University of Alberta

Do morphological properties of affixes affect the perception and processing of novel English suffixed words?

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

Master of Science

Department of Linguistics

Edmonton, Alberta
Spring 2006

Your file Votre référence
ISBN: 0-494-13895-5
Our file Notre référence
ISBN: 0-494-13895-5

Abstract

This thesis sets out to determine whether native speakers of English are sensitive to selectional restrictions, suffix family size, and suffix productivity in the combinations of real words and derivational suffixes in novel pseudowords. Approximately 150 participants took part in three lexical decision experiments, a category decision task, and two offline questionnaires. Single syllable nouns and verbs served as bases to which suffixes were attached. There appeared to be a small effect of suffix productivity, with higher error rates observed for suffixes with higher productivity values. There was no main effect of selectional restrictions; potential effects may be been too small to be observed in real word data. Results suggest that English speakers are influenced by suffix family size when processing novel words and that semantic interpretability may be an important factor in determining the status of a potential word.

Acknowledgements

I would like to thank my family and friends for their support during the course of this thesis. It was always appreciated. Thanks go as well to my current supervisor, Dr. Gary Libben, for his invaluable advice and aid, and to Dr. Ron Smyth at the University of Toronto for introducing me to the field of Psycholinguistics and for his continued enthusiasm and support. I also wish to acknowledge the financial support provided by a SSHRC CGS Master's Scholarship and the opportunities provided by the "Words in the Mind, Words in the Brain" research project. It was at my first student conference, supported by this project, where the seeds of this research were first planted.

Table of contents Page

List of Tables Page

List of Figures

List of Abbreviations

A: Adjective

ANOVA: Analysis of Variance

FS: Family Size

HFS: High Family Size

LD: Lexical Decision

LFS: Low Family Size

ms: milliseconds

N: Noun

RT: Reaction Time

SFS: Suffix Family Size

V: Verb

CHAPTER 1

INTRODUCTION AND OVERVIEW

Just as sentences are made up of words, words can be made up of smaller units that have traditionally been called *morphemes*. In traditional descriptions, morphemes are the smallest unit of linguistic meaning, and while one word can consist of just one morpheme, morphemes can also be combined to form complex words. They can be formed by the combination of a root or stem with another unit, either an affix (inflectional or derivational) or another root (compounding). This combination does not occur randomly. Morpheme combination follows language-specific structural patterns. These patterns can be realized terminologically as *selectional restrictions*. If selectional restrictions are not violated, then a word can be considered morphologically legal, or allowable in a language. If selectional restrictions are violated, then a resulting word could be considered morphologically illegal, because it does not follow regular patterns of morpheme combination, and one would not expect such items to be produced regularly by speakers. For example, in English, the suffixes *-ation* and *–ment* both combine with verbs to form nouns with a general meaning of 'state' or 'action' (Quirk & Greenbaum, 1973: 436-441). Real words with these suffixes include *colourization* (*colourize+ation*) and *amusement* (*amuse+ment*). The non-occurring items ?*colourizement* (*colourize+ment*) and ?*amusation* (*amuse+ation*) should equally be possible, according to broad selectional restrictions that specify the lexical category of the base word. A word like **birdment* or **birdation* violates the selectional restrictions of these affixes, because *bird* is a noun and not a verb.

The suffixes *–ation* and *–ment* can be used to illustrate another point: some suffixes are expressed far more often than others and in more environments. The suffix *–ation* occurs in a far higher number of words than *–ment*, by a simple count. In other words, *-ation* has a higher family size than *–ment* because it occurs in more words. Regardless of how frequently any of those words occur, it has a greater number of family members. The notion of morphological family size captures the notion that some suffixes in a language are more active than others. That is, they occur in more morphologically complex words. There is another notion of morphological activity that may also be relevant to the mental representation of morphemes and of complex words. This is the notion of morphological productivity.

