University of Alberta

An Experimental Investigation of Complexity-Based Ordering

by

Laura Gene Helena Teddiman

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

Spring 2012
Edmonton, Alberta

Dedication

To my parents.

Abstract

This dissertation examines the phenomenon of suffix ordering in English from a psycholinguistic perspective. Key to this work is an examination of *Complexity-Based Ordering*, a theory of affix ordering that combines both selectional restrictions and processing constraints. Complexity-Based Ordering provides a hierarchical rank for each suffix in English by combining suffix-specific combinatorial restrictions with general principles of processing complexity (e.g., ease of parsing, relative root and derived frequencies). Suffixes higher on the hierarchy are expected to be easier to parse out from the word and should be attached outside of suffixes of lower rank. The first paper examines lexical decision and naming latencies to base+suffix+suffix words (e.g., *hope*+*ful*+*ly*), finding roles for root, base, and word frequencies as significant predictors of response latencies. Effects of Rank, however, are absent. The second paper presents a lexical decision experiment with an additional eye-tracking component, revealing a time-course for lexical access, with the root+suffix1 (*hope*+*ful* in *hopefully*) frequency appearing as a significant predictor of fixation durations before whole word frequency. The final paper presents an eye-tracking study of words in sentence context with an additional Event-Related Potential (ERP) component. In this experiment, the Rank of the second suffix becomes a useful predictor. When the second suffix in a base+suffix+suffix word is of low Rank, higher processing costs are reflected in longer response times. In all three experiments, a role for a new frequency measure, the suffix pair frequency, is revealed. The effects of Rank, as determined by the Complexity-Based Ordering hierarchy, are absent during single-word recognition tasks (lexical decision, naming), but are prevalent during sentence reading, highlighting the role of sentential context and predictability during language processing.

Acknowledgement

This research was supported in part by a Doctoral Fellowship (#752-2007-1358) awarded to the author by the Social Sciences and Humanities Research Council of Canada.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the following dissertation, I present three papers designed to explore the nature of the processing of suffix pairs in English. The ultimate goal of this dissertation was to test the predictivity of Complexity-Based Ordering, a recently proposed theory of affix ordering for reader comprehension, in English. Each chapter is formatted as an independent article. Although they may be read independently, I view these papers as a continuing journey through experimental methodologies targeting the same type of stimuli, that is, English words containing two suffixes. In the first paper, I use data from the English Lexicon Project (Balota et al., 2007) to examine the processing of root+suffix+suffix words in both lexical decision and naming. The second experiment proceeds along the path of visual word recognition, adding an eye-tracking component to a lexical decision experiment with an expanded set of doubly-derived suffixed words. The third experiment adds a sentence context to our study of visual word recognition with eye-tracking, with an additional event-related potential (ERP) measure for analysis.

1.1 Theoretical Background

The issue of suffix ordering in English has been a lively area of discussion for morphologists. In English, there are many more potential suffix combinations than are attested in everyday speech. For example, Hay and Plag (2004) took 15 English suffixes and examined the attested suffixes pairs that could result from their combination, and found that of the 210 possible combinations, only 36 were used. The reasons underlying the relatively sparse number of combinations are not clear, but several theories have been put forward to explain this patterning. These theories fall into three basic categories: stratum-oriented models (e.g., Kiparsky, 1982), selectional restriction-based models (e.g., Fabb, 1988), and psycholinguistically motivated models based on processing restrictions (e.g., Hay 2000, 2002). *Complexity-Based Ordering* combines processing restrictions and selectional restrictions to formulate a hierarchy to explain which suffixes are more and less likely to occur outside of others (Hay & Plag, 2004).

Stratum-oriented models, like other affix ordering models, begin with the observation that certain affix pairs do not occur (e.g., Siegel, 1974). In stratum-oriented models, affixes are grouped into Classes, or Levels, based on their phonological characteristics and/or their etymology (e.g., Kiparsky, 1982; Giegerich, 1999). For instance, Kiparsky (1982) assigned affixes to different classes according to their phonological behaviour and their distance from the root. Class I suffixes, or those in the first stratum, are often Latinate borrowings into English. An example division of Class 1 and Class 2 suffixes comes from Spencer (1991: 79):

> Class I: +ion, +ity, +y, +al, +ic, +ate, +ous, +ive, +able, +ize
>
> Class II: #ness, #less, #hood, #ful, #ly, #y, #like,#ist, #able, #ize

The addition of a Class I suffix is often accompanied by a phonological change in the base word (e.g., by altering the stress pattern of the base), resulting in a less transparent derivation from base to product in phonology, with opacity in meaning also more likely when compared to derivations using Class 2 suffixes. Native Germanic suffixes make up the bulk of the second stratum. They typically do not alter the phonology of the base, and critically are not expected to occur inside of Class 1 suffixes. Proponents of Lexical Phonology (also called Lexical Morphology by Katamba, 1994) hold that all phonological processes associated with the first suffix of a suffix pair should be completed before the second suffix is added to its base, resulting in a cyclical application of phonological rules (Katamba, 1994).

Proposed strata in stratum-oriented models can be useful in that they often reflect different phonological properties, with most Class 1 suffixes beginning with vowels. This in turn affects parsing and processing of the suffixed word (Hay, 2002). However, the basis upon which stratum classification is made is not consistent between authors (e.g., Aronoff & Fuhrhop, 2002) or, indeed, consistent by affix. For instance, in order to explain its behaviour across English words as a whole, the suffix -*able* must be both a Class 1 and Class 2 suffix simultaneously (c.f. Giegerich, 1999: 21-52).

Perhaps the biggest problem with stratum-oriented models is that they cannot be used to predict which suffixes will co-occur beyond the most basic tenet that Class 1 suffixes should not occur following Class 2 suffixes. They make no claims as to how suffixes should behave within a Class level. Worse, they cannot be reliably used, even at their most basic, to predict all suffixation behaviour. For example, the Class 1 suffix *-ic* should not occur following the Class 2 suffix *-ist*, as that requires a Class 2 suffix to precede a Class 1 suffix, but the suffix pair *-ist-ic* is common.

Experimental research into the processing costs associated with different strata has not yielded clear evidence for differentiating between classes. While an initial lexical decision experiment by Vannest and Boland (1999) suggested that there might be processing differences between Class 1 and Class 2 suffixes, their results were based only three suffixes (Class 1: *-ity*, *-ation*; Class 2: *-less*). In this experiment, they found root frequency effects for words suffixed with *-less*, where shorter reaction times were recorded for words with high root frequency. They did not find root frequency effects for words suffixed with *-ity* or *-ation*, instead finding whole-word frequency effects for the Class I suffixes that were not observed for *-less*. From these results, they hypothesized that words affixed with the Class 1 suffixes *-ity* and *-ation* were stored as whole units, so that the root was not accessed independently during processing, whereas words containing Class 2 *-less* were decomposed during lexical decision into the root and its suffix. This finding did not survive further testing with an expanded suffix list in a second experiment. Vannest and Boland's (1999) results did support frequency as an important element in processing morphologically complex words, as observed elsewhere (e.g., Beauvillain, 1996; Bertram & Hyönä, 2003). However, while they manipulated the frequencies of the root and whole word in their experiments, they did not take into account the relative frequencies of the root versus the whole word containing the root (e.g., the frequency of *hope* versus that of *hopeful*). A later study presented by Hay (2001) showed that a derived word occurring more frequently than its root is likely to be analyzed as a single unit. Likewise, a derived word that occurs less frequently than its root is likely to be decomposed. In the former case, one would expect whole word frequency effects during initial processing, and in the latter we expect root frequency effects. When a suffix occurs in derived words that are analyzed as whole-word units, it is less likely to be a relevant processing unit because it is not recognized as one. That is, if a complex word is being accessed as a single unit, its components do not need to be parsed to recognize it. A suffix in such a word will not be as salient as one that is frequently observed in complex words that are recognized via their constituent morphemes. The likelihood that a complex word will be decomposed can be estimated using a parsing ratio that compares the number of words with a particular suffix that are parsed to the number of words containing that same suffix that are not parsed (Hay & Baayen, 2002).

The second main approach to explaining affix ordering looks to the combinatorial properties of specific morphemes. In this approach, *selectional restrictions* constrain the ways in which bases and affixes combine. Selectional restrictions are specific to each morpheme, and can be based on the phonological, morphological, semantic, or syntactic properties of either the affix (e.g., Fabb, 1988) or the base (e.g., Plag, 2004). Since affix-specific information is required to describe affixational behaviour regardless of whether or not one takes a stratum-oriented approach, proponents of a selectional restrictions account hold that stratum-oriented restrictions are redundant and therefore unnecessary in accounts of affix ordering (Fabb, 1988; Plag 1996, 2002).

Under a selectional restriction account of affix ordering, there is no need for lexical strata, as affix membership in a given Class cannot reliably predict all potential affix pairs, and nor can it exclude combinations that do not exist (Fabb, 1988). Fabb (1988) identifies four suffix classes: suffixes that only combine with unsuffixed words, suffixes that only combine with one other suffix, suffixes that freely combine with others, and problematic suffixes (suffixes that combine with a few other suffixes, but not freely). These general patterns are observed beyond the particulars of a single suffix. Plag (1996) adds semantic constraints (e.g., blocking), contending that *in addition* to specific selectional restrictions, there are more general constraints on affixation that are not specific to a given affix. He also presents an argument for base-driven selectional restrictions, where the base essentially chooses the suffix, rather than the properties of the suffix limiting its combination with a given base. For example, verbs ending with the suffix -*ize* will quite frequently take the suffix -*ation* when they are derived into nouns (e.g., *colourize* and *colourization*). Plag (1996) argues that this pattern is not based on restrictions defining what -*ation* can combine with, but rather that words ending with -*ize* preferentially select -*ation*.

The last theoretical approach to affix ordering is based on the ease of language processing during uptake of the linguistic signal, holding that those affixes that are most easily parsed should occur outside of affixes that cannot be easily parsed (Hay, 2002). This is based on the idea that suffixes are more likely to combine with bases that are perceived to have a simpler internal morphological structure. Structurally complex words (i.e., those with two or more morphemes) are perceived as less complex when their morphemes are less parsable (and therefore more likely to be perceived as a single unit). Processing costs are reliant both on the relative frequencies of the morphemes involved and on their the phonotactic properties (Hay, 2000, 2002). When a complex word is more frequent than its root, it is more likely to be processed as a whole unit (Hay, 2001). Similarly, the phonological shape of a suffix can make it more or less likely to be recognized as a morphological unit. Suffixes beginning with consonants are more easily identifiable as units in the speech stream. Take, for example, the word *hopeful*, which is composed of two morphemes, *hope* and *-ful*. The morpheme boundary occurs across the phonetic sequence [pf], which is easily identified as such because [pf] does not frequently occur as a morpheme-internal sequence in English. When a suffix begins with a vowel, identifying the morpheme boundary can become much more difficult, and therefore such suffixes are less easily parsed. Under this hypothesis, Hay (2002) argues that affix order is a consequence of speaker sensitivity to phonotactic probabilities and parsability across morpheme boundaries, morpho-phonological changes to the base triggered by affixation, and the frequencies of the root, suffix, and whole-word. This hypothesis captures one of the most important observations from stratum-oriented models without needing to overlay a level structure. In stratum-oriented models, many Class 1 suffixes begin with vowels while many Class 2 suffixes begin with consonants, although this is not exclusive. Class 2 suffixes generally occur after Class 1 suffixes. Unlike stratum-oriented models, processing-based models can also bring into account differences caused by the particular forms of a given affix.

Processing-based accounts are not without criticisms. For instance, Plag (2002) contends that they cannot account for combinations such as *-al-ize-ation* (e.g., in *formalization*), *-able-ity*, and *-al-ist*, or those suffixes that benefit from base-driven restrictions. By means of example, the productive suffix combination *-able-ity* presents an unexpected suffix sequence because *-ity* changes the stress of the base word and *-able* does not, meaning that *-able* is phonologically more easily parsed than *-ity*. In a strict processing account, we would expect *-able* to occur after *-ity* because it is more readily parsed. A second problem identified by Plag (2002) is that while we expect those suffixes beginning with consonants to be the most productive because they are the most easily parsed and recognized (and therefore, available for use), this is not the case.

The theory under investigation in this dissertation, Complexity-Based

Ordering, is essentially a combination of a processing constraints approach and suffix-specific selectional restrictions (Hay & Plag, 2004). In Complexity-Based Ordering, suffixes can be ranked according to how easily they can be parsed out during uptake (e.g., as in Hay 2002), with added constraints obtained from observed restrictions on combinations. This generates a rank hierarchy, where lower ranked suffixes are expected to be more difficult to parse than suffixes higher in the hierarchy. Suffixes lower on the hierarchy are not expected to occur outside of suffixes higher on the hierarchy. While this ranking captures the same behaviour targeted by stratum-oriented models, it is also gradable because it is based on probabilistic values. Additionally, the probabilistic nature of the scale can both make strong predictions about how suffixes should be ordered and can allow for small deviations from the scale when a suffix's selectional restrictions strongly favor word formation outside the hierarchy (Plag & Baayen, 2009: 124).

Complexity-Based Ordering has received some initial experimental support. Plag and Baayen (2009) conducted a study using lexical decision and naming latencies from the English Lexicon Project (Balota et al., 2007) to determine whether suffix rank (CO-Rank) was a useful predictor of reaction times for bimorphemic suffixed words. While CO-Rank was not predictive for individual words, it was a significant predictor for median reaction times at the level of the suffix. At the extremes of the hierarchy, suffixes showed shorter response latencies than at the middle of the hierarchy, which provides evidence for both lexical storage (for suffixes low in the hierarchy) and parsing (for suffixes high in the hierarchy). Suffixes in the middle of the hierarchy are biased towards neither storage nor parsing, and so are not as efficiently processed. In general, storage was faster than computation. A major strength of Plag and Baayen's (2009) study is the inclusion of other lexical variables, such as word frequency, root frequency, and word length, using linear mixed effects models in a regression analysis to examine the contributions of predictors to processing. In the current dissertation, we take a similar approach, using regression analyses rather than factorial designs to capture gradient effects of continuous predictors.

1.2 Dissertation Synopsis

The first paper that I will present examines the processing of trimorphemic words extracted from the English Lexicon Project (Balota et al., 2007). In these words, composed of a root and two suffixes, actual suffix order is revealed. This is in contrast to Plag and Baayen (2009), which focused on bimorphemic words containing a single suffix. The paper here presents an analysis of response latencies from both a naming task and a standard lexical decision task. In addition to Rank Ordering, as obtained from Plag and Baayen (2009), word frequency, morpheme frequency (root, suffix1, suffix2), and frequencies of intermediate morpheme combinations (root+suffix1, suffix1+suffix2) are also considered as predictors of response latencies. This chapter further presents a comparison of linear models for both experiments to mixed linear effects models in which *suffix* is included as a random-effect factor.

In the second paper, eye-tracking is added to a standard lexical decision task in order to investigate the processing of our target words. Eye-tracking provides a more detailed time-course for processing, as we are not solely relying on reading + decision time as in standard lexical decision. Instead, time spent on individual fixations can be recorded and then these fixation durations can be analyzed in order to determine which predictors influence the length of time spent on information uptake. We also obtain information about total reading time and the number of fixations on a word, both of which are influenced by lexical predictors. In this chapter, we analyze the results using linear mixed-effects models.

In the third paper, our target words are presented in sentences, and eye-movements are tracked as participants progress through the them. In this case, we focus on the target trimorphemic words. Participants were asked to read sentences that were presented, in full, on a computer screen, one sentence at a time. While experimental conditions will never be fully natural, sentence reading more closely approximates the way in which we normally encounter words than lexical decision tasks do. To this eye-tracking experiment, we added an Event-Related Potential (ERP) component, where the electrophysiological response of the brain, measured at the scalp, is recorded during sentence reading, and where the presentation of target items is linked to eye-fixations. The ERP recording provides further temporal granularity, which can be added to that offered by eye-movements, allowing us a more detailed window into lexical processing. Results in this experiment are modelled using generalized additive models (GAMs).

Results are summarized and their implications discussed in the Conclusion.

1.3 References

[1] Aronoff, M. & Fuhrhop, N. (2002). Restriction suffix combinations in German and English: Closing suffixes and the Monosuffix Constraint. *Natural Language and Linguistic Theory, 20*, 451-490.

[2] Baayen, R. H, Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

[3] Balling, L. & Baayen, H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, 23(7-8), 1159-190.

[4] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

[5] 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*, 801-820.

[6] Bertram, R. & Hyönä, J.(2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48*, 615-643.

[7] Fabb, N. (1988). English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory, 6*, 527-539.

[8] Giegerich, H. (1999). *Lexical strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

[9] Hay, J. (2002). From speech perception to morphology: Affix ordering revisited. *Language, 78*(3), 527555.

[10] Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics, 39*(6), 1041-1070.

[11] Hay, J. (2000). *Causes and Consequences of Word Structure.* PhD Dissertation, Northwestern University.

[12] Hay, J. & Baayen, H. (2002). Parsing and productivity. In G.E. Booj and J. van Marle (Eds.), *Yearbook of morphology 2001* (pp. 203-255). Dordrecht: Kluwert.

[13] Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory, 22*, 565-596.

[14] Katamba, F. (1994). *English words.* London/New York: Routledge.

[15] Kiparsky, P. (1982). Lexical phonology and morphology. In I. Yang (Ed.), *Linguistics in the morning calm*(pp. 3-91). Seoul: Hanshin.

[16] 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.

[17] Plag, I. (2004). Syntactic category information and the semantics of derivational morphological rules. *Folia Linguistica, 38*(3-4), 193-225.

[18] Plag, I. (2002). The role of selectional restrictions, phonotactics and parsing in constraining suffix ordering in English. In Geert Booij and Jaap van Marle. (Eds.), *Yearbook of Morphology 2001*(pp. 285-314). Dordrecht: Kluwer.

[19] Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Berlin/New York: Mouton de Gruyter.

[20] Plag, I. (1996). Selectional restrictions in English suffixation revisited: a reply to Fabb (1988). *Linguistics, 34*(4), 769-798.

[21] Plag, I., & Baayen, H. (2009). Suffix ordering and morphological processing. *Language, 85*(1), 106-149.

[22] Vannest, J. & Boland, J.E. (1999). Lexical morphology and lexical access. *Brain and Language, 68*, 324-332.

Chapter 2

Suffix Pair Processing in the English Lexicon Project

2.1 Introduction

Suffix ordering is a phenomenon in English that has been studied from a theoretical viewpoint in great detail, but that has received little attention from a psycholinguistic perspective. Suffix ordering as a phenomenon in English refers to the observation that the total number of observed suffix combinations is far fewer in number than the total possible suffix combinations. For example, *-ful+ness* is an attested combination, as seen in words like *colourfulness*, but we do not see combinations like *-ful+ity*, although this combination could arguably perform the same function. In this study, we examine the processing speeds of morphologically complex words having a root+suffix+suffix structure in lexical decision and naming experiments.

