was as strong as stem priming in terms of magnitude. Stem priming effects are interpreted as a reflection of repeated access to the stem representation in the current literature (for a discussion, see Marslen-Wilson, 2007). Furthermore, the results of the following experiment (Experiment 2) excluded the possibility that the affix priming effects obtained in this experiment resulted from the phonological overlap between the prime and the target: Derivatives of productive affixes such as *darkness* or *rearrange* did not prime pseudo-affixed words such as *harness* or *recent*, which only phonologically overlap with the real derivatives, and even caused a significant interference effect. Marslen-Wilson et al. (1996) argued that these results, taken together, indicate that productive English derivational affixes have an equally independent lexical status as stems and that our mental lexicon is combinatorial in nature.

At this point, however, it should be noted that all the studies described so far dichotomously distinguished affix productivity into productive and (relatively) unproductive affixes, instead of using measures on a continuous scale (such as type counts *V*, the hapax-conditioned degree of productivity $\mathscr{P}^*$, or the category-conditioned degree of productivity $\mathscr{P}$), although affix productivity is characterized as a gradual phenomenon rather than an absolute notion. To the best of my knowledge, only one experimental study has investigated the correlation between such an affix productivity measure and the morphological decomposability of given derived words; this was Kuperman et al.'s (2010) study, which did not find evidence for such a correlation. Some studies have assessed affix productivity using a continuous measure, but only to (somewhat arbitrarily) classify productivity as highly productive, moderately productive, and so forth. (e.g., Ford et al., 2010; Järvikivi, Bertram, & Niemi, 2006).

Corpus data analyses, however, have demonstrated that such continuous measures of affix productivity are correlated with some indices of the decomposability of an affix. For example, Hay (2003) showed that two indices of the decomposability of an affix, *parsing ratio* and *proportion of word types with low-probability junctural phonotactics*, are highly correlated with the category-conditioned degree of productivity $\mathcal{P}$. First, the parsing ratio of a given affix is defined as the proportion of its derivative types for which the base frequency is higher than the whole-word frequency (i.e., the proportion of its derivative types whose *relative frequency* is larger than 1).[14] On the assumption that morphological parsing occurs for derivatives for which the base frequency is higher than the whole-word frequency (see Section 2.2), Hay (2003) proposed this ratio as an index for the likelihood that an affix is easily parsed (see also Hay & Baayen, 2002). The proportion of word types with low-probability junctural phonotactics for an affix denotes the ratio of the affix's derivative types in which improbable phoneme transition across a morphemic boundary occurs at the juncture between the base and the affix. For instance, the /pf/ transition that occurs in *pipeful* is unlikely to occur within a morpheme or a simplex word in English. Based on this observation, Hay (2003) hypothesized that derived words containing such improbable phoneme transitions at the morphemic juncture facilitate morphological parsing. This hypothesis seemed to be borne out: Native English speakers judged derived words with the less probable junctural phonotactics (e.g., *pipeful*) as more complex than those with more probable junctural phonotactics (e.g., *bowlful*).[15] On the basis of this result, Hay (2003) suggested the proportion of derivative types with low-probability junctural phonotactics as another index of the morphological decomposability of an affix.

---

[14] Hay and Baayen (2002) calculated parsing ratios using a computational model for morphological segmentation, and their parsing ratios were lower than those of Hay (2003). They also found, however, a significant correlation between their parsing ratio and the measure $\mathcal{P}$.

[15] See Bertram, Pollatsek, and Hyönä, 2004 for additional experimental evidence that phonotactic cues enhance morphological parsing.

An important finding of Hay's (2003) study is that these two indices of the morphological parsability of affixes are highly correlated with the logarithmically transformed potential productivity $\mathscr{P}$, explaining a large portion of the observed variance in affix productivity when a multiple regression model of $\mathscr{P}$ was fitted with the two indices.[16] Hay and Baayen (2002, 2003) also discuss correlations between other continous productivity measures (i.e., *V* and $\mathscr{P}$*) and other indices of affix parsability (e.g., the average probability of the junctures created by each affix).

3.4 Summary

This chapter discussed the properties that previous studies have established as most influential in the processing of derived words, namely, whole-word frequency, base frequency, and the productivity of morphological constituents. In sum, previous findings suggest that derived words of low whole-word frequency and high base frequency are more likely to be processed via morphological decomposition than those of high whole-word frequency and low base frequency. In addition, derived words containing a productive base or a productive affix are more likely to be morphologically decomposed than those containing an unproductive base or an unproductive affix. In Chapter 4, I present a lexical priming experiment designed to test whether late L2 learners are able to process derived words via morphological decomposition like native

---

[16] To be precise, instead of using parsing ratios as in her other analyses (Hay & Baayen, 2002), for this particular correlation analysis Hay (2003) used *non-parsing* or *whole-word ratios* (i.e., the proportions of derivative types for which the whole-word frequency exceeds the base frequency), which can be calculated by subtracting the parsing ratio from one.

speakers. An important goal of this experiment is to compare L1 and L2 processing of derived words with regard to effects of the factors discussed in this chapter. That is, to what extent do the factors that are significant in the L1 context influence L2 processing of derivatives in similar ways?

Chapter 4 Experiment 1: Lexical Priming Experiment

Using one of the overt paradigms, namely, the fully-visible priming paradigm, I first investigate whether L2 learners of English can process derived words in English via morphological decomposition at the central representation level. As discussed in Chapter 2, no study has explored L2 processing of derived words at the central level of lexical representation, using such paradigms. Furthermore, I compare L1 and L2 processing of derived words in more detail. I examine how whole-word frequency, root frequency, and morphological productivity affect the morphological decomposability of English derived words in L1 and to what extent the findings from native English speakers apply to L2 learners. As mentioned in Chapter 2, no study has investigated how these factors influence the decomposability of derived words in L2.

The L2 learner group investigated in this study is native Korean-speaking late learners of English who are of advanced proficiency. They were chosen as the participants in this study for the following reasons. First, the derivational morphology of Korean is similar to that of English: As in English, derivation is a major way of coining lexical items in Korean, and Korean has several hundred derivational affixes (prefixes and suffixes), many of which correspond at least roughly to English affixes (see Sohn, 2001 for an overview and detailed information on derivational morphology in Korean, as well as examples). For example, Korean has negative prefixes that are comparable to English negative affixes such as the following (Sohn, 2001, p. 221):

(4) *pul-* ‘non-, in-, un- or ir-’: *pul-kanung* ‘impossibility’

mu- ‘non-, -less, or ir-’: *mu-chakim* ‘irresponsibility’

mol- ‘non-, -less, no’*: mol-sangsik* ‘senselessness’

Other examples include the followings (Sohn, 2001, pp. 219–231):

(5) *-cek* ‘-ic’: *pi-kwahak-cek* ‘unscientific’

-hwa ‘-ization’: *pi-kwunsa-hwa* ‘demilitarization’

-i/-li ‘-ly’: *sayloi* ‘newly’

-key ‘-ly’: *kupha-key* ‘hurriedly’

-(u)m ‘-ness’: *kippu-m* ‘happiness’

-i ‘-ness’: *chwuw-i* ‘coldness’

-(k)ay, -key ‘-er’: *cip-key* ‘tweezer’

ta- ‘multi-’: *ta-mokcek* ‘multi-purpose’

Furthermore, there is some empirical evidence that Korean speakers decompose derived words in Korean (Kim, Wang, & Ko, 2011). Therefore, if Korean learners are found to be incapable of morphological decomposition in English, their incapability cannot be attributed to their L1 background. Rather, it would suggest qualitative differences between L1 and L2 representation of derivational morphology.

Regarding statistical techniques, I utilize mixed-effects modeling (Baayen, Davidson, & Bates, 2008), as this experiment considers a number of variables and interactions among the variables. Mixed-effects models make it possible to take account of all potentially relevant variables at the same time, and thus, to explore interactions among all the variables as well. Furthermore, these models provide information about the extent to which each variable influences other variables and how independently of other variables each contributes to the phenomenon of interest. Such multivariate analysis cannot practically be carried out using traditional Analyses of Variance (ANOVAs) with factorial designs, in which typically only one or two variables are manipulated at a time, while all other potentially relevant variables are controlled for. For these reasons, an increasing number of studies on morphological processing have begun to use mixed-effects modeling instead of ANOVAs with factorial designs (e.g., Baayen et al., 2007; de Zeeuw, Verhoeven, & Schreuder, 2012; Gor & Jackson, 2013; Kuperman et al., 2008, 2010; Mulder, Dijkstra, Schreuder, & Baayen, 2014; see Bertram, Hyönä, & Laine, 2011, for discussion for this issue).

In summary, the experiment presented in this chapter (Experiment 1) addresses the two following research questions:

RQ1. Are Korean-speaking advanced late learners of English able to decompose English derived words into their morphological constituents at the central lexical representation level?

RQ2. How similarly or differently is the processing of derived words by L1 speakers and these L2 learners affected by the major factors in the morphological decomposability of derived words in L1 English?

It is difficult to make predictions for the first research question, because the current literature provides only a handful of studies on L2 processing of derived words, and their results are not consistent, as discussed in Chapter 2. Also, recall that most of the studies employed masked priming paradigms that have been assumed to target the surface (i.e., orthographic) lexical representation level, rather than unmasked priming paradigms that has been argued to reflect the central lexical representation levels (including semantic and syntactic representation). However, because derivation is a major way of coining words in Korean, as in English, and there is psycholinguistic evidence that Korean speakers indeed decompose derived words in Korean, they are expected to be able to decompose at least certain types of derived words. Therefore, if these learners turn out to be incapable of morphological decomposition in online lexical comprehension, their incapability may be attributed to some qualitative difference between L1 and L2 processing and/or representation of derived words.

The second research question is even more exploratory than the first research question, since no previous study has investigated it before. The influence of each of the frequency and productivity factors and their interactions on morphological decomposition in L2 contexts has not been explored yet. As discussed in Chapter 2, previous L2 studies only focused on the question of whether or not learners are able to process derived words via morphological parsing in a given context where various lexical and distributional properties are controlled.

4.1 Method

4.1.1 Participants

Fifty-five native speakers of English and 35 native Korean-speaking late learners of English participated in this experiment. All of the native speakers were undergraduate or graduate students at a university in Washington D.C. The L2 speakers were late learners of English, most of whom began to learn English around the age of 13 (*M* = 11.11, *SD* = 1.79) in classroom learning contexts in Korea. Most of them came to the United States in their late teens or at an older age (*M* = 21.94, *SD* = 6.00). All the learners were currently attending universities as undergraduate or graduate students, or had already earned a bachelor's or a higher degree from universities in the United States. In addition, most of them had lived in the United States for at least four years (*M* = 6.17, *SD* = 2.09). In this sense, they were considered advanced learners of English overall. A summary of the 35 advanced learners' language background is provided in Table 9.

Table 9. Background information of the advanced Korean ESL group (*n* = 35)

	M	*SD*	*Min-Max*
Age at time of study	28.49	5.87	19–37
Age at onset of learning English (in classroom learning contexts)	11.11	1.79	8–13
Age at time of arrival at the U.S.	21.94	6.00	14-33
Years of residence in the U.S.	6.17	2.09	2–10[17]

Because vocabulary size appears to be one of the most relevant aspects of L2 proficiency in this study, the L2 learners' vocabulary size was measured and compared to that of the native speakers, employing Nation and Beglar's (2007) Vocabulary Size Test. This test consists of 140

[17] Two of the L2 participants had lived in the United States for only relatively short periods of time (i.e., about two years). However, these participants majored in English language or English education in college in Korea, and most of the courses that they took in college in Korea were provided in English. Therefore, they were considered advanced learners of English, and data from these learners were included in the data analysis.

questions in total, and a learner's vocabulary size is calculated by multiplying the number of correct answers by 100. For example, if the learner gets 70 items correct, then it is assumed that he/she knows 7,000 of the most frequent 14,000 word families in English. [18] According to Nation and Beglar (2007), learners need to know 8,000 word families to adequately understand fairly complex spoken and written texts (e.g., newspapers) without assistance, and competent non-native doctoral students have a vocabulary size of around 9,000 word families.

The results of the vocabulary size test showed that the L2 participants are highly advanced learners of English in terms of their vocabulary size. The lowest score was 98, and the learners' mean score was 111.77 (SD = 7.94). Although this score is significantly lower than the native English speakers' mean score, 124.92 (SD = 7.25), $t(86) = 8.02$, $p < 0.001$, it suggests that the learners know about 11,000 word families, which is quite a large vocabulary size for L2 learners according to Nation and Beglar's (2007) score index, introduced above. Table 10 below gives a summary of the vocabulary size test results.

[18] Here, a word family consists of a root and its inflected and derived forms. For instance, a word family, *nation*, includes words such as *nations*, *nationalism*, *nationalisms*, *nationalized*, *nationalization*, etc. The underlying assumption of an assessment based on word families is that L2 learners who know at least one of the family members could deduce the meanings of other family members they encounter when reading and listening, using morphological knowledge (see Nation, 2006).

Table 10. Mean vocabulary size scores of the three participant groups (*max* = 140)

	Vocabulary size test scores			
	N	*M*	*SD*	*Min-Max*
English native speakers	53	124.92	7.25	109 - 138
Korean L2 learners	35	111.77	7.94	98 - 133

4.1.2 Materials

4.1.2.1 Critical stimuli

From the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995, CD-ROM), 72 prime-target sets served as critical stimuli in this priming experiment. Each set consisted of one target and two types of primes: (i) a prime morphologically related to the target and (ii) a prime unrelated to the target. The target was a morphologically simplex word,[19] such as *polite*. The morphologically related (i.e., test) prime was a semantically transparent suffixed[20] word derived from the target (e.g., *politeness*). Only semantically transparent derived words were used

[19] Here, morphologically simplex words are defined as words that are not further analyzed into a free-standing root and an affix. Thus, targets included words such as *predict*, which can be further decomposed into the prefix *pre-* and the bound root *-dict*. According to Marslen-Wilson et al. (1994), bound roots, which do not have a consistent meaning, do not have their own representation, and thus bound affixed forms are represented as if they were monomorphemic.

[20] According to the current literature, prefixed words and suffixed words are processed differently in some ways (e.g., Beauvillain, 1996; Colé, Beauvillain, & Segui, 1989). This study focuses on suffixed words leaving prefixed words for future research.

in this experiment, because semantic transparency has been argued to be one of the major determinants of the morphological decomposability of given derived words. Many morphological processing models take semantic transparency into consideration (e.g., Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995), and a large number of empirical studies have demonstrated that only semantically transparent derivatives are morphologically decomposed in processing (e.g., Longtin et al., 2003; Marslen-Wilson et al., 1994; Rastle et al., 2000). This morphological relationship between the target and the prime was determined using the *New Oxford American English Dictionary*: A prime was considered morphologically related to the target when it was listed in the dictionary as a derivative of the target, or when its meaning was defined using the target word itself or words that were used to define the target. For most of the morphologically related prime-target pairs (76% or 55 out of 72), the orthographic form of the target was fully retained in the orthographic form of the derived word. Although for the rest of the pairs (24% or 17 out of 72), the derived word showed base allomorphy, all the allomorphic alternations were common orthographic changes that result from suffixation in English: (i) final *e* deletion as in *hostility* (8 out of 16); (ii) change from the final *y* to *i* as in *merciless* (7 out of 16); and (iii) final consonant duplication as in *cancellation* (2 out of 16). Such regular allomorphic alternations have been reported not to disrupt morphological segmentation in online lexical processing (McCormick et al., 2008).

Crucially, the morphologically related primes were derivatives of 10 different suffixes whose productivity in terms of the measures V, $\mathscr{P}^*$, and $\mathscr{P}$ varies widely, as demonstrated in Table 11. The inclusion of suffixes of distinct productivities enables looking into potential effects of affix productivity on priming effects. For each of 8 suffixes other than -*able* and -*less*, 8 derivatives were selected. For each of the suffixes -*able* and -*less*, only 4 derivatives were selected, because derivatives of these suffixes provided a limited range of relative frequency: Their relative frequency was in general very high.

Table 11. Productivity of suffixes used in this study (ordered in *V*)

Suffix	Type count (*V*)	Hapax-conditioned degree of productivity ($\mathscr{P}^*$)[a]	Category-conditioned degree of productivity ($\mathscr{P}$)[b]
-ly	1189	.1063916 (346)	.0015480 (-6.47)
-ness	810	.0630381 (205)	.0102855 (-4.58)
-ity	435	.0129151 (42)	.0009676 (-6.94)
-ation	364	.00067651 (22)	.0005759 (-7.08)
-able	248	.0073801 (24)	.0014352 (-6.55)
-ment	220	.0039975 (13)	.0002451 (-8.31)
-ous	200	.00492 (16)	.0008655 (-7.05)
-less	191	.0061501 (20)	.0029176 (-5.84)
-ence	166	.0021525 (7)	.0002467 (-8.31)
-ance	112	.0015375 (5)	.0002551 (-8.27)

[a] The number of hapaxes (V1) in parentheses

[b] The log-transformed value in parentheses

The whole-word frequency of related primes was distributed in a wide range (10–2697) in order to examine a potential whole-word frequency effect on morphological priming. The frequency of targets widely varied (18–5386) as well, such that the ratio of target (i.e., base) frequency to prime (i.e., whole-word) frequency was dispersed (0.15–244.82). In addition, the family sizes of the targets varied in a range (1–14) in order to explore a potential root productivity effect on morphological priming. In this study, only semantically transparent derived words or compounding words were counted as a member of morphological family, because it has been reported that family size effect in lexical decision is stronger when only semantically transparent members were included than when semantically nontransparent ones are included as well. Relevant properties of related primes and targets are summarized in Table 12, and the frequency distributions of primes and targets and the family size distribution of targets are visualized in Figure 3.

Table 12**.** Stimulus properties for the primes and targets (standard deviation in parentheses)

		Related Prime (e.g., *politeness*)	Target (e.g., *polite*)
Number of syllables	*Mean*	3.28 (.45)	2.00 (0)
	Min-Max	3–4	2–2
Number of letters	*Mean*	9.49 (1.27)	5.90 (.95)
	Min-Max	7–12	4–8
Frequency	*Mean*	315.18 (431.06)	1057.68 (1134.72)
	Min-Max	10–2697	18–5386
Relative frequency	*Mean*	11.69 (31.22)	n/a
	Min-Max	0.15–244.82	n/a
Family size	*Mean*	n/a	4.32 (3.03)
	Min-Max	n/a	1–14

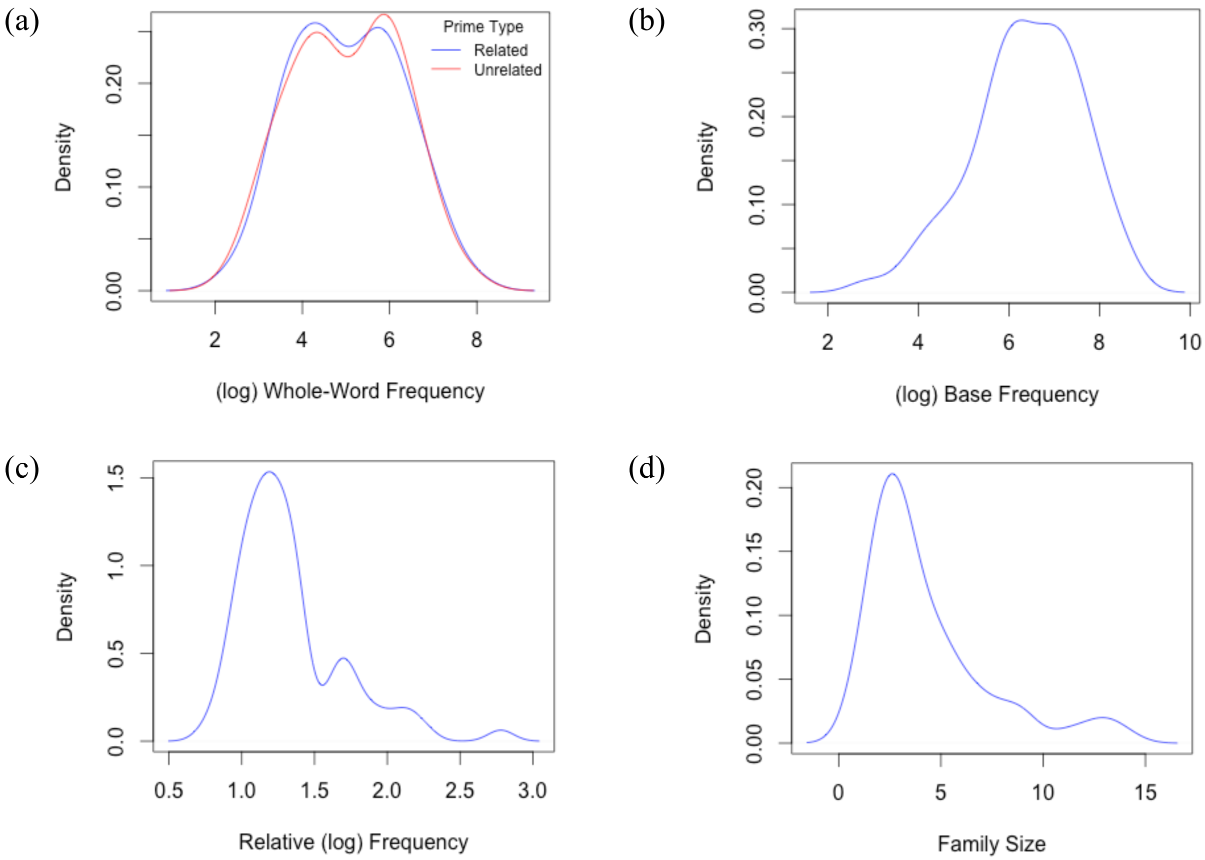


Figure 3. Frequency and family size distributions of primes (derivatives) and targets (bases)

The phonological length of the related primes (and targets) was controlled such that the related primes were relatively long both phonologically and orthographically. The reason for selecting long derived words was that some studies suggest that an important factor in lexical processing of morphologically complex words is the length of those words: Increasing word length leads to lexical access via the constituent morphemes, inhibiting lexical access via the full form (e.g., Beauvillain, 1996; Bertram & Hyönä, 2003; Niswander, Pollatsek, & Rayner, 2000; Niswander-Klement & Pollatsek, 2006). To be specific, all the targets were always two-syllable words, and most of the related primes were three-syllable words (M = 3.28, SD = .45); only the related primes with the suffixes -*ity*, -*ation*, and -*able* (20 items in total) contained four syllables, because these suffixes had two syllables in themselves. As a result, the orthographic length (i.e., the number of letters) of the related primes was also consistently long (M = 9.49, SD = 1.27). The phonological and orthographic length of the related primes was comparable to that of the derived words that Niswander-Klement and Pollatsek (2006) used as long words (phonological: M = 3.14, SD = 0.86; orthographic: M = 8.76, SD = 1.63).