The term morphological productivity captures the fact that some affixes can be used to create new words more easily than others. As an example of morphological productivity, consider the difference between the English suffixes *–ity* and *–ness*. Although both these derivational suffixes create nouns from adjectives (e.g., *happy+ness=happiness*, *legal+ity=legality*), the suffix *–ness* can be used much more easily by speakers to create new words. So, for example, if the word *dope* is used innovatively as an adjective (as is the trend in recent popular music), then the novel form *dopeness* will be much more acceptable than *dopity* because the suffix *–ness* is more productive. Indeed, a current Google search of English websites, yields 89,000 matches for *dopeness*, but only 530 for *dopity*.

The three characteristics of English suffixes introduced above seem to suggest that some novel words will seem more word-like for reasons that are primarily related to the properties of their suffixes. We would expect, for example, that a new word that violates the selectional restrictions of its suffix, contains a suffix that is unproductive and has a low morphological family size, will be treated as less word-like by native speakers of English. This expectation is based on the hypothesis that native speakers are indeed sensitive to these variables in their processing of novel English words. The goal of this thesis is to test this hypothesis. In the chapters that follow, I report on a series of experiments that investigate whether native speakers' judgments of novel words are influenced by the morphological family size and productivity of suffixes and by the extent to which novel combinations of base and suffix satisfy the selectional restrictions of the target suffix.

The thesis is organized in the following manner: In Chapter 2, I summarize the relevant literature on the processing of suffixed words with special attention to the roles of selectional restrictions, family size, and productivity. In Chapter 3, I report on a series of online and off-line experiments that were employed as a pilot investigation in order to determine which types of experimental paradigms could best be employed and to determine whether experimental participants are aware of morphological constraints on word formation in English. This is followed by a report of the main experiments in Chapter 4 and a General Discussion in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

Complex words in the Mind

Morphologically complex words have been the subject of a wide range of research, much of it aiming to understand how existing complex words are processed by speakers. For example, evidence suggests that very frequent complex words can be processed more like single units (e.g., see Meunier & Segui, 1999), regardless of their descriptive structure.

As mentioned, words can be composed of multiple units of meaning, called morphemes. Other useful terms include *root*, and *affix*. In this work, *root* and *base* will be used to refer to the monomorphemic lexical item which serves as the base of a target item. Affixes, of which suffixes are one type, are bound morphemes that attach to roots and perform various functions (e.g., indicating verbal agreement or changing a noun into a verb), and they can be either inflectional or derivational. Inflectional affixes tend to be grammatical (e.g., marking possession, pluralization) and do not change the class of the root, whereas derivational affixes can change the lexical category of the root and generally also change the meaning of the root in some substantial way. There is debate over whether inflectional and derivational processes should be considered separate, or if they are part of a larger continuum (c.f. Bybee, 1985). Clahsen, Sonnenstuhl and Blevins (2003) report findings from a priming study suggesting that there is a distinction between inflection and derivation. Based on their work with productive derivational affixes, the authors suggest the existence of separate mechanisms underlying inflectional and derivational behaviour. They claim that 'productive derivation' is modulated by rules, such that neologisms must follow them. Such rules could be realized as selectional restrictions.

Selectional Restrictions

There has not been much investigation into the legality of root+suffix complex words. Libben (1993) examined reactions to violations in nonwords where morphologically legal and illegal nonwords were created by attaching real prefixes and suffixes to nonsense bases (in the form prefix+base+suffix). Subjects were found to be sensitive to morphological legality during this experiment, taking longer to reply to nonword+suffix trials where the suffix was illegally attached to the root, as defined by the morphological structure of the whole pseudoword. A case study of a patient with aphasia revealed similar results (Libben, 1994). The patient made more errors when naming illegal words than when naming morphologically legal words of the same length. These studies suggest that individuals are sensitive to the internal morphological structure of words as well as word formation processes. Both of these studies used nonwords to avoid difficulties arising from the lexical properties of the root (e.g., frequency) and to bypass difficulties arising from varying levels of semantic interpretability of combinations of real words and real affixes. However, real words and bases are more natural to speakers