From a theoretical perspective, suffix ordering restrictions have most commonly been explained as the result of either level ordering (e.g., Kiparsky, 1982) or suffix-specific selectional restrictions (Fabb, 1988; Plag, 1996), which constrain how specific suffixes combine with bases. Hay (2002) introduces psycholinguistically motivated processing constraints. Hay and Plag (2004) combine selectional restrictions and processing constraints in a theory called *Complexity-Based Ordering*. According to this theory, suffix order is constrained by processing complexity, such that suffixes that are more easily parsed out of a word should occur outside of affixes that are more difficult to parse, in addition to suffix-specific constraints. Parsability is taken to be influenced by the phonotactic probabilities of bigrams across the morpheme boundary and frequencies of the roots and whole words involved. Hay (2001) presents evidence that the relative frequencies of a root and a word derived from that root influence how words are processed. When the derived word is more frequent than the root, it is more likely to be accessed as a whole stored form, whereas if the root is more frequent than the derived word, the word is more likely to be decomposed during comprehension. For this reason, root and derived word frequencies can influence the overall parsability of a base+suffix combination.

Plag and Baayen (2009) present one of the first psycholinguistic investigations of Complexity-Based Ordering, using naming latencies and lexical decision reaction times extracted from the English Lexicon Project (Balota et al, 2007). In this investigation, they focused on bimorphemic words consisting of a single root and a single suffix, and in doing so examined the premise upon which Complexity-Based Ordering is built. Across lexical decision and naming experiments, CO-Rank (the relative ranking of suffixes) was predictive for reaction times at the level of the suffix, but not for individual items. That is, values in the CO-Rank hierarchy were predictive for response latencies when latencies were averaged for each suffix. Their investigation yielded other significant predictors, including word length, number of syllables, lexical neighbours, and root and whole word frequencies, among others.

For these bimorphemic words, suffixes that were in the middle of the CO-Rank hierarchy were responded to more slowly than those suffixes at either end of the hierarchy. This result was interpreted as reflecting a higher processing cost for the mid-rank suffixes. Suffixes that have a very low rank in the CO hierarchy are those less easily parsed, and so are more likely to be stored in memory, allowing fast retrieval and resulting in faster reaction times. For those suffixes falling in the mid to upper range of the hierarchy, parsing becomes prevalent and more effort goes into processing, resulting in higher processing costs and longer response latencies. For suffixes at the very top of the CO hierarchy, there is a slight reduction in reaction latencies, reflecting efficient parsing. For the most part, however, parsing is costly and as a suffix becomes a more separable unit, there is an increase in processing costs.

In the current paper, we build on Plag and Baayen's (2009) investigation of Complexity-Based Ordering by addressing the processing of words containing a single root and two suffixes, to more directly assess the predictive power of CO-Rank and other variables for suffix ordering in the processing of English words.

2.2 Methods

As in Plag and Baayen (2009), the source of our data is the English Lexicon Project (Balota et al, 2007). This is a database of experimental results providing naming latencies, lexical decision latencies, and error rates for 40,481 words and 40,481 nonwords, presented in isolation to native speakers of English. Naming latencies have been collected from 444 participants and consist of 1,125,880 measurements, while lexical decision data has been collected from 816 participants for 2,752,698 reaction time measurements. From this database, 568 root+suffix+suffix words and their item means in both experiment types were extracted based on a complete morphological breakdown in the CELEX Lexical Database (Baayen, Piepenbrock & Gulikers, 1995). Using CELEX as a source, we collected the following set of frequency measures to be used as predictors in models of naming and lexical decision latencies: token frequencies for whole words, root+suffix sequence frequencies, suffix pair frequencies, and type and token frequencies for suffixes. For ease of discussion, we will call the intermediate combination of the root and suffix1 the "base." We likewise included the CO-Rank order from the proposed Complexity-Based Ordering hierarchy, in addition to word length, number of syllables, and number of phonemes. Reaction times and frequencies were logged, and high correlations between frequency measures (root, base, whole word frequency) were decorrelated by residualizing the base and whole word frequencies on root frequency. Table 2.1 summarizes the stimulus properties that were significant in our analysis. Within the 568 words extracted from the English Lexicon Project, there were 385 roots, 72 suffixes in the first suffix position, and 48 in the second suffix position (163 suffix pairs). Section A contains structural information such as word length. Section B contains lexical frequency predictors such as root and base frequency. Section C contains suffix-related variables.

The predictors in Section C are Suffix1 token frequency, Suffix2 token frequency, Suffix1 type frequency, Suffix2 type frequency, and Suffix CO-Rank for both suffixes 1 and 2. The type frequency of a suffix is the number of words that a suffix occurs in, but not the summed frequency of those words. Finally, the CO-Rank for a suffix is the rank of a suffix in the Complexity-Based Ordering hierarchy. Our rank order data is derived from Plag and Baayen (2009). Suffix rank has been calculated using the y-axis values from a graphical representation, where the most highly ranked suffixes are located at the bottom of the graph. As a result, in the following paper, suffixes that are low in the Complexity-Based Ordering Hierarchy have higher numerical

Table 2.1: Lexical predictors used in the analyses of lexical decision and naming latencies. Section A contains predictors related to the form of the word. Section B contains lexical frequency predictors. Section C contains suffix-specific predictors.

	Variable	Mean	Min	Max
A	Word Length	10.77	6	18
	Number of Syllables	3.77	2	6
B	Root Frequency	711.65	0	483429
	Base Frequency	221.64	0	7694
	Word Frequency	35.27	0	4701
C	Suffix1 Token Frequency	10550.84	41	119275
	Suffix2 Token Frequency	18474.43	4	155742
	Suffix1 Type Frequency	85.08	2	817
	Suffix2 Type Frequency	228.74	2	1191
	Affix Pair Frequency	419.86	1	11284
	Suffix1 CO-Rank	4140.72	675	7434
	Suffix2 CO-Rank	2542.73	193	6952

values in our analyses, so that an increase in the model corresponds to a decrease in CO-Rank.

2.3 Results

2.3.1 Naming Latencies

The data has been analyzed using both linear modelling and mixed effects modelling, and the results of both are included for discussion. Analyses were carried out using the statistical computing software R (R Development Core Team, 2010). The *stats* package used to run linear regression (ordinary least squares, ordinary least squares) is native to the core R program. We used the lme4 package (Bates, Maechler & Bolker, 2011) for linear mixed effects models. The main difference for the present purposes between linear modelling and linear mixed effects modelling is that we can include *suffix* as a random-effect factor in the latter, which allows us to take into account suffix-specific variability.

In addition to the lexical variables extracted from CELEX, we also included *voicing* as a variable in the analysis of the naming latencies. This variable brings into the model an acoustic property of the first segment, namely, whether it is voiced or not. Words beginning with voiceless sounds tend to trigger the voice key later than words with voiced initial segments, thereby creating an artificial difference in naming latency. The factor *voicing* controls for this voice key artefact. Only those characteristics which were found to be predictive will be discussed.

In what follows, two models for the naming latencies are presented, one without (ordinary least squares) and one with suffix as random-effect factor (mixed effects), in that order. Each of the sections of the tables below represents a different subset of predictors. Section A includes predictors capturing aspects of a word's form, such as length and number of syllables. Section B includes frequency information for the root and root + suffix combinations. Section C includes suffix specific information such as suffix type frequency and CO-Rank. Section D includes information about interactions between predictors. These sections are used throughout this paper for all our analyses.

Table 2.2 lists the significant predictors in our linear regression model. Among Section A predictors, both word length (measured in number of letters) and number of syllables influenced naming latencies, such that longer words and words with more syllables were associated with longer response latencies. Voicing was also a significant predictor, with words beginning with voiceless sounds taking longer to be registered than their voiced counterparts, as expected.

Section B included frequency measures for the root, the base (the intermediate derivation), and the whole word. These predictors were also involved in interactions (Section D). In this model, higher type frequency of the second suffix is predictive of faster naming latencies (Section C).

Section D outlines two interactions that influence naming latencies. First, the frequency of the root and base interact (Figure 2.1), such that the fastest response latencies are recorded when the base frequency is low and the root frequency is high. In general, faster response latencies are recorded when one of the two frequencies in the interaction is high, and the other is low. Here, the main pattern we see is that a low root frequency hurts response times to a greater extent at lower values of base frequency. The slowest response latencies are recorded when both root and base frequencies are low, which is expected. Second, the whole word frequency also interacts with the base frequency (Figure 2.2). The effect of word frequency is greatest for high base frequency. Effects of base frequency are most clearly present for high word frequency and absent for low word frequency.

The mixed effects model fitted to the naming data from the English Lex-

Table 2.2: Significant predictors for naming latency using the best linear regression model.

		Estimate	SE	*t* value	*p*
	Intercept	6.572	0.033	196.175	<0.0001
A	Word Length	0.01	0.003	3.389	0.0008
	Number of Syllables	0.027	0.006	4.587	<0.0001
	Initial Voicing (voiceless)	0.03	0.008	3.722	0.0002
B	Word Frequency	-0.03	0.003	-11.384	<.0001
	Root Frequency	-0.018	0.002	-8.622	<.0001
	Base Frequency	-0.036	0.007	-4.905	<.0001
C	Suffix2 Type Frequency	-0.007	0.003	-2.851	0.0045
D	BaseFreq*RootFreq	0.003	0.001	3.253	0.0012
	BaseFreq*WordFreq	-0.003	0.001	-2.373	0.018

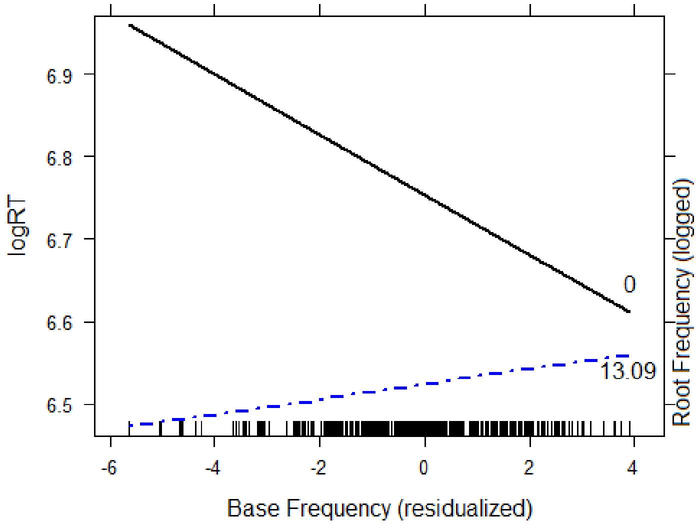

Figure 2.1: Root Frequency*Base Frequency interaction in naming using the regression model. Responses are fastest when the Base Frequency is low and the Root Frequency is high.

icon Project shares a number of similarities with the linear model. Table 2.3 lists the coefficient estimates for each predictor in the linear mixed-effects regression model. The same effects were observed in Section A as in the linear regression model, with longer words and words with more syllables predicting longer naming latencies, and words beginning with voiceless segments also

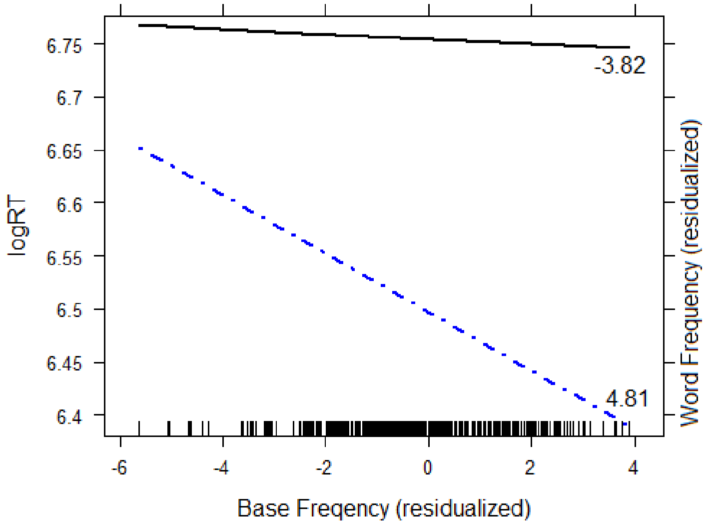

Figure 2.2: Base Frequency*Word Frequency interaction in naming using a linear regression model. Words with high Word and Base Frequencies show the fastest response times, with the effect diminishing as Word Frequency goes down.

having longer recorded latencies. This model included random intercepts for root (s.d. 0.0673), suffix1 (s.d. 0.0352) and suffix2 (s.d. 0.0324).

Likewise, in Section B, higher values for word frequency, root frequency, and base frequencies are predictive of shorter reaction times. Naming laten-

Table 2.3: Significant predictors in the linear mixed-effects regression model of naming latencies.

		Estimate	SE	*t* value	*p*
	Intercept	6.613	0.039	171.71	0.0000
A	Word Length	0.009	0.004	2.3	0.0219
	Number of Syllables	0.02	0.004	2.45	0.0148
	Initial Voicing (voiceless)	0.035	0.009	3.72	0.0002
B	Root Frequency	-0.015	0.003	-6.1	0.0000
	Base Frequency	-0.015	0.003	-4.91	0.0000
	Word Frequency	-0.023	0.003	-8.3	0.0000
C	Affix Pair Frequency	-0.008	0.003	-2.85	0.0046

cies decrease as these lexical frequencies increase. However, there were no interactions between root frequency, base frequency, or word frequency in our best model for naming latencies using this method (lmer).

In section C, instead of individual suffix effects, we instead see here an effect of the suffix string as a whole, where more frequent suffix pairs were produced more quickly during naming. This is unlike the linear regression (OLS) model.

Between the linear regression and linear mixed effects models, the model that best fits the data is the simpler mixed-effects model. By means of comparison, the AIC value for the mixed effects model is lower than that of the linear regression model (linear mixed-effects model: -1096.376, linear regression: -1086.353).

2.3.2 Lexical Decision Latencies

In the linear regression analysis, length and number of syllables are significant predictors for lexical decision reaction times, with higher values of both predicting longer reaction times (Table 2.4). In section B, root frequency, base frequency, and word frequency all predict faster reaction times when their values are higher. In section C, higher type frequencies of both the first and second suffix are predictive of faster response latencies. However, there is an interaction between word length and affix pair token frequency (Figure 2.3), where higher frequencies are more facilitatory for longer words. Longer words were, in general, responded to more slowly, as one might expect, but this effect was modulated by the frequency of the suffix pair at the end of the word. When the frequency of the suffix pair was higher, response latencies decreased.

Table 2.4: Predictors in the Linear model of lexical decision latencies.

		Estimate	SE	*t* value	*p*
	Intercept	6.57	0.038	172.386	<.0001
A	Word length	0.049	0.007	6.812	<.0001
	Number of Syllables	0.014	0.006	2.232	0.026
B	Root Frequency	-0.012	0.002	-5.404	<.0001
	Base Frequency	-0.006	0.002	-2.433	0.01527
	Whole Word Frequency	-0.036	0.003	-11.841	<.0001
C	Suffix1 Type Frequency	-0.01	0.003	-2.997	0.00285
	Suffix2 Type Frequency	-0.017	0.003	-6.451	<.0001
	Affix Pair Frequency	0.02	0.011	1.723	0.08537
D	Length*AffixPairFreq	-0.002	0.001	-2.8173	0.03019

The mixed effects model of lexical decision does not include an interaction (Table 2.5), but adds coefficients for the random effects of root (s.d. 0.0661), suffix1 (0.0243), and suffix2 (0.0490). Predictors in sections A and B are the same. In this model, higher values for Affix Pair Frequency are predictive of faster response latencies, as are the type frequencies of both suffixes individually. We also see a type frequency effect of the second suffix,

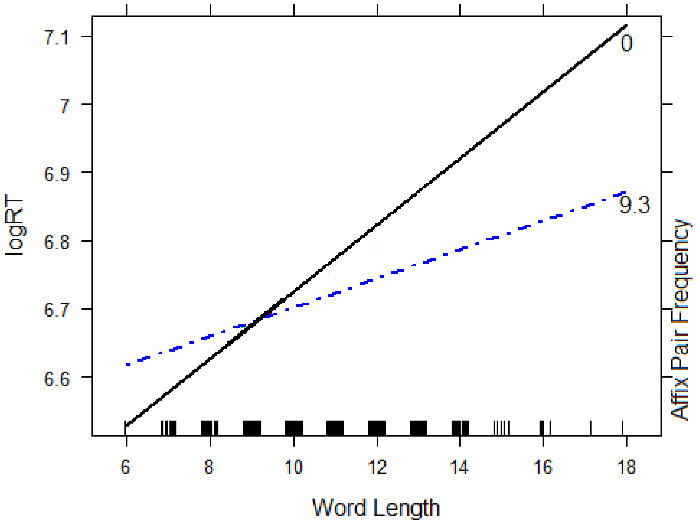

Figure 2.3: Word length * Affix Pair Frequency interaction in a linear model of lexical decision latencies. Higher frequencies of the Affix Pair facilitate processing, and this effect is more pronounced for longer words.

where faster response latencies are recorded for words with a higher Suffix2 Type Frequency. When type frequency is high, the number of potential bases a suffix combines with is high. That is, the suffix has been experienced in a larger number of contexts, and by extension, it will be expected in more conditions than a suffix with a low type frequency. When a decision must

Table 2.5: Significant predictors in the linear mixed-effects regression model of lexical decision reaction times.

		Estimate	SE	*t* value	*p*
	Intercept	6.602	0.049	133.44	<.0001
A	Word Length	0.022	0.004	5.54	<0.0001
	Number of Syllables	0.027	0.009	3.04	0.0025
B	Root Frequency	-0.011	0.003	-4.32	<.0001
	Base Frequency	-0.01	0.003	-3.05	0.0024
	Word Frequency	-0.031	0.003	-10.28	<.0001
C	Suffix2 Type Frequency	-0.024	0.006	-3.56	0.0004
	Affix Pair Frequency	-0.01	0.003	-3.36	0.0008

be made about the lexicality of a suffixed word, words with such productive suffixes are likely to be quickly identified and accepted. A suffix with a low type frequency is not used in a larger number of different words, so although it may occur frequency in one word (and thus have a high token frequency), participants are not as likely to be familiar with its use and so recognition of words containing it is slower.

Between the linear model and the linear mixed effects model for lexical decision latencies, the linear model is the better of the two (AIC -1027.058 for lm vs. -1003.406 for lmer).

2.4 Discussion

The models presented here represent those models that best fit the data, and while several similarities held across models, they were not constrained by a specific set of predictors. As a result, some predictors are not present in all models. The initial impetus for this investigation was a desire to test the Complexity-Based Ordering hierarchy on suffix pairs, building on Plag and Baayen's (2009) examination of root + suffix combinations. CO-Rank was not a significant predictor in any of our models, in either naming or lexical decision, at the level of the individual word. According to Complexity-Based Ordering, suffixes with lower ranks are more difficult to parse than highly ranked suffixes. This prediction could indicate that speech planning is slowed when confronted with less easily parsed suffixes in the middle of a word. It is possible that, in the mixed effects model, the use of the first and second suffix as random effects might encompass the information encoded in CO-Rank. We extracted the random intercepts of Suffix1 and Suffix2 from the mixed effects model and compared them to the CO-Rank ordering. There was no correlation between coefficients and CO-Rank. Given that type frequencies for suffixes did not pattern with CO-Rank, it is unlikely that CO-Rank conveys information about type frequency (or suffix family size). In this study, we therefore do not find support for CO-Rank ordering as a psycholinguistic variable that aids in our understanding of mutlimorphemic processing.