On the other hand, the unrelated prime was a morphologically complex word consisting of a morphologically simplex base word unrelated to the target morphologically, semantically, and orthographically (phonologically) and a different suffix from the suffix that the morphologically related prime contained (e.g., *censorship* as the counterpart of *politeness*). The unrelated priming condition was included as a baseline for measuring priming effects in the other two priming conditions. That is, the reaction time difference between the unrelated and morphologically related conditions was considered the morphological priming effect. Latent Semantic Analysis (LSA; Landauer, Foltz, & Laham, 1998; http://lsa.colorado.edu) scores for critical prime-target sets confirmed that the unrelated primes were semantically unrelated to the targets, whereas the (morphologically) related primes are semantically strongly related to their targets. LSA is a corpus method for extracting and representing the semantics of a given word by analyzing all contexts in which the word does and does not appear. LSA enables a close approximation of human judgment on semantic similarity between words, according to Landauer et al. (1998). Since Rastle et al. (2000) demonstrated that LSA scores on the semantic similarity between morphologically related words were highly correlated with participants' subjective ratings, many studies have used this method to measure the semantic transparency of derived words (e.g., Diependaele et al., 2011; Feldman et al., 2009; Lavric et al., 2007; Rastle et al., 2004). LSA scores range from -1 to 1, and the LSA scores of 1, 0, and -1 indicate that the meanings of a word pair are identical, unrelated, and opposite, respectively.

The mean LSA score of the related prime-target pairs (*M* = .40, *SD* = .19) was comparable to the mean score of synonym pairs such as *surprise-astonishment*[21] (*M* = .42, *SD* = .04), *t*(54.11) = -.40, *p* = .690, whereas the LSA score for unrelated prime-target pairs was close to zero (*M* = .04, *SD* = .03). The LSA scores for the two types of prime-target pairs significantly differ from each other, *t*(71) = 16.64, *p* < .001). With regard to orthography, the number of letters in the same position shared between the unrelated prime and the target was minimized.

For each critical prime-target set, the unrelated primes were matched with the morphologically related primes in whole-word frequency (*M* = 311.13, *SD* = 414.02, *Min-Max*: 11–2593; see also Figure 3) and length (phonological: *M* = 3.32 syllables, *SD* = .55, *Min-Max*: 2–5; orthographic: *M* = 9.49 letters, *SD* = 1.27, *Min-Max*: 7–12). A paired-samples *t*-test shows that the frequencies of the two types of primes did not significantly differ from each other, *t*(71) = -.48, *p* = .634. For each critical set, the related and the unrelated prime were exactly matched in number of letters, and in most cases in number of syllables as well (79% or 57 out of 72 pairs). The full list of critical prime-target sets is provided in Appendix A.

Two counterbalanced experimental lists were created by rotating prime type (i.e., morphologically related and unrelated) using a Latin square design. Each list contained 72 critical prime-target pairs, with 36 targets being preceded by a morphologically related prime and 36 by an unrelated prime. Each target appeared only once in each version. Participants were randomly assigned to one of the two lists.

[21] These synonym pairs were selected and used as filler items. For details, see Section 5.2.2 below.

4.1.2.2 Filler stimuli

In addition to the 72 critical prime-target pairs, 120 filler pairs were included in each experimental list. They consist of (i) 9 synonym pairs (e.g., *surprise-astonishment*), (ii) 9 orthographically similar pairs (e.g., *even-eventful*), and (iii) 90 word-nonword pairs. The 18 real-word pairs were included to decrease the percentage of morphologically related real-word pairs. The synonymous pairs were selected from the *Oxford American Writer's Thesaurus* (Moody, 2012), and the orthographically related pairs were selected by consulting Morris et al.'s (2007) experimental stimuli. The targets and primes of these pairs are morphologically simplex words and suffixed words, respectively, as the targets and primes of the critical sets are. Among the 90 word-nonword pairs, 45 pairs consist of a nonword target and an orthographically related suffixed prime (e.g., *narrowness*-*carrow*), while the other 45 pairs consist of a nonword target and an orthographically unrelated suffixed word (e.g., *spiritless*-*centeg*). The nonword targets for the orthographically unrelated pairs were selected among the 40,481 nonwords that the English Lexicon Project website (Balota et al., 2007) provides. The nonwords of the English Lexicon Project were constructed by exchanging one or two letters of existing English words, making sure that they conform to the lexical phonotactics of English; the position of the letter exchange differed across different words. These nonword targets were paired with orthographically unrelated existing English suffixed words. The nonword targets for the orthographically related pairs were generated in two steps: among the English Lexicon Project nonwords, those generated from existing English suffixed words (e.g., *carrowness* from *narrowness*) were first selected, and then their suffixes were removed (e.g., *carrowness* → *carrow*); the original English suffixed words (e.g., *narrowness*) served as the primes for these nonword targets. All primes and targets in filler stimuli pairs are matched with those in critical stimuli sets in length (filler primes: *M* =

3.19 syllables, *SD* = 0.39, and *M* = 9.2 letters, *SD* = 1.17; filler targets: 2.00 syllables, *SD* = 0, and *M* = 6.01 letters, *SD* = .80). The full list of filler item sets is provided in Appendix B.

In sum, each of the experimental lists contained 72 critical prime-target pairs and 108 filler prime-target pairs; 180 pairs in total. Among the targets of these 180 pairs, 90 targets were existing English words, while the other 90 targets were nonwords; all primes were existing English suffixed words. The proportion of morphologically related pairs among the 180 pairs was 20% (36 pairs).[22]

4.1.3 Procedure

As mentioned earlier, this study employed the fully-visible priming paradigm. In each trial, a cross for visual fixation (+) was first presented in the middle of the screen for 500 ms. This cross was replaced by the visual prime in lower case, which disappeared after 270 ms. For the native speakers, the target, a string of letters in upper case, was presented in the middle of the screen as soon as the prime disappeared. For the advanced L2 learners, the target was preceded by a 100 ms display of a blank screen in order to provide the learners with additional processing time, because it takes 50 ms to 150 ms longer for advanced L2 speakers to fully process and recognize a prime than for native speakers (see Jiang, 1999, 2011, Chapter 3). Participants decided whether the letter string was an existing word in English or not, and their decision times were measured from the onset of the target display. The target disappeared as soon as a lexical decision was made or after a deadline of 2500 ms. Between each trial, there was a 1000 ms interval, during which a blank screen was displayed. Primes and targets were presented in 20

[22] Remember that the 72 critical prime-target sets were counterbalanced using a Latin square design, and consequently, in 36 pairs among the 72 critical prime-target pairs, the target is preceded by an unrelated prime.

point Geneva font, and appeared in white on a black background. Prime-target pairs were presented in a different randomized order for each participant. For test material presentation and data collection, SuperLab version 4.0, a software developed by Cedrus Corporation, was used.

Participants were tested individually. They first filled out a questionnaire about their language background. Then the participants were seated in front of a computer and given oral instructions for the main experiment. They were informed that for each trial, a word would appear in the middle of the screen and quickly disappear, and then a letter string would be presented. They were asked to decide as quickly and accurately as possible whether the letter string was an existing word in English or not by pressing one of two designated keys ("YES" and "NO") on the keyboard. The positions of keys for word and nonword responses changed depending on participants' hand orientation, so that they could easily reach the key for words ("YES") with their main hand. The experiment began with 10 practice trials, with five word targets and five nonword targets that were not included in the critical experimental item set. The total duration of the main experiment was approximately 10–15 minutes. After the experiment, participants (both native and non-native speakers) took the vocabulary size test, which took about 20–30 minutes. Native speakers took the monolingual version (in English), and non-native speakers took the English-Korean version. The test items (i.e., vocabulary) in the two versions were exactly the same.

4.2 Results

The data analysis excluded data from two native English speakers whose response accuracy in the lexical decisions fell below 70%. All the other participants' (53 native speakers and 35 L2 learners) response accuracy was higher than 80% (native speakers: 83.33–99.44%; L2 learners: 81.11–97.22%), and the data from these participants were all analyzed. The reaction time (RT) data analysis included only RTs from critical items whose lexical decision responses were correct. This procedure excluded 3.37% (126 out of 3744 tokens) of the native speaker data and 4.44% (112 out of 2520 tokens) of the L2 learner data. The remaining RT data went through two additional data trimming procedures to reduce the influence of outliers, following Baayen (2008). First, the RT data were logarithmically transformed to reduce the skewing in the distribution of RTs. Second, outliers for each participant were identified in his/her quantile-quantile plot and then removed. This procedure made every participant's RTs more or less follow a normal distribution with no outliers. This procedure affected 0.48% (18 out of 3618 tokens) of the native speaker data and 0.12% (3 out of 2408 tokens) of the L2 learner data.

Table 13 provides the two participant groups' means of the raw RTs and error rates in the morphologically related and unrelated conditions, and Figure 4 visually represents the two groups' RTs in the two conditions.

Table 13. Mean RTs (in msecs), *SD*s (in parentheses), and error rates (in %) in Experiment 1.

	Native English speakers		Korean L2 learners	
	RTs	Errors	RTs	Errors
Related	685 (152)	1.87	683 (181)	1.51
Unrelated	765 (165)	4.89	796 (198)	7.39
Difference	80		113	

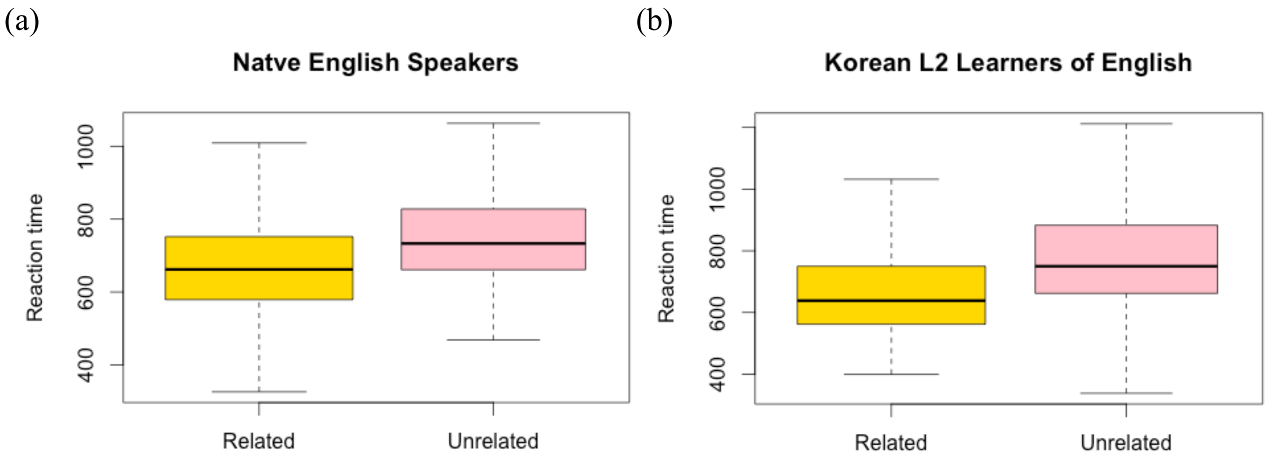


Figure 4. RTs in the related and unrelated conditions

Table 13 first shows that in the related condition, the L2 learners' error rate was comparable to the native speakers' (L1: 1.87; L2: 1.51), although in the unrelated condition, the L2 learners' error rate was higher than the native speakers' (L1: 4.89; L2: 7.39). Importantly, not only in the L1 group but also in the L2 group, the mean RT in the related priming condition was significantly less than that in the unrelated priming condition (L1: $t(3553.054) = -14.94$, $p < .0001$; L2: $t(2347.19) = -14.66$, $p < .0001$). Unexpectedly, the numeric value of the RT difference between the related and unrelated conditions was even larger for the L2 group than for the L1 group. This unexpected result can be interpreted as a floor effect. Note that the L2 group's mean RT in the unrelated condition is longer than that of the L1 group, but the two groups' mean RTs in the related condition are almost the same. This result suggests that the native speakers did not show a larger morphological priming effect than observed, simply because in the related condition, their lexical decision times reached the fastest lexical decision time possible.

This result suggests that although the L2 learners' lexical decisions were slower than those of the native speakers, the morphological priming effect was sufficiently large that both groups' lexical decision times were reduced to the ceiling point (i.e., the fastest lexical decision time possible).

In order to examine each language group's performance in detail, each group's log-transformed RT data were analyzed by fitting a linear mixed-effects model (Baayen, Davidson, & Bates, 2008) for each group, using the lmerTest function from the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014) in the software R (R Development Core Team 2014). The models included as fixed effect variables (1) Priming Type (related vs. unrelated (reference level)), (2) Whole-Word Frequency, (3) Base Frequency, (4) Family Size, and (5) Affix Type Frequency (V). Note that these models included V as a representative measure of affix productivity, not the other two affix productivity measures, $\mathscr{P}$ and $\mathscr{P}^*$. This is because the likelihood ratio tests revealed no significant difference among the three models including each one of the three affix productivity measures. Whole-word frequency, base frequency, family size, and V were logarithmically transformed and centered before they were subjected to the models to reduce the skewing of their distributions. All interactions of Priming Type and other fixed factors were assumed as well. A random intercept was included for participants, and a random intercept and a random slope for priming type were included for items. Inclusion of additional random slopes did not improve model fit, and no significant difference was found when additional random slopes were included. Therefore, a model only with random slopes for Priming Type for items is reported.

4.2.1 Native English Speakers

The results of the model for the native speakers are reported in Table 14.

Table 14. Results of the linear mixed-effects model for the native speakers

Fixed effects	Estimate	Std. Error	df	*t*-value	*p*-value
(Intercept)	6.6200	0.0186	67.53	356.5310	.0000***
Priming	-0.1140	0.0070	158.90	-16.2420	.0000***
BF	-0.0387	0.0092	59.36	-4.1900	.0000***
FS	-0.0263	0.0101	57.81	-2.6060	.0116*
Priming×BF	0.0172	0.0086	163.60	2.0000	.0472*
Priming×WWF×BF×FS	-0.0312	0.0099	158.40	-3.1570	.0019**

Note. Priming = Priming Type (related vs. unrelated); WWF = Whole-Word Frequency; BF = Base Frequency; FS = Family Size; V = Affix Type Frequency; ***$p < .001$, **$p < .01$, *$p < .05$

As Table 14 shows, the coefficient of the Priming Type variable was significantly negative and larger in its absolute value than the significant coefficients of any other fixed variable or interaction (β = -.1140; $t(158.90) = -16.24$, $p < .0001$). This result indicates that a robust morphological priming effect occurred for the native speaker group independently of other fixed variables. In addition, the significant positive coefficents of Base Frequency and Family Size (β = -.0387; $t(59.36) = -4.19$, $p < .0001$ and β = -.0263; $t(57.81) = -2.61$, $p = .0116$, respectively) indicate that native speakers' RTs decreased as the frequency and family size of the target increased. That is, the native speakers' lexical decisions were facilitated when the target had a higher frequency and a larger family size, as in previous studies (e.g., Grainger, 1990; Schreuder & Baayen, 1997, among others). Futhermore, Base Frequency—but not Family Size—emerged in a significant interaction with Priming Type (β = .0172; $t(163.60) = 2.00$, $p = .0472$), signifying that the priming effect was significantly qualified by base frequency. The positive coefficient of this interaction indicate that the priming effect was reduced as the base frequency of the target increased. This result was a surprise in that, according to the current literature, high base frequency facilitates morphological decomposition during online processing, as discussed earlier.

This unexpected direction of the interactions, however, appears to have arisen from a type of ceiling effect. Take a look at the plots in which the two interactions are visualized in Figure 5.

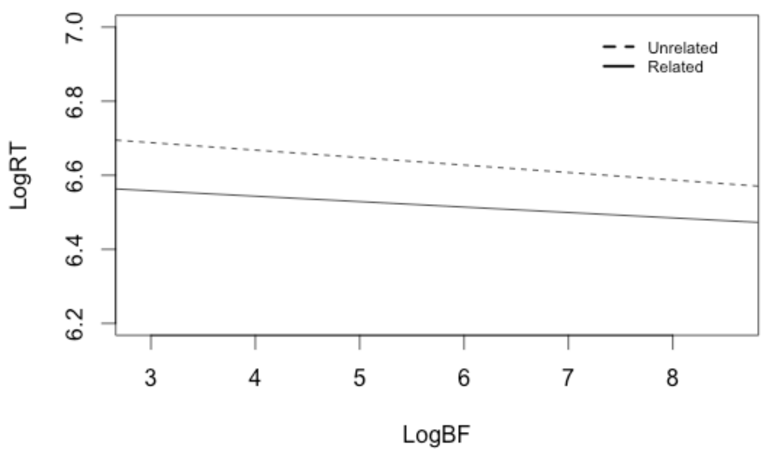


Figure 5. Interaction of Priming Type with Base Frequency

priming effects—which are calculated by subtracting the RT in the related condition from the RT in the unrelated condition—clearly emerged throughout the whole base-frequency range. Thus, it was not the case that the significant interactions were found because the priming effect was found only with primes (i.e., derived words) of low base frequency, and not with primes of high base frequency. Rather, Figure 5 demonstrates that priming effects were strong independently of Base Frequency, but the size of priming effects attenuated as Base Frequency increased, because the slope of the line for RTs in the related condition is more gentle than the slope of the line for

RTs in the unrelated condition. It seems that the RTs in the unrelated condition were directly affected by Base Frequency (as expected), whereas the RTs in the related condition was much less affected by it, because the priming effect was large enough to bring the RTs down to near the ceiling point (i.e., the fastest RT reachable) without the facilitative effect of Base Frequency.

The main effects of the other predictors (i.e., Whole-Word Frequency and *V*) did not reach significance. Nor did either of them per se significantly interact with Priming Type, suggesting that they did not directly affect the priming effect. All the other interactions but the Priming Type × Whole-Word Frequency × Base Frequency × Family Size interaction failed to reach significance. The significant negative coefficient of the Priming Type × Whole-Word Frequency × Base Frequency × Family Size interaction indicates that the priming effect enlarged when the primes had a higher whole-word frequency, a higher base frequency, and a larger family size (β = -.0312; t(138.40) = -3.16, p = .0019). That is, whole-word frequency had an indirect facilitatory effect on morphological priming via an interaction with other factors. This interaction will be addressed in the Discussion section (4.3).

4.2.2 L2 Learners

The results of the model for the L2 learners are summarized in Table 15. As in the report of native speakers' results, only significant fixed factors are presented.

Table 15. Results of the linear mixed-effects model for the L2 learners

Fixed effects	Estimate	Std. Error	df	*t*-value	*p*-value
(Intercept)	6.6540	0.0244	46.20	272.4720	0.0000***
Priming	-0.1583	0.0097	2288.00	-16.2610	0.0000***
BF	-0.0516	0.0124	97.10	-4.1570	0.0001***
Priming×BF	0.0257	0.0119	2291.00	2.1630	0.0307*
Priming×WWF×BF	0.0299	0.0101	2290.00	2.9710	0.0030**
Priming×BF×FS	-0.0458	0.0185	2287.00	-2.4810	0.0132*
Priming×BF×FS×V	-0.0423	0.0257	2289.00	-1.6420	0.0107*

Note. Priming = Priming Type (related vs. unrelated); WWF = Whole-Word Frequency; BF = Base Frequency; FS = Family Size; V = Affix Type Frequency; ***$p < .001$, **$p < .01$, *$p < .05$

As in the native group, the coefficient of Priming Type was statistically reliable and far greater in its absolute value than the significant coefficients of any other fixed variables or their interactions in the L2 group ($\beta = -.1583$; $t(2288) = -16.26$, $p < .0001$). This result indicates that a strong morphological priming effect occurred for the L2 learners independently of the other factors, just as in the L1 group. L2 learners' lexical decision times on the targets (i.e., bases) were significantly shorter in the morphologically related condition compared to in the unrelated condition, independently of the frequency or morphological productivity factors, like the native speakers'. In addition, the main effect of Base Frequency was found to be significant, as in the native speaker group, indicating that the L2 learners' lexical decision times significantly shortened as the frequency of targets increased ($\beta = -.0516$; $t(97.10) = -4.15$, $p < .0001$). In other words, the L2 learners' lexical decisions were facilitated for targets of higher frequency as the native speakers' were. Furthermore, the coefficient of the interaction between these two fixed variables was significantly positive (β = .0257; t(2291) = 2.16, p = .0307), indicating that, for the L2 group, the priming effect significantly diminished as base frequency increased, just as it did for the L1 group. This interaction is visualized in Figure 6.

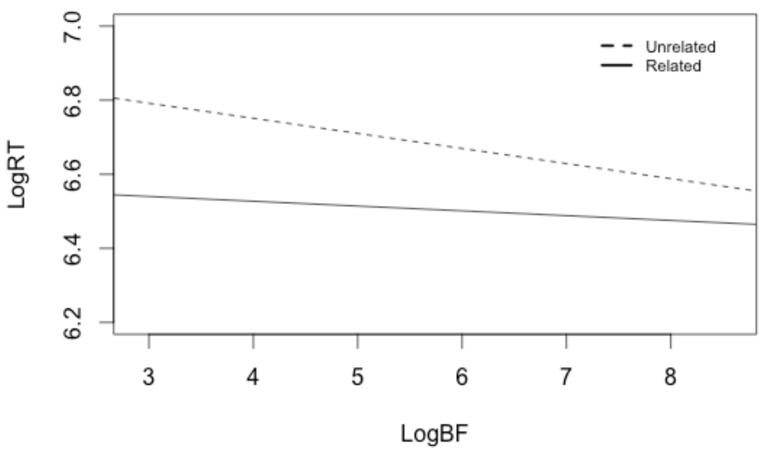


Figure 6. Interaction between Priming Type and Base Frequency

Figures 5 and 6 show that the priming effect was modulated by base frequency in a similar way in the two participant groups. The priming effect was strong throughout the whole base frequency range, and the RTs were much less affected by base frequency in the related condition than in the unrelated condition. As discussed earlier, in the data analysis on the L1 group, this discrepancy seems to have occurred because the priming effect was large enough to decrease RTs to near the ceiling point in the related condition. In addition, the negative coefficient of Family Size approached significance (β = -.0262; t(97.20) = -1.92, p = .0581), suggesting that there was a trend for the L2 learners for lexical decision latencies to be reduced as the family size of the target increased. Family Size, however, did not significantly interact with Priming Type, as in the L1 group. The main effects of the other fixed variables (i.e., Whole-Word Frequency and *V*) both failed to reach significance, and neither of them significantly interacted with Priming Type, as in the L1 group.

Several interactions of Priming Type with other predictors, however, were found to be statistically significant for the L2 group, unlike for the L1 group. As Table 15 demonstrates, the Priming Type × Base Frequency × Whole-Word Frequency, Priming Type × Base Frequency × Family Size, and Priming Type × Base Frequency × Family Size × *V* interactions were significant. The significant positive coefficient of the Priming Type × Base Frequency × Whole-Word Frequency interaction (β = .0299; $t(2290) = 2.97$, $p = .0030$) indicates that as whole-word frequency increased, the priming effect decreased to a larger extent with the increase of base frequency (recall that the priming effect decreased as base frequency increased, β = .0257; $t(2291) = 2.16$, $p = .0307$). This result suggests that high whole-word frequency has an inhibitory effect on morphological priming via an interaction with another factor, although it does not directly affect the priming effects by itself (recall that the interaction of Whole-Word Frequency with Priming Type was not significant). In addition, the negative coefficients of the Priming Type × Base Frequency × Family Size interaction show that the priming effect enlarged as base frequency and family size increased (β = -.0458; $t(2287) = -2.48$, $p = .0132$). Furthermore, the negative coefficients of the Priming Type × Base Frequency × Family Size × *V* interaction indicate that the increased priming effect with a higher base frequency and a larger family size further increased as the affix type count increased (β = -.0423; $t(2289) = -1.64$, $p = .0107$). Taken together, these results suggest that high base frequency, large family size, and large affix type count (or high affix productivity) had facilitative effects on morphological priming through interactions among themselves.

In sum, both native speakers and L2 learners exhibited significantly reduced lexical decision times on targets (i.e., bases) in the morphologically related condition compared to in the unrelated condition, showing robust morphological priming effects. Importantly, in both groups, these priming effects occurred independently of the frequency and morphological productivity factors. However, some group differences were observed as well. Specifically, the two groups showed different patterns with regard to how the priming effect was affected by interactions of the frequency and morphological productivity factors. All these interactions will be discussed in detail in the following section.

4.3 Discussion

In order to better understand L2 representation of derived words in English, the present study investigated L1 and L2 processing of such words. Two specific research questions were addressed. The first question was whether advanced late L2 learners of English are able to decompose derived words into their morphological constituents at the central lexical representation level, like native English speakers. The second question was how similarly or differently L2 and L1 processing of derived words is affected by whole-word frequency, base frequency, and morphological productivity. These factors have been claimed to be major determinants of the morphological decomposability of given derived words in the current literature. This study found a robust priming effect not only in the native speaker group but also in the L2 group, suggesting that advanced late L2 learners are able to morphologically decompose derived words in English. Furthermore, the results showed that both groups' priming effects were affected by the major determinants of morphological decomposability in similar ways, suggesting that L1 and L2 representations of English derived words are determined in similar ways. However, the results of this study also demonstrated that the two groups' priming effects were influenced differently by interactions of the frequency and morphological productivity factors, suggesting some potential differences between L1 and L2 in the processing of derived words in English. The following subsections discuss these issues in detail.

4.3.1 L1 and L2 Representation of Derived Words in English

Not only for the native speakers but also for the L2 learners, the present study found robust priming effects for prime-target pairs of a semantically transparent derived word and its base in English, using the fully-visible priming paradigm. To the best of the authors' knowledge, this is the first finding of a morphological priming effect for L2 learners with this experimental paradigm. As discussed earlier, previous studies on L2 processing of derived words have consistently employed the masked priming paradigm, which has been generally assumed to target only the surface (i.e., orthographic) level of lexical representation. As a result, the scope of the L2 research on derived word processing has been limited to the surface lexical representation. Recall that, on the other hand, the fully-visible priming paradigm reflects central lexical representation, and that priming effects for morphologically related prime-target pairs observed in this paradigm are interpreted in the literature as a reflection of repeated access to the central representation shared by the prime and the target. Thus, the strong priming effects found in this study suggest that semantically transparent English derived words are represented in combinatorial form (e.g., [polite][-ness]) not only in L1 but also in L2 lexical representation at the central level, and thus that central level processing of derived words is similar in the L1 and L2. In this sense, the current study lends support to the idea that the representation and processing of derived words may not fundamentally differ between L1 and advanced L2 speakers.

Here it should be emphasized that the findings of weaker or no morphological priming effect for L2 speakers in previous studies that used masked priming paradigms (e.g., Clahsen & Neubauer, 2010; Silva & Clahsen, 2008) do not necessarily contradict the strong morphological priming observed in this study. Recall that the overt and the masked paradigm tap into different levels of lexical representation. Therefore, it might be the case that L1 processing and L2 processing of derived words occur through distinct mechanisms in the domain of surface-level lexical access, even if they occur through the same mechanism in central-level processing, as Clahsen, Felser, Neubauer, Sato, and Silva (2010) suggested. The idea, however, needs to be further tested in order to be supported, as studies on L2 processing of derived words are scarce in general, and their results are not consistent even when they use the same experimental paradigm, as discussed earlier.

In addition, the results of this study suggest that L1 representation and L2 representation of derived words in English are determined in a similar way. The L1 group showed strong priming effects independently of whole-word frequency, base frequency, family size, and affix type frequency (*V*). Recall that the main effect of Priming Type was significant, and its coefficient was much greater than the significant coefficients of any of the other fixed variables and interactions (see Table 14). Furthermore, this effect was not modulated by Whole-Word Frequency, Family Size and *V*, and the significant interaction of Priming Type with Base Frequency turned out to just reflect a ceiling effect. Taken together, these results suggest that in L1 English, semantically transparent derived words are represented in combinatorial form independently of such frequency and morphological productivity factors. This idea seems to fit best with Schreuder and Baayen's (1995) Meta model, in which (degree of) semantic transparency is the major determinant of the representation of morphologically complex words. The data are also in accordance with Marslen-Wilson et al.'s (1994) model in which only semantically transparent derived words are represented in combinatorial form.

Crucially, the L2 learners in this study exhibited patterns that were very similar to the native speakers'. As in the L1 group, the main effect of Priming Type was significant, and its coefficient was much larger than the significant coefficients of any of the other fixed variables and their interactions (see Table 15). Moreover, this effect was not directly influenced by other fixed variables except for Base Frequency, and the significant interaction between Priming Type and Base Frequency simply reflected the same type of ceiling effect as in the native group. These results suggest that in L2 English, semantically transparent derived words are represented in combinatorial form independently of the frequency and productivity factors, as in L1. Taken together, the processing patterns of the L1 and L2 groups suggest that semantic transparency appears to be the major determinant of how English derived words are represented in both L1 and L2.

However, it should be pointed out that the late L2 learners' native-like processing observed in the current study does not necessarily exclude the possibility that at earlier stages of L2 acquisition, L1 representation and L2 representation of derived words at the central level may be fundamentally different from each other (e.g., Ullman, 2005). Recall that the L2 learners of this study are highly advanced learners of English. It might be the case that at the early stage of English learning, derived words in English were holistically represented in these learners' mental lexicon, without any morphological structure retained, and such representations gradually have developed to be native-like (i.e., in combinatorial form) as the learners' overall proficiency increased.

Lastly, it is left to future research to explore whether native-like combinatorial representation of derived words in L2 is attainable for L2 learners whose native language considerably differs from the target language in its morphological system (e.g., Chinese-speaking learners of English whose L1 is a typical pure isolating language, whereas the target language is relatively much less isolating). As the L2 learners in this study were native speakers of Korean, whose derivational morphology is very similar to that of English, the results of this study do not tell much about how language differences in morphological systems affect the development of L2 lexical knowledge.

4.3.2 Differences between L1 and L2 Processing of Derived Words in English

Although the L1 and L2 groups processed derived words similarly in several important ways, the groups also exhibited some differences in their processing patterns. Specifically, the two groups showed different patterns in how the priming effect was affected by interactions of the frequency and morphological productivity factors. First, for the native speakers, the priming effect was not influenced by any of the interactions except Whole-Word Frequency × Base Frequency × Family Size. Recall that the significant interaction of priming effect with the three-way interaction implied an indirect facilitatory effect of high whole-word frequency on morphological priming. This result is difficult to interpret within processing models in the current literature, which fairly consistently assume that high whole-word frequency has an inhibitory rather than facilitatory effect on morphological priming, as discussed earlier. In this sense, it does not seem unreasonable to consider the possibility that the significant four-way interaction is an artifact of the complex statistical model and/or the complex experimental design involving many variables, although no conclusion can be drawn at this point. Future research is definitely needed. That being said, without the four-way interaction, the picture for native speakers is clear. The native speakers' priming effect was only affected individually by base frequency and family size, and these effects turned out to reflect a ceiling effect. That is, robust morphological priming occurred independently of all the frequency and morphological productivity factors, suggesting that for semantically transparent derived words, the morphological decomposition process occurs independently of these factors.

In contrast, the L2 learners' priming effect was modulated by several interactions of the frequency and morphological productivity factors (i.e., Base Frequency × Whole-Word Frequency, Base Frequency × Family Size, and Base Frequency × Family Size × Affix Count), although their effect sizes were much smaller than the effect size of Priming Type per se. As discussed in the Result section, these results suggest that on the one hand, high whole-word frequency had an inhibitory effect, and on the other hand, high base frequency, high base productivity, and high affix productivity had facilitative effects on morphological priming via interactions with other factors, although each of the factors alone did not have such effects. Because the direction (i.e., facilitative vs. inhibitory for morphological priming) of each of the factors was as expected based on previous research on the processing of derived words, it appears that these indirect effects on morphological priming cannot be simply considered as experimental or statistical artifacts. Therefore, the results of the L2 learners seem to suggest that in L2 English, the frequency and morphological productivity factors play at least ancillary roles in the morphological decomposition process, unlike in L1 English.

What does this difference tell us? One explanation is that it might reflect slower or less efficient L2 lexical processing compared to L1 processing, although full understanding of such a difference would require further research. Native speakers' morphological decomposition process may be very efficient such that the process consistently takes a very short time for derived words of any degree of whole-word or base frequency or morphological productivity. On the other hand, the morphological decomposition process in L2 may be relatively less efficient, perhaps occurring at different speeds depending on frequency and morphological productivity factors. In other words, the L2 morphological decomposition process might occur relatively faster when the whole-word frequency is low and when the base frequency and morphological productivity are high, or relatively slower in the opposite case.

4.4 Limitations and Future Research

The priming effect found in the L2 group was interpreted as *morphological* (rather than semantic or orthographic). In many experiments, controls are included to rule out whether these priming effects are due only to semantic or orthographic relationships between words that are not part of their morphological structure. While these types of pairs were not included here, we do not believe that these factors are responsible for the obtained effects. One reason for this interpretation is that the learners' priming effect was larger than the native speakers'. As discussed in Section 2.2, orthographically related prime-target pairs yield no priming effect or even inhibitory effects, and semantically related pairs yield much weaker priming effects. Therefore, assuming that the native speakers' priming effect found in this study is morphological in nature (following previous studies; see Section 2.2), it seems plausible to interpret the L2 learners' priming effect as morphological too. In addition, some of the results of this study appear to further support such an interpretation. Recall that the important findings of this study include that the priming effect decreased as the whole-word frequency of the morphologically related primes (i.e., derived words) increased, and the priming effect increased as the affix productivity increased. As discussed in Section 4.3.2, these results can be understood within current morphological processing models if we interpret the priming effect as morphological. In contrast, these results become difficult to interpret if we assume that the priming effect is just semantic or orthographic (and that derived words are represented holistically in the L2). First, according to this assumption, the morphologically related primes (e.g., *politeness*) and their targets (e.g., *polite*) do not share any semantic or orthographic information about the affix (e.g., -*ness*), because such information would be only part of the primes. Therefore, the affix productivity of the primes would be expected to have no effect on the processing of their targets. Second, if the priming effect were just semantic or orthographic in nature, high whole-word frequency (i.e., the frequency of the primes) would facilitate the processing of the target and consequently increase priming effects, because the representations of the primes and targets would share semantic and orthographic information, and this shared information would have a high resting activation level due to the frequent access to the primes. However, the present study found a facilitative effect of high affix productivity and an inhibitory effect of high whole-word frequency.

In this sense, it seems plausible to interpret the priming effect found in the L2 group in this study as morphological to some extent and not as purely semantic and/or orthographic. Nonetheless, in order to demonstrate more clearly that the priming effect is morphological (and thus L2 learners are able to decompose derived words), a future study will need to examine priming effects for other types of prime-target pairs, including only semantically or orthographically related pairs (e.g., *surprise-astonishment* or *brothel-broth*). Note that the present study investigated priming effects only for semantically transparent morphologically related prime-target pairs, without these other types of pairs as controls. Further research is needed to investigate whether priming effects for morphologically related pairs are significantly larger than priming effects for only semantically or orthographically related pairs.

4.5 Conclusion

Overall, the results reported in the present study provide some evidence that at the central level, English derived words are represented in combinatorial form both in L1 and L2, and thus, they are morphologically decomposed during online processing both in L1 and L2, although such a decomposition process appears to be less efficient in L2. However, the present data do not exclude the possibility that such native-like representation and processing in L2 is limited to highly advanced L2 learners. Furthermore, it is left to future study to explore whether such native-like representation and processing in L2 is attainable for L2 learners whose native language considerably differs from the target language in its morphological complexity. Lastly, for a complete examination of morphological decomposition in the L2, a future study design needs to include other types of prime-target pairs such as only semantically and orthographically related pairs in addition to semantically transparent pairs.

In the next chapter, I shift the focus from derivational morphology to inflectional morphology. To better understand L2 syntactic representation, Experiment 2 investigates how late L2 learners process inflection in the L2. As in Experiment 1, I compare L1 and L2 processing in order to examine in what ways they are similar or different, and I discuss the implications of the similarities and differences for syntactic representation in the L2.

Chapter 5 Difficulties in L2 Inflection and L2 Processing of Plural Inflection in English[23]

This chapter discusses the background literature on L2 processing of inflectional morphology. First, the relevant theoretical approaches to L2 inflection are discussed in detail, and then previous studies on L2 processing of plural inflection in English are thoroughly reviewed.

5.1 Representational Deficit Account vs. Performance Deficit Account

As mentioned in Section 1.2, existing literature provides two major contrasting theoretical accounts of L2 learners' difficulty mastering L2 inflection: the Representational Deficit Account (RDA)[24] and the Performance Deficit Account (PDA). First, the view of the RDA is well represented in Hawkins and Chan's (1997) Failed Functional Features Hypothesis (FFFH), according to which post-critical-period L2 learners cannot acquire abstract grammatical features that are not instantiated in their L1. Proponents of the FFFH (e.g., Franceschina, 2005;

---

[23] Many parts of Chapters 5 and 6 also appear in my article, "L2 Processing of Plural Inflection in English," published in 2015 in *Language Learning* (DOI: 10.1111/lang.12100).

[24] In this dissertation, representational deficits include any sort of deficit in L2 mental grammar, not limited (as in Hawkins & Chan, 1997 or Tsimpli & Dimitrakopoulou, 2007) to absence of certain grammatical features (due to incomplete acquisition). Thus, in this dissertation, representational deficits include lack of syntactic knowledge such as agreement of grammatical features (e.g., number or gender) as well, following Jiang (2004) and Wen et al. (2010).

Hawkins & Hattori, 2006; Hawkins & Liszka, 2003; Tsimpli & Dimitrakopoulou, 2007) claim that the learner cannot attain the target inflection that these L2 grammatical features underlie, because grammatical features that are not present in a learner's L1 grammar cannot be integrated into his or her L2 grammar. For example, Hawkins and Liszka (2003) take advanced Chinese ESL learners' struggle with past-tense marking in English as evidence that they cannot acquire the [±past] feature, which their L1 purportedly lacks.

In addition, Jiang et al.'s (2011) Morphological Congruency Hypothesis states that L2 learners' successful acquisition of a particular inflectional morpheme in TL heavily relies on whether their L1 has a morpheme corresponding or equivalent to the target. Using an online reading comprehension method, they tested sensitivity to missing plural inflection in QPs such as **several of the rare coin* with two groups of ESL learners. One group consisted of highly advanced Russian ESL learners, whose L1 is considered to be morphologically congruent with English in terms of nominal plural marking, while the other was of Japanese ESL learners with comparable proficiency, whose L1 is not. The results showed that only the Russian group was sensitive to the plural error.

However, proponents of the PDA provide evidence against the RDA which suggests that, ultimately, late L2 learners can develop native-like morphosyntactic knowledge of inflection in the TL, whether the target representation is present in their L1 or not (Foote, 2011; Hopp, 2010, 2013; Keating, 2009; Wen et al., 2010). In other words, they can eventually acquire it as their proficiency increases, even though they may lack target-like inflectional representation when their L2 proficiency level is low. For instance, Hopp (2010) found that although the L2 inflectional processing mechanism can be computationally less efficient due to L1 influence, when L2 learners' general proficiency becomes high enough, they can attain native-like knowledge of L2 inflection. He tested native German speakers and adult advanced and near-native L2 speakers of German on their ultimate attainment of case and subject-verb agreement inflection in two types of offline grammatical judgment tasks (GJTs) —with and without time pressure—and in an online self-paced reading task. He recruited late L2 learners of three different L1 backgrounds (English, Dutch, and Russian) to see whether ultimate attainment of L2 inflection is influenced by L1–L2 differences. Only German and Russian case-mark on full nouns, although German, English, Dutch, and Russian all employ subject-verb agreement; in English and Dutch, only pronouns are case-marked. In the GJT with no time pressure and the self-paced reading task, all near-native L2 groups revealed native-like sensitivity to both case and subject-verb agreement inflection in German, although the advanced groups did not show such sensitivity. In addition, the L1–L2 difference effects for the English and Dutch groups were detected only in the GJT with time pressure, whereas there was no difference between the native speakers and the Russian group. These results suggest that although L1-L2 differences in inflection systems can make L2 learners inefficient to some extent in processing target inflections, the learners can ultimately attain native-like inflectional knowledge when their overall proficiency reaches certain levels.