Although we do not observe effects of rank, our results do provide an interesting account of trimorphemic word processing. In particular, we find that the frequency of the suffix pair as a whole (e.g., *fulness* in *hopefulness*) makes a difference during reading. When the string is more frequent, reaction times are reduced.

2.4.1 Word-based effects

Word-based effects are those effects relating to the root or combinations of the root with one or both suffixes. There were very clear effects of root frequency, base frequency, and whole word frequency in every model, with higher frequencies being predictive of faster response latencies. These effects were robust, appearing in both the lexical decision and naming experiments, under both types of analyses. Similar results were reported by Plag and Baayen (2009) for the root frequency and derived frequency, although they did not report any interactions. Also robust were effects of length and syllable number, with higher values of both predicting longer response latencies.

The fact that there are multiple morphemes in each target word allows us to examine combinatorial frequency effects of the root and both suffixes. We saw an interaction between Root Frequency and Base Frequency in the linear model for naming latencies, with naming latencies for low root frequency becoming longer at lower values of base frequencies. If the base is a high frequency word, then it may have its own, independent, representation in the lexicon that is separate from the root alone (e.g., Hay, 2001). Although the words have the same root, the base may not require decomposition to be recognized, and at higher frequencies of the base, this recognition speeds naming. When the root and base are both highly frequent, then the base (an independent word) and the root compete for resources. When the base is low frequency, the lexical system does not encounter any difficulties from lexical retrieval upon encountering suffix1 after reading the root, as the word is parsed and decomposed without interference from other lexical items. Similar results of competition and interference have been reported elsewhere. For instance, Kuperman et al. (2009) found that for compounds, lexical decision latencies increase when both the frequency of the left constituent and the whole compound are high. Similarly, Balling and Baayen (2008) showed that for response latencies to auditory stimuli, high suffix frequencies were inhibitory when the word frequency was also high.

In the linear model for naming latencies, we had a second interaction between whole word and base frequencies, where faster response times were associated with high frequencies of both. In this case, there is facilitation. The difference may stem from both the overall frequencies of the whole words versus bases, as well as suffix ordering itself. First, the whole word frequencies were typically lower than those of the base words, so an interference effect like the one just described was not likely to be observed. With one suffix already in place (Suffix1), there are a limited number of suffixes that can follow it. Before the final suffix has been fixated upon, the reader can still tell that the target word is longer than the base combination (root+suffix1), and when the whole word is highly frequent, the final suffix may be predicted before it is directly viewed.

2.4.2 Suffix Effects

There was variation in which suffix-related variables were predictive, and where they were of predictive value. Suffix2 Type Frequency and Affix Pair Token Frequency were both significant predictors in three of four models. Suffix token frequencies were tested during the modeling process, but were not as successful as suffix type frequencies in their predictions.

Suffix type frequencies can be thought of as a relatively simple measure of productivity. When a suffix has a high type frequency, it occurs with many distinct words, and by extension, may be more readily used to create new words than suffixes that do not occur in as many existing words. Likewise, individuals are likely to have encountered suffixes with high type frequencies across more words, so that they are more familiar with them and their interactions with other morphemes. In experimental terms, individual sensitivity to previous occurrences is reflected in faster reaction times to more familiar items or parts of items.

The frequency of the affix pair as a whole was found to be a significant predictor in the mixed effects lexical decision model, in addition to effects related to individual suffixes. This is interesting, as it indicates that more frequent affix pairs might be prone to processing as single units, which could also help to explain the lack of effect for CO-Rank in suffix pairs.

2.5 Conclusions

In this series of analyses, we have examined naming latencies and lexical decision reaction times for trimorphemic English containing a root and two suffixes. Latencies were extracted from the English Lexicon Project and modeled using both linear models and mixed effects linear models. While there were some differences between the significant predictors in the linear and mixed effects models, the main frequency effects of the root, base, and word were stable across all models. Suffix-related frequency effects were significant in all models, with both type frequencies (primarily for Suffix2) and token frequencies (Affix Pair Frequency) being used during reading and naming.

The proposed CO-Rank was not found to be a significant predictor in any of our reported models. Differences between the models may, to a certain extent, be attributed to the random effects of "Suffix" that can be accounted for using mixed effects modeling but not in simple linear models. "Suffix1" and "Suffix2" were used as random effects in our mixed effects models, meaning that variation attributable to suffix identity was included in the model. However, across both types of models, the CO-Rank fails to be predictive of response latencies in trimorphemic words.

Although effects of CO-Rank were absent, we did find clear effects of the Root, Base, Whole Word, and suffixes in Type Frequencies for Suffix1 and Suffix2 (although the latter was more prevalent) as well as in the Affix Pair Frequency, which we introduce here as a significant predictor for trimorphemic words. The existence of frequency and type effects for all of our morphemes in this experiment suggests that all three units in our root+suffix+suffix words have lexical representations in a classical morphological model, as do some instances of the Base and Suffix1+Suffix2 combinations. The Affix Pair Frequency effect was separate from frequency effects for Suffix1 and Suffix2, and this suggest that the suffix pair is also represented at some level of analysis (but see Baayen et. al, 2011, for a two-layer naïve discriminative learning model that accurately reported various lexical frequency effects in lexical decision). The frequency and type effects presented here are at home within a model of lexical processing that emphasizes speaker sensitivity to morphological units during language use.

2.6 References

[1] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

[2] Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*, 438-482

[3] Baayen, R. H, Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

[4] Balling, L. W. & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes, 23*, 1159-1190.

[5] Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-38.

[6] Fabb, N. (1988). English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory, 6*, 527-539.

[7] Hay, J. (2002). From speech perception to morphology: Affix ordering revisited. *Language, 78*(3), 527 - 555.

[8] Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics, 39*(6), 1041 - 1070.

[9] Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory, 22*, 565-596.

[10] Kiparsky, P. (1982). Lexical phonology and morphology. In I. Yang (Ed.), *Linguistics in the morning calm*(pp. 3 - 91). Seoul: Hanshin.

[11] Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R.H. (2009). Reading of polymorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Journal of Experimental Psychology: HPP, 35*, 876-895.

[12] Plag, I. (1996). Selectional restrictions in English suffixation revisited: a reply to Fabb (1988). *Linguistics, 34*(4), 769-798.

[13] Plag, I.,& Baayen, H. (2009). Suffix ordering and morphological processing. *Language, 85*(1), 106-149.

[14] R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0.

Chapter 3

Processing of trimorphemic derived words: Evidence from eye-tracking

3.1 Introduction

A major question for researchers of multimorphemic lexical processing is how (and which) morphemes are accessed in the lexicon. The present study addresses the processing of trimorphemic words in English. Single-route models of lexical access encompass both obligatory decomposition (bottom-up processing) and full listing accounts (direct, top-down processing). In single-route theories proposing obligatory decomposition, words are automatically broken into their component morphemes as they are encountered (Taft & Forster, 1975; Taft, 2004). Top-down (full listing) approaches hold instead that the word is recognized first as a whole, although it can be broken down during further processing (e.g., Giraudo & Grainger, 2001). Dual-route models combine decomposition and whole-word recognition (e.g., Baayen & Schreuder, 1999), and are often formulated in terms of a race, where both routes are available to incoming complex words but where one route will be preferred depending on lexical factors such as root and whole word frequencies.

Numerous studies of morphological processing have examined the importance of lexical frequency, manipulating root and whole word (surface) frequencies to tease apart the route morphemes must take to be accessed in the lexicon. Lexical decision experiments have revealed effects of root frequency when reading suffixed words (e.g., Taft, 1979; Cole, Beauvillain, & Segui 1989). When the whole word frequency, or surface frequency, is kept constant, but the frequency of the root is varied, more frequent roots are responded to more quickly than less frequent roots. Taft (1979) also found that when the root frequency remains constant, but the whole word frequency is varied, words with higher frequencies are processed more quickly, as reflected in reaction times. Similar results have been obtained in Italian using phonologically neutral, productive suffixes (Burani & Caramazza, 1987).

The contribution of suffixes in lexical processing has also received attention. Results from priming studies support an early phase of morphological decomposition, where suffixes are recognized at a level that is independent of semantic interpretability and purely orthographic effects (Marslen-Wilson & Bozic, 2008; Longtin & Meunier, 2005). Suffix properties such as type and token frequency have been found to influence reaction times (Laudanna & Burani, 1995; Burani & Thornton, 2003). However, whether a suffix will come into play during processing appears to be sensitive to factors such as suffix allomorphy (and confusability), the relative frequencies of morphemes, and phonological factors. In the case of the former, suffixes that have many allomorphs or that have more than one use (e.g., comparative *-er* vs. agentive *-er*) are less likely to be used as independent units, as they are not as salient to speakers (Järvikivi, Bertram, & Niemi, 2006). Bertram et al. (2000) argued that when different suffixes sharing the same form (i.e., homonyms) were of similar frequency, the lexical system would store those suffixed words in memory. Storage of such forms would reduce processing difficulties arising from the selection and/or identification of the appropriate suffix.

The general idea underlying affix salience is that the more a suffix "sticks out," or is obvious to a speaker, the more likely it is to be used as a morphological processing unit. In addition to homonymy, suffix productivity and length also factor in to the level of saliency of an affix (Järvikivi et al., 2006). Suffix parsability, including whether or not the addition of a suffix changes the phonology of the root, will also affect how easily a suffix can be recognized as an independent unit (Hay, 2002). The relative frequencies of the root and the derived form have roles to play here, where a word is more likely to be decomposed during processing when the root is more frequent than its derived form (Hay, 2001). Conversely, when the derived form is more frequent than its root, then it is more likely to be treated in the same way as a monomorphemic word, which means that the affix combining with the root will not be readily acknowledged. Hay's work shows that speakers are sensitive to relative frequencies even when a word is comparatively uncommon.

Although there is strong evidence that morphemes (roots and affixes) can be accessed during word recognition, the time course of recognition is not always clear. Lexical decision reaction times necessarily come at the end of word reading, for instance. Eye-tracking methodology has been added to traditional behavioural experiments to better understand the time course of lexical access. Beauvillain (1996) manipulated the root and surface frequencies of suffixed words in French in an eye-tracking study using a semantic relatedness task. In the first experiment, the surface frequency (whole word frequency) was either high or low while the roots were kept constant (and so maintained a constant root frequency). Higher surface frequencies were associated with shorter fixation durations on the whole word, with this effect emerging at the second fixation. The whole word frequency did not appear to influence duration at first fixation. In the second experiment, the root frequency was manipulated while the whole word frequency was kept constant. The first fixation was shorter for words with more frequent roots. Niswander, Pollatsek & Rayner (2000) also found that the first fixation duration on derived words was affected by the frequency of the root, but not the whole word frequency. Effects for whole word frequency were readily apparent for the total gaze duration on the word. For both effects, shorter durations were reflective of higher frequencies. These results, from English, were similar to results found in compound processing for Finnish (Bertram & Hyönä, 2003) and Dutch (Kuperman, Schreuder, Bertram & Baayen, 2009).

Effects of word length are made clear through the eye-tracking record. Bertram and Hyönä (2003) examined the processing of Finnish compounds, finding that higher frequencies of the first constituent facilitated the duration of the first fixation on a word in an eye-tracking study. This effect was absent for shorter words (7-9 letters), but prevalent in longer words. For longer compounds, they reported that higher frequency left constituents facilitated faster gaze durations (whole word and at the first fixation). A follow-up study targeting longer compounds (12 - 18 letters) confirmed this effect of length (Hyönä, Bertram & Pollatsek, 2004). Similarly, Niswander-Klement and Pollatsek (2006) examined the processing of prefixed words in English inside of a sentence context. For prefixed words, they again found whole-word frequency effects for short words, but an effect of the root frequency for longer prefixed words in total gaze duration. They interpret their findings as supportive of a dual-route model wherein longer words are more prone to decomposition, due to the need for more than one fixation to fully recognize a given word. Similar effects for suffix length have been observed in Dutch derived words (Kuperman, Bertram & Baayen, 2010).

Kuperman, Bertram and Baayen (2008) proposed a model of morphological processing based on principles of information theory that they called PROMISE, the PRObabilistic Model of Information Sources, to account for experimental results that proved problematic for other models. In their study, they examined the processing of trimorphemic Finnish compounds in a sentence reading task where participant eye-movements were recorded using an eye-tracker. Target items were noun-noun compounds, some of which also contained a derivational suffix as part of the first noun. They found effects of frequency (compound, first constituent), first constituent family size and word length at the time of the first fixation. Said another way, there are effects of both the whole word (compound frequency) and the left constituent at the same time. A strict full listing model of lexical access would not predict any effect of the first constituent until after the whole word has been recognized, while obligatory decomposition models would not predict an effect of whole word frequency until after the individual constituents had been recognized. Their findings led the authors to suggest that readers "make inferences about the compound's identity as soon as they have available any (potentially incomplete) information about a word" (pg. 1108). We expect, then, that readers (and by extension, listeners) begin to process lexical information before they have enough information to identify a word, and that this process is both dynamic and integrative, in that progress in word identification must be updated as new information becomes available. Frequency and family size effects from the right constituent support the continual uptake of information.

The general purpose of the present study was to examine the processing of suffix+suffix combinations in English, with the goal of understanding the factors influencing lexical processing. For example, we hypothesized that root frequency will play a role in fixation durations as is observed in bimorphemic processing (e.g., Beauvillain, 1996; compounds: Kuperman, 2008). We made use of the lexical decision paradigm with an eye-tracking component. Eye-tracking as a methodology allows greater insight into the time course of lexical processing than lexical decision or naming experiments alone. Rather than focus on a single summed reaction time, we can extract information from each fixation on a word from eye-tracking data. We expect that there will be some role of decomposition in our target words, because 1) they are relatively low frequency, and 2) they are, on average, fairly long and are likely to require more than one fixation to be fully read.

In English we have an additional point of interest in examining suffix combinations. Suffix ordering has been the source of much debate among morphologists. In English, there are many more possible suffix combinations than are actually attested, so a key theoretical question is why this should be the case. Most treatments of suffix ordering in English have been theoretical, and fall into two basic categories. On the one hand, there are level-ordering theories, which hold that there are two (or more) lexical strata in English, and that the levels (strata) combine in specific ways (e.g., Kiparsky, 1982; Giegerich, 1999). The other holds that restrictions on suffix behaviour are affix-specific, and these are called selectional restrictions (e.g., Fabb, 1988).

Hay and Plag (2004) developed a psycholinguistically motivated theory of affix ordering, called Complexity-Based Ordering, which holds that suffix ordering is modulated by a combination of processing constraints and suffix-specific selectional restrictions. Plag and Baayen (2009) tested the predictive value of the Complexity-Based Ordering hierarchy for lexical decision reaction times to bimorphemic words. They found that the ranking was not predictive for individual words, but was predictive for mean reaction times for each suffix. We include the hierarchy for further testing in our investigation of root+suffix+suffix processing.

3.2 Methods

3.2.1 Participants

Twenty-six participants took part in this experiment. Participants were drawn from the Department of Linguistics subject pool, and completed this experiment for course credit (6 percent of the final grade for summer students, 3 percent of the final grade during the regular school year). Age ranged between 18 - 25. One participant was removed for excessive errors ($> 30\%$).

3.2.2 Equipment

Eye movements were recorded using the Eyelink II head-mounted eye-tracking system (SR Research, Ltd., Mississauga, Ontario, Canada). This system runs on two computers: a host computer, which runs the eye-tracking equipment, and a presentation computer, which presents the experiment to participants. The host computer was a Dell Inspirion laptop running the Eyelink II proprietary software. The presentation computer was a Dell desktop running Windows XP. Participants were seated at an optimal viewing distance from the presentation screen. Cameras focusing on the pupils were adjusted such that they had a clear view of the pupils but did not impair vision. A third infrared camera was used to track head position relative to the presentation screen. The system was calibrated every ten words during the experiment. Both eyes were calibrated for tracking, and the best was selected for recording. Wherever possible, the system used both pupil detection and corneal reflections to capture eye-movements and eye position, with a sample rate of 250 Hz. All participants had normal to corrected vision.

Behavioural data were recorded to the presentation computer through the use of a MS Sidewinder game controller.

3.2.3 Stimuli

Critical stimuli consisted of trimorphemic words composed of a root and two suffixes (e.g., *hope+ful+ly*). There were 650 target stimuli in this experiment, in addition to 650 nonwords. These nonwords were created by taking trimorphemic words of the same structure as the target items, and altering one morpheme to a non-existing morpheme (e.g., *hope+ful+ak*). One third had non-existing roots, one third non-existing first suffix, and the final third had a non-existing final suffix. This variation was included to encourage participants to look at every item completely before making a decision as to their word or nonword status. All participants saw all items.

For each word, we included as predictors for our results the whole word frequency (lemma), as extracted from the CELEX Lexical Database (Baayen, Piepenbrock, & Gulikers, 1995), the root frequency (the frequency of the first morpheme) and the frequency of the root plus the first suffix. For ease of discussion, the Root+Suffix1 combination will be referred to as the "base" for the duration of this paper. Two more frequency variables were also investigated, the frequency of the suffix pair at the end of the word (suff1 + suff2) and the discontinuous frequency of the root and the second suffix frequency (root + suff2). The latter was motivated by the possibility that, if readers are making inferences about the identity of the viewed item, they might also be sensitive to existing patterns that are not sequential. In this experiment, where we use three morphemes, it is unclear whether effects of frequency will be purely sequential, or whether expectations about upcoming elements based on previous experience can affect processing even when morphemes are not next to each other.

Target words ranged from 6 to 17 letters in length, with a mean length of 11 letters. Word frequency ranged from 0 to 2224 in the Celex Lexical Database, with a mean of 43. The frequency of the root ranged from 0 to

35438 (mean 1671), and the base frequency ranged from 0 - 7171 (mean 305). All frequencies were log transformed to reduce the effect of extreme values. Base and Word frequency were residualized on the Root to address correlations between predictors.

For each individual suffix, we included suffix token frequencies and suffix type frequencies. Suffix type frequency has also been called suffix family size, and refers to the number of words containing a given suffix, but not the summed frequencies of those words. We included as a variable the family size of the root and suffix rank in the Complexity-Based Ordering hierarchy, with limited results for the first and none at all for the second. In our investigation of fixation durations, we also included as a variable whether the fixation under investigation was also the last fixation, before the lexical decision was made.

3.2.4 Procedure

Participants were seated comfortably in front of a computer screen before being fitted with an Eyelink II head mounted eye-tracker. The system uses a headband that is tightened or loosened as necessary to best fit the participant's head, on which three cameras are held. The eye cameras can be maneuvered to capture eye movements, and the third camera is an infrared camera on the eye-tracker that is used with a set of sensors on the presentation computer to monitor head movements with reference to the screen. Fitting took approximately 10 minutes. They were then asked to complete a simple lexical decision task.

The eye-tracker was calibrated for eye position at the beginning of the experiment and every ten complete trials afterwards. We used a 3-point calibration for this experiment. For each calibration, participants were asked to follow a dot on the screen to three locations. Additionally, a drift correction was made prior to each trial. The drift correct used before every trial doubled as a fixation point to centre the eye in the optimal viewing position for each word. This was to ensure that the first fixation was always inside of the target item. Even so, sometimes very short fixation durations (< 35 ms) were recorded at the first fixation, and these were removed as quick refixations.

Half of the items viewed by participants were existing words, and half were nonwords. They were asked to indicate whether or not an item was a word in English by pressing a button on an MS Sidewinder game controller. If the item was not a word, they were asked to press the left trigger button, and if the item was a word, then they were asked to press the right trigger button.

There were three programmed breaks of 5 minutes during the experiment. Participants were also able to take additional breaks if they were required. The entire experimental session lasted between one and 1.5 hours.

3.3 Results and Discussion

Results were analyzed using linear mixed effects models, using the lme4 package (Bates & Sarkar, 2007) for the statistical computing software R (R Development Core Team, 2011). In each model, we kept the Subject (participant), Root, and Suff1 and Suff2 as random effects. Our results have been broken down into three basic categories. Our dependent variables in this experiment were fixation durations (times spent on a particular fixation, e.g., the first), total fixation duration on a word, number of fixations, and lexical decision response times. First, we examine the total number of fixations per target item and total fixation duration per target item, followed by an analysis of lexical decision reaction times. These three measures reflect the reading times of the entire target item. We then examine a breakdown of fixation durations and consider the durations of each of the first to fourth fixations. Beginning with the second fixation duration, the number of total observations per word falls. Linear mixed effects models are known to be robust for missing values. In addition, the number of participants represented at each fixation number remains stable, meaning that each participant viewed at least one word with four separate fixations. Similarly, all target items were viewed using three fixations by at least one participant.

In each of the following subsections, we report the mixed effects models that best fit the data. In these models, we only report on those predictors that were significant at the 5% level.

3.3.1 Number of Fixations

Most words received between 3-5 fixations (Table 3.1). The number of fixations increased with increased word length. Higher lexical frequencies of the root, the base, the affix pair, the second suffix, and the whole word were associated with fewer fixations. Interestingly, the discontinuous frequency of the root and the second suffix (*hope* and *-ly* in *hopefully*) was also associated with fewer fixations at higher frequencies (Table 3.2).

There was an interaction between the Affix Pair Token Frequency and the frequency of the discontinuous Root+Suffix2 frequency (Fig 3.1). The most fixations were made when The Affix Pair token frequency and the

Table 3.1: Overview of fixation counts for trials and words. *Number of trials* refers to the number of individual trials which received the total number of fixations listed in the first column. The *% Words* column lists the percentage of words in the experiment that receive the number of fixations given in the first column.

Total # of Fixations	Number of Trials	% Words
1	125	16.5
2	1708	81.8
3	4077	95.1
4	2888	95.9
5	1338	82.1
6-7	963	97.0
8-9	263	34.0
10-11	64	9.2
12-13	32	4.4
14-17	13	2.1

Table 3.2: Predictors for Total Number of Fixations, based on 11875 observations

	Estimate	SE	*t* value
(Intercept)	1.24	0.09	13.11
WordLength	0.05	0.002	21.83
Base Frequency	-0.03	0.002	-14.67
AffixPairTokenFreq	-0.01	0.002	-5.41
Root+Suff2	-0.03	0.01	-4.35
Root Frequency	-0.03	0.002	-13.79
Word Frequency	-0.02	0.003	-6.87
Suff2Freq	-0.02	0.01	-2.39
AffixPairTokenFreq*Root+Suff2	0.005	0.001	4.10

Root+Suffix2 frequencies were both high. The lowest number of fixations were made when the Affix Pair token frequency was low and the Root+Suffix2 frequency was high. This effect might be due to expectations for the last suffix stemming from the identity of the root. If the root brings with it expectations for the third suffix, then a highly frequent affix pair with its own representation might be expected to interfere with that expectation, even if the ending suffix is the 'correct' suffix.

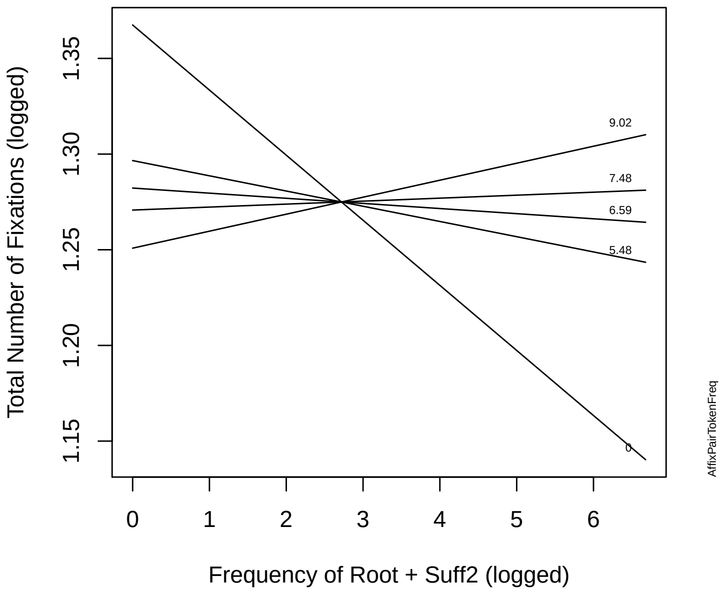

Figure 3.1: Interaction between Root+Suff2 and Affix Pair Token Frequency in total fixation duration.

3.3.2 Total Fixation Duration

The total fixation duration is the summed time participants spent reading a given item before making their lexical decision (Table 3.3), and as such, the best model for total fixation duration shares many similarities with that for lexical decision reaction times. The mean total fixation duration was 909 ms. The root frequency, base frequency, whole word frequency, and affix pair token frequency were all facilitatory, with higher frequencies for each predictor being associated with shorter total fixation durations. Word length was inhibitory, with longer words showing longer total fixation durations. As participants proceeded through the experiment, they read more quickly. Total fixation durations were also shorter when the token frequency of the second suffix was high. This effect was separate from the frequency of the affix pair as a whole. However, we find no effect of the frequency of the first suffix. Effects of the first suffix may be absorbed by the base frequency, especially where the base combinations are more frequent than the root alone, and therefore more likely to be accessed as a single unit (Hay, 2001, 2002). This, in turn, means that the first suffix is less likely to be as readily parsed as the second. When CO-Rank was considered in our model, it was not predictive for total fixation duration. Although we did not find effects of Complexity-Based Ordering Rank here, the principle of parsing differences in different morphemic slots remains plausible.

There was an observed interaction between the root and base (Fig 3.2). The longest fixation durations were observed when the root frequency and the

Table 3.3: Predictors for Total Fixation Duration, based on 11875 observations

	Estimate	SE	*t* value
(Intercept)	7.02	0.09	77.48
Base Frequency	-0.06	0.01	-6.52
Root Frequency	-0.03	0.003	-10.56
Word Frequency	-0.02	0.004	-4.38
AffixPairTokenFreq	-0.01	0.003	-4.30
WordLength	0.02	0.00	5.59
Suff2Freq	-0.02	0.01	-2.36
Trial Index	-0.05	0.002	-22.57
Base Frequency*RootFreq	0.004	0.001	2.59

base frequency were both low. The shortest fixation durations were observed when both frequencies were high. This facilitation effect was strongest when the Root Frequency was high and Base Frequency was low, and was negligible when both were high. This is somewhat different from interactions described by Kuperman et al. (2008) for compound constituents and by Winther-

Balling and Baayen (2008) for suffixes and whole word frequency in Danish, where an interference effect was reported when both frequencies were high. In our data, it is possible that the frequency of the base is overtaking that of the Root as the main processing unit, which is why we see large gains for low frequency roots and not for high frequency roots. In a dual-route model of word recognition this delayed processing may still be taken as evidence for interactions between the whole word route and parsing route. This is more likely to be the case for higher frequency roots than those at the lowest values in our data. That is, overall evidence suggests that the two access/retrieval routes for multimorphemic words are not independent of each other.

3.3.3 Lexical Decision Reaction Times

The best model to describe the reaction time data had seven predictors and one interaction, in addition to the random factors. Table 3.4 provides the intercept and coefficients for each predictor. Lexical decision reaction times correlated strongly with total fixation duration (r=0.94).

Higher lexical frequencies of the word and affix pair facilitate response times in lexical decision. We see again that participants became faster during the task, with Trial number being positively associated with faster reaction

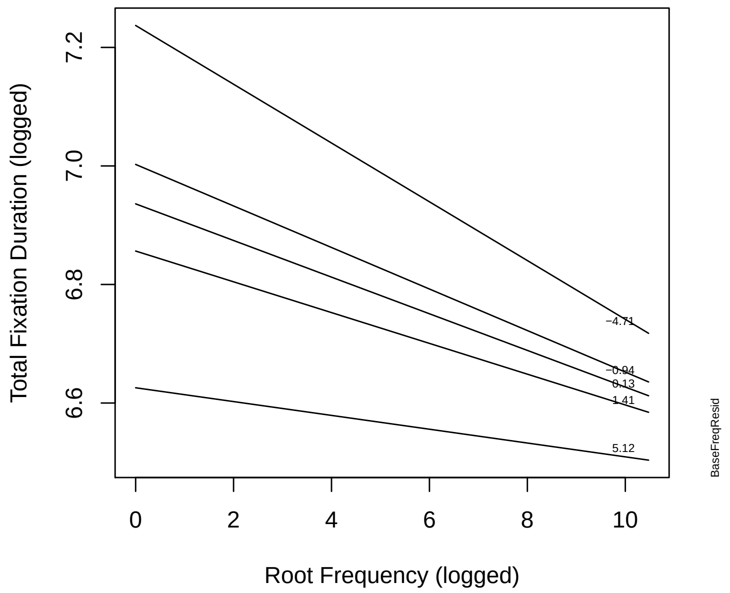

Figure 3.2: Interaction between Root and Base Frequency in total fixation duration.

times. This may reflect task familiarity as this admittedly long experiment proceeds.

An interaction between the root frequency and the base frequency was also observed (Fig 3.3), similar to our results from total fixation duration. Reaction times were longest when both the root frequency and the base frequency were low. Reaction times were fastest when the frequency of the root and the base were high. The facilitatory effect of the base frequency

Table 3.4: Predictors for Lexical Decision Reaction Times, based on 11875 observations

	Estimate	SE	*t* value
(Intercept)	7.55	0.14	55.17
Trial Index	-0.0001	0.000007	-20.75
WordLength	-0.09	0.02	-3.77
WordLength^2	0.005	0.001	5.04
Word Frequency	-0.02	0.003	-4.93
Root Frequency	-0.04	0.002	-12.72
Base Frequency	-0.08	0.01	-8.38
AffixPairTokenFreq	-0.01	0.003	-4.81
RootFreq*Base Frequency	0.01	0.001	3.84

was more pronounced for low frequency roots.

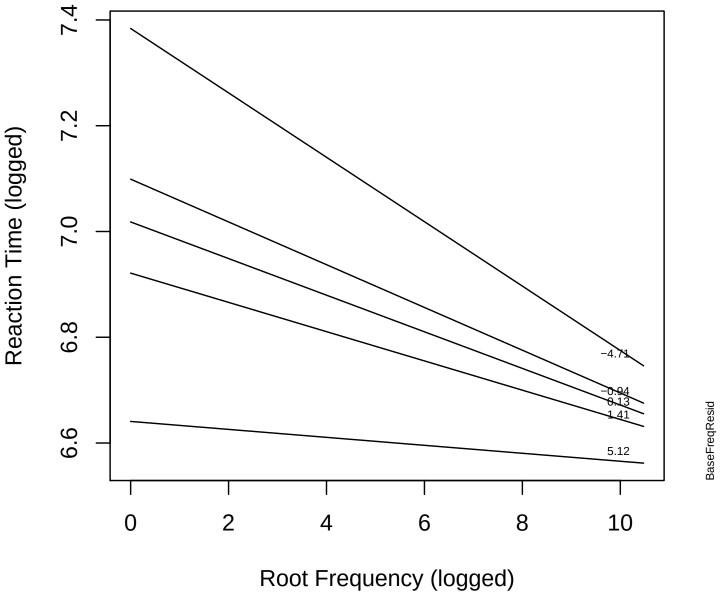

Figure 3.3: Interaction between Root and Base frequencies in lexical decision reaction times.

3.3.4 Fixation Durations

3.3.5 First Fixation Duration

The mean fixation duration for the first fixation was 262 ms. Duration of the first fixation was best predicted by a model containing an interaction between word length and type frequency of the second suffix (Table 3.5). As the experiment went on, the amount of time spent on the first fixation became longer.

The first and most obvious frequency-related finding is that we have morpheme frequency effects at the first fixation, where the whole word is unlikely to have been adequately perceived. The frequency of base (but not the root alone) sped processing at higher frequencies. At first glance, an effect for the base, but not the root, may seen contradictory. However, recall Hay's (2001) study of relative root and whole word frequencies, where derived words with a higher frequency than their roots are more likely to be stored/processed as whole word forms, and less likely to be decomposed during processing. An analysis of our target items showed that our bases, on average, had a higher frequency than the root alone, meaning that many of our base items were predisposed to be viewed as a whole. That is not to say that the root cannot become important later during processing (it does), only that at the first fixation, the emphasis for these words was on the intermediate base form. Additionally, while the experiment was set up such that participants viewed the root during their first fixation on a word, this does not preclude viewing effects of the first suffix during first fixation. Given that we do not see effects of the first suffix as a single unit (i.e., not outside of the base frequency), it is likely that the base is being accessed as a single unit at this point during reading.

Table 3.5: Predictors at First Fixation Duration, based on 11875 observations

	Estimate	SE	*t* value
(Intercept)	5.60	0.04	134.59
Trial Index	0.02	0.002	7.15
WordLength	-0.00	0.002	-1.69
Suffix2 TypeFreq	-0.07	0.01	-4.70
Base Frequency	-0.004	0.002	-2.51
WordLength*Suffix2 TypeFreq	0.01	0.001	4.38

The second suffix type frequency interacted with word length. This result is intriguing, and suggests that participants made use of their parafoveal view to extract information about the second suffix. For shorter words, a higher suffix2 type frequency resulted in a shorter first fixation duration, while longer durations were associated with smaller type frequencies (Figure 3.4). For the shortest words, it may be the case that lexical information from all morphemes is available for uptake at first fixation. For longer words, an extended first fixation duration may be related to planning eye-movements to the rest of the word while simultaneously paring down the number of items in the lexicon that can be selected based on a preview of the final suffix. A suffix with a higher type frequency is a suffix that attaches to more bases (i.e., that has more family members) than one with a low type frequency, with a higher number of potential combinations. This remains true even when the root is viewed at first fixation. In sentence reading, longer first fixations are reported when a following word (n+1) is more predictable (Kliegl, Nuthmann, & Engbert, 2006). This successor effect is interpreted as anticipated recognition of the next word (n+1), where information that is predictable from prior context is retrieved while fixated on the target, leading to a longer fixation on the target item. A suffix with a larger family is also, in a simple sense, more productive and more likely to be part of a word than a suffix with a smaller family, and more predictable during reading where a single word requires more than one fixation. This suggests that participants readily make use of available information as they read our target words.

3.3.6 Second Fixation Duration

At the second fixation duration, we start to see effects of the variable "Last-Fix," which refers to cases where the current fixation was also the last fixation

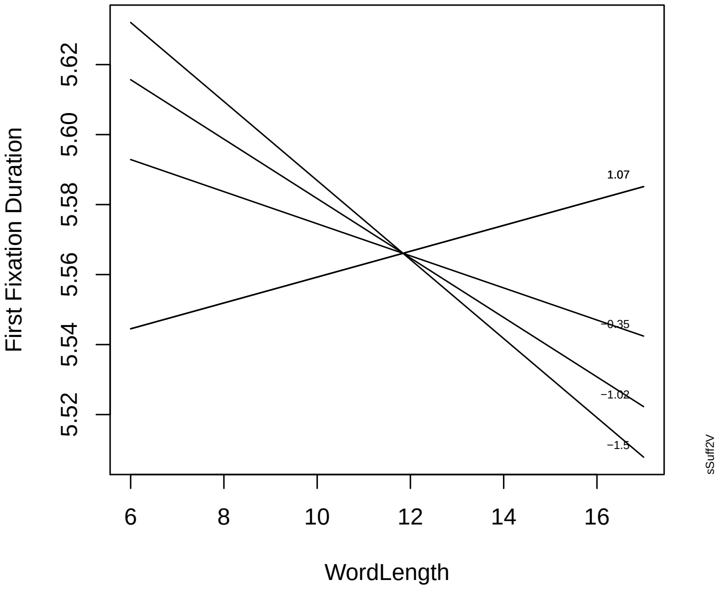

Figure 3.4: Interaction between word length and the suffix type frequency of the second suffix at the first fixation duration.

before the lexical decision was made. Mean fixation duration for the second fixation duration was 209 ms. The duration of the second fixation was facilitated by high root and base frequencies. We also observed an effect of the discontinuous frequency of the first and the third morphemes (Table 3.6).

At this second fixation, the frequency of the root begins to have an effect. It may be the case that the identity of the root becomes relevant as the integration of the whole word comes together.

Table 3.6: Predictors at Second Fixation Duration, based on 11707 observations

	Estimate	SE	t value
(Intercept)	5.78	0.04	143.38
LastFixTRUE	-0.35	0.06	-6.28
Trial Index	0.0007	0.004	0.21
WordLength	-0.03	0.002	-13.62
Root Frequency	-0.02	0.002	-6.89
Base Frequency	-0.01	0.003	-5.39
LastFixTRUE*Trial Index	-0.05	0.01	-5.36
LastFixTRUE*WordLength	0.06	0.01	10.95

We observed interactions between LastFix and Word Length, where longer words would receive longer second fixation durations if the subject was on their final fixation of a word (Figure 3.5.). The second fixation duration was shortest when the word was long, but the second fixation was not the final fixation. For shorter words, there was little difference between non-final and final second fixation durations (the difference grew larger as word length increased). We also observed an interaction between LastFix and Trial Number (Fig 3.6), where the time spent at the second fixation was shorter if it was the final fixation. There was no difference in time spent at second fixation when it was not the final fixation as the experiment progressed.