Another piece of evidence for the PDA comes from studies demonstrating that L2 learners who exhibit insensitivity to target inflection under heavy processing loads may reveal sensitivity under more manageable processing loads (e.g, Gillon Dowens, Vergara, Barber, & Carreiras, 2010; Keating, 2010). These studies suggest that even if learners have native-like knowledge of target inflection, this knowledge can be obscured as processing loads increase. Thus, it seems that learners' insensitivity to target inflections that previous studies have found cannot always be attributed to their defective knowledge of the inflection. For example, in an eye-tracking experiment on gender agreement processing in Spanish, Keating (2009) showed that adult English-speaking learners with advanced proficiency in Spanish are sensitive to gender agreement violations when they have manageable processing loads, even though their L1 lacks gender agreement. The task was an online reading comprehension task, where the participants read Spanish sentences involving gender agreement between nouns and adjectives. In one condition, the noun and the adjective were adjacent to each other in a determiner phrase (DP), as in (6)a. In the other conditions, they were linearly and structurally distant from each other across a verb phrase (VP) as in (6)b, or across a complementizer phrase (CP) as in (6)c:

(6) Example experimental sentences from Keating (2009)

a. *[$_{DP}$ Un trabajo$_{masc.}$ aburrida$_{fem.}$] es ideal para alguien que no tolera el estrés.

A work boring is ideal for someone who not tolerates the stress

b. *[$_{DP}$ Una casa$_{fem.}$] [$_{VP}$ es bastante pequeño$_{masc.}$ cuando tiene sólo una habitación].

A house is quite small when has only one room

c. *[$_{IP}$ Un refresco$_{masc.}$ [$_{VP}$ tiene muy buen sabor [$_{CP}$ cuando [$_{VP}$ está fría$_{fem.}$ y no caliente]]]].

A soft-drink has very good taste when is cold and not hot

The reading time and regression proportion data showed that the advanced learners were sensitive to gender disagreement in the adjacent condition, whereas they were not in the distant conditions. Based on the widely held assumption that longer-distance dependencies tax computational resources for sentence processing more than shorter-distance dependencies (Gibson, 1998), Keating concluded that gender agreement is acquirable for adult learners, even when abstract gender is not instantiated in the L1; however, due to processing limitations (e.g., the relatively poor working memory capacity for an L2; see McDonald, 2006; Service et al., 2002 for examples), the processing burdens from long-distance dependencies may obscure their knowledge.

Furthermore, the observation that even native speakers' sensitivity to inflectional morphology also varies as processing loads increase (e.g., Deutsch, 1998; Foote, 2011; Hopp, 2010; Keating, 2010; Wagers, Lau, & Phillips, 2009) suggests that inflectional variability (under relatively heavy processing burdens) is not an L2-specific phenomenon but a general psycholinguistic phenomenon. Thus, considering processing limitations, such as learners' low working memory capacity in their L2, compared to their L1 (e.g., McDonald, 2006; Service et al., 2002), the variability in L2 inflectional morphology does not necessarily have to be attributed to learners' representational deficits. Instead, it seems that inflectional variability in an L2 simply shows that L2 processing is susceptible to the effects of heavy processing loads due to learners' limited computational resources, rather than due to representational differences between the L1 and L2, at least for highly advanced learners.

Foote's (2011) study appears to lend support to this idea. She examined early and late bilinguals of English (L1) and Spanish (L2) on their sensitivity to Spanish subject-verb number agreement and noun-adjective gender agreement violations in two conditions. Their scores on a standardized proficiency test suggested that all learners had advanced or native-like proficiency in Spanish. In one condition, the two elements in number or gender agreement dependency were adjacent to each other as in (7a) and (7b). In the other condition, they were linearly and structurally distant from each other, being separated by intervening words, as in (7c) and (7d):

(7) Example experimental sentences from Foote (2011)

a. *Veo que tu padre $_{\text{3rd pers. } \textit{sing.}}$ son $_{\text{3rd pers. } \textit{pl.}}$ de Texas.

See $_{\text{1st pers. sing.}}$ that your father are from Texas

b. *Dicen que el libro $_{\textit{masc.} \text{ sing.}}$ blanca $_{\textit{fem.} \text{ sing.}}$ está en esa mesa.

Say $_{\text{3rd pers. pl.}}$ that the book white is on that table

c. [$_{\text{DP}}$ *El reloj $_{\text{3rd pers. } \textit{sing.}}$ del hombre] [$_{\text{VP}}$ son $_{\text{3rd pers. } \textit{pl.}}$ de Suiza].

The watch of-the man are from Switzerland

d. [$_{\text{DP}}$ *El pollo $_{\textit{masc.} \text{ sing.}}$ del taco] [$_{\text{VP}}$ esta rica $_{\textit{fem.} \text{ sing.}}$ pero picante].

The chicken of-the taco is tasty but spicy

The task was a word-by-word self-paced sentence reading task. The reading time data revealed that early and late bilinguals were sensitive to both types of agreement violations. Like native speakers, they slowed down in response to agreement mismatches regardless of misagreement types or distance of the agreement dependency. Furthermore, not only in the bilingual groups but also in the native group, participants' sensitivity to agreement violations became attenuated (i.e., their reading time increase was statistically significant, but numerically smaller) when agreement sources and targets were distant from each other.

5.2 The Shallow Structure Hypothesis

As discussed in the previous section, existing research on the variability of L2 inflection has focused on the issue of whether learners can develop native-like representations of target inflections. However, some studies (e.g., Foote, 2011; Keating, 2009, 2010) have also yielded findings relevant to another hypothesis on L2 processing, namely, the Shallow Structure

Hypothesis (SSH) of Clahsen and Felser (2006b). According to this hypothesis, L2 sentence processing fundamentally differs from L1 sentence processing, in that L2 learners compute only shallow representations that lack detailed syntactic information during online sentence processing, regardless of their proficiency levels and the degree of L1–L2 similarity. Specifically, the SSH states that learners only construct representations that capture predicate-argument structures, but "lack hierarchical detail and more abstract syntactic structure" (Clahsen & Felser, 2006b, p. 32); instead, learners over-rely on lexical-semantic, pragmatic, and relevant non-linguistic knowledge to build up semantic representations of sentences.

Although the SSH was initially proposed to account for some behavioral differences between native speakers and learners regarding ambiguity resolution and filler-gap dependency processing (Clahsen & Felser, 2006b), its scope later became extended to include the processing of grammatical feature checking such as subject-verb number agreement or determiner-noun number/gender agreement (Clahsen & Felser, 2006a), as Keating (2009) points out. More specifically, the SSH makes the strong claim that L2 learners are not capable of processing *nonlocal* checking of grammatical features, because their representation during online sentence processing lacks even very basic properties of configurational structure such as c-command (Clahsen & Felser, 2006a), on which successful feature agreement relies under current minimalist assumptions (Chomsky, 1995). In the literature, 'nonlocal' seems to be interpreted as across a phrase (or phrases), and 'local' as within a phrase (e.g., Keating, 2009), although Clahsen and Felser (2006a) do not provide a precise definition of these terms. Clahsen and Felser only loosely define 'local' as "between closely adjacent constituents" (2006a, p. 111).

Current literature, however, provides evidence that L2 learners —at least advanced ones—can successfully process not only local checking of grammatical features (e.g., Keating, 2009; Sagarra & Herschensohn, 2010), but also nonlocal checking (e.g., Foote, 2011; Jiang et al., 2011; Keating, 2010; Wen et al., 2010). For instance, in Keating's (2010) study, advanced English-speaking learners of Spanish were sensitive to gender agreement violations between a noun and a predicative adjective occurring across a VP (e.g., *[$_{IP}$ La casa$_{fem.}$ [$_{VP}$ es viejo$_{masc.}$…]] 'The house is old'). Also, in Foote's (2011) study, late bilinguals of English (L1) and Spanish (L2) exhibited sensitivity to subject-verb and gender agreement errors that occurred across a VP (see above for example sentences). These results challenge the SSH and lend support to the idea that L2 learners can compute fairly detailed syntactic representations of incoming sentences that include hierarchical phrase structure and grammatical features required for successful feature checking.

5.3 L2 English Plural Inflection

5.3.1 Conflicting Results of Previous Studies

Although many studies have provided evidence that late L2 learners can eventually attain target-like knowledge of inflections even when the L1 lacks equivalent ones, the picture is still not clear with respect to the L2 acquisition of plural inflection in English. On the one hand, Jiang (2007) showed that advanced Chinese ESL learners, whose L1 differs considerably from English in plural inflection (for detailed discussion of these differences, see Lardiere, 2009), are not sensitive to plural inflection errors. In a self-paced reading task, advanced adult Chinese ESL learners did not reveal sensitivity to missing plural inflection in quantifier phrases (QPs) such as **several of the rare coin* (i.e., their reading time did not increase). Furthermore, highly-advanced or near-native Japanese ESL learners in Jiang et al.'s (2011) study were also not sensitive to the same type of plural error. Japanese, like Chinese, differs substantially from English with regard to plural inflection (e.g., Nakanishi & Tomioka, 2004).

On the other hand, Wen et al. (2010) found that advanced Chinese and Japanese ESL learners are sensitive to plural inflection errors. Pointing out that the Quantifier phrase (QP) structure used in Jiang (2007) (e.g., [$_{QP}$ *several* [$_{PP}$ *of* [$_{DP}$ *the rare* **coin/coins*]]]) is syntactically complex, Wen et al. proposed that the learners in Jiang's (2007) study might indeed possess native-like knowledge of plural inflection, but this knowledge might have become obscured as the higher syntactic complexity of the QP structure overloaded their working memory, preventing the successful processing of such knowledge. In order to test this hypothesis, they investigated adult Chinese and Japanese ESL learners' sensitivity to DP-internal number agreement violations in English with syntactically simple DPs such as [$_{DP}$ *these* [$_{NP}$ *beautiful houses/*house*]]. The results show that advanced learners were sensitive to DP-internal number agreement mismatches, although intermediate learners were not. In an online reading comprehension task, the advanced learners took significantly longer, as did the native English speakers, to read nouns in number mismatch conditions (e.g., *these beautiful *house*) compared to the same nouns in number agreement conditions (e.g., *these beautiful houses*). Based on this finding, Wen et al. claimed that the insensitivity found in Jiang's L2 learners likely reflects their limited L2 processing capacity, rather than deficits in their knowledge.

5.3.2 Different Types of Plural Inflection in English

The difference in the syntactic complexity of Jiang's (2007) QP structure and Wen et al.'s (2010) DP structure seems to be an important factor in the apparently contradictory results of the two studies, but a thorough analysis of the phrase structures suggests that there might be another important factor. Specifically, the plural inflection in the two structures involves different types of feature checking. Although Wen et al. (2010) seem to have assumed that number agreement is underlying syntactic knowledge in the plural inflections of the nouns in both the DP and the QP, it appears that number agreement occurs only in the DP:

(8) Example DPs from Wen et al. (2010)

a.	this	beautiful	house	/*houses
	[-plural]		[-plural]	[+plural]
b.	these	beautiful	houses	/*house
	[+plural]		[+plural]	[-plural]

(9) Example QPs from Jiang (2007)

a.	many	of the houses	/*house
	$\begin{bmatrix}\text{+plural}\\ \text{+count}\end{bmatrix}$	[+plural]	[-plural]
b.	one	of the dinosaurs	/*dinosaur
	$\begin{bmatrix}\text{-plural}\\ \text{+count}\end{bmatrix}$	[+plural]	[-plural]

As is illustrated in (8) and (9), within the DP, plural inflection is required for [+plural] number agreement between the demonstrative and the noun, but in the QP, the plurality of the quantifier does not seem to be relevant to number agreement. In particular, (9b) clearly shows that the quantifier and the noun in the QP are not in number agreement dependency. Note that although the quantifier *one* and the noun *dinosaurs* in (9b) bear contradicting number features, this phrase is perfectly grammatical. In the QP, the plural inflection is required when the partitive is headed by a [+count] quantifier, whether it is [+plural] or [-plural].

This difference between the DP and QP structures can be accounted for by unique semantic and syntactic characteristics of the QP structure, which is a type of partitive construction. Semantically, the partitive denotes a subset of a larger discourse-relevant set, meaning "what proportion of" (Stickney, 2009, p. 41). Thus, the partitive contains two discourse-relevant entities. For instance, the QP *several of the students* introduces two sets of students: (a) a set of more than several students (a superset) and (b) a set of several students that belongs to the superset (a subset). Syntactically, the partitive has the structure of one noun phrase (NP) within another, and therefore contains two noun phrases (NPs) (Hoeksema, 1996; Jackendoff, 1977; Stickney, 2009), as shown in (10). According to Jackendoff (1977), the head of the larger NP (i.e., UNIT) is phonologically null and semantically interpreted as "something like unit(s)" (p. 110):

(10) $[_{QP}$ several $[_{NP}$ (UNIT) $[_{PP}$ of $[_{DP}$ the $[_{NP}$ students]]]]]

These semantic and syntactic characteristics of the partitive explain why the quantifier and the second noun (e.g., *students*) in this structure are not in number agreement dependency. As mentioned, the partitive contains two discourse-relevant entities in a superset-subset relation. The quantifier denotes the number of members in the subset, while the DP bearing the second noun (e.g., *the students*) constitutes the superset. Thus, the quantifier cannot denote the number of members referred to by the DP, because the number of members in a subset is smaller than that in its superset (except for cases where a subset denotes its entire superset as in *all of the* *students*). In addition, the structure in (10) suggests that the quantifier should agree in number with the null noun, not with the overt noun. The complement of the quantifier (i.e., *several*) is the larger NP (i.e., *UNIT of the students*), which is headed by the null noun *UNIT*, not the DP containing the overt noun *students*.

Why, then, is the plural inflection always required on the overt noun in the partitive when it is headed by a quantifier with the feature [+count], as the example phrases in (9) show? This phenomenon can be explained by category-selectional (c-selectional) feature checking in the partitive construction (see Adger, 2003, for c-selectional feature checking). In the partitive construction of English, [+count] quantifiers (e.g., *several*) take as their complement (i.e., c-select) a partitive NP containing a [+plural] noun (e.g., *UNIT of the students*). To be specific, a [+count] quantifier (e.g., *several*) c-selects an NP headed by [+count] UNIT, and this UNIT c-selects a prepositional phrase (PP) headed by the preposition *of*, which contains a [+plural] noun. In other words, in the partitive construction, when the [+count] feature of the UNIT is successfully checked by a [+count] quantifier, the [+plural] feature of the second noun is checked by the [+count] UNIT. These linked c-selectional feature-checking dependencies in the partitive construction are illustrated in (11):

(11) $[_{QP}$ several $_{\{+count\}}$ $[_{NP}$ (UNIT) $_{\{+count\}}$ $[_{PP}$ of $[_{DP}$ the $[_{NP}$ students $_{\{+plural\}}$]]]]]

In sum, the apparently contradictory results of Jiang (2007) and Wen et al. (2010) cannot be accounted for by L2 processing limitations alone, as these two studies tested two different parts of English plural inflection knowledge. The plural marking in Wen et al.'s simple DP structure is related to number agreement, whereas that in Jiang's partitive structure is related to c-selectional restrictions. Therefore, without further evidence, we cannot exclude the possibility that the insensitivity of Jiang's participants simply reflects a deficit of knowledge. In other words, it might be the case that L2 learners can acquire only part of the knowledge of plural inflection (i.e., DP-internal number agreement).

In the next chapter, I present an experiment designed to test this possibility. Specifically, using an online reading comprehension task, I examine whether advanced Korean learners of English as a second language (ESL) are sensitive to missing plural inflection even when successful nonlocal checking of grammatical features is required for such sensitivity. I also explore whether they are sensitive to the structural distance of feature-checking dependency in order to further test the SSH.

Chapter 6 Experiment 2: Self-Paced Reading Experiment

The experiment presented in this chapter reexamines whether late L2 learners can attain native-like knowledge of English plural inflection when the L1 lacks an equivalent inflection, employing a word-by-word self-paced reading comprehension task. In this task, a participant is asked to read each sentence for comprehension, word by word, as quickly as possible; the participant's reading time (RT) for each word is recorded. The rationale of this paradigm in L2 inflection processing is that participants' slowdowns in reading words with inflection errors (as compared to the same words without inflection errors) reflect their sensitivity to errors, and this sensitivity demonstrates that they possess the relevant linguistic knowledge. This paradigm is appropriate for testing learners' integrated knowledge of inflection in that it minimizes the involvement of metalinguistic knowledge (see Jiang, 2004, 2007, for a discussion of this advantage). Because of this advantage, self-paced reading tasks have been extensively used in inflection processing research (e.g., Foote, 2011; Hopp, 2010; Jiang et al., 2011; Sagarra & Herschensohn, 2010).

A self-paced reading task was used here to test advanced Korean ESL learners' sensitivity to plural inflection errors in Wen et al.'s (2010) simple DP structure and Jiang's (2007) partitive structure, with the goal to investigate whether these learners can acquire the two different types of feature checking related to English plural inflection (i.e., one relevant to number agreement and the other relevant to c-selectional restrictions):

(12) Simple DP structure used by Wen et al. (2010)

[$_{DP}$ these $_{\{+plural\}}$ [$_{NP}$ beautiful houses $_{\{+plural\}}$]]

(13) Partitive structure used by Jiang (2007)

[$_{QP}$ many $_{\{+count\}}$ [$_{NP}$ (UNIT) $_{\{+count\}}$ [$_{PP}$ of [$_{DP}$ the [$_{NP}$ houses $_{\{+plural\}}$]]]]]

In Korean, although it is a classifier language, nouns can be pluralized by a plural marking suffix (i.e., *-tul*), as in most cases of nominal plural marking in English. Plural marking on nouns, however, is usually optional in Korean (Baek, 2002; Kim, 2005). Note that the sentence in (14) (an example from Hwang, 2012, p. 35) is grammatical, regardless of the presence or absence of the plural marker *-tul*.

(14) ku kos-ey-nun salam(*-tul*)-i manh-ta.

DEM place-LOC-TOP person(-PL)-NOM be.a lot-DEC

'There are many people in the place.'

However, plural marking with *-tul* is strongly preferred in certain contexts. Plural referents often need to be plural marked when they are specific (see Hwang & Lardiere, 2013 and Lardiere, 2009 for detailed discussion on the relation between the plural marker *-tul* and specificity). For example, plural marking is strongly preferred when a noun has already been introduced as plural in the discourse (and is thus specific). This characteristic of Korean plural marking is illustrated in (15) (from Nemeto, 2005, p. 399).

(15) a. pakkath-ey haksayng-i sey-myeng iss-ta.

outside-in student-NOM three-CL exist-DECL

'There are three students outside.'

b. haksayng(-*tul*)-un acwu ttwungttwungha-ta.

student-*PL*-TOP very be.fat-DECL

'The students are very fat.'

c. *haksayng-un acwu ttwungttwungha-ta.

student-TOP very be.fat-DECL

'The students are very fat.'

Nonetheless, plural marking is optional in the partitive construction in Korean, although the noun in the partitive must be specific (see Ladusaw, 1982, for a discussion on semantic constraints on nouns in partitives).

(16) The partitive construction in Korean

a. haksayng(-*tul*) cwung yele myeng

student(-*PL*) among several CL

'several of the students'

b. ku-uy haksayng(-*tul*) cwung yele myeng

he-GEN student(-*PL*) among several CL

'several of his students"

This optionality of plural marking in the Korean partitive construction indicates that in Korean, the grammatical number feature [+plural] of the superset noun (e.g., *haksayng(-tul)* in [16]) does not need to be checked by the [+count] quantifier (e.g., *yele* 'several' in [16]), unlike in English. In other words, Korean partitives do not involve such linked c-selectional feature-checking dependencies as the dependencies that English partitives involve (see [11]).[25]

In addition, Korean lacks DP-internal number agreement equivalent to that of English.

[25] It seems that the only case in which the superset noun of a Korean partitive should be plural-marked is when the superset noun is modified by a demonstrative such as *ce* 'that' or *i* 'this'.