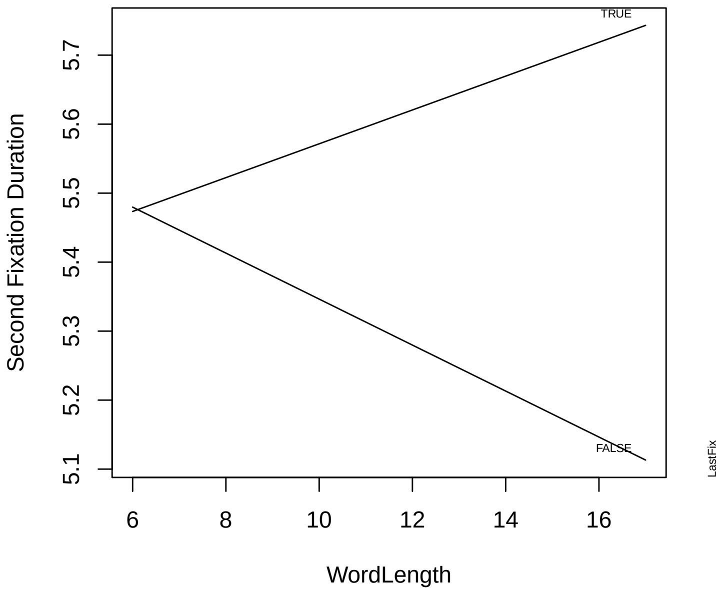

Figure 3.5: Interaction between Word length and LastFix at the second fixation duration.

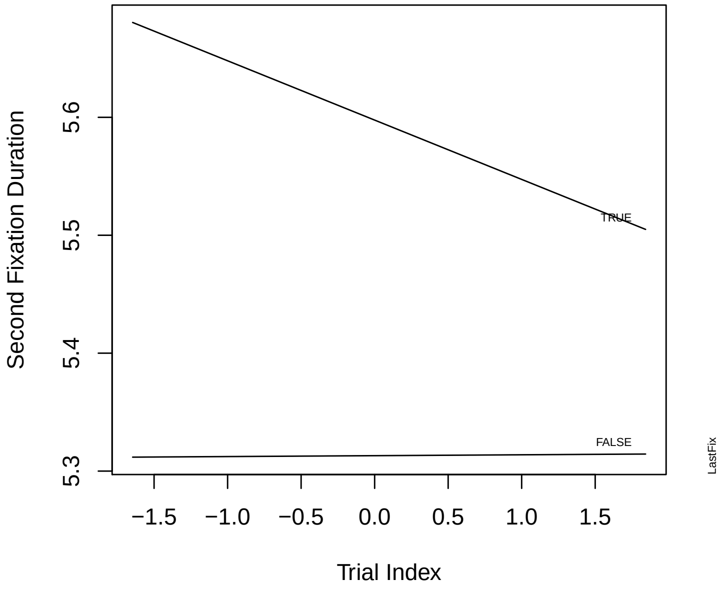

Figure 3.6: Interaction between Trial Index and LastFix at the second fixation duration.

3.3.7 Third Fixation Duration

The mean fixation duration at the third fixation was 205 ms. At the third fixation, the frequency of the suffix pair as a whole emerges as a significant predictor for fixation duration, in addition to the frequency of the base, with both speeding fixation duration (Table 3.8). That is, for the affix pair frequency, the more often we encounter two suffixes together, the more they may become "fused", or expected together. If this can be true for high frequency multimorphemic words, then there is in principle no reason why lexical memory should not also be able to store combinations of frequent affixes.

As observed elsewhere, participants became faster as the experiment progressed. Participants were faster on the third fixation when it was their final fixation in the word. This is visualized in Fig 3.7, which shows the interaction between trial number and whether the third fixation was the last.

There was an interaction between trial index and whether participants were on their last fixation, such that they were faster during the last fixation as the experiment progressed.

3.3.8 Fourth Fixation Duration

By the fourth fixation, it is likely that we are moving away from strictly morphological effects and moving into decision-based effects. Even so, a number of lexical predictors still significantly contribute to the fourth fixation duration. The number of fixations has dropped to 5713 from the 11875 recorded observations for first fixation duration. The words with with four fixations (or more) tend to come from the lower frequencies available in the

Table 3.7: Interaction of Trial number and LastFix at the third fixation duration.

	Estimate	SE	*t* value
(Intercept)	5.39	0.03	165.70
Base Frequency	-0.02	0.003	-4.65
AffixPairTokenFreq	-0.01	0.002	-2.41
Trial Index	-0.01	0.01	-1.41
LastFixTRUE	-0.09	0.01	-8.22
Trial Index*LastFixTRUE	-0.05	0.01	-4.76

Table 3.8: Predictors for Third Fixation Duration, based on 9992 observations

experiment (mean Word Frequency of 13 vs. 46 for the entire experiment).

At the fourth fixation, the frequencies of the whole word, base and root are all contributing to processing speeds, with higher values of each being predictive of shorter fixation durations (Table 3.9). The mean fixation duration at the fourth fixation duration was 202 ms. Participants were much faster when this was the last fixation in their reading of the word, and this effect was more pronounced as participants got further into the experiment,

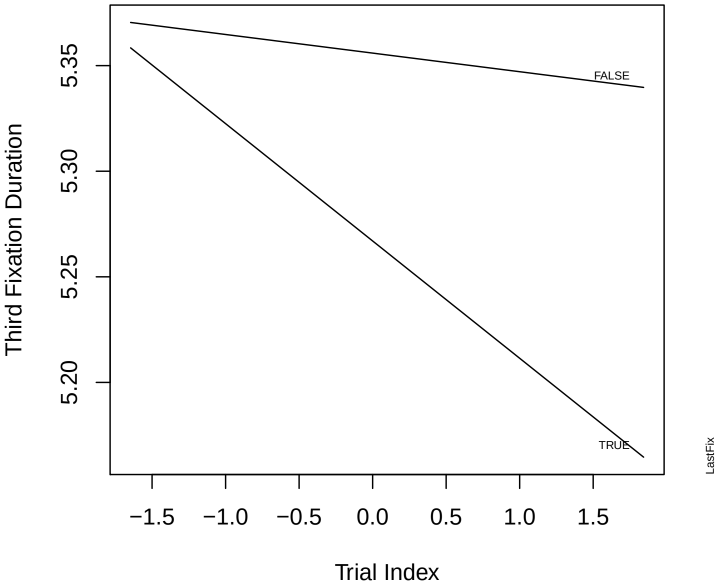

Figure 3.7: Interaction between Trial Index and LastFix at the third fixation duration.

as was observed for the third fixation.

What we observe at the fourth fixation duration is a breakdown of the morphemes making up the word. It is also interesting that we are only now observing effects of whole word frequency. The fourth fixation is relatively late, when compared to the studies on compounds recently completed by Kuperman et al. (2009). While it is possible that this late emergence of whole word frequency is a task-related effect, other lexical decision experiments with

Table 3.9: Predictors at Fourth Fixation Duration, based on 5713 observations of 613 items

	Estimate	SE	*t* value
(Intercept)	5.54	0.06	89.85
Suff1Freq	-0.01	0.01	-2.25
Word Frequency	-0.02	0.01	-3.78
Base Frequency	-0.01	0.005	-2.63
Root Frequency	-0.01	0.004	-3.73
Trial Index	-0.01	0.01	-0.61
LastFixTRUE	-0.37	0.01	-25.97
Trial Index*LastFixTRUE	-0.04	0.01	-3.16

similar experimental set-ups did not encounter this late effect. It may be that the relative obscurity of some of our target items paired with overall word length predisposed participants to decomposition, and so only later do we see effects of the whole word.

3.4 General Discussion

In this experiment, native speakers of English completed a standard lexical decision task while the movements of their eyes were tracked. Target items in this experiment were trimorphemic English words containing two suffixes. In our fixation durations, we observed that the frequency of the root does not appear as a significant predictor until the second fixation, whereas the base is a significant predictor of fixation duration in the first through fourth fixations. This finding may be a result of the composition of our target items. There may be a preference for the base to be processed as a unit before further morphological breakdown, due to the relative frequencies of the root versus the base. This particular result argues against obligatory decomposition upon first encountering a word (i.e., at first fixation). This result is particularly interestingly in light of the fact that we find effects of suffix facilitation for word processing at the first fixation, which indicates that morphological decomposition is possible at first fixation, but not obligatory across the board.

Broadly speaking, these results support previous eye-tracking studies where root frequency is facilitatory at the first fixation, and whole word frequency is facilitatory for the total fixation duration, but is not present at the first fixation (e.g., Niswander, Pollatsek & Rayner, 2000). Our results are slightly more complicated than that, as it is the base, or the first derivation, that is present as a facilitating effect at the first fixation, and not the root. We propose that this effect is caused by the relative frequencies of the base combinations (root+suffix1) when compared to the roots alone (c.f. Hay, 2001). The relative frequency of the base versus the root (base frequency/root frequency) was not, however, a significant predictor in and of itself. Our results are better modeled when each constituent is given weight during analysis.

In addition to the more common word frequency effects (root, base, word), we also observed suffix effects. At the first fixation, early effects of suffix productivity and family size were observed, and in later processing stages we see effects both of the affix pair as a whole and of the second suffix. We also found an effect of the discontinuous frequency for the first and last morpheme in our trimorphemic words. If we assume that frequency effects arise from mental representations, then disjunctive structures also require some level of representation or specification. Our data could be taken as a sign of human sensitivity to familiar structures, and the ability to make use of the information that sensitivity imparts (c.f. Bod, 2009, for an account of a dataoriented parsing model that takes previous training/experience into account). There appear to be benefits to working with more familiar structures, even when the morphemes comprising them are not continuous.

The eye-movement record also provides evidence of the cascading nature of the uptake of information. For example, information about the affix pair as a whole is not available at first fixation, but becomes so as reading progresses. As reading progresses, the affix pair frequency becomes a significant predictor.

We did not find evidence for the Complexity-Based Ordering hierarchy in results from this experiment. Future research must contend with one more presentation type before rejecting Complexity-Based Ordering as a psycholinguistically valid theory of affix ordering, namely, presentation inside of a sentence. Lexical decision experiments are highly constrained in their presentation of target material, and the task is not one normally performed outside the lab. Each word is presented without context, and in our experiment, participants benefited when they were able to identify morphemes during reading. Our nonwords contained two existing morphemes and one pseudomorpheme (a non-existing, phonologically possible form), to encourage complete viewing of each item before making a lexical decision.

Participants may have emphasized individual morphemes during reading, and therefore privileged decomposition, as recognition of each morpheme was required to reject nonwords. Such a setup may bias participants to be more attentive to easily accessible morpheme information (e.g., morpheme frequencies). A contrast between our results and those of Kuperman et al. (2008) for compounds is the lack of a word frequency effect in our data at the first fixation. It may be that, our words, being derived, were more compositional than those used in Kuperman et al.'s (2008) study. The meaning of the compounds used may have been less predictable than that generated through the use of derivational affixes. This, coupled with the nature of our nonword stimuli, may be responsible for the differences in our results.

We did find that lexical properties of both the first and second suffix were relevant in our models. Of the two, the properties of the second suffix were more reliably associated with processing times. Larger family sizes for the second suffix were associated with shorter fixation durations at the first fixation, an effect that was greater when the target word was short. Higher frequencies of Suffix2 were predictive of shorter total fixation durations as well as fewer total fixations. It may be the case that the second suffix, as it is at the end of the whole word, is more perceptually salient in reading that the embedded first suffix.

Our results add to a growing body of research that points us in the direction of a complex and interactive model of lexical processing. As with the results of Kuperman et al. (2008) and Winthers-Balling (2008), the existence of interactions between constituents within the same fixation in our data poses difficulties for models of lexical access that postulate 1) full-listing, 2) obligatory decomposition (before whole word recognition), and 3) dual-route models that do not allow for interaction between the two processing routes. Top-down, full-listing hypothesis cannot easily account for the data presented here, as whole-word frequency effects do not emerge until the fourth fixation, or relatively late in processing, with base effects emerging earlier during reading. Obligatory decomposition models fair slightly better, but still must contend with the lack of root frequency effect at the first fixation. Dual-route models that propose independent processing streams for whole word recognition and decomposition cannot account for the interactions between morphemes and morpheme combinations (words or otherwise). Within dual-route models, we can have balances (or imbalances) between parsing and whole-word retrieval, and within which we may see a complex picture that includes sensitivity for relative frequencies as well as a role for conjunct and disjunct frequencies. The interactions we observed suggest that processing routes are not neatly compartmentalized, but that they instead can interfere with each other. Our results are best suited to probabilistic theories of lexical processing, which can accommodate not only the frequency and family size effects of individual morphemes, but also the frequencies of constituent pairs. Given that our task was lexical decision, the role of greater context and possibly predictive behaviour of our tested variables is not available or accessible. The next step in an investigation of suffix pair processing should include a more naturalistic language task to more closely approximate the language of an average reader.

3.5 References

[1] Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5*(3), 436-461.

[2] Baayen, R. H, Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

[3] Baayen, R. H. & Schreuder, R. (1999). War and Peace: Morphemes and full forms in a non-interactive activation parallel dual-route model. *Brain and Language, 68*, 27-32.

[4] Bates, D. M., & Sarkar, D. (2007). The lme4 library. http://lib.stat.cmu.edu/R/CRAN/.

[5] 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*, 801-820.

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

[7] Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., & Hyönä, J. (2000). Affixal Homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language. *Cognition, 74*, B13-B25.

[8] Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science, 33*, 752-793.

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

[10] Burani, C.,& Thornton, A.M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In H. R. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 157-208). Berlin: Mouton de Gruyter.

[11] Cole, 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*, 1-13.

[12] Cole, P., Segui, J. & Taft, M. (1997). Words and morphemes as units for lexical access. *Journal of Memory and Language, 37*, 312-330.

[13] Fabb, N. (1988). English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory, 6*, 527-539.

[14] Giegerich, H. (1999). *Lexical strata in English: Morphological causes, phonological effects*. Cambridge: Cambridge University Press.

[15] Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review, 8*, 127-131.

[16] Hay, J. (2002). From Speech Perception to Morphology: Affix-ordering Revisited. *Language, 78*(3), 2002, 527-555.

[17] Hay, J. (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics , 39*(6), 1041-1070.

[18] Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory, 22*, 565-596.

[19] Hyönä, J., Bertram, R.& Pollastek, A. (2004). Are compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory and Cognition, 32*(4), 523-532.

[20] 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.

[21] Kiparsky, P. (1982). Lexical phonology and morphology. In I. Yang (Ed.), *Linguistics in the morning calm* (pp. 3 - 91). Seoul: Hanshin.

[22] Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*, 13-35.

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

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

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

[26] Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B. (Ed.), *Morphological Aspects of Language Processing: Cross-Linguistic Perspectives* (pp. 345-364). Hillsdale: Erlbaum.

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

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

[29] Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected words during reading. *Language and Cognitive Processes, 14*(4/5), 389-420.

[30] 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*, 685-702.

[31] Plag, I.& Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language, 85*, 106-149.

[32] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[33] Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology, 57A*(4), 745-765.

[34] Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition, 7*(4), 263-272.

[35] Taft, M. & Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

[36] Vannest, J. & Boland, J.E. (1999). Lexical Morphology and Lexical Access. *Brain and Language, 68*, 324-332.

[37] 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.

Chapter 4

Morphological processing of triconstituent words in sentence reading: Evidence from Eye-tracking and ERPs

Single-word recognition tasks such as lexical decision have been illuminating in the study of morphological processing, but cannot, by their very nature, speak to word processing during sentence reading. While not entirely natural in an experimental setting, sentence reading tasks more closely approximate the ways in which speakers use language on a daily basis. In the present study, we examine the processing of trimorphemic words in English as they are read in sentence context, through the use of both eye-tracking and event-related potentials.

Studies of single word processing have revealed by-now familiar effects of frequency, wherein items, typically words, with higher frequency show reduced reaction times in lexical decision and fixation durations in the eye-movement record. Faster response latencies and shorter fixation durations are taken to be indicative of faster processing. In the case of multimorphemic words, individual morpheme frequencies add complexity to lexical decision data. Root frequency, for instance, has been found to influence the processing of derived words in lexical decision (Taft, 1979; Burani & Caramazza, 1989). Beauvillain (1996) found similar results in an eye-tracking study. In this study, eye-movements were recorded during a semantic relatedness task where participants first read the target item, a French derived word, and then made a decision as to whether a second word that followed was semantically related to the first. Their results showed that at the first fixation duration, the frequency of the root of a derived suffixed word influenced reaction times, but that the frequency of the whole word did not. Higher frequency roots were responded to more quickly than those of lower frequency. Effects of whole word frequency appeared in subsequent fixations and in total gaze durations, with higher frequencies associated with shorter durations.

As in single word reading, lexical frequency has been found to affect reaction times and gaze durations during sentence reading. Niswander, Pollatsek and Rayner (2000) found facilitatory effects of root frequency at first fixation duration during sentence reading, with whole word frequency becoming facilitatory at the second fixation. This time course of effects suggests that, at the very least, the root is available and processed before the whole word is recognized, and mirrors results from single word processing studies. Similar results have been reported for Finnish compounds read in sentence context (Hyönä &Pollatsek, 1998; Pollatsek, Hyönä & Bertram, 2000). Bertram and Hyönä (2003) further investigated the role of word length while reading Finnish compounds. In one experiment, they manipulated the frequency of the first constituent while keeping whole word frequency stable and found that the first constituent frequency affected the length of the first fixation duration for long words (mean 12.8 letters) but not for short words (mean 7.7 letters). In a second experiment, they varied the frequency of the whole word, finding that whole word frequency was facilitatory for first fixation duration for short compounds, but that effects of whole word frequency only appeared later during reading for longer compounds. They interpreted their results as supportive of a dual-route model where the length of short compounds allowed for complete information extraction early in gaze fixations.

With respect to event-related potential research, much of the work related to morphological structure has focused on inflection in grammatical versus ungrammatical utterances (e.g., O'Rourke & Van Petten, 2011; Allen, Badecker & Osterhout, 2003; Friederici, Pfeifer & Hahne, 1993), with many studies relying on the detection of anomalies in the linguistic stimulus. The classical N400 and P600 effects are found to semantically anomalous and ungrammatical sentence elements, respectively (Kutas & Hillyard, 1983; Coulson, King, & Kutas, 1998). N400 effects have also been found to non-anomalous words in sentence context, but this is modulated by both frequency and by sentence position (Van Petten & Kutas, 1990). Van Petten and Kutas (1990) showed that lower frequency words generated larger N400s during reading, with this effect growing smaller as the amount of preceding sentence context grew larger. That is, frequency effects as realized through the ERP signal were strong at the beginning of a sentence, but weakened through the sentence as context was created. Other ERP studies have used the priming paradigm to provide evidence that the parsing of morphologically complex words is achieved very quickly during reading, on the order of 200 - 250 ms (Morris, Grainger, & Holcomb, 2008; see also de Vega, Urrutia, & Dominguez, 2010, for parsing at 200 ms for Spanish verb agreement), and that this is a structural segmentation that is effectively blind to the semantics of the words in question (Lavric, Clapp, & Rastle, 2007).