(1) ce haksayng*(-*tul*) cwung yele myeng

those student(-*PL*) among several CL

'several of those students'

Therefore, it appears that in Korean, plural marking on nouns in partitives is required simply by demonstratives, rather than by such c-selectional feature checking as English partitives involve.

(17) a. i/ce haksayng

this/that student

'this/that student'

b. i/ce haksayng-*tul*

this/that student-*PL*

'these/those students'

As shown in (17), demonstratives in Korean are realized as the same form whether they modify singular or plural nouns. Moreover, in contrast to QPs in English, QPs in Korean lack internal number agreement as well. Note that nouns in Korean are not necessarily pluralized when they are modified by non-numeric (weak) quantifiers such as *manhun* 'many' or *motun* 'all', as in the examples in (18) (Lardiere, 2009, p. 204).

(18) manhun/motun kenmul(-*tul*)

many/all building(-*PL*)

'many/all buildings

Furthermore, unlike in English, pluralized nouns cannot be preceded by a numeric quantifier in Korean, unless the nouns have the feature [+human] (Lardiere, 2009). Even for [+human] nouns, the plural marking is not obligatory.

(19) a. twu chayk

two book

'two books'

b. *twu chayk-*tul*

two book-*PL*

'two books'

c. twu haksayng(-*tul*)

two student(-*PL*)

'two students'

Note that *chayk* 'book', which has the feature [-human], cannot be plural marked by the suffix -*tul* when the numeral quantifier *twu* 'two' precedes it, while *haksayng* 'student', which does have the feature [+human], can be pluralized when the same numeral quantifier precedes it. Also, notice that the plural marking on [+human] noun *haksayng* 'student' is optional, as (19c) shows.

To summarize, Korean DPs (and QPs) lack internal number agreement, unlike English DPs (and QPs), and Korean partitives do not involve such c-selectional [+plural] feature checking as English partitives involve. Thus, Korean learners' sensitivity to the plural errors in Wen et al.'s (2010) DP structure and Jiang's (2007) partitive structure—if such sensitivity is indeed observed—cannot be attributed to L1 transfer.

In addition, investigating advanced L2 learners' (in)sensitivity to the plural inflection in these two structures also allowed for testing of the SSH. As discussed earlier, the SSH predicts that nonlocal checking of grammatical features is not possible in L2 sentence processing, since representations in L2 processing are shallow, lacking even very fundamental properties of sentence structure such as c-command. Thus, according to the SSH, L2 learners cannot be sensitive to the plural errors in either Wen et al.'s (2010) simple DP structure or Jiang's (2007) partitive structure, as the errors can be detected only via successful nonlocal checking of the [+plural] feature. As shown in (12), in Wen et al.'s DPs, the feature checking occurs across a phrase (i.e., NP), and the two elements relevant to the feature checking (i.e., the demonstrative and the noun) are not adjacent to each other, but separated by an adjective. Also, as (13) demonstrates, in Jiang's (2007) partitives, the [+plural] feature of the second noun (e.g., *houses*) is checked by another noun UNIT across three phrases (i.e., a PP, a DP, and an NP).

Furthermore, the partitive structure bears a longer feature-checking dependency than the DP. The [+plural] feature checking in the partitive occurs across a *structurally* longer distance than that in the DP.[26] As mentioned above, the [+plural] feature checking in the partitive occurs across three X-phrase (XP) nodes, while that in the DP occurs across only one XP node. Therefore, for native speakers, it is expected that feature checking in the partitive requires more computational resources than that in the DP, and thus that the processing of plural inflection in the partitive becomes delayed, compared to that in the DP. That is, native speakers are expected to slow down at a later point when they encounter a plural inflection error in the partitive, compared to when they encounter such an error in the DP. However, if the SSH is correct, we do not expect this to be the case for L2 learners because their representations would lack the structural details that make the DP and partitive distinct in the structural distance of feature-checking dependency, and consequently in processing costs. On the other hand, if learners construct hierarchically structured representations, contrary to the SSH, their processing of the

[26] In this paper, structural distance between two elements in a sentence is defined as the number of intervening XP nodes between them, following Keating (2010) and O'Grady et al. (2003).

plural inflection will be delayed in the partitive compared to in the DP, as expected for native speakers.

In order to more precisely investigate L2 learners' sensitivity to structural distance, Wen et al.'s (2010) simple DP structure and Jiang's (2007) partitive structure were modified. As (12) and (13) show, the structures differ not only in structural distance but also in linear distance (i.e., the number of intervening words) between the two elements in feature-checking dependency. In Jiang's partitive, there are two intervening words between UNIT and the second noun (e.g., *houses*), while in Wen et al.'s DP, there is only one intervening word between the demonstrative (e.g., *these*) and the noun (e.g., *houses*). Thus, even if readers display delayed inflection processing with the partitive structure compared to the DP structure, this difference cannot be attributed only to the difference between the structural distances of the two structures.

To factor out the linear distance effect, another intervening adjective was inserted between the demonstrative and noun in the DP structure:

(20) Simple DP structure used in the present study

[$_{DP}$ those $_{\{+plural\}}$ [$_{NP}$ long Latin words $_{\{+plural\}}$]]

(21) Partitive structure used in the present study

[$_{QP}$ many $_{\{+count\}}$ [$_{NP}$ (UNIT) $_{[+count]}$ [$_{PP}$ of [$_{DP}$ her [$_{NP}$ books $_{\{+plural\}}$]]]]]

As (20) and (21) show, the modification made the linear distances between the two elements in the feature-checking dependency in the DP and partitive structure the same, without changing the basic structures of the phrases. With this modification, delayed processing of the plural inflection errors in the partitive can be attributed to the longer structural distance of the phrase without the confound of a shorter linear distance.

In sum, in order to test the RDA and the SSH, the study presented in this chapter addresses two research questions:

RQ3. Do late advanced Korean ESL learners display native-like sensitivity (i.e., slowdown in reading) to plural inflection errors in the simple DP structure, in which plural marking is required for number agreement?

RQ4. Do late advanced Korean ESL learners display native-like sensitivity to plural inflection errors in the partitive structure, in which plural marking is required to satisfy c-selectional criteria?

For these research questions, the RDA and the SSH make the following predictions. Regarding RQ3, both the RDA and the SSH predict that the late advanced Korean ESL learners will not display native-like sensitivity in the DP structure. Regarding RQ4, both the RDA and the SSH predict that the Korean ESL learners will not display native-like sensitivity in the partitive structure. Also, the SSH further predicts that native speakers' RT increase (or slowdown of reading speed) due to plural inflection errors will be observed with a *delay* when they read the partitive structure, compared to when they read the DP structure, because the partitive structure involves a structurally longer feature-checking dependency than the DP structure, and consequently plural inflection processing in the partitive structure takes more time than that in the DP structure. On the other hand, for L2 learners, the SSH predicts that such a delay will not be observed, because L2 learners' representations of the two structures are syntactically shallow such that the structures are not distinguishable by the structural distance of the feature-checking dependency.

6.1 Method

6.1.1 Participants

Thirty-five Korean ESL learners[27] and 19 native English speakers participated in this study. All of the participants were undergraduate or graduate students from universities around Washington, D.C. All of the ESL learners were initially exposed to English in classroom learning contexts in Korea (mostly at the age of 13[28]), and most of them came to the United States in their late teens or at an older age. Although all of these 35 participants considered themselves advanced learners, their English proficiency was independently measured by Schultz's (2006) C-test, which was used by Wen et al. (2010). In the C-test, participants were provided with two short English paragraphs containing 40 blanks where parts of words had been left out (e.g., *These are a refle___ of th___ trying to under___ us*). Participants were asked to fill in the blanks by completing the words (one point each). Among the 35 learners, 13 scored only half of the total possible score (i.e., 20) or less (M = 18.92, SD = 1.12), and 22 scored 26 or

[27] This learner group does not include any of the 35 L2 learners of the Experiment 1 (the lexical priming experiment)

[28] The advanced ESL group included learners who began to learn English relatively early (four learners who started learning before 10). However, the RT patterns of the advanced group did not change when the four learners were excluded and thus the RT data from these learners were included in the data analysis.

above (*M* = 30.68, *SD* = 2.90).[29] The score of 26 was chosen as a cut-off point for being considered an advanced learner because it seemed to be a natural cut-off given the quite large gap between 20 and 26.[30] Thus, only the 22 learners who scored 26 or above were grouped as advanced, while the 13 lower-scoring learners were grouped as intermediate. The high mean score of the advanced group (30.95) signified their high proficiency in English. However, an independent samples *t*-test showed that the proficiency of the advanced L2 group was significantly lower than that of the native speaker group (36.05), $t(39) = 6.92$, $p < 0.001$, $d = 2.25$, indicating that the proficiency of these learners was not at the near-native level. Table 16 below gives the results of the C-test.

Table 16. Mean C-test scores of the three participant groups (max = 40)

	C-test scores			
	N	*M*	*SD*	*Min-Max*
English native speakers	19	36.05	1.87	33 - 39
Korean advanced ESL learners	22	30.95	3.45	26 - 35
Korean intermediate ESL learners	13	18.92	1.12	17 - 20

The target dataset reported in this article included the RT data from the advanced learners, which was compared to the RT data from the native speakers; the RT data from the intermediate

[29] The C-test scores were not significantly correlated with either the learners' starting age of learning English ($r = .26$, $p > .05$) or their years of residence in an English-speaking country ($r = -.25$, $p > .05$).

[30] Wen et al. (2010) used the score of 28 as cut-off point, but they did not provide a rationale for their decision.

learners was not of major interest in the present study, due to a concern that any insensitivity they showed to plural inflection might simply reflect underdeveloped knowledge. The data of the intermediate group is separately reported in Appendix C. A summary of the 22 advanced learners' language background is provided in Table 17.

Table 17. Background information of the advanced Korean ESL group (n = 22)

	M	*SD*	*Min-Max*
Age at time of study	29.09	3.75	23–39
Age at onset of learning English	11.55	2.15	6–13[31]
Years of residence in the U.S.	6.02	3.52	1–16

6.1.2 Materials

As mentioned earlier, this study used both Wen et al.'s (2010) simple DP structure and Jiang's (2007) partitive structure with some modifications. These two modified structures are referred to as the "simple structure" and the "partitive structure," respectively. An example set of test sentences is provided in Table 18. As this table shows, a test sentence consisted of 10 words. Each of them constituted a region of primary interest, in terms of RTs, and each of these word-regions was individually presented to a participant.

[31] As a reviewer for the Language Learning article pointed out, the advanced ESL group included learners who began to learn English relatively early (four learners who started learning before 10). However, the RT patterns of the advanced group did not change when the four learners were excluded and thus the RT data from these learners were included in the data analysis.

Table 18. Example set of test sentences

Conditions		Region									
		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
S	GR	Kevin	memorized	those	long	Latin	words	in	just	ten	seconds.
	UGR	Kevin	memorized	those	long	Latin	word	in	just	ten	seconds.
P	GR	Mary	donated	many	of	her	books	to	the	public	library.
	UGR	Mary	donated	many	of	her	book	to	the	public	library.

Note. S=Simple, P=Partitive, G=Grammatical, U=Ungrammatical

There were two independent variables, namely, the Structure variable (simple vs. partitive) and the Grammaticality variable (grammatical vs. ungrammatical), which yielded four experimental conditions: (1) Simple Grammatical, (2) Simple Ungrammatical, (3) Partitive Grammatical, and (4) Partitive Ungrammatical. In the simple conditions, sentences contained a DP/QP that consisted of a determiner/quantifier (e.g., *these/those*, *many*, *several*, and cardinal plural numbers such as *three*), two following adjectives, and a noun.[32] On the other hand, sentences in the Partitive conditions had a partitive that consisted of a quantifier (e.g., *both*, *many*, *several*, and cardinal plural numbers), a preposition *of*, a determiner/possessive, and a noun. The nouns in these two structures always had to be marked by the plural morpheme *-s* in order to be grammatical. In the grammatical conditions, the noun in the simple DP/QP and the partitive was properly marked by the plural *-s* suffix; in the ungrammatical conditions, the required plural

[32] Although Wen et al. used only demonstratives (i.e., *this/these* and *that/those*), in this study both demonstratives and quantifiers were used. A reviewer for the Language Learning article raised the question of whether the simple DPs headed by a demonstrative (e.g., *those long Latin words*) and the simple QPs headed by a quantifier (e.g., *several long Latin words*) involve the same syntactic structures and agreement processes. In the study, these two types of phrases were grouped together and compared with partitive QPs (e.g., *several of her books*), because they have two important commonalities that distinguish them from partitive QPs. First, both of them involve number agreement (here, I assume that plural marking in the simple QP is the spell-out of a formal agreement feature [+plural], which is valued by the [+plural] feature of the quantifier, following Lardiere, 2009). Second, the structural distance of the number agreement dependency in the two structures is the same. Compare the structural distances of the structures: [$_{DP}$ those [$_{NP}$ [$_{AP}$ long] [$_{AP}$ Latin] words]] and [$_{QP}$ several [$_{NP}$ [$_{AP}$ long] [$_{AP}$ Latin] words]] (or [$_{DP}$ [$_{QP}$ several [$_{NP}$ [$_{AP}$ long] [$_{AP}$ Latin] words]]]).

inflection was missing.[33] In all conditions, the simple DP/QP and partitive began at Region 3 and ended at Region 6. Hence, the plural inflection error always occurred at Region 6 in both structures. This made Region 6 the critical region of the experiment.

Sixteen grammatical sentences were first composed for each of the two structural conditions (i.e., 16 grammatical sentences with a simple DP/QP and 16 grammatical sentences with a partitive). From these 32 sentences, 32 ungrammatical sentences were generated by omitting the obligatory plural *-s* suffixes in Region 6. The 32 sentence sets of a grammatical and an ungrammatical version were counterbalanced (using a Latin square design) across two lists such that each participant would see only one of the versions of each item. Participants were randomly assigned to one of the two lists. Each list consisted of 16 sentences of the Simple

---

[33] A reviewer for the Language Learning article pointed out that because test sentences with a plural *-s* were always grammatical, while those without an *-s* were always ungrammatical, grammaticality and presence/absence of *-s* are confounded in this experiment. He/she further commented that due to this confound, the results of this study may not be directly comparable to those of Wen et al. (2010) and Jiang (2007). It should first be noted that only Wen et al.'s study, but not Jiang's study, lacks this confound. It seems that this difference between Wen et al.'s and Jiang's studies can be explained by the difference in their research questions. Since Wen et al. investigated whether L2 learners can acquire DP internal number agreement (i.e., not only [+plural]–[+plural] but also [-plural]–[-plural] agreement), they had to include not only item pairs such as *these houses* and **these house* but also those such as *this house* and **this houses*. On the other hand, because Jiang examined whether L2 learners can acquire native-like knowledge of obligatory nominal *plural* inflection, he did not have to include item pairs such as *much of the flour* and **much of the flour*s, because these pairs test knowledge of c-selectional [-plural]/[-count] feature checking, rather than knowledge of plural marking. As this study also focuses on L2 learners' knowledge of obligatory plural marking—specifically, their knowledge about number agreement and c-selectional feature checking relevant to obligatory plural marking—item pairs such as *this house*/**houses* or *much of the flour*/**flour*s were not tested.

condition (8 grammatical and 8 ungrammatical) and 16 sentences of the Partitive condition (8 grammatical and 8 ungrammatical). In addition to these 32 test sentences, the lists included 32 filler sentences, which were all grammatical. Like the test sentences, the filler sentences each contained 10 word-regions. No filler sentence had the simple DP/QP or the partitive, although some included pluralized nouns (e.g., *The grocery store in Arlington County sells only organic vegetables*). The 64 sentences in each list were presented in random orders to participants. Each of the test and filler sentences was immediately followed by a yes/no comprehension question to ensure that participants were paying attention to the meaning of the sentences. The answer to half of the questions was "yes," and the answer to the other half was "no." A complete list of the target sentences and comprehension questions can be found in Appendix D.

6.1.3 Procedure

Participants were tested individually. They first filled out a questionnaire about their language background, and took the C-test with a 10-minute time limit. Then they were seated in front of a computer and asked to begin by reading the written instructions. Participants were told that they should read each sentence as quickly as possible, but carefully enough so that they could correctly answer a following reading comprehension question. The experimental sentences appeared on the screen word by word. Before each sentence was presented, an asterisk appeared on the screen to focus participants' attention. Participants could see the first word of each sentence by pressing the spacebar, which they also pressed in order to see the next word. The sentence appeared from left to right in a single line. Each time they pressed the spacebar, the previous word immediately disappeared. Importantly, each word appeared and disappeared on a blank screen, so that participants could not anticipate the length of upcoming words or predict the structure of the sentence. When they reached the last word of the sentence, which was marked by a following period, they had to press the spacebar once again to see a yes/no comprehension question, which was included to ensure that participants processed each sentence for meaning. Participants answered the question by pressing one of two designated keys. The experiment included a practice session consisting of eight practice trials. Including the practice session, the entire reading comprehension task lasted approximately 15–20 min. For test material presentation and data collection, SuperLab version 4.0, a software developed by Cedrus Corporation, was used.

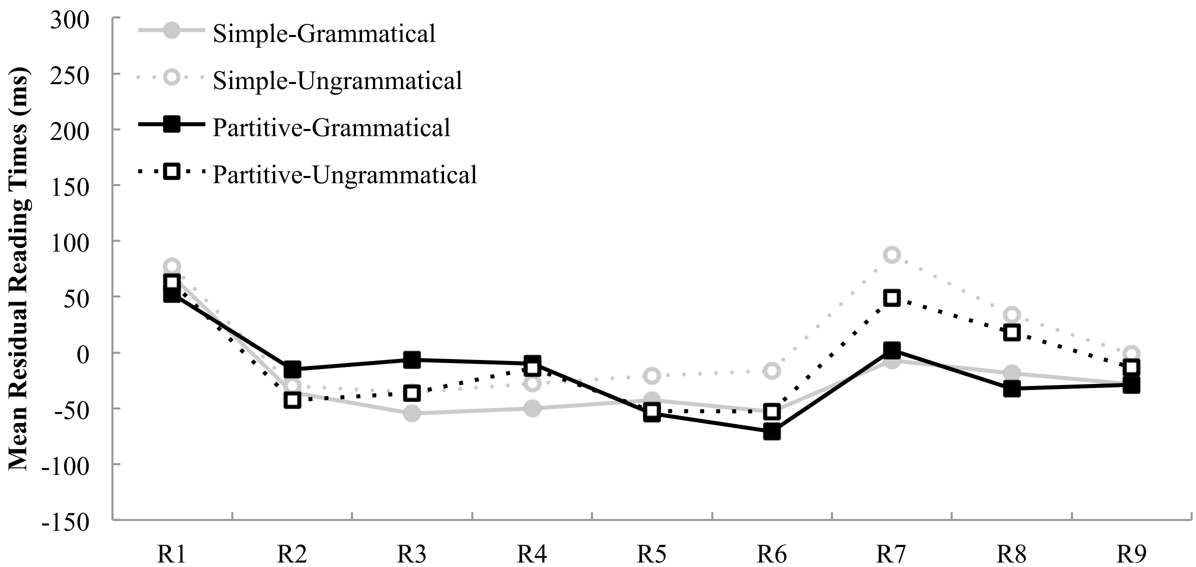
6.2 Results

6.2.1 Comprehension Question Accuracy

Overall, both the native and ESL groups showed very high accuracy rates (98% and 96%, respectively) for the sentence comprehension questions. This result indicates that they were attending to comprehension during the task.

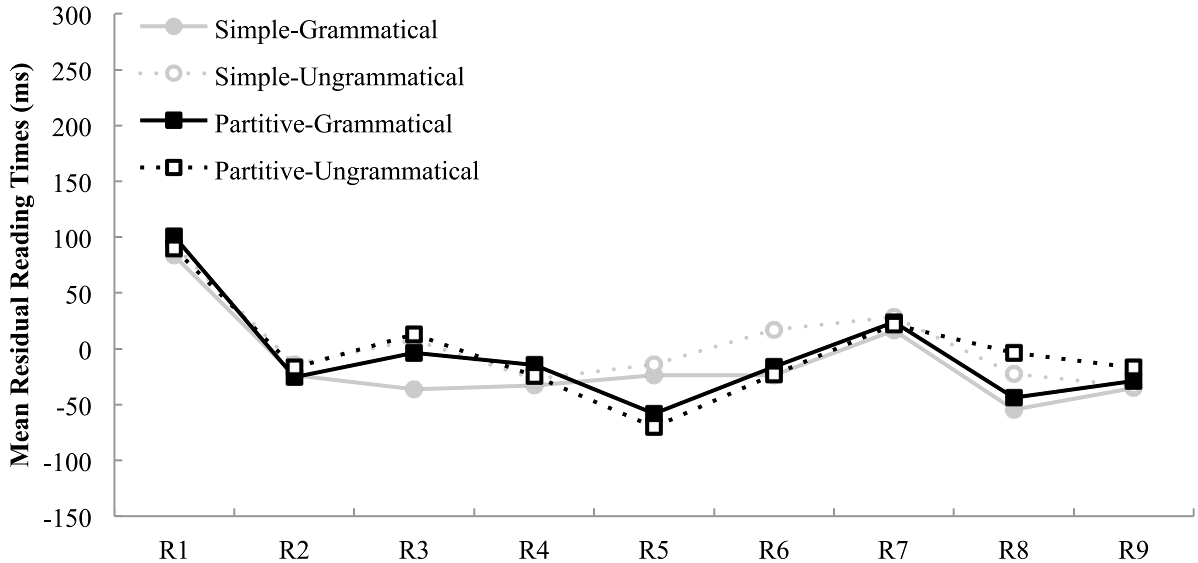
6.2.2 Self-paced Reading

The RT data analyses excluded RTs from experimental trials whose comprehension questions were answered incorrectly (3%). The remaining RT data went through a two-step data trimming procedure. First, any RT longer than 2000 ms was defined as an outlier and discarded. Then, for each participant, any RTs longer or shorter than his/her mean RT ±2.5 standard deviations were replaced with his/her mean RT plus or minus 2.5 standard deviations. This procedure affected 1.4% of the native speaker data and 1.7% of the learners' data. The raw RTs were then transformed to residual RTs (Ferreira & Clifton, 1986) in order to diminish RT differences resulting from word length (i.e., the number of letters) and reduce RT variances among participants (see Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). For this transformation, the best linear fit between word length and RTs was first computed for each participant. Then, at each word, the predicted RT from the participant's linear fit was subtracted from the actual measured RT. All the following statistical analyses were conducted with these residual RTs. The native speaker and ESL learner mean residual RT profiles are illustrated in Figure 7 and Figure 8, respectively.



S_GR	Kevin	memorized	those	long	Latin	words	in	just	ten …
S_UGR	Kevin	memorized	those	long	Latin	word	in	just	ten …
P_GR	Mary	donated	many	of	her	books	to	the	public …
P_UGR	Mary	donated	many	of	her	book	to	the	public …

Figure 7. Mean residual RT profile of native English speakers



S_GR	Kevin	memorized	those	long	Latin	words	in	just	ten …
S_UGR	Kevin	memorized	those	long	Latin	word	in	just	ten …
P_GR	Mary	donated	many	of	her	books	to	the	public …
P_UGR	Mary	donated	many	of	her	book	to	the	public …

Figure 8. Mean residual RT profile of Korean ESL learners

The RT data analyses were carried out for the critical region (Region 6) and two post-critical regions (Regions 7 and 8) both by participant and by item. The two post-critical regions were included in the RT analyses, because the previous studies that used similar experimental materials (Jiang, 2007; Jiang et al. 2011; Wen et al., 2010) observed spillover effects in these regions (see Mitchell, 1984, for discussion of spillover effects). For each of these regions, data from both native and L2 groups were submitted to three-way mixed analyses of variance (ANOVAs) with Group (native vs. ESL) × Grammaticality (grammatical vs. ungrammatical) ×

Structure (simple vs. partitive) as the target factors. In the participant analyses, Group was a between-participant factor, while Grammaticality and Structure were within-participant factors. In the item analyses, Structure was a between-item factor, whereas the other two factors were within-item factors. The primary purpose of these analyses was to see how similarly or differently the native group and the ESL group behaved in processing the missing plural inflections. ANOVA statistics for the critical and post-critical regions are presented in Table 19.

Table 19. Omnibus repeated measures analysis of variance.

	By participant					By item				
	df	MS	F_1	*p*	η^2	*df*	MS	F_2	*p*	η^2
Region 6										
Group**	1	14470.31	10.36	.003	.21	1	44920.16	10.49	.003	.26
Grammaticality**	1	21136.36	6.88	.012	.15	1	11578.05	6.78	.014	.18
Structure*	1	19381.29	5.59	.023	.13	1	3952.77	3.58	.068	.11
Group×Grammaticality	1	862.99	.28	.599	.01	1	838.40	.39	.536	.01
Group×Structure	1	868.89	.25	.619	.01	1	1115.99	.26	.613	.01
Grammaticality×Structure**	1	10615.49	5.86	.020	.13	1	13612.29	7.97	.008	.21
3-way interaction	1	2274.22	1.26	.269	.03	1	2348.78	1.10	.304	.04
Region 7										
Group	1	30998.90	1.20	.280	.03	1	3185.52	1.27	.268	.04
Grammaticality**	1	59656.86	13.29	.001	.25	1	38247.21	16.43	.000	.35

Structure	1	2723.67	.74	.394	.02	1	660.82	.78	.394	.02
Group×Grammaticality**	1	47073.80	10.49	.002	.21	1	34787.64	14.35	.001	.32
Group×Structure	1	3551.05	.97	.331	.02	1	2854.94	1.14	.294	.04
Grammaticality×Structure	1	10436.39	2.73	.105	.07	1	5955.91	2.56	.120	.08
3-way interaction	1	4085.56	1.07	.31	.03	1	2467.27	1.02	.321	.03
Region 8										
Group**	1	10594.20	8.55	.006	.18	1	30969.38	23.74	.000	.44
Grammaticality**	1	73709.15	17.28	.000	.31	1	64098.11	40.37	.000	.57
Structure	1	118.93	.36	.852	.00	1	.56	.00	.977	.00
Group×Grammaticality	1	2799.02	.66	.423	.02	1	1878.23	.90	.350	.03
Group×Structure*	1	8461.02	2.52	.121	.06	1	6042.30	4.63	.040	.13
Grammaticality×Structure	1	420.67	.15	.70	.00	1	582.09	.37	.55	.01
3-way interaction	1	245.25	.09	.765	.00	1	240.57	.12	.737	.00

*Significant at .05 level in either participant or item analysis

**Significant at .05 level in both participant and item analysis

Furthermore, each group's residual RT data were separately analyzed by planned paired-samples *t*-tests, since the major interest of this study was to determine how the learners reacted to plural inflection errors in each structural condition. Specifically, for each region, residual RTs for the grammatical and ungrammatical sentences in each structural condition were compared within each group. Here, the alpha level was not adjusted (i.e., it remained set at .05), because these comparisons were orthogonal to each other (see also Jiang et al., 2011; Keating, 2009). Each group's mean residual RTs for each grammaticality condition in each structural condition along with RT differences between the ungrammatical and the grammatical sentences are presented in Table 20.

Table 20. Residual RT means (standard deviations) of the English native speaker and Korean ESL groups for grammatical and ungrammatical sentences at four regions within a sentence (ms)

		Simple DP/QP				Partitive QP			
		Kevin memorized those long				*Mary donated many of*			
	Region	R5 Latin	R6 word(s)	R7 in	R8 just…	R5 her	R6 book(s)	R7 to	R8 the…
English native speakers (n = 19)	Grammatical	-43(47)	-53(38)	-7 (22)	-19 (48)	-55 (30)	-71 (54)	2 (50)	-32 (34)
	Ungrammatical	-19(37)	-17(43)	91(97)	31 (60)	-53 (29)	-52 (38)	48(75)	19 (65)
	Difference	24	36*	98*	50*	2	19	46*	51*
Korean ESL learners (n = 22)	Grammatical	-25 (49)	-23 (51)	16 (65)	-54 (51)	-58 (74)	-16 (88)	23 (57)	-44 (71)
	Ungrammatical	-15 (50)	19 (70)	26 (68)	-26 (73)	-71 (49)	-22 (64)	21 (62)	-4 (75)
	Difference	10	42*	10	28	-13	-6	-2	40*

*Significant at .05 level in paired t-tests (two-tailed) in both participant and item analyses

At Region 6 (the critical region), the mixed ANOVAs revealed a significant main effect for Group, $F_1(1, 39) = 10.36$, $p = .003$, $\eta^2 = .21$, $F_2(1, 30) = 10.49$, $p = .003$, $\eta^2 = .26$. Both native and ESL groups read this region more quickly than expected, but the native speakers did so to a significantly larger degree than the learners (-48 ms vs. -10 ms). Crucially, a significant main effect of Grammaticality was found, $F_1(1, 39) = 6.88$, $p = .012$, $\eta^2 = .15$, $F_2(1, 30) = 6.78$, $p = .014$, $\eta^2 = .18$, but no significant interaction between Grammaticality and Group was found, $F_1(1, 39) = .28$, $p = .599$, $\eta^2 = .01$, $F_2(1, 30) = 0.391$, $p = .536$, $\eta^2 = .01$. As a group, the native speakers and ESL learners took less time than expected to read Region 6 in both Grammatical and Ungrammatical conditions, but to a significantly smaller degree in the Ungrammatical condition than in the Grammatical condition (-18 ms vs. -41 ms). This result shows that both native speakers and learners were sensitive to the plural inflection errors, and that the two groups' levels of sensitivity to the errors were comparable. In addition, a main effect for Structure reached significance in the participant analysis, $F_1(1, 39) = 5.59$, $p = .023$, $\eta^2 = .13$, and approached significance in the item analysis, $F_2(1, 30) = 3.07$, $p = .068$, $\eta^2 = .11$. The participants took less time than expected in both structural conditions, but to a significantly larger degree in the Partitive condition than in the Simple condition (-40 ms vs. -19 ms). Crucially, the interaction between Grammaticality and Structure was significant, $F_1(1, 39) = 5.86$, $p = .020$, $\eta^2 = .13$, $F_2(1, 30) = 7.97$, $p = .008$, $\eta^2 = .21$, but the Grammaticality × Structure × Group interaction was not, $F_1(1, 39) = 1.26$, $p = .269$, $\eta^2 = .03$, $F_2(1, 30) = 1.11$, $p = .304$, $\eta^2 = .04$. These results suggest that the two groups' sensitivity to the plural inflection errors was modulated by the given structures (i.e., Simple and Partitive) in a similar way.

Subsequent planned paired-samples *t*-tests provided further information on the reliable interaction between Grammaticality and Structure. The *t*-tests demonstrated that both native speakers and L2 learners were sensitive to plural inflection errors only in the Simple condition. That is, both groups significantly slowed down to read ungrammatical sentences in the Simple condition; the native speakers read faster than expected in both Grammatical and Ungrammatical conditions, but to a significantly smaller degree in the Ungrammatical condition than in the Grammatical condition (-17 ms vs. -53 ms), $t_1(18) = 3.70$, $p = .002$, $d = .85$, $t_2(15) = 2.82$, $p = .013$, $d = .71$; the ESL learners read faster than expected in the Grammatical condition (-23 ms), but they read slower than expected in the Ungrammatical condition (19 ms), $t_1(21) = 2.40$, $p = .026$, $d = .520$, $t_2(15) = 2.97$, $p = .009$, $d = .69$x. In contrast, neither group exhibited this grammaticality effect in the Partitive condition. Both native speakers and learners read faster than expected both in the Grammatical condition and in the Ungrammatical condition to comparable degrees (Native: -71 ms vs. -52 ms, $t_1(18) = 1.41$, $p = .176$, $t_2(15) = .77$ $p = .451$; ESL: -16 ms vs. -22 ms, $t_1(21) = -.30$, $p = .765$, $t_2(15) = -.83$, $p = .422$). Moreover, *t*-test results identified a major source of the main effect of Structure found in the ANOVA results. The residual RTs in the Simple condition were larger than those in the Partitive condition, mainly because participants slowed down in ungrammatical sentences only in the Simple condition.

At Region 7 (the first post-critical region), the ANOVA found a main effect of Grammaticality again, $F_1(1, 39) = 13.29$, $p = .001$, $\eta^2 = .25$, $F_2(1, 30) = 16.43$, $p < .001$, $\eta^2 = .35$. As a group, the native speakers and ESL learners spent longer reading this region than expected in both Grammatical and Ungrammatical conditions, but to a significantly larger degree in the Ungrammatical condition than in the Grammatical condition (43 ms vs. 9 ms). This time, however, the effect was qualified by the interaction between Grammaticality and Group, $F_1(1, 39) = 10.49$, $p = .002$, $\eta^2 = .212$, $F_2(1, 30) = 14.35$, $p = .001$, $\eta^2 = .324$. Taken together, these results suggest that at this region, there was some difference between the native speakers and the L2 learners in the level of sensitivity, although they revealed sensitivity to the plural inflection errors as a group.

The source of the significant Grammaticality × Group interaction was revealed by subsequent paired-samples *t*-tests. At Region 7, the native speakers slowed down when reading ungrammatical sentences in both structural conditions. In the Simple condition, they read a bit faster than expected in the Grammatical condition (-7 ms), but they read much more slowly than expected in the Ungrammatical condition (91 ms), $t_1(18) = 4.49$, $p < .001$, $d = 1.38$ t_2 $(15) = 5.43$, $p < .001$, $d = 1.46$; in the Partitive condition, they read as fast as expected (2 ms) in the Grammatical condition, but they read much more slowly than expected in the Ungrammatical condition (48 ms), $t_1(18) = 2.22$, $p = .039$, $d = .52$ $t_2(15) = 2.49$, $p = .025$, $d = .63$. On the other hand, the learners did not slow down in either of the structural conditions. In both conditions, they read grammatical and ungrammatical sentences faster than expected to comparable degrees (Simple: 16 ms vs. 26 ms, $t_1(21) = .48$, $p > .640$, $t_2(15) = .41$, $p = .691$; Partitive: 23 ms vs. 21 ms, $t_1(21) = -.11$, $p = .916$, t_2 $(15) = -.18$, $p = .860$). These results indicate that at this region, the native speakers were sensitive to missing plural inflection in both structural conditions, but such sensitivity was not detected for the learners in either of the conditions.

Lastly, at Region 8 (the second post-critical region), the ANOVA for the residual RTs also yielded a main effect of Grammaticality, $F_1(1, 39) = 17.28$, $p < .001$, $\eta^2 = .31$, $F_2(1, 30) = 40.37$, $p < .001$, $\eta^2 = .57$. Participants were much faster than expected when reading grammatical sentences (-37 ms), while they were slower than expected when reading ungrammatical sentences (5 ms), showing sensitivity to the plural inflection errors. In addition, a significant main effect of Group was found, $F_1(1, 39) = 8.55$, $p = .006$, $\eta^2 = .18$, $F_2(1, 30) = 23.74$, $p < .001$, $\eta^2 = .44$. The native speakers read this region as fast as was expected (0.41 ms), while the ESL learners read more quickly than was expected (-32 ms). However, the interaction between these two factors did not reach significance, $F_1(1, 39) = .66$, $p = .423$, $\eta^2 = .02$, $F_2(1, 30) = .90$, $p = .350$, $\eta^2 = .03$, indicating that, overall, all participants were sensitive to errors. Furthermore, the interaction between Structure and Group was significant only in the item analysis, $F_1(1, 39) = 2.52$, $p = .121$, $\eta^2 = .06$, $F_2(1, 30) = 4.63$, $p = .040$, $\eta^2 = .13$, suggesting that the native and ESL groups' RTs were modulated by the given structure in different ways.

Subsequent paired-samples *t*-tests revealed the source of the significant interaction between Structure and Group. Although both groups tended to slow down at this region when reading ungrammatical sentences in both structural conditions, the extent of the ESL group's slowdown due to the inflection error in the Simple condition was relatively small (28 ms), and not statistically significant, $t_1(21) = 1.52$, $p = .144$, $t_2(15) = 1.92$ $p = .074$; in contrast, their slowdown in the Partitive condition was relatively large (40 ms) and statistically significant, $t_1(21) = 2.15$, $p = .043$, $d = .46$, t_2 $(15) = 2.22$ $p = .042$, $d = .60$. On the other hand, the native speakers' slowdown in both structural conditions was relatively large and statistically reliable (Simple: 50 ms slowdown, $t_1(18) = 2.66$, $p = .016$, $d = .80$, $t_2(15) = 4.07$, $p = .001$, $d = 1.04$; Partitive: 51 ms slowdown, $t_1(18) = 2.98$, $p = .008$, $d = .71$ $t_2(15) = 4.87$, $p < .001$, $d = 1.15$). Consequently, the two groups showed different patterns of reading speed in the two structures. The learners read faster than expected in both structural conditions, but to a larger degree in the Simple condition (-40 m) than in the Partitive condition (-24 ms). On the other hand, the native speakers read slower than expected in the Simple condition (6 ms), whereas they read faster than expected in the Partitive condition (-6 ms).

In sum, both native speakers and ESL learners showed reliable RT differences between grammatical and ungrammatical sentences at the critical and/or post-critical regions in both structural conditions, signifying their sensitivity to the missing plural inflection, although some group differences were found. Also, their slowdown patterns in the two structural conditions differed considerably, suggesting that plural inflection processing is affected by the structural distance of feature-checking dependency. The native speakers immediately slowed down when they encountered the inflection error in the simple DP/QPs (i.e., at Region 6). In contrast, their slowdown due to the error in the partitive QPs was delayed until they reached Region 7. Like the native speakers, the ESL learners immediately reacted to the errors in the simple DP/QPs; they slowed down at Region 6 in the Simple condition in response to the plural errors. Their reaction to the errors in the Partitive condition was also similar to the native speakers' in that it was delayed as well. The learners did not slow down until they reached Region 8. The slowdowns of the two language groups were comparable in magnitude (Native: 46 ms at Region 7 and 51 ms at Region 8; ESL: 40 ms at Region 8), although their reaction was more delayed than that of native speakers.

6.3 Discussion

In order to test the RDA and the SSH, the present study investigated (a) whether Korean ESL learners, whose L1 plural inflection is quite different from that of English, are sensitive to English plural inflection errors during online sentence processing, and (b) how the structural distances of the feature-checking dependencies in the structures affect their processing of plural inflection. The results suggest that (1) during sentence processing, Korean ESL learners are sensitive to plural inflection in English, despite the L1–L2 difference, and (2) both L1 and L2 processing of plural inflection in English can be delayed when the inflection involves a longer feature-checking dependency. These findings are interpreted as evidence against the RDA and the SSH. The following subsections discuss these issues in detail.

6.3.1 L2 Representation of Plural Inflection in English

Despite the substantial difference of the English and Korean plural-inflection systems, the late Korean ESL learners were sensitive to missing plural inflection both in the simple structure and in the partitive structure. Like the native speakers, they slowed down in response to the inflection errors in both structures. As previously discussed, successful processing of the plural inflections in the simple structure and the partitive structure requires knowledge of DP/QP internal number agreement and c-selectional [+plural] feature checking, respectively; however, Korean lacks such grammatical feature checking processes. Therefore, the Korean ESL learners' sensitivity to the plural errors in the two structures—which reflects their successful acquisition of the two different types of feature checking—argues against the RDA. According to the RDA, the Korean ESL learners cannot acquire such feature-checking processes, because they are not instantiated in Korean.

Here it should be emphasized that the ESL learners were sensitive not only to the plural error in the simple structure, but also to that in the partitive structure. To the best of my knowledge, no previous study has shown that L2 learners can attain native-like knowledge of plural inflection in the partitive when the equivalent structure of their L1 does not involve the c-selectional checking of the [+plural] feature, since Jiang (2007) and Jiang et al. (2011) failed to observe highly advanced Chinese and Japanese ESL learners' sensitivity to missing plural inflection in the same structure. Hence, the Korean ESL learners' sensitivity to plural inflection (errors) found in the current study adds to evidence against the RDA.

How then can then the insensitivity of Jiang's (2007) and Jiang et al.'s (2011) participants to L2 inflections be explained, especially since those participants were highly advanced ESL learners as well? A review of test materials suggests that some inconsistencies in the experimental items might have led to the conflicting results between those two studies and the present investigation. First, although most of the partitives in the studies conducted by Jiang and colleagues consisted of four words, some had more than four words. For instance, some partitives had an additional word before the noun at the end (e.g., *several of the board member(s)*). In the present study, every partitive consisted of four words. Note that an additional word can be expected to increase the cost of processing the plural inflection in the partitive, because it increases the linear distance of the feature-checking dependency, as (22) shows. Therefore, the sensitivity of Jiang and Jiang et al.'s L2 learners might have just been obscured by the extra processing load caused by the additional words in their partitives, as shown in (22):

(22) [$_{QP}$ several [$_{NP}$ (UNIT) $_{\{+count\}}$ [$_{PP}$ of [$_{DP}$ the [$_{NP}$ board members $_{\{+plural\}}$]]]]]

Existing literature seems to lend support to this possibility, although it is left to future research to test it. For example, Keating (2010) showed that the linear distance between the noun and the adjective in gender agreement dependency affects advanced L2 learners' sensitivity to violations of gender. The language of interest was Spanish, and the learners' L1 was English. The results show that the learners of Spanish failed to display sensitivity when the linear distance was relatively long (i.e., four or seven words), but did show sensitivity when it was short (i.e., one word). Furthermore, the learners' sensitivity was modulated by their memory span. Keating's interpretation of these results was that L2 learners' insensitivity to gender violations in the long linear distance conditions reflects imposed computational overload, rather than representational deficits. Thus, future research should investigate the effects of linear distance and working memory on native speaker and L2 learner sensitivity to the plural inflection error in the partitive structure, in order to verify the current interpretation of Jiang's (2007) and Jiang et al.'s (2011) results.

6.3.2 The SSH Revisited

The results of this study also provide evidence against the SSH. First, they counter the prediction of the SSH that L2 learners cannot process nonlocal agreement of grammatical features because L2 learners' representation is syntactically shallow. As discussed earlier, the plural inflection in the simple and the partitive structures involves nonlocal grammatical feature checking; in both structures, the two elements in the feature-checking dependency are distant from each other, not only linearly but also structurally (see [20] and [21]). However, the ESL learners in this study displayed their sensitivity to the plural errors in both structures, suggesting that they do compute representations with rich syntactic detail, which are required for successful nonlocal feature checking, during online sentence processing.

In addition, the finding of structural distance effects on L2 processing provides further evidence against the SSH. This is because the structural distance of the feature-checking dependency was longer in the partitive structure than in the simple structure, whereas linear distance was controlled for (see [20] and [21]). Therefore, it was expected that the relatively long structural distance of the partitive would impose more computational burden on the native speakers than the relatively short structural distance of the simple DP/QP, and consequently that their processing of plural inflection would become delayed when reading partitives, compared to when reading simple DP/QPs. This expectation was borne out: native speakers' sensitivity to plural inflection errors (i.e., slowdown in reading) emerged as soon as they encountered the errors (i.e., at Region 6) when reading simple DP/QPs, but it surfaced at a later point (e.g., at Region 7) when reading partitives. Crucially, this delayed sensitivity was observed with the advanced Korean ESL learners as well. As the native speakers, the learners slowed down immediately (i.e., at Region 6) when they encountered the inflection error in the simple structure, but they slowed down somewhat later (i.e., at Region 8) when they encountered the error in the partitive structure.

This delayed sensitivity of the L2 learners stands against the SSH, suggesting that nonnative speakers—at least advanced ones—do construct hierarchically structured representations during real-time language comprehension, like native speakers. If learners compute shallow representations that are not hierarchically structured, as the SSH proposes, their processing cannot be influenced by structural distance, because a speaker's sentence processing can be influenced by structural distance only when the parser computes hierarchically structured representations.

6.3.3 Differences between L1 and L2 Processing

As discussed in the previous section, the Korean ESL learners in this study behaved very similarly to the native speakers in terms of inflection processing, suggesting that L1 and L2 processing are not fundamentally different from each other. However, this study also found a quantitative difference that suggests a limitation of L2 processing: when reading partitives, the Korean ESL learners' sensitivity to missing plural inflection emerged one word region later than for the native speakers; while the native speakers began to slow down at Region 7, the learners did not slow down until they reached Region 8 (See Table 20). Such late sensitivity has been reported in previous studies as well (Keating, 2010; Sabourin & Stowe, 2008; see also Gillon Dowens et al., 2010). For example, in an eye-tracking experiment on gender agreement in Spanish, Keating (2010) found that Spanish native speakers showed their sensitivity to gender violations (i.e., increase in RT) both in first-pass and second-pass reading, while advanced anglophone learners did not show such sensitivity until second-pass reading. Moreover, in an Event-Related Potential (ERP) study of gender agreement in Dutch, Sabourin and Stowe (2008) observed that although native German-speaking learners of L2 Dutch exhibited a clear P600 effect in reaction to gender violations, it peaked later compared to Dutch native speakers. These researchers interpreted this late sensitivity as a quantitative difference between L1 and L2 processing—rather than a qualitative one—that reflects the relatively slower L2 processing speed.

This study also observed another quantitative difference between L1 and L2 processing, although this interpretation of the difference must await future research. The L2 learners showed relatively more attenuated sensitivity than the native speakers: Their sensitivity was relatively smaller in magnitude and relatively more transient in duration. As Table 20 shows, in the Simple condition, the native speakers' maximum RT increase in response to missing plural inflection was 98 ms, whereas the learners' was only 42 ms. Table 5 also shows that the learners' slowdown was maintained only at one region (at Region 6 in the Simple condition and at Region 8 in the Partitive condition), whereas the native speakers' slowdown was continuously observed over more than one region (from Region 6 to Region 8 in the Simple condition, and from Region 7 to Region 8 in the Partitive condition). Although this attenuated L2 sensitivity possibly reflects limitations of L2 processing, it is left to future research to investigate what this phenomenon really signifies. To the best of my knowledge, no study of L2 inflection processing has discussed this type of attenuated sensitivity of L2 learners in detail, although the phenomenon has also been observed in other studies (Jiang, 2007; Jiang et al., 2011); see also Gillon Dowens et al., 2010).

6.4 Conclusion

Although the results of this study show that the Korean ESL learners were sensitive to plural inflection in the partitive structure, we cannot exclude a possibility that these learners have a non-target-like representation such as (23):

(23) [$_{QP}$ several$_{\{+plural\}}$ [$_{PP}$ of [$_{DP}$ the [$_{NP}$ toys$_{\{+plural\}}$]]]]

In other words, it might have been the case that the L2 learners were sensitive to missing plural inflection in the partitive, not because they have native-like knowledge of the c-selectional checking of the [+plural] feature, but because they misapplied their number agreement knowledge to this structure. The present study cannot refute this alternative explanation. In order to exclude this possibility, a future design should examine how L2 learners react to plural inflection in partitives where the quantifier and noun clearly disagree with each other in number, such as *one of the toys*. If learners in fact misapply number agreement knowledge to the partitive structure, in such a study they would consider the grammatical phrase *one of the toys* ungrammatical, and the ungrammatical phrase *one of the toy* grammatical.

Furthermore, while the simple DP/QP structure and the partitive structure are clearly distinguished in terms of the structural distance of the feature-checking dependency, it should be noted that these two structures may differ in other important aspects as well (e.g., semantic complexity, usage frequency, etc.). For example, the semantic complexity of the simple DP/QPs and partitives of this study may differ at least to some extent. First, excluding the determiner/quantifier, the simple DP/QP contains three content words (i.e., two adjectives and one noun), which is expected to increase the semantic processing load; in contrast, excluding the quantifier, the partitive contains only one content word, along with two functional words (i.e., the preposition *of* and a determiner), and thus its semantic processing load is expected to be minimal. In this sense, the simple DP/QPs appear to be more semantically complex than the partitives. However, at the same time, the partitives seem to be more semantically complex in that they involve the concept of the subset-superset relation (see Section 5.3.2). Although in this study, I assumed that these semantic complexity differences cancel each other out (and thus that the two structures are comparable in semantic complexity), I acknowledge the possibility that a more thorough and precise measure would find the simple DP/QP and partitive structures to differ in semantic complexity. The existence of this possibility suggests that my interpretation of the participants' delayed sensitivity in the partitive construction may need to be modified depending on the results of future research that will take into account not only syntactic (or structural) complexity but also other important aspects of syntactic constructions, including their semantic complexity.

Two methodological shortcomings might have influenced the participants' performances in the present design. First, the number of filler sentences was relatively small: a ratio of 1 to 1. As pointed out by a reviewer of an earlier article on this study (Song, 2015), typically, fillers outnumber experimental items by a ratio of 2 or 3 to 1, since the lack of adequate number fillers may make participants consciously aware of the aim of study. Second, the experimental and filler items were presented to each participant in a completely random order rather than a pseudorandom order. As the same reviewer pointed out, the consecutive appearance of grammatical errors of the same type, which may occur in a completely random-order presentation, may also lead participants to become aware of the purpose of the study. Due to these two limitations, one cannot completely exclude the possibility that some Korean ESL learners in this study might have tapped into their classroom knowledge of how to form English plural inflections.

It should also be pointed out that we need more research testing late L2 learners' sensitivity to inflection that involves nonlocal (i.e., non-adjacent) feature checking. The current literature does include some studies that provide evidence for such sensitivity, but the number of such studies is still very small (e.g., Foote, 2011; Keating, 2010; Wen et al., 2010). In order for the idea that L2 learners' syntactic representations do not lack syntactic detail to be strongly supported, future research needs to more thoroughly investigate L2 sensitivity to inflection in nonlocal contexts. Ideally, such sensitivity should be tested using a variety of grammatical features (e.g., gender, number, person, etc.) and types of nonlocality (i.e., linear vs. structural).

Overall, the results reported in this study suggest that native Korean-speaking later ESL learners can eventually attain target-like knowledge of plural inflection in English, even if their L1 lacks an equivalent inflection. This provides evidence against the RDA, or the position that L2 learners necessarily have a ceiling in their acquisition of target inflectional morphology because of a representational deficit (e.g., Franceschina, 2001; Hawkins & Liszka, 2003; Jiang et al., 2011; Sabourin & Stowe, 2008). Furthermore, the results are in line with the idea that L2 learners, at least at the advanced level of proficiency, do compute hierarchically structured representations with rich syntactic detail during online sentence processing, although their processing of the target inflection may not be as efficient as that of native speakers. These findings lend support to the idea that L2 inflection processing is not fundamentally different from

L1 inflection processing, which argues against the SSH, or the position that L2 learners construct syntactically less detailed representations and are incapable of native-like processing of inflection that involves nonlocal checking of grammatical features (Clahsen & Felser, 2006a, 2006b). Finally, quantitative but not qualitative differences between L1 and L2 processing were observed, in that the L2 learners' sensitivity to inflectional ungrammaticality was relatively smaller in magnitude and relatively more transient in duration than that of the native speakers. This attenuated L2 sensitivity has been reported in previous studies but its theoretical importance has gone unnoticed. I suggest attenuated sensitivity of L2 learners to inflectional ungrammaticality possibly reflects limitations of L2 processing and is a phenomenon whose theoretical significance is worthy of future systematic attention.

Chapter 7 Conclusion

In this dissertation, I have investigated one of the major unsettled issues in the field of second language acquisition—namely, whether L2 learners can develop mental representations (or grammatical knowledge) similar to native speakers', or can only develop mental representations that are fundamentally different from native speakers'—focusing on derivational and inflectional morphology in the L2. I presented a morphological priming experiment and a self-paced reading experiment that were designed to investigate L2 processing of suffixed derived words in English and L2 processing of plural inflection in English, respectively. Overall, the results of these two experiments suggested that L2 learners can eventually develop native-like representations, at least in certain lexical and syntactic domains, as their overall L2 proficiency increases, although L2 lexical and syntactic processing might be less efficient or slower than L1 processing.

The two major results of this dissertation research and their theoretical implications are summarized in Sections 7.1 and 7.2. To conclude, Section 7.3 sketches some possible extensions of the research to further address the question of whether L2 learners can eventually develop lexical and syntactic representations that are qualitatively similar to those of native speakers.

7.1 Lexical and Syntactic Representations in L2

The dissertation first presented a morphological priming experiment. The results showed that late L2 learners, at least highly advanced ones, process derived words in English in very similar ways to native speakers of English (Experiment 1). In the experiment, native speakers demonstrated robust morphological priming effects for semantically transparent derived words in

English. The priming effects were independent both of the whole-word and base frequency and of the base and affix productivity of the words, suggesting that native speakers morphologically decompose these words independently of such lexical or distributional factors. Crucially, this pattern was replicated by highly advanced Korean-speaking late L2 learners of English. Like native speakers, the L2 learners exhibited strong morphological priming effects independently of the lexical or distributional factors. These results lend support to the position that late L2 learners can eventually develop native-like representations of derived words in English in which the morphological structure of the words is analyzed.

The dissertation then shifted its scope of investigation to L2 processing of inflectional morphology, and presented a self-paced reading experiment (Experiment 2). The results of this experiment suggested that late L2 learners can eventually acquire particular aspects of the knowledge of plural inflection in English (i.e., DP/QP-internal number agreement and c-selectional [+plural] feature checking in partitives), even if their L1 lacks such knowledge, providing evidence against the RDA (e.g., Franceschina, 2001; Hawkins & Liszka, 2003; Jiang et al., 2011; Sabourin & Stowe, 2008). The Korean-speaking late L2 learners of highly advanced proficiency in the experiment clearly exhibited their sensitivity to missing plural inflection in both simple DP/QPs and partitive QPs, like native speakers. More specifically, the L2 learners significantly slowed down at or after word-regions containing a plural inflection error, as native speakers did. In addition, the results provide some evidence that L2 learners, at least at the advanced level of proficiency, construct hierarchically structured representations with syntactic detail during real-time language comprehension as native speakers do, arguing against the SSH (Clahsen & Felser, 2006a, 2006b). Not only for the native speakers but also for the advanced L2 learners, the processing of plural inflection in English was delayed when the inflection involved a structurally longer feature-checking dependency, signifying all participants' sensitivity to structural distance.

Taken together, the results of the two experiments suggest that at least in certain lexical and syntactic domains, native-like representation is attainable by late L2 learners.

7.2 Differences between L1 and L2 Processing

As summarized in the previous section, the L2 learners and native speakers in the two experiments demonstrated very similar patterns in the processing of derived words and plural inflection in English, suggesting that L1 and L2 processing do not fundamentally differ from each other. Nonetheless, the experiments also found some quantitative, rather than qualitative, differences between L1 and L2 processing that appear to suggest some limitations of L2 processing. First, in the morphological priming experiment, the magnitude of the L2 learners' morphological priming effect was affected to a small extent by the interactions between the whole-word and base frequency and the base and affix productivity of given derived words, whereas the native speakers' priming effect was not influenced by such interactions. In the self-paced reading experiment, the L2 learners' sensitivity to inflectional ungrammaticality was relatively smaller in magnitude and relatively more transient in duration than that of the native speakers. In addition, with relatively complex syntactic structures, the L2 learners showed delayed sensitivity compared to the native speakers. In this dissertation, all these differences between L1 and L2 processing were interpreted as indicating relatively less efficient or slower L2 processing compared to L1 processing. Further research, however, is required for a more precise understanding of these differences.

7.3 Future Avenues of Study

Although this dissertation provides further evidence for the idea that L2 learners can develop native-like knowledge, currently, this idea is still supported only in some lexical and syntactic domains. For example, no study has yet explored whether the individual morphemes of multi-morphemic words are represented in hierarchical structure in L2 such that, for instance, the internal structures of *unkindness* and *unreadable* (i.e., [[un-kind]-ness] and [un-[read-able]]) are distinguished, as is assumed to be the case in L1 (e.g., Lieber, 1980). Also, the research on how compound words or prefixed derived words are represented in L2 is almost non-existent. It is left to future research to explore whether L2 learners can develop mental representations qualitatively similar to those of native speakers in such unexplored domains as well.

Appendices

Appendix A: Critical Prime-Target Pairs and their Characteristics

Critical prime-target pairs and their characteristics

Related prime	Unrelated prime	Target (base)	Log WWF of related prime	Log base frequency	Relative log frequency	Family size
attainable	militarism	attain	3.14	5.46	1.74	2
manageable	alcoholism	manage	3.81	7.74	2.03	9
obtainable	capitalize	obtain	4.01	7.10	1.77	1
predictable	territorial	predict	5.21	6.24	1.20	6
acceptance	exhibition	accept	6.02	8.22	1.37	4
annoyance	hazardous	annoy	4.70	5.75	1.22	2
assistance	ridiculous	assist	6.01	5.89	0.98	3
attendance	projection	attend	5.47	7.20	1.32	4
disturbance	scholarship	disturb	5.42	6.66	1.23	2
endurance	vibration	endure	4.67	5.81	1.24	4
resistance	biological	resist	6.40	6.61	1.03	8
variance	colonist	vary	3.26	6.92	2.12	13

alteration	brightness	alter	4.80	6.56	1.37	3
cancellation	tranquillity	cancel	4.23	5.67	1.34	1
combination	electricity	combine	6.59	7.08	1.07	3
consultation	championship	consult	5.48	6.33	1.16	5
declaration	progressive	declare	5.51	6.90	1.25	4
expectation	imperialism	expect	5.00	8.54	1.71	5
expiration	neglectful	expire	2.30	4.34	1.89	2
invitation	electronic	invite	6.08	7.07	1.16	2
dependence	corruption	depend	5.43	6.35	1.17	14
difference	industrial	differ	7.90	6.01	0.76	7
existence	wonderful	exist	7.10	7.77	1.09	12
inference	modernize	infer	4.50	4.36	0.97	2
insistence	accidental	insist	5.14	7.24	1.41	3
occurrence	contractor	occur	4.65	7.71	1.66	1
persistence	crystallize	persist	4.44	6.06	1.36	3
preference	mysterious	prefer	6.00	7.29	1.21	5

equality	alliance	equal	6.30	6.99	1.11	11
hostility	collector	hostile	5.68	6.17	1.09	3
humidity	arguable	humid	3.50	3.95	1.13	3
mobility	consumer	mobile	6.38	5.61	0.88	7
morality	delivery	moral	5.67	6.94	1.22	13
punctuality	familiarize	punctual	2.83	2.89	1.02	2
rapidity	adorable	rapid	3.58	6.45	1.80	2
sincerity	violation	sincere	4.47	5.04	1.13	5
effortless	lesbianism	effort	3.47	7.80	2.25	3
fatherless	burdensome	father	3.09	8.59	2.78	9
merciless	adulthood	mercy	4.17	5.66	1.35	5
motionless	journalism	motion	5.04	6.57	1.30	3
frequently	attraction	frequent	6.94	5.98	0.86	5
honestly	tropical	honest	5.80	6.47	1.12	6
mentally	historic	mental	5.71	6.89	1.21	2
overtly	glorify	overt	3.93	4.70	1.20	2

properly	northern	proper	6.82	7.02	1.03	4
quietly	payment	quiet	6.89	7.38	1.07	9
stupidly	borrower	stupid	3.85	6.52	1.69	2
timidly	agonize	timid	3.56	4.93	1.39	4
achievement	corporation	achieve	6.57	7.66	1.17	1
agreement	professor	agree	7.13	8.14	1.14	8
alignment	terrorize	align	4.04	3.78	0.94	4
announcement	characterize	announce	5.60	7.20	1.28	3
assessment	reflection	assess	6.00	6.03	1.01	2
involvement	desperately	involve	6.04	8.24	1.36	2
postponement	hierarchical	postpone	3.85	5.35	1.39	1
punishment	marvellous	punish	6.44	6.04	0.94	2
awareness	infection	aware	6.07	7.48	1.23	3
clumsiness	pressurize	clumsy	3.26	5.38	1.65	2
eagerness	planetary	eager	4.43	6.04	1.36	3
emptiness	digestion	empty	4.51	7.24	1.60	6

gentleness	parenthood	gentle	3.87	6.48	1.67	3
happiness	formation	happy	6.22	7.91	1.27	7
laziness	theorize	lazy	3.58	7.72	2.15	2
politeness	censorship	polite	4.51	5.97	1.32	5
dangerous	financial	danger	7.29	7.44	1.02	3
furious	density	fury	5.57	5.75	1.03	3
glamorous	symbolism	glamor	4.54	4.88	1.07	3
humorous	donation	humor	4.55	6.06	1.33	4
mountainous	endorsement	mountain	3.76	7.31	1.94	6
rigorous	seasonal	rigor	4.43	4.23	0.96	2
scandalous	inflection	scandal	3.76	5.52	1.47	3
vigorous	survivor	vigor	5.39	4.90	0.91	3

*WWF = whole-word frequency

Appendix B: Filler Prime-Target Pairs

Filler prime-target pairs

Filler type	Prime	Target
Semantically related word-word pairs	aptitude	talent
	astonishment	surprise
	comfortable	cozy
	difficulty	problem
	obligation	duty
	occupation	career
	rediculous	absurd
	specialist	expert
	valuable	precious
Orthographically related word-word pairs	candidacy	candid
	colonize	colon
	commander	comma
	demonstrable	demon
	entertainer	enter
	eventful	even
	extractor	extra
	rationalism	ratio
	villainous	villa
Orthographically related word-nonword pairs	adversely	idverse
	amazing	abaze
	atonement	abone
	believable	melieve
	cartoonist	lartoon
	certainly	fertain
	cleverly	flever
	clinical	flinic
	debatable	gebate
	defendant	nefend
	defiance	mefy
	deportation	meport
	deprivation	meprive
	description	lescript
	disruptive	fisrupt
	divinely	hevine
	effective	effoct
	election	edect
	excusable	ebscuse
	exertion	edert
	fanciful	gancy
	fashionable	mashion

	featureless	deature
	fellowship	hellow
	fertilize	gertile
	humanism	duman
	ignition	idnite
	invader	ibvade
	inventive	idvent
	legalize	cegal
	licensee	licente
	limitless	himit
	machinery	fachine
	marginal	bargin
	marketable	harket
	masterful	baster
	measurable	heasure
	narrator	marrate
	narrowness	carrow
	nervousness	bervous
	nobleness	joble
	objective	oc ject
	observer	ofserve
	respondent	hespond
	retirement	letire
Orthographically related word-nonword pairs	appearance	junny
	appliance	kingo
	brotherhood	jower
	civilize	massel
	cognizance	measlos
	coherence	lutton
	colorless	neddle
	convergence	humbup
	conversation	nairy
	coverage	iglay
	creative	kentor
	deterrence	lormant
	emergent	knotny
	envious	hantry
	expensive	jatal
	feverish	loward
	flavorless	midden
	formalize	acclict
	indulgent	lother
	information	durdle
	leverage	oftuse
	majority	ifvent
	membership	hallet

minority	ojive
motherless	haddie
murderous	offinse
numberless	abong
partnership	bumbo
powerless	abhar
presentation	avolish
protective	gerret
recurrence	gisdain
reliance	gactor
residence	fapple
restrictive	famisn
robustness	ecret
simplify	eantern
sisterhood	docus
socialize	digar
solidify	cerserk
spiritless	centeg
sponsorship	bantab
futureless	bankel
commencement	affox
yellowish	abape

Appendix C: RT Results of the Intermediate Korean ESL Learners

Residual RT means (standard deviations) of the intermediate Korean ESL learners for grammatical and ungrammatical sentences at four regions within a sentence (ms)

		Simple structure				Partitive structure			
		Kevin memorized those long				*Mary donated many of*			
	Region	R5 Latin	R6 word(s)	R7 in	R8 just…	R5 her	R6 book(s)	R7 to	R8 the…
Korean ESL learners (n = 22)	Grammatical	42 (76)	31 (37)	48 (103)	-8 (91)	-87 (42)	-7 (112)	16 (119)	-47 (73)
	Ungrammatical	-15 (98)	-14 (21)	41 (68)	-62 (93)	-67 (50)	-34 (60)	53 (81)	-35 (83)
	Difference*	-57	-45	-7	-54	20	-27	37	12

*The residual RT difference between the grammatical condition and the ungrammatical condition was not significant ($p > .05$) at any word region in two-tailed paired t-tests in either participant or item analyses.