Of particular interest in this study is the processing of the suffix pair (e.g., -*fulness* in *hopefulness*) both as a whole and in its parts. In English, suffix ordering has presented an interesting challenge to morphologists, as the number of existing suffix combinations is much smaller than the possible number of combinations. For example, -*fulness* is well attested, but -*fulity* is not, even though -*ity* can be used to create nouns in the same way as -*ness*. Several hypotheses have been put forward to explain this discrepancy. The most commonly proposed fall into two broad camps: level ordering or selectional restrictions. Level ordering hypotheses hold, in general, that certain suffixes must occur outside of other suffixes because of phonological characteristics or language of origin (e.g., Giegerich, 1999). Selectional restrictions refer to specific properties of the affix that limit the bases to which it can be applied (e.g., Fabb, 1988; but see also Plag, 2004 for a base restriction account).

Complexity-Based Ordering, a recent theory of affix ordering, combines suffix-specific selectional restrictions with processing constraints to form a parsability hierarchy upon which suffixes can be ranked (Hay & Plag, 2004). Plag and Baayen (2009) used data from the English Lexicon Project (Balota et. al, 2007) to test predictions made by Complexity-Based Ordering. They examined lexical decision and naming latencies for bimorphemic words with a selection of English suffixes, and found that the relative rank of the suffix along the Complexity hierarchy influenced response latencies. Suffixes that were very low on hierarchy, and thus that were difficult to parse and largely unproductive, showed the fastest response latencies. The authors proposed that, for the lowest ranked suffixes, memory for stored items facilitated reaction times. The computation required to parse out suffixes higher on the hierarchy was costly, which was reflected in longer response latencies.

Plag and Baayen (2009) worked with single suffixed words. To extend this research, Teddiman and Baayen (this thesis) examined the processing of trimorphemic words in a lexical decision experiment with an eye-tracking component. While they did not find an effect of suffix rank, higher suffix family size facilitated processing during reading, as did higher base frequencies and higher frequencies of the affix pair as a whole. Root frequency was facilitatory at some fixations and in total fixation measures and lexical decision, but was not consistently significant.

The present study extends this research by adding sentence context to our eye-tracking study, as well as an ERP component. We focus on multiply derived trimorphemic words in a grammatical sentence context. The overall goals of this research are to 1) increase understanding of morphological processing in a more natural setting with greater ecological validity than is possible in lexical decision, and 2) to specifically test the explanatory value of Complexity-Based Ordering as a theory of suffix ordering in English.

4.1 Methods

In this experiment, eye-movements and electroencephalograms (EEGs) were simultaneously recorded during sentence reading. Our analysis begins with the onset of the target word, and does not include an analysis of parafoveal effects from the previous word.

4.1.1 Participants

Fifty-one participants took part in this experiment, of whom 39 produced usable data in both eye-tracking and event-related potentials. Participants were recruited from Introductory Linguistics courses at the University of Alberta via the Department of Linguistics participant pool. Students received one course credit per half hour for the completion of this experiment. Each experiment lasted a maximum of 1.5 hours.

4.1.2 Stimuli

The target items in this experiment were triconstituent words in English composed of a root and two suffixes. These words were taken from a larger set of words examined in an eye-tracking lexical decision experiment (Teddiman & Baayen, this thesis). Two-hundred and fifty target words were randomly selected from the original base+suffix+suffix list. Sentence contexts for each target word were extracted from the Corpus of Contemporary American English (COCA), a 400-million word corpus of current American English (Davies, 2008-). Example sentences include *We struck up an* ***acquaintanceship*** *over the phone* (target word: *acquaintanceship*) and *He left the team because he was upset with* ***cliquishness*** *among his teammates* (target word: *cliquishness*).

A set of lexical statistics were collected for each target word using the CELEX Lexical Database (Baayen, Piepenbrock & Gulikers, 1995). These included the frequency of the whole word, the frequency of the base, the frequency of the root alone, root family size, and word length. Suffix-related predictors were also collected, including the frequency of each suffix, the type frequency (or family size) of each suffix, the frequency of the affix pair as a whole, and the rank of each suffix along the Complexity-Based Ordering hierarchy. Values for Complexity-Based Ordering rank (CO-Rank) were derived from Plag and Baayen (2009). Note that suffixes lower on the hierarchy as associated with high numerical values. To compensate for collinearity among lexical frequencies, the base and whole word frequencies were residualized on the root frequency. All frequency variables were log transformed and centered. Table 4.1 provides a list of our predictor variables along with their untransformed numerical ranges and means.

4.1.3 Equipment and Procedure

Eye-tracking was conducted using the EyeLink 1000 (SR Research, Mississauga, Ontario). The desktop setup consisted of a chin-bar attached to a

Table 4.1: Predictor values

Predictor	Min. Value	Max. Value	Mean
Root Frequency	0	20675	1110
Base Frequency	0	3862	268
Word Frequency	0	2224	47
Family size	1	46	8
Word Length	6	17	11
Suffix1 Frequency	0	119274	18046.48
Suffix2 Frequency	7	155741	78559.91
Suffix1 CO-Rank	1641	7434	3742.86
Suffix2 CO-Rank	675	6952	2065.8
Suffix1 FamSize	1	601	191.49
Suffix2 FamSize	1	1190	695.76
Affix Pair Frequency	0	8305	1585

table with a camera set up 60cm away, in front of the computer monitor. At this distance, the EyeLink 1000 has a gaze tracking range of 32 degrees in the horizontal plane. Eye movements were captured at a 1000 Hz refresh rate, and were tracked using both pupil recognition and corneal reflections. EEGs were recorded using the BioSemi ActiveTwo EEG system with a 32 AG/AgCl electrode cap. Recordings were made from all active electrodes (Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, PO4, O1, Oz, O2). Activity from eye-movements was recorded by placing electrodes at the canthi of both left and right eyes, and by placing electrodes above and below the left eye. Two reference electrodes were also placed at the mastoids (right and left).

The experiment was programmed in Experiment Builder (SR Research). The presentation software ran on Windows XP. A three-computer set-up was necessary to run the experiment, with one computer running the experiment and recording eye-tracking data, another controlling the eye-tracker, and a third recording the ERP data.

During an experimental session, participants were first outfitted with the ERP equipment, consisting of an appropriately sized electrode cap (based on head circumference), electro-conductive gel, and electrodes. They were then seated inside of a sound booth, and asked to use a chin-rest to maintain their head position during the experiment. When the participant was comfortably seated, the BioSemi system was turned on, and the DC Offset levels were observed. If any electrodes showed poor connectivity, measures to improve connectivity were taken (e.g., adding gel, moving hair). Setup time ranged between 20 - 40 minutes.

Once the equipment was set up, participants began the sentence reading task. The eye-tracker was calibrated using a 9 point calibration before the experiment began. During calibration, participants were asked to follow a dot across the screen to nine different points. The system was recalibrated every ten trials afterwards. Before each sentence, participants saw a fixation point ($*$) on the left side of the screen. This fixation point also acted as a drift correct for each trial. Sentences were presented to participants one at a time, with each sentence appearing the middle of the screen over a single line. Sentences were between seven to fifteen words long, and were presented in size 16 Courier New font. Participants advanced through the experiment at their own pace by pressing a button on a Microsoft Sidewinder game controller when they had finished reading a sentence. They were informed that they would be asked simple comprehension questions about some sentences, in order to encourage attentive reading during the task. The experiment lasted between 45 - 60 minutes, not including setup time.

Data for the ERP recordings were sampled at 8,102 Hz. In order to be of manageable size for analysis, ERP data was downsampled to 128 Hz. The signal was re-referenced to the average of the mastoid references, and a bandpass filter was applied from 0.5 to 30 Hz. This pre-analysis work with the ERP data was completed in Brain Vision Analyzer (1.05). Signal disruption from eye-movements and blinks were corrected in R (R Development Core Team, 2011), through the use of the *icaOcularCorrection* package (Tremblay, 2010). The EEG signal was not averaged prior to analysis.

4.2 Results and Discussion

Eye-tracking and ERP results were analyzed with generalized additive models (GAMs) in R using the mgcv package (Wood, 2011). In particular for ERP data, the use of GAMs allows us to examine the effects of numerical predictors, such as word frequency, over time while allowing for non-linearities. Subject (participant), Root, and Suff1 and Suff2 were coded as random effects during the modeling process for eye-tracking data. We are not, in this particular study, interested in the processing of morphological violations, and instead wish to address a set of continuous frequency predictors in natural sentence reading. This goal does not lend itself well to the same statistical methodologies as have frequently been used in the past to study event related potentials to linguistic stimuli. We use generalized additive models to take into account predictors that are useful in modeling our results continuously through our target items.

4.2.1 Eye-tracking

4.2.2 Fixation Count

On average, each target item received 3 fixations (mean: 2.99), with a range of 1 - 23 fixations. Higher frequencies of the base and of the whole word resulted in fewer fixations on target items. Words that were longer received more fixations. Words that occurred later in the sentence (variable: TargetWordPosition) also received fewer fixations. Table 4.2 lists each predictor that reached significance along with its associated AIC value. Higher AIC values represent a greater contribution to the explanatory power of the model.

Predictors that were modeled by smooths are given in the second section of the table. The Trial number was the most powerful predictor for number of fixations. Participants also showed relatively more fixations at the beginning of the experiment, with this number dropping and then leveling off near the end (Figure 4.1). This pattern likely reflects a practice effect, where participants became better at the experiment as they continued through it. Among lexical predictors, the Base Frequency was the most powerful, followed by the Word Frequency and the Affix Pair Frequency. The model also included Subject and Root as random factors.

The CO-Rank of the second suffix was a significant predictor for number of fixations, with suffixes lower on the hierarchy, i.e. those that are less parsable, being associated with a higher number of fixations.

4.2.3 Total Fixation Duration

Turning to the total fixation duration on our target items, we find that higher frequencies of the base and whole word give rise to shorter fixation times (Table 4.3). Likewise, when the target word appeared later in a sentence, it had a shorter overall fixation time. Longer words elicited longer total fixation durations. Average total fixation duration was 808 ms, with a range of 82

Table 4.2: Predictors for total number of fixations

Predictor	Estimate	SE	t value	p	AIC
Base Frequency	-0.08	0.05	26.26	0.0001	10.35
Word Frequency	-0.07	0.01	-5.75	0.0001	5.2
Suffix2 CO-Rank	0.03	0.01	2.44	0.0146	1.17
Word Length	0.16	0.01	11.07	0.0001	19.46
Target Word Position	-0.02	0.004	-4.53	0.0001	7.15

Predictor	edf	F	p	AIC
Affix Pair Frequency	3.52	14.28	0.0059	5.02
Trial	2.27	88.91	0.0001	129.22

ms to 9004 ms. When the second suffix was less parsable (lower on the CO-Rank hierarchy), it was associated with a longer total fixation duration. Participants were slower at the beginning of the experiment than at the end, although this effect leveled off at the end of the experiment. Again, this is evidence of a practice effect, where participants became more comfortable with the experimental task as the experiment progressed. Finally, the frequency of the affix pair also influenced total fixation duration. In general, higher

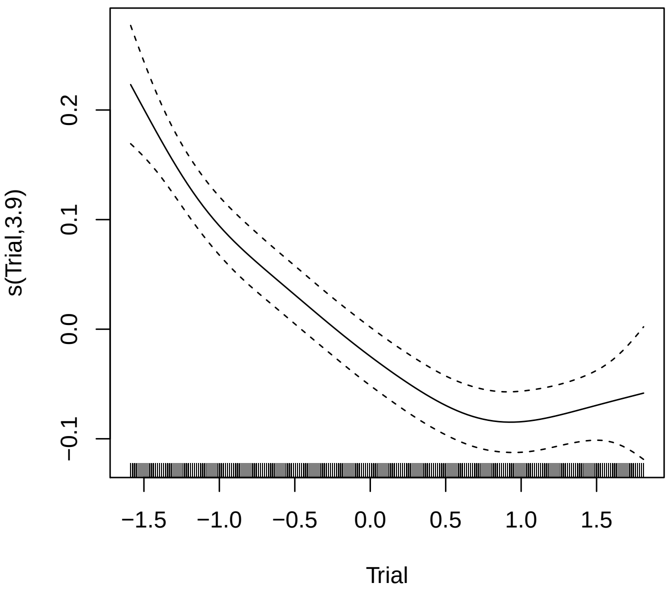

Figure 4.1: Trial effects for number of fixations.

frequencies resulted in shorter durations, with longer durations for lower frequencies as as a main trend (Figure 4.2). The behaviour of the Affix Pair frequency for total fixation duration on a word was similar to its behaviour for total number of fixations.

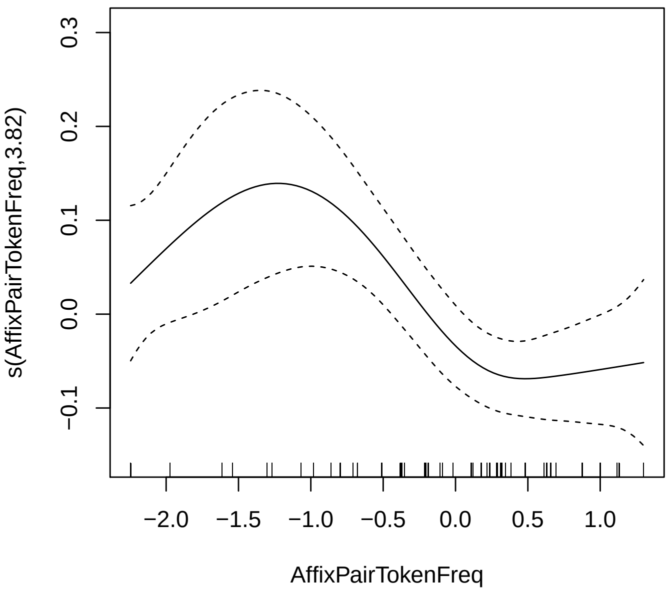

Figure 4.2: Affix Pair Frequency effects for total total fixation duration on a word.

4.2.4 Gaze Durations

4.2.5 First Fixation Duration

Before results from the first fixation were analyzed, outliers that were shorter than 81.45 ms and longer than 1096.63 ms were removed, resulting in a loss of 166 items, or 3% of the data. The mean length of the first fixation was 258 ms.

Table 4.3: Predictors for total fixation duration

Predictor	Estimate	SE	*t* value	*p*	AIC
Base Frequency	-.011	0.02	-6.92	.0001	10.94
Word Frequency	-0.09	0.01	-6.36	0.0001	5.45
Suffix2 CO-Rank	0.06	0.02	3.69	0.0002	5.24
Word Length	0.15	0.02	8.93	0.0001	10.01
Target Word Position	-0.02	0.005	-3.85	0.0001	4.72

Predictor	edf	F	*p*	AIC
Affix Pair Frequency	3.83	3.88	0.0035	6.81
Trial	2.67	35.27	0.0001	106.53

First fixation duration was influenced by the frequency of the root, the frequency of the base, word length, the trial number and the rank of the second suffix along the Complexity-Based Ordering hierarchy. The base frequency was facilitatory for fixation duration, with higher frequencies being associated with shorter fixations. When a target item was presented later in a sentence, the first fixation duration was longer than when it appeared near the beginning of the sentence. An interaction between the CO-Rank of suffix2 and trial number was modeled using a tensor product (Figure 4.3). Later in the experiment, lower CO-Rank of suffix2 gave rise to longer fixation durations, although this did not reach significance in a model without an interaction (p = 0.06).

When the frequency of the Root was low, first fixation duration was in general longer than when the root frequency was high. However, this effect leveled off for the highest frequencies (Figure 4.4). In general higher frequencies of the base combination were associated with shorter first fixations. This effect was markedly stronger for words with the highest base frequencies.

Within our data, the base frequency was the most powerful predictor. Table 4.4 provides the AIC values associated with the addition of each new predictor in a stepwise model comparison. The model also included Subject (F = 15.50, p<0.0001) as a random-effect factor. Including the Root as a random-effect factor, however, did not lead to a significant improvement in the model fit.

4.2.6 Second Fixation Duration

Before results from the second fixation were analyzed, durations shorter than 55 ms and longer than 1097 were removed as outliers, resulting in a loss of

Table 4.4: Significant predictors at First Fixation Duration. Listed with each predictor is the AIC gain from adding a predictor to the model. Larger AIC values correspond to greater contribution in explaining the data. *edf* refers to the estimated degrees of freedom.

Predictor	Estimate	SE	t value	p	AIC
Target Word Position	0.01	0.002	3.68	0.0002	10.39

Predictor	edf	F	p	AIC
Root Frequency	6.35	4.23	<0.0001	12.63
Base Frequency	7.76	7.11	<0.0001	45.24
Trial*Suffix2 CO-Rank	11.29	2.01	0.0161	8.40

73 items, or 2.3% of the data. The length of the second fixation duration ranged from 56 ms to 986 ms (mean 248 ms).

The duration of the second fixation was decreased by high base frequencies, high whole word frequencies, and the frequency of the affix pair as a whole (e.g., the -*fulness* in *hopefulness*). The rank of the second suffix was inhibitory, with suffixes of lower rank associated with longer gaze durations. The Subject (F = 6.206, p<0.0001) was useful as a random-effect factor in

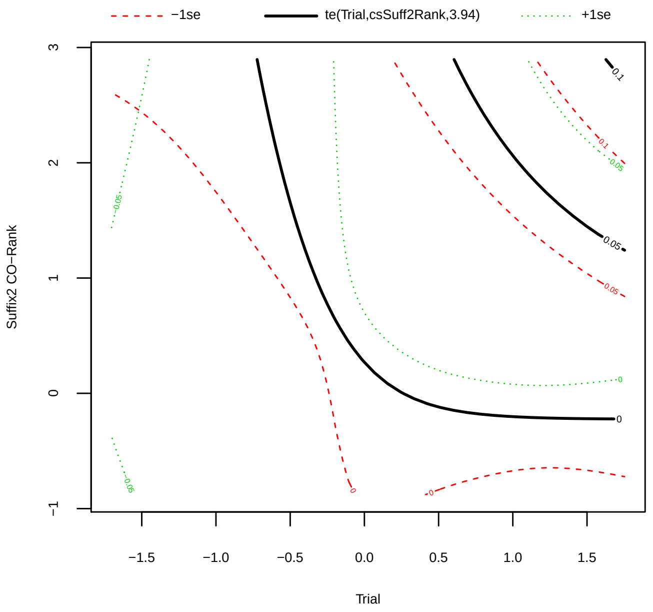

Figure 4.3: Contour graph of the interaction between CO-Rank and Trial at the first fixation duration. Where the lines are closer together, the effect is "steeper" along the surface. The dotted lines are confidence lines for the contour lines. In this graph we see greater activity in the upper right corner, which corresponds to trials later in the experiment and to suffixes with a lower CO-Rank.

the model. The Root Frequency was again nonlinear, following the same pattern as observed for the first fixation. Table 4.5 provides a breakdown of predictors and the usefulness of the significant predictors in the GAM model

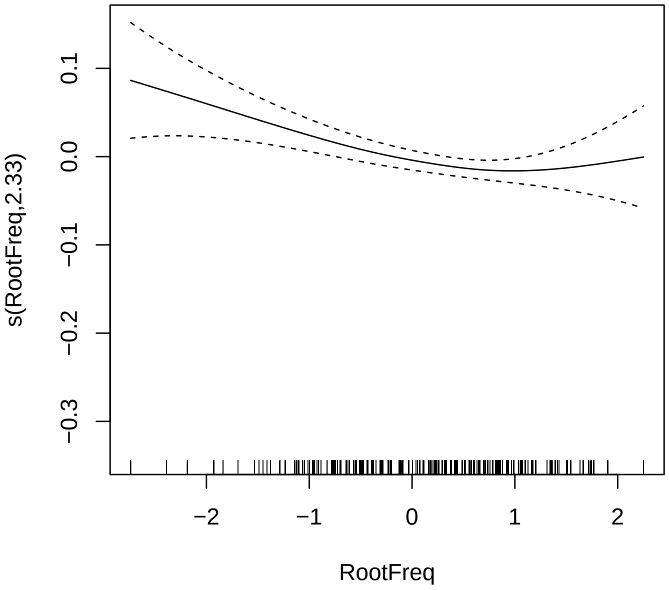

Figure 4.4: Root Frequency at First Fixation.

by the AIC gains for each. Although not large, the addition of each variable resulted in a better model. At the second fixation, whole word frequency has become a more powerful predictor than base frequency, although base frequency is still a significant predictor in its own right. The rank of the second suffix also becomes more salient at the second fixation. Recall that in our analysis, the higher the numerical value of the CO-Rank, the more difficult it is to parse. Since rank is becoming more strongly inhibitory at the second

Table 4.5: Predictors at Second Fixation Duration

Predictor	Estimate	SE	t value	p	AIC
Base Frequency	-0.32	0.01	-3.52	0.0004	4.42
Word Frequency	-0.02	0.01	-2.74	0.0061	4.71
Affix Pair Frequency	-0.02	0.01	-2.42	0.015	3.59
Suffix2 CO-Rank	0.02	0.01	2.55	0.0108	3.35

Predictor	edf	F	p	AIC
Root Frequency	1.72	3.46	0.0311	3.43

fixation, this result implies that as word reading continues, processing times increase for those words with less parsable final suffixes.

4.2.7 Regression Duration

In addition to examining the characteristics of individual fixations, we were also interested in looking at the regressions participants made during reading. The regression duration was defined as the amount of time spent on a word after a regressive saccade has been made. The Regression duration tells a different story than the first and second fixation durations. There was no

Table 4.6: Predictors for regression duration

Predictor	Estimate	SE	t value	p	AIC
Word Length	22.22	4.11	6.14	<0.0001	33.29
Target Word Position	-5.16	1.39	-3.71	0.0002	22.05
Predictor	edf	F	p	AIC	
Base Frequency	2.85	10.97	<0.0001	33.87	
Suffix1 Frequency	4.71	2.83	0.0106	10.42	

effect of root frequency on the regression duration, but there was an effect of both base and suffix1 frequency (Table 4.6). Subject ($F = 4.48$, $p < 0.0001$) was included in the model as a random-effect factor. This is the only place where we see an effect of the first suffix alone (i.e., not as a part of a larger morphemic unit). That this suffix frequency emerges after first pass reading suggests that when participants looked back within a word, they may have been engaged in more decompositional behaviour. We can hypothesize that the rarity of some of our target items was cause for greater reading effort by some of our participants, and unfamiliar words encountered in the text might require further processing to be fully understood after the whole sentence had been viewed.

4.2.8 Eye-tracking Summary

Results from eye-tracking data provide evidence for representations of individual morphemes and morpheme combinations, in addition to effects related to the whole word. Throughout the data, the base frequency contributed to shorter fixations and lower fixation counts. This was true even at the first fixation, concurrent with separate facilitation related to the frequency of the root alone. Word frequency is not a significant predictor at the first fixation, but becomes facilitatory at the second fixation duration, and remains significant for total fixation duration and total fixation count. Also emerging at the second fixation duration is a facilitatory effect of the Affix Pair frequency. Root frequency plays a role at our individual fixation durations, but its effects are not felt for total fixation duration or fixation count. Finally, we see an inhibitory effect of the CO-Rank of the second suffix clearly from the second fixation duration and total fixation count and duration. This indicates a role later in processing, although there are signs of its effect as early as the first fixation.

4.2.9 Event-Related Potentials

Our analysis of ERP data was built using results from eye-tracking as a guide, and only those predictors that were significant during the eye-tracking analysis were used in the ERP analysis. The eye-tracking record was used to locate the point in time at which the target word was first fixated upon. First fixation was the event from which we traced event-related potentials.

Analysis of the ERP data was carried out with R and the mgcv package (Wood, 2006). The ERP signal was decomposed into a series of additive components, each of which is a function of time. The first component is the grand average waveform. The second component comprises, for each subject, the temporal adjustment waveform function, which when added to the grand average waveform results in subject specific average waveforms. Functionally, these adjustment waveforms capture subject specific variability just as by-subject random intercepts and slopes in mixed effect models capture subject specific variability. The third component comprises temporal adjustment waveforms for the items, functionally comparable to by-item random intercepts and slopes. The fourth component captures changes in the microvoltage over the course of the experiment (compare the *Trial* variable in the eye-movement analysis). The remaining components in the model capture non-linear interactions of *Time* by lexical distributional predictors. These interactions are conceptualized as wiggly microvoltage surfaces and are modeled with the help of tensor products.. For technical and computational reasons, the ERP signal for a given target word was analyzed for four overlapping epochs: 0 - 300ms, 200 - 500ms, 400 - 700ms, 600 - 900ms.

The eye-tracking data clearly showed that frequency effects are present during normal sentence reading. The predictors that we analyzed in our ERP data were based on our eye-tracking results. They were Root Frequency, Base Frequency, Word Frequency, Affix Pair Frequency, Word Length, and Suffix2 CO-Rank. Figure 4.5 presents typical activation patterns for each predictor at a single electrode, primarily in the 600 - 900 ms time slice following the first fixation on the target word; 400 - 700 ms for Affix Pair Frequency. Where we discuss locations of activation, we refer to the location of the electrodes and their placement on the scalp, and not to neuroanatomical regions.

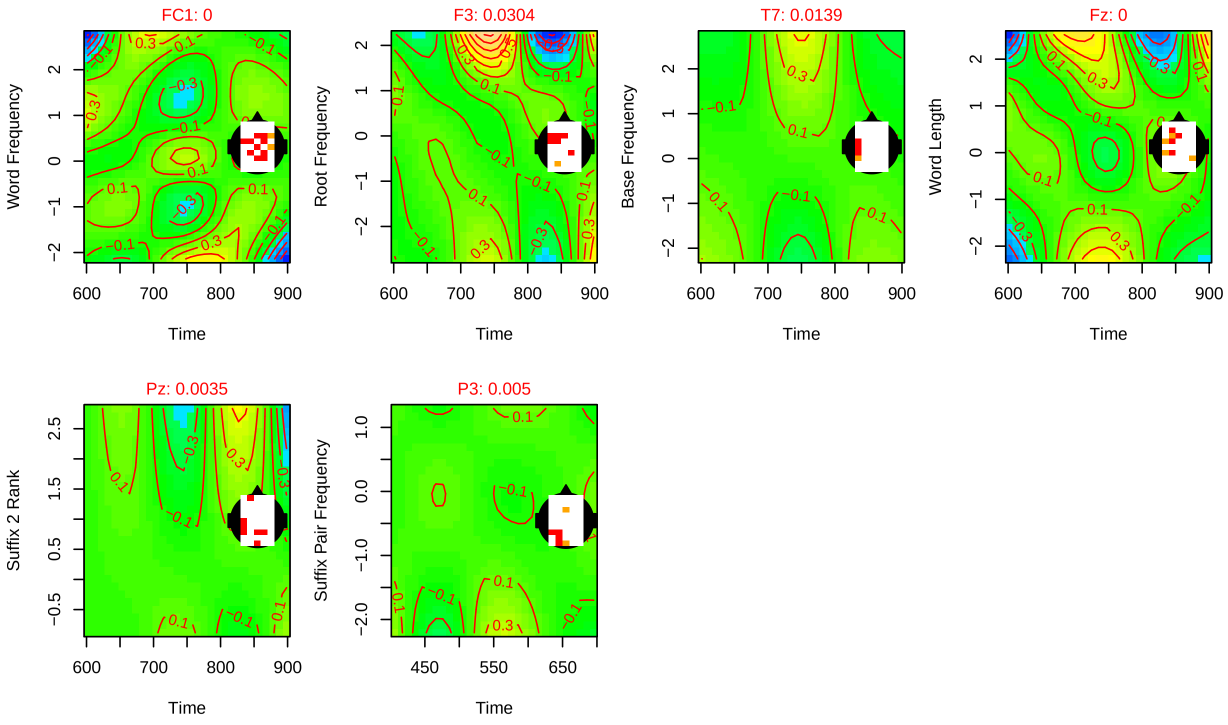

Figure 4.5: Topographic maps of brain potentials to significant predictors from the eye-tracking component of this study. Panel 1: Word Frequency, Panel 2: Root Frequency, Panel 3: Base Frequency, Panel 4: Word Length, Panel 5: Suffix 2 CO-Rank, Panel 6: Affix Pair Frequency

Word Frequency Panel 1 of Figure 4.5 shows the effects of whole word frequency. Time is given on the x-axis of the plot, with scaled and centered word frequency on the y-axis. On the x-axis of the plot, time is given in milliseconds. The z-axis, represented by both contour lines and by shades of gray, corresponds to the voltage recorded at each electrode in microvolts. Light and dark values of gray correspond to positive and negative oscillations. If viewed in colour, green represents voltages hovering around 0. Yellow, orange, and red correspond to positive values, with red being the highest, while varieties of blue correspond to negative voltages. Inset into each graph is a representation of a head. Each dark gray (in colour: red) square represents an electrode at which ERP results were significant. Light gray (in colour: orange) squares represents marginal significance.

Activity related to word frequency was recorded from frontal, central, and central parietal locations. There are low amplitude hints of an effect of word frequency in the 200 - 500 ms time window at two right parietal electrodes (P4, P8). Word frequency shows a clear and topographically widespread sinusoid frequency oscillations in the 600 - 900 ms time window. These are theta oscillations (5Hz), which have been associated with semantic and lexical retrieval (Bastiaansen et al., 2005), so this is not an unexpected result when examining lexical processing during sentence reading. We also see a phase shift in oscillations dependent on frequency, where they begin earlier for higher frequencies. Oscillations for high word frequency begin reliably at 600 ms post onset (i.e., after word reading has begun on the target word). Effects for high word frequency appear stronger earlier, and fade out at about 775 ms. For low frequency words, weak oscillations appear slightly later, becoming stronger at around 725 ms (at 3-5 Hz) and remaining strong until the end of our analysis period (900 ms). Greater amplitudes, as seen at the edges of our frequencies (highest and lowest), may be taken to show that the lexical system is sensitive to the probabilities with which an item will occur. These results tie in well with observations that high frequency facilitates processing, as is observed in reaction time studies. The effects of word frequency are gradient, even here.

Root Frequency. Root frequency effects are shown in Panel 2 of Figure 4.5. In our data, Root frequency effects are only well-attested in the final time window, where we observe a convex-concave pattern for low and high frequencies in the theta range of oscillations, primarily over left frontal regions. High frequency roots showed greater engagement than low frequency, particularly between 700 - 850 ms. In our eye-tracking data, high root frequency was associated with shorter first and second fixation durations, and the ERP results follow that same pattern. We might extend this observation to encompass the idea that higher root frequencies relate to more recognizable roots, and so the greater engagement in ERPs reflect greater activation activity in the lexical system.

Base Frequency. Typical results for Base Frequency (*hopeful* in *hopefully*) are shown in Figure 4.5, Panel 3. Most of the recorded ERP activity is topographically restricted to sites in the left hemisphere (FC5, T7, CP5). Oscillations for both high and low frequencies begin at approximately 600 ms, with higher frequencies receiving more pronounced activity beginning at 700 ms. Again, there are signs here of a probability distribution effect, where low probability events (i.e., not mid-range, ”normal” values) require more cognitive effort.

Word Length. Typical effects of Word Length are visualized in Panel 4 of Figure 4.5. Electrodes recording relevant behaviour were primarily located over the left hemisphere, with more activity towards the frontal electrodes. We again see a convex-concave pattern for high and low word lengths, indicating that values nearer to the extremes engage the processing system to a greater degree. There are hints of this pattern in the 400 - 700 time window.

Shorter words show oscillations starting at 550 ms (not shown) and taper off at 850 ms. Oscillations for longer words begin at approximately 600 ms and continue until the end of our time window (900 ms). At intermediate word lengths, we also see smaller oscillations that end at our 900 ms mark. Greater word length is associated with longer reaction times both in our data and elsewhere.

Suffix2 CO-Rank. Panel 5 of Figure 4.5 shows typical effects of Suffix2 CO-Rank. There is a very small effect of rank in the 0 - 300 ms time window at two frontal sites (Fz, FC1), which might correspond to early rank effects that we observe in the eye-movements. In the final, 600 - 900 time window, we see theta oscillations for lower ranked suffixes primarily over left parietal and occipital regions that start at about 625 ms and remain until the end of the window. At 700 ms, the range of lower ranked suffixes included in this behaviour increases. Greater activity for lower CO-Ranked suffixes suggests that the system is engaged in efforts to processes less parsable items. A low Suffix2 CO-Rank implies a low Suffix1 CO-Rank, as the second suffix is usually higher on the hierarchy than the first for most words. Our results indicate that having two unproductive suffixes next to each other in a word makes for more difficult processing.

Most of the recorded is located in the left hemisphere, scattered over frontal and parietal regions. Similar patterns of activation are visible over different sites, with lower CO-Rank for the second suffix showing greater activity. Figure 4.6 shows electrode sites on the 32 electrode cap for the 600 - 900 window. Time is on the x-axis and scaled CO-Rank is on the y-axis (recall that suffixes of low rank in the hierarchy are represented with higher numerical values in our analysis). Voltage (in microvolts) is represented on the z-axis through the use of contour lines. When the lines are closer together, the surface is steeper. Directionality is given by the numbers inset into the contour lines. Each rectangle represents an electrode site on the scalp. At the top of each rectangle, the name of the electrode is given, along with a *p*-value. When activity is significant, the electrode label is dark. The top of the figure corresponds to the front of the head, with left and right sides corresponding to left and right scalp positions. BE, in the bottom right corner, records electrical potentials related to eye-movements. There is some overlap in location with base frequency.

Affix Pair Frequency. Effects of affix pair frequency are shown in Panel 6 of Figure 4.5. Most of the recorded electrical activity was in response to low frequency pairs, beginning at 450 ms and continuing through to 700

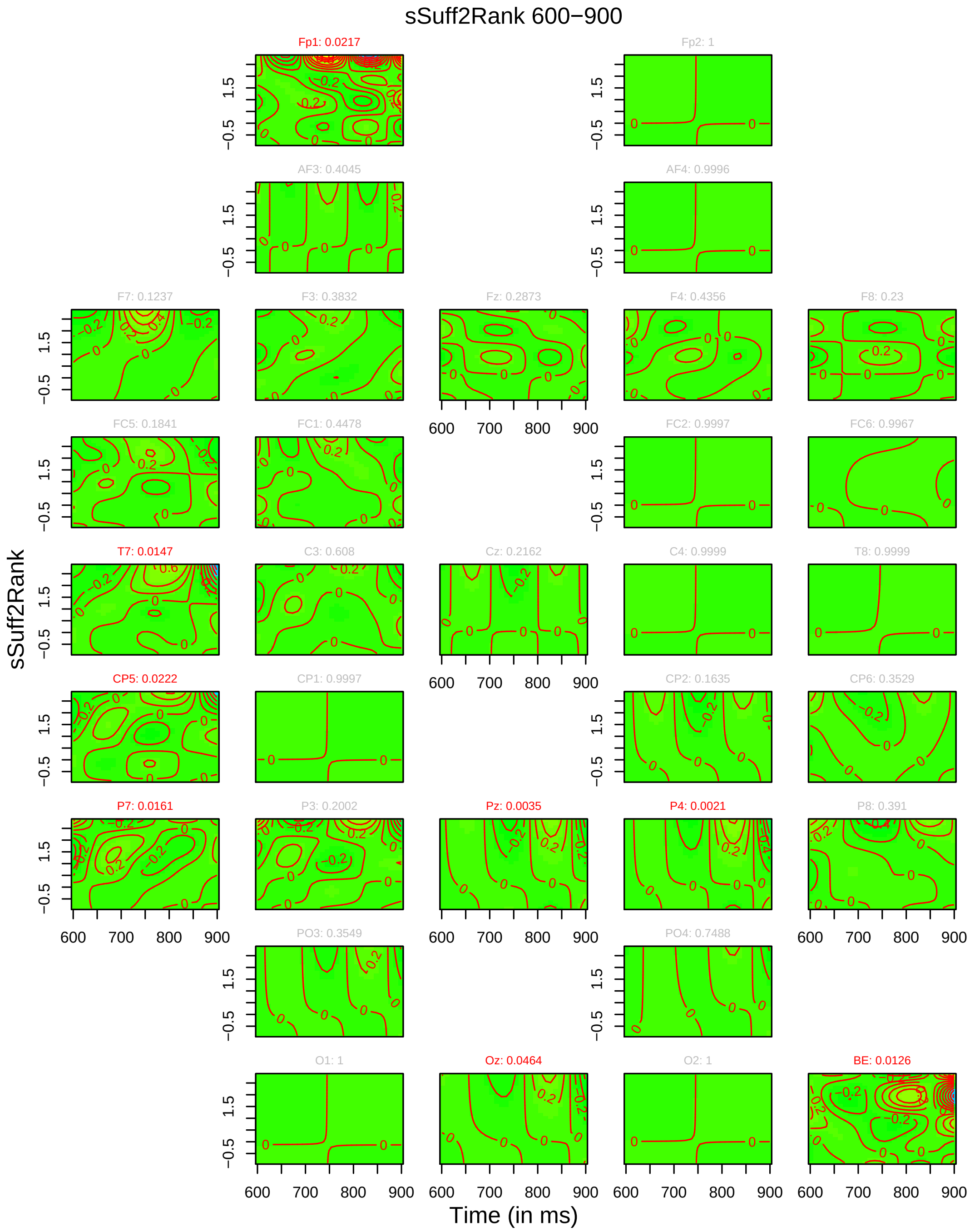

Figure 4.6: Rank effects of the second suffix in the 600 - 900 ms time window.

ms, from primarily left parietal regions. There were minimal oscillations to intermediate and high frequency forms, both starting at 500 ms and ending close to 700 ms.