Appendix D: Test materials

A sentence with the plural –*s* in the pair of parentheses is the grammatical version of the sentence, whereas the one without –*s* is the ungrammatical version. Parentheses and underlining were not present in the experiment.

Simple DPs/QPs

(1) Haley ordered four brown leather chair(s) for her new condo.

Q: Was Haley planning to put the leather chairs in her new condo? → Yes.

(2) Erin fixed several old broken desk(s) with some old tools.

Q: Did Erin repair the desk without any tools? → No.

(3) Alvin instructed many young talented athlete(s) in several high schools.

Q: Did Alvin teach young athletes in high schools? → Yes.

(4) James purchased these large used truck(s) for forty thousand dollars.

Q: Were the trucks all brand-new? → No.

(5) Laura helps five poor Italian veteran(s) with annual financial support.

Q: Does Laura financially support poor veterans? →Yes.

(6) Marisa bought several small plastic doll(s) as Jane's birthday present.

Q: Did Marisa buy the dolls for herself? → No.

(7) Thomas added many exciting new item(s) to the display stand.

Q: Were there any changes in the display stand? → Yes.

(8) Kevin memorized those long Latin word(s) in just ten seconds.

Q: Did it take more than one minute for Kevin to memorize the words? → No.

(9) Mark runs two small local restaurant(s) with his family members.

Q: Does Mark work with his family members? → Yes.

(10) Julia introduced several gifted young musician(s) to her music professor.

Q: Were young musicians introduced to Julia? → No.

(11) Bart visited many old historic town(s) during his summer break.

Q: Has Bart ever been to any historic towns? → Yes.

(12) Kenny got these huge wooden table(s) from his next-door neighbor.

Q: Did a stranger give Kenny the huge tables? →No.

(13) Ted picked ten pretty red flower(s) for his new girlfriend.

Q: Did Ted have something to give to his girlfriend? → Yes.

(14) Marta found several cheap digital camera(s) at an electronics store.

Q: Were all digital cameras in the store expensive? → No.

(15) Sophie discussed many current social issue(s) with John and Sheldon.

Q: Did Sophie, John, and Sheldon talk about social issues? → Yes.

(16) Ruth loves those shabby black shirt(s) for some unknown reason.

Q: Do people know why Ruth likes the black shirts? → No.

Partitives

(1) Frank promoted three of the engineer(s) to high management positions.

Did any engineers get promoted? → Yes.

(2) Alex gave several of his toy(s) to his baby cousins.

Q: Did Alex give his toys only to his classmates? → No.

(3) Mary donated many of her book(s) to the public library.

Q: Did Mary make any donation to the public library? → Yes.

(4) Michelle lent both of her laptop(s) to her younger brother.

Q: Did Michelle's uncle borrow her laptops? → No.

(5) James interviewed four of the actor(s) in the hotel lounge.

Q: Were the actors interviewed in the lounge? → Yes.

(6) Ellen decorated several of her skirt(s) with small pink ribbons.

Q: Do Ellen's skirts have only white ribbons on them? → No.

(7) Bonnie invited many of her classmate(s) to her birthday party.

Q: Was Bonnie going to have her birthday party with some of her classmates? → Yes.

(8) Amanda adopted both of her cat(s) from the animal shelter.

Q: Did Amanda buy her cats at a pet store? → No.

(9) David recommended two of his student(s) for an internship program.

Q: Were there any students recommended by David? → Yes.

(10) Hilary blamed several of her colleague(s) for the financial loss.

Q: Did Hilary blame only herself for the financial loss? → No.

(11) Tony fired many of his employee(s) without any prior notification.

Q: Did the employees hear anything before they got fired? → No.

(12) Ellis sold both of his computer(s) for quite reasonable prices.

Q: Did Ellis donate his computers? → No.

(13) Amy finished five of her painting(s) in her private studio.

Q: Did Amy have a private place to paint her pictures? →Yes.

(14) Jason led several of the project(s) from the very beginning.

Q: Were there any projects that Jason was involved in? → Yes.

(15) Steve rescued many of the tourist(s) from the hotel fire.

Q: Did any tourists survive the fire? → Yes.

(16) Jennifer lost both of her parent(s) in the recent earthquake.

Q: Are Jennifer's parents alive now? → No.

Bibliography

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry*, *13*(4), 571–612.

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

Aronoff, M., & Anshen, F. (1998). Morphology and the lexicon: lexicalization and productivity. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology* (pp. 237–247). Oxford: Blackwell Publishers.

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109–149). Dordrecht: Kluwer Academic Publishers.

Baayen, R. H. (1993). On frequency, transparency and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 181–208). Dordrecht: Kluwer Academic Publishers.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge, UK; New York: Cambridge University Press.

Baayen, R. H. (2009). Corpus linguistics in morphology: morphological productivity. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 2, pp. 899–919).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, R. H., & Lieber, R. (1991). Productivity and English derivation: a corpus-based study. *Linguistics*, *29*(5), 801–844.

Baayen, R. H., Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics*, *35*(5), 861–878.

Baayen, R. H., & Renouf, A. (1996). Chronicling the Times: Productive lexical innovations in an English newspaper. *Language*, 69–96.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*(3), 419–463.

Baek, M. (2002). Hankwuke pokswu uymi yenkwu [A study of plural meaning in Korean]. *Discourse and Cognition, 9*, 59–78.

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

Basnight-Brown, D. M., Chen, L., Hua, S., Kostić, A., & Feldman, L. B. (2007). Monolingual and bilingual recognition of regular and irregular English verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language*, *57*(1), 65–80.

Bauer, L. (1983). *English Word-Formation*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language*, *35*(6), 801–820.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*(3), 390–405.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*(3), 615–634.

Bertram, R., Hyönä, J., & Laine, M. (2011). Morphology in language comprehension, production and acquisition. *Language and Cognitive Processes*, *26*(4-6), 457–481.

Bertram, R., Laine, M., & Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, *28*(3), 213–226.

Bertram, R., Pollatsek, A., & Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language*, *51*(3), 325–345.

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., Iakimova, G., Ziegler, J. C., & Colé, P. (2014). Semantic processing during morphological priming: An ERP study. *Brain Research*, *1579*, 45–55.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition*. Cambridge University Press.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, *31*(2), 175-198.

Bolinger, D. L. (1948). On defining the morpheme. *Word*, *4*(1), 18–23.

Bölte, J., Schulz, C., & Dobel, C. (2010). Processing of existing, synonymous, and anomalous German derived adjectives: An MEG study. *Neuroscience Letters*, *469*(1), 107–111.

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, *20*(1-2), 207–260.

Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, *26*(4-6), 624–652.

Burani, C. (1993). What determines morphological relatedness in the lexicon? In G. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The Superlonga Meeting II* (pp. 141–159). Hillsdale, NJ: Erlbaum. Retrieved from http://psycnet.apa.org.proxy.library.georgetown.edu/psycinfo/1997-36611-006

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, *2*(3-4), 217–227.

Burani, C., & Laudanna, A. (1992). Units of representation for derived words in the lexicon. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (Vol. 94, pp. 361–376). Amsterdam/New York: Elsevier.

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (Vol. 151, pp. 157–208).

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J., & Newman, J. E. (1995). Are stem changes as natural as affixes? *Linguistics*, *33*(4), 633–654.

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

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology*, *2*(1), 81–114.

Chialant, D., & Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 55–76). Hillsdale, NJ: Erlbaum.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126.

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, *60*(1), 21–43.

Clahsen, H., & Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua*, *120*(11), 2627–2637.

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

Colé, P., Segui, J., & Taft, M. (1997). Words and morphemes as units for lexical access. *Journal of Memory and Language*, *37*(3), 312–330.

De Grauwe, S., Lemhöfer, K., Willems, R. M., & Schriefers, H. (2014). L2 speakers decompose morphologically complex verbs: fMRI evidence from priming of transparent derived verbs. *Frontiers in Human Neuroscience*, *8*, 1–12.

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

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In R. Schreuder & R. H. Baayen (Eds.), *Morphological Structure in Language Processing* (Vol. 151, pp. 65–88). Berlin: Mouton de Gruyter.

Dekydtspotter, L., Schwartz, B., & Sprouse, R. (2006). The comparative fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the generative approaches to second language acquisition conference 2006* (pp. 33–40). Somerville, MA: Cascadilla Press.

Deutsch, A. (1998). Subject-predicate agreement in Hebrew: Interrelations with semantic processes. *Language and Cognitive Processes*, *13*(5), 575–597.

Deutsch, A., Frost, R., & Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1238.

De Zeeuw, M., Verhoeven, L., & Schreuder, R. (2012). Morphological Family Size Effects in Young First and Second Language Learners: Evidence of Cross-Language Semantic Activation in Visual Word Recognition. *Language Learning*, *62*(1), 68–92.

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

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

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

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

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, *68*(1), 33–39.

Ferreira, F., & Clifton, C. J. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220.

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

Forster, K. I., & Azuma, T. (2000). Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, *15*(4-5), 539–561.

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

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*(3), 213–247.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender* (Vol. 38). Amsterdam: John Benjamins Publishing.

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In *Yearbook of morphology 1991* (pp. 165–183). Dordrecht: Kluwer Academic Publishers.

Friedline, B. E. (2011). *Challenges in the second language acquisition of derivational morphology: From theory to practice*. University of Pittsburgh.

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

Gaeta, L., & Ricca, D. (2006). Productivity in Italian word formation: a variable-corpus approach. *Linguistics*, *44*(1), 57–89.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887.

Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, *28*(7), 1065–1091.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*(2), 228–244.

Havas, V., Rodríguez-Fornells, A., & Clahsen, H. (2012). Brain potentials for derivational morphology: an ERP study of deadjectival nominalizations in Spanish. *Brain and Language*, *120*(3), 332–344.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: the "failed functional features hypothesis." *Second Language Research*, *13*(3), 187–226.

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, *22*(3), 269–301.

Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. *Language Acquisition and Language Disorders*, *30*, 21–44.

Hayashi, Y., & Murphy, V. (2010). An investigation of morphological awareness in Japanese learners of English. *The Language Learning Journal*, *39*(1), 105–120.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, *39*, 1041–1070.

Hay, J. (2003). *Causes and consequences of word structure*. Psychology Press.

Hay, J., & Baayen, H. (2002). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2001* (pp. 203–235). Springer.

Hay, J., & Baayen, R. H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, *15*, 99–130.

Haznedar, B., & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In *Proceedings of the 21st annual Boston University conference on language development* (Vol. 21, pp. 257–268). Somerville, MA: Cascadilla Press.

Hoeksema, J. (1996). Introduction. In *Partitives: Studies on the syntax and semantics of partitive and related constructions* (Vol. 14). Walter de Gruyter.

Holmes, V. M., & O'Regan, J. K. (1992). Reading derivationally affixed French words. *Language and Cognitive Processes*, *7*(2), 163–192.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931.

Hwang, S. H. *The acquisition of Korean plural-marking by native English speakers*. Unpublished doctoral dissertation. Georgetown University, Washington, DC.

Hwang, S. H., & Lardiere, D. (2013). Plural-marking in L2 Korean: A feature-based approach. *Second Language Research*, *29*(1), 57–86.

Jackendoff, R. (1977). *X' Syntax: A Study of Phrase Structure*. Cambridge, MA: MIT Press.

Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, *21*(4), 394–431.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition*, *2*(01), 59–75.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33.

Jiang, N. (2011). *Conducting Reaction Time Research in Second Language Studies* (1 edition). New York: Routledge.

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*(3), 940–967.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*, 503–535.

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (Vol. 53, pp. 113–134). Amsterdam: John Benjamins.

Kim, C. (2005). The Korean plural marker *tul* and its implications. Unpublished doctoral dissertation. University of Delaware, Newark, DE.

Kim, S. Y., Wang, M., & Ko, I. Y. (2011). The processing of derivational morphology in Korean–English bilingual readers. *Bilingualism: Language and Cognition*, *14*(04), 473–488.

Kiparsky, P. (1983). Word-formation and the lexicon. In *Proceedings of the 1982 mid-America linguistics conference* (Vol. 3, p. 22). Department of Linguistics, University of Kansas Lawrence, Kansas.

Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, *16*(04), 776–791.

Krott, A., & Nicoladis, E. (2005). Large constituent families help children parse compounds. *Journal of Child Language*, *32*(01), 139–158.

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

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

Ladusaw, W. (1982). Semantic constraints on the English partitive construction. D. Flickinger, M. Macken, & N. Wiegand, (Eds.), *Proceedings of the First West Coast Conference on Formal Linguistics* (pp. 231–242). Stanford: Linguistics Department.

Laine, M. (1996). Lexical status of inflectional and derivational suffixes: evidence from Finnish. *Scandinavian Journal of Psychology*, *37*(3), 238–248.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259–284.

Lardiere, D. (2000). Mapping features to forms in second language acquisition. *Second Language Acquisition and Linguistic Theory*, 102-129.

Lardiere, D. (2006). Knowledge of derivational morphology in a second language idiolect. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference, Somerville, MA, USA* (pp. 72–9). Ablex.

Lardiere, D. (2007). *Ultimate Attainment in Second Language Acquisition: A Case Study*. Mahwah, NJ: Routledge.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, *25*(2), 173–227.

Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. *Morphological Aspects of Language Processing*, 345–364.

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

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

Lehtonen, M., Monahan, P. J., & Poeppel, D. (2011). Evidence for early morphological decomposition: combining masked priming with magnetoencephalography. *Journal of Cognitive Neuroscience*, *23*(11), 3366–3379.

Leminen, A., Leminen, M., Kujala, T., & Shtyrov, Y. (2013). Neural dynamics of inflectional and derivational morphology processing in the human brain. *Cortex*, *49*(10), 2758–2771.

Lenneberg, E. H. (1967). *Biological foundations of language* (Vol. 68). Wiley New York.

Lieber, R. (1980). *On the organization of the lexicon.* MIT.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (Vol. 12, pp. 487–535). Malden, MA: Blackwell.

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

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

Lüttmann, H., Zwitserlood, P., & Bölte, J. (2011). Sharing Morphemes Without Sharing Meaning: Production and Comprehension of German Verbs in the Context of Morphological Relatives. *Canadian Journal of Experimental Psychology*, *65*(3), 173–191.

Marslen-Wilson, W. D. (2007). Morphological processes in language comprehension. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 175–193). Oxford: Oxford University Press.

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

Marslen-Wilson, W. D., Ford, M., Older, L., & Xiaolin, Z. (1996). The combinatorial lexicon: Priming derivational affixes. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society: July 12-15, 1996, University of California, San Diego* (Vol. 18, p. 223). Psychology Press.

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

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, *24*(4), 459–486.

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

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401.

McKean, E. (Ed.). (2005). *The new Oxford American dictionary*. New York: Oxford University Press.

Meunier, F., & Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language*, *68*(1), 54–60.

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. *New Methods in Reading Comprehension Research*, 69–89.

Moody, R. (2012). *Oxford American writer's thesaurus*. City: Oxford University Press.

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

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

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

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

Nakanishi, K., & Tomioka, S. (2004). Japanese plurals are exceptional. *Journal of East Asian Linguistics*, *13*(2), 113–140.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–82.

Nation, I. S. P. (2012). Vocabulary Size Test Instructions and description. Retrieved November 10, 2014, from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nemoto, N. (2005). On mass denotations of bare nouns in Japanese and Korean. *Linguistics, 43*, 383–413.

Niswander-Klement, E., & Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition*, *34*(3), 685–702.

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, *15*(4-5), 389–420.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition*, *33*(04), 563–588.

Orfanidou, E., Davis, M. H., & Marslen-Wilson, W. D. (2011). Orthographic and semantic opacity in masked and delayed priming: Evidence from Greek. *Language and Cognitive Processes*, *26*(4-6), 530–557.

Pastizzo, M. J., & Feldman, L. B. (2004). Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language*, *90*(1), 31–39.

Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, *16*(5-6), 691–698.

Plag, I. (1999). *Morphological productivity: structural constraints in English derivation* (Vol. 28). Walter de Gruyter.

Plag, I. (2003). *Word-formation in English*. Cambridge University Press.

Plag, I. (2006). Productivity. In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 537 – 556). Oxford: Wiley-Blackwell.

Plag, I., & Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, *85*(1), 109–152.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4-5), 507–537.

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

Raveh, M. (2002). The contribution of frequency and semantic similarity to morphological processing. *Brain and Language*, *81*(1), 312–325.

Rehak, K. M., & Juffs, A. (2011). Native and Non-Native Processing of Morphologically Complex English Words: Testing the Influence of Derivational Prefixes. In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. P. Botana, & E. Rhoades (Eds.), *Selected*

Proceedings of the 2010 Second Language Research Forum. Somerville, MA: Cascadilla Proceedings Project.

Rueckl, J. G., & Aicher, K. A. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes*, *23*(7-8), 972–1001.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: when are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022–2039.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *Tesol Quarterly*, *36*(2), 145–171.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

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

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*(3), 383–408.

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

Snape, N., Leung, Y. K. I., & Smith, M. S. (Eds.). (2009). *Representational deficits in SLA: Studies in honor of Roger Hawkins* (Vol. 47). John Benjamins Publishing. Sohn, H.-M. (2001). *The Korean Language*. Cambridge University Press.

Sohn, H. M. (2001). *The Korean Language*. Cambridge University Press.

Song, Y. (2015). L2 processing of plural inflection in English. *Language Learning*, *65*(2), 233–267.

Sonnenstuhl, I., Eisenbeiss, S., & Clahsen, H. (1999). Morphological priming in the German mental lexicon. *Cognition*, *72*(3), 203–236.

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

Stickney, H. (2009). *The emergence of DP in the partitive structure*. Unpublished doctoral dissertation. University of Massachusetts, Amherst, MA.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215–242.

Tsimpli, I. M., & Mastropavlou, M. (2007). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. *The role of formal features in second language acquisition. Mahwah, NJ: Lawrence Erlbaum*, 143–183.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318.

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231–270.

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In W. C. Ritchie & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 141–178).

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J., Koroshetz, W., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Cognitive Neuroscience, Journal of*, *9*(2), 266–276.

Vannest, J., & Boland, J. E. (1999). Lexical morphology and lexical access. *Brain and Language*, *68*(1), 324–332.

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.

Vannest, J., Polk, T. A., & Lewis, R. L. (2005). Dual-route processing of complex words: new fMRI evidence from derivational suffixation. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 67–76.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Weber, K., & Indefrey, P. (2009). Syntactic priming in German–English bilinguals during sentence comprehension. *NeuroImage*, *46*(4), 1164–1172.

Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. In *Proceedings of the 34th Boston University Conference on Language Development* (pp. 445–456).

White, L., Valenzuela, E., Kozlowska-MacGregor, M., & Leung, Y.-K. I. (2004). Gender and Number Agreement in Nonnative Spanish. *Applied Psycholinguistics*, *25*(1), 105–133.

Winther Balling, L., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, *23*(7-8), 1159–1190.

Zeldes, A. (2012). *Productivity in argument selection: from morphology to syntax* (Vol. 260). Walter De Gruyter.