This frequency does not match that of the other predictors reported on here, where higher frequency items showed oscillatory behaviour in the brain signal earlier than words with lower frequency. Here, we see that lower frequency pairs require more cognitive effort to process.

4.3 General Discussion

This joint eye-tracking and event-related potential experiment examined properties of gaze during the reading of and neural reactions to multimorphemic words in English, when presented in a sentence context. Beginning with our eye-tracking results, we see an unfolding story of lexical access for our particular target items. During the first fixation, we see effects of both the root and the base, but not of the whole word frequency. Word frequency becomes predictive at the second fixation and onward. The majority of our target words required at least two fixations due to their length, so we hypothesize that word frequency is not becoming available until the majority (if not all) of the word has been viewed. These results echo those of Niswander et al. (2000) for the reading of derived words in sentence context. Likewise, the frequency of the affix pair was a significant predictor at the second fixation, when it is likely to have been in the visual field, and in measures of total fixations, but not earlier during reading. The frequency of the root was a significant predictor for the first and second fixation durations, but not the total fixation duration on a word or total number of fixations. Root activity during early fixations speaks to initial word access that is effectively "drowned out" in measures that capture the entire time-course of a word's processing. The frequency of the base was one of the most powerful predictors in our experiment, and was present concurrently with effects of root frequency, word frequency and Suffix2 CO-Rank in both eye-tracking and in the evoked potentials. Our eye-tracking results clearly indicate that individual morphemes in our target items are useful in processing before the whole word has been identified. As the effects were present at the same time and across multiple measures, this provides evidence for a processing system that is constantly updating input and using the information available to access words as it becomes available.

Perhaps the most interesting result is that we found significant effects of the suffix rank of the second suffix. When the second suffix was lower on the CO-Rank hierarchy, that is, when it is more difficult to parse, fixation durations were increased. Lower ranked suffixes occurring in the second suffix position appear to draw more resources than more highly ranked suffixes. This effect also interacted with trial, becoming larger as the experiment went on. That this only happens with the second suffix suggests that, by position, participants may expect more parsable items, and that the inclusion of lower ranked suffixes requires more effort to process. As they proceed through the experiment, they may become more attuned to the type of reading the task requires and more familiar with the inclusion of longer words and their parsing in general, and so the apparent difficulty associated with low ranked suffixes becomes more pronounced. CO-Rank is unambiguously a significant predictor in eye-tracking for our target items when they are read in sentence context. This finding is contrary to previous word (Teddiman &Baayen, this thesis), where no effects of CO-Rank were found in a lexical decision task. The effect of CO-Rank appears only to emerge in the more natural sentence reading task, rather than in lexical decision. In our sentence presentation, context allows readers to generate expectation about what will follow, including words and morphemes (e.g., a word ending with the morpheme *-ize* will occur as a verb in a sentence). Lexical decision tasks lack this greater context. Furthermore, lexical decision tasks may emphasize particular processing routes (e.g., decompositional), which may not be as strongly preferred in ordinary reading.

Although we find CO-Rank effects for the second, or outermost, suffix, we do not see any effect of this predictor for the first suffix. When the second suffix is of low rank, the first suffix will also be of low rank, because higher ranked suffixes occur outside of lower ranked suffixes. This means that in words with a second suffix that has a low CO-Rank, the affix pair is made up of a sequence of suffixes that are both difficult to parse, and it is here that we see the greatest cognitive effort, as reflected in fixation durations and the ERP data.

In our ERP data, we hoped to find theta oscillations, as they have been previously reported to reflect semantic retrieval (Bastiaansen et al., 2005). We found theta oscillations for our lexical predictors, most strongly for Word Frequency and Root Frequency, primarily in the last time window (600 - 900 ms). The effects most clearly visible in the ERP are those for Word Frequency, Base Frequency, Word Length, and Suffix2 CO-Rank. These lexical effects emerge in the ERP data later than in the eye-tracking data. The time course of our results differs from most ERP word recognition tasks in that we do not observe effects of frequency as early as 100 ms after stimulus onset (e.g., Sereno, Rayner & Posner, 1998) or between 200 - 250 ms (Morris et al. 2008, de Vega et al., 2010), although our pattern of simultaneous activation for different lexical predictors in later stages of processing is similar to previous findings (e.g., Hauk et al., 2006). There are a few critical differences in our methodology that may influence this discrepancy; namely, our participants read sentences in a normal left-to-right fashion, with a single sentence presented for each trial. This is contrasted with the Rapid Serial Visual Presentation method that is typically used during ERP studies of sentence reading. Our target items were also, on average, much longer than those used in previous studies (average 10 letters vs. 4-6). Given that our participants read each sentence in a left-to-right fashion, with the entire sentence available for each trial, it is entirely likely that some processing for previous parts of the sentence was still ongoing when participants fixated on target words. A delay in the processing of words that are fixated upon during normal sentence reading, relative to processing during single word recognition or during RSVP, may reflect ongoing cognitive work as words are being read quickly in a grammatical sentence. We have a few hints in our data of earlier

ERP effects, but they are not strong or topographically well distributed. It is likely that the ERP signal is overlaid with other processes that do not have time to resolve in ordinary reading between one word to the next. The noise generated from other processes may obscure early lexical effects that are present in reading. However, given that this presentation method for ERP experiments is in its infancy, more research using this technique will be necessary before solid conclusions can be made about lexical processing during normal reading. For the moment, we can say that for regular sentence reading, the eye-movement record is more sensitive to early lexical effects.

4.4 References

[1] Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes 18*(4), 405-430.

[2] Baayen, R. H, Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

[3] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

[4] Bastiaansen, M., Van Der Linden, M., Ter Keurs, M., Dijkstra, T., &Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience, 17*, 530-541.

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

[6] Bates, D. M., & Sarkar, D. (2007). The lme4 library. http://lib.stat.cmu.edu/R/CRAN/.

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

[8] Coulson, S., King, J. W., & Kutas, M. (1998). Expect the Unexpected: Event-related Brain Response to Morpholsyntactic Violations. *Language and Cognitive Processes, 13*(1), 21-58.

[9] Davies, Mark. (2008-). *The Corpus of Contemporary American English: 425 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/.

[10] Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research 1*, 183-192.

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

[12] Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics, 39*(6), 1041-1070.

[13] Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory, 22*, 565-596.

[14] Hyönä, J. & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1612-1627.

[15] Kutas, M. & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition 11*(5), 539-550.

[16] 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.

[17] 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.

[18] Munte, T.F., Say, T., Clahsen, H., Schiltz, K. & Kutas, M. (1999). Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Cognitive Brain Research, 7*, 241-253.

[19] 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.

[20] O'Rourke, P. L. & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research, 1392*, 62-79.

[21] Penke, M., Weyerts, H., Gross, M., Zander, E., Munte, T.F., & Clahsen, H. (1997). How the brain processes complex words: an event-related

potential study of German verb inflections. *Cognitive Brain Research, 6*, 37-52.

[22] Plag, I. (2004). Syntactic category information and the semantics of derivational morphological rules.*Folia Linguistica, 38*(3-4), 193-225.

[23] Plag, I. & Baayen, R.H. (2009). Suffix ordering and morphological processing. *Language, 85*, 106-149.

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

[25] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[26] Sereno, S. C., Rayner, K. & Posner, M. I. (1998). Establishing a timeline of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport, 9*, 2195-2200.

[27] Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition, 7*(4), 263 - 272.

[28] Tremblay, A. (2010). Independent Components Analysis (ICA) based eye-movement correction. R package version 1.2.

[29] Van Petten, C. & Kutas, M. (1990). Interactions between sentence context ad word frequency in event-related brain potentials. *Memory & Cognition, 18*(4), 380-393.

[30] de Vega, M., Urrutia, M. & Dominguez, A. (2010). Tracking lexical and syntactic processes of verb morphology with ERP. *Journal of Neurolinguistics, 23*, 400-415.

[31] Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1):3-36.

Chapter 5

General Discussion and Conclusions

Throughout this dissertation, we have explored the nature of suffix pair processing in English using a variety of psycholinguistic methods. In particular, we were interested in the predictive value of the suffix rank derived from the Complexity-Based Ordering theory of affix ordering. We did not find evidence for rank in either the naming or lexical decision tasks, but when our target words were embedded in a sentence context, effects of rank became prevalent. Results also showed that in addition to expected root and whole word frequency effects, participants were sensitive to the frequency of the suffix pair and to individual suffix type and token frequencies. In what follows, I summarize our results and discuss their implications.

Across all experiments, effects of word length, Root Frequency, Base Frequency, and Word Frequency were consistently present. Table 5.1 provides an overview of which predictors were useful in which experiment. This table does not include structural predictors specific to a given experiment (e.g., "sentence position" for our sentence reading experiment). As expected, longer words elicited longer response latencies in lexical decision, as well as more and longer fixations during word reading. For both eye-tracking studies, the time spent on the first and second fixations was shorter for longer words, although the total fixation duration for longer words was also higher than for shorter words. Trial information was not available for data extracted from the English Lexicon Project (Balota et al., 2007), but for both of our eye-tracking studies, participants generally became faster as the experiment progressed.

Both base and word frequencies facilitated processing, although word frequency was not predictive at the first fixation duration for either experiment. The emergence of a whole word frequency effect later during word reading, as recorded as part of the eye-movement record, is in line with previous research

Table 5.1: Presence of predictor effects by experiment (overview)

Predictor	Naming	LD	LD (Eye-tracking)	Sentence reading
Word length	Yes	Yes	Yes	Yes
Root Frequency	Yes	Yes	Yes	Yes
Base Frequency	Yes	Yes	Yes	Yes
Word Frequency	Yes	Yes	Yes	Yes
Root+Suff2 Frequency	No	No	Yes	No
Affix Pair Frequency	Yes	Yes	Yes	Yes
Suffix1 Frequency	No	No	No	Yes
Suffix2 Frequency	No	No	Yes	No
Suffix1 Type Frequency	No	Yes	Yes	No
Suffix2 Type Frequency	No	Yes	Yes	No
Suffix2 CO-Rank	No	No	No	Yes
Trial	NA	NA	Yes	Yes

on suffixed words (Beauvillain, 1996; Niswander, Pollatsek & Rayner, 2000), and highlights the potential for segmentation during word recognition. However, there are a number of studies that find word frequency effects at the first fixation. Kuperman et al. (2008) focused on compound words of a similar length to our targets, and found compound frequency effects in addition to effects of the left constituent at first fixation. A key difference between the present work and Kuperman et al. (2008) is the nature of the stimuli. The target words used in this dissertation tend to be infrequent and their meanings tend to be fairly transparent and easily derived, so it is unlikely that many will be accessed as a whole word. Compounds, on the other hand, may have more specific meanings than can be generated based on their parts alone. Of the two types of words, the low frequency, semantically transparent derived words are less likely to be accessed as whole forms. Another major difference is the number of fixations participants required to read the target word. Kuperman, Bertram, and Baayen (2010) investigated the processing of bimorphemic derived words in Dutch, when those words were presented in a sentence context. Again, they found effects of word frequency at the first fixation. However, most words in their study only received a single fixation (83%), with 1% of the target items receiving three or more fixations. This is remarkably different from our results, where very few items received only a single fixation. Our words, on average, were 11 letters long, compared to the mean of 8 for Kuperman et al. (2010). With just one fixation, word frequency is available to emerge immediately because access to the entire word is possible without a second fixation. Finally, although we did not find a *word* frequency effect at the first fixation, we did find a *base* frequency effect. The base (root+suffix1) was, in many cases, also a free-standing word. So in effect, we do have a partial word frequency effect (from the base) at first fixation very much like that reported by Kuperman et al. (2010), but due to the length of our items, the whole word cannot be identified at the first fixation.

Effects of individual suffix type and token frequencies were found primarily in our analysis of the lexical decision data. Suffix type frequencies were predictive of fixation durations and reaction time latencies in both the English Lexicon Project lexical decision data and in the eye-tracking lexical decision data, with higher type frequencies being associated with faster reaction times and shorter fixations. This may be due to the nature of the lexical decision task, where a specific decision about lexicality must be made. In our lexical decision experiment with an eye-tracking component, we wanted to ensure that participants read each word in full before making their decision. To do that, we used nonwords that had a pseudo-morphological structure as foils, where one element was a non-morpheme in English. As participants could not predict where to look to determine if an item was a word or a nonword, fixations across the entire word were assured. However, this setup also makes decomposition and recognition of individual morphemes a useful strategy. This may be the reason we see a facilitatory effect of the first suffix at the fourth fixation, for instance. This strategy on the part of participants is not inherently bad, but it does highlight one of the limitations of using lexical decision as the sole task to examine lexical processing.

Taking the sentence reading experiment as the most naturalistic of our data, we see a time course that emerges during the reading of our complex words. Most of our items in both eye-tracking experiments required 2 or more fixations to read. This in turn means that part of the word (i.e., the base or the root alone) is read and is at least partially available for processing before the entire word has been fixated upon. At the first fixation, individual morphemes and morpheme groups that occur early in the word are activated, crucially including the combination of the root and first suffix (the base). At the second fixation, the whole word becomes identifiable, and word frequency effects emerge. Word recognition is also faster if the suffix+suffix affix pair is more frequent, which also becomes apparent at the second fixation. The presence of the affix pair frequency effect tells us is that, although composed of two morphemes, the affix pair is being treated as a single unit itself, separate from the properties of the individual suffixes that compose it. When the frequency of a given suffix pair is higher, response latencies and fixation durations are faster and shorter.

In sentence reading, few single-suffix frequency effects play a role in lexical processing (we contrast this with lexical decision, where participants may find it helpful to parse individual suffixes). When a word receives regressions *after* first pass reading, then we see effects of the word internal first suffix, which corresponds well with the idea of returning to a word that has caused some difficulty in greater sentence processing (e.g., because it is not well known).

5.1 The Role of CO-Rank

The CO-Rank effect for the second suffix is clearly observed when in a sentence context, and not at all in single word recognition. During sentence reading, participants have a sentential context in which to place the target item, and so have a better idea of what word (or what kind of word) will be coming next. Even if a speaker has not seen a word before, the greater context will support its interpretation. When the second suffix was lower on the CO-Rank hierarchy, participants spent longer on the target words, indicating that the second suffix was less parsimonious than a suffix of a higher rank, and that they presented greater difficulty to the processing system. Said another way, the expectation that a suffix at the end of a word will be of a relatively high CO-Rank was not met by all words, and when it was not met, longer processing times resulted. Likewise, a greater amount of time was spent at the first fixation when the CO-Rank of the second suffix was lower, even though it was unlikely that the second suffix could be fully perceived. This suggests that even the possibility of encountering a lower ranked suffix slows down the reader, a delay which is continued as the identity of the suffix is realized. However, without a sentence context, there can be no expectations with which the processing system can work.

We have clearly found effects of CO-Rank, as it is defined by Hay and Play (2004) and Plag and Baayen (2009), but we have only found them for the second suffix. This presents us with a quandary. There may be structural and physiological reasons for the lack of an effect at Suffix1. First, if the frequency of the base (root+suffix1) is greater than that of the root alone (Hay, 2001), adding a second suffix is, in terms of processing, more like adding a first suffix to a monomorphemic word. Second, the second suffix is likely to require a second (or third) fixation to be fully perceived during reading, and in many cases, as when the CO-Rank is high, it is easily parsed out from the rest of the word and therefore is perceptually very salient, whereas the first suffix is not. However, the lack of an effect for the first suffix may be part of a larger question: is CO-Rank really a measure of affix ordering?

The mean CO-Rank of the second suffix in our reading experiment was higher than the first, meaning that, in general, the second suffix was more easily parsed than the first. We expect this pattern, as suffixes in Complexity-Based Ordering are ranked according to their parsability and selectional restrictions, where suffixes of higher rank are expected to attach outside of suffixes of lower rank. We see our strongest effects when the CO-Rank of the second suffix is low, which in turn means that the first suffix is also of low rank (e.g., strengthen, root: strong, suffixes: -th and -en). Two low ranked suffixes are evidently difficult to process, as observed in both fixation durations and in our ERP results. However, we would expect to find effects of CO-Rank on the first suffix, if the rank is constraining the suffix that will follow during processing. For example, if the CO-Rank of the first suffix is high, then the CO-Rank of the second suffix should also be high, which means that there are fewer suffixes to choose from for the next item, and reaction times should be faster. Conversely, we would expect longer reaction times when the CO-Rank of suffix1 is low because the number of candidates for the following suffix is high. However, there is no evidence for any role of the CO-Rank of the first suffix. It may be that our base items are too readily analyzed as single units for CO-Ranks of the first suffix to appear in sentence reading. That possibility, however, will require further research.

5.2 Future Directions

The next steps in our consideration of suffix ordering should be to expand on our dataset to include multiply suffixed words with more than two suffixes in a sentence context. This would allow us to determine whether all suffixes beyond the first can make use of CO-Rank as a processing tool, or whether this is an effect that only appears for the final suffix in a string. If it is the case that CO-Rank is only relevant for the final suffix, then while a useful predictor, it would not reflect an ordering process. It would also be worthwhile to investigate the processing of plausible novel words containing suffix pairs, to overcome potential interference from existing base (root+suffix1) frequencies. Novel words would allow us to directly test whether the first suffix in a pair shows rank effects when it is not a part of a base that is prone to whole-word access. If the CO-Rank of the first suffix is not predictive under such conditions, then CO-Rank cannot not be reflective of ordering.

Finally, while Complexity-Based Ordering originated in the quest to explain English suffix ordering, the principles of parsing and selectional restrictions as determinants of affix ordering should be applicable cross-linguistically. If Complexity-Based Ordering is to be considered a general psycholinguistic theory of affix ordering, then it stands to reason that it needs to be tested in other languages. Results from this dissertation stress the importance for any such experimentation to make use of a sentence context.

5.3 Final Words

This dissertation began with the intention of testing the theory of Complexity-Based Ordering. Along the way, several other interesting results manifested, chief among them the role of the affix pair frequency in addition to, and often instead of, individual suffix frequencies. CO-Rank was not found to be predictive for reaction time latencies or fixation durations when words were presented alone, as in lexical decision. However, it was prevalent as a predictor in sentence reading. This, more than anything, reminds us that attention must be paid to task demands, as our linguistic systems are sensitive to context as well as to the specific stimuli with which they are presented. Had we been constrained to single word presentations we would not have discovered the usefulness of CO-Rank, and indeed, would have rejected it outright. The role of Complexity-Based Ordering was found in the more naturalistic task, sentence reading, which provided greater context for the linguistic system to work with. This speaks to the great complexity of language as a whole, and to the sensitivity of our language processing systems, which are capable of exploiting every informative advantage to aid in our ongoing decoding of the linguistic signal.

5.4 References

[1] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

[2] 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*, 801 - 820.

[3] Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics, 39*(6), 1041-1070.

[4] Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural language and Linguistic Theory, 22*, 565-596.

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

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

[7] Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected words during reading. *Language and Cognitive Processes, 14*(4/5), 389-420.

[8] Plag, I.& Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language, 85*, 106-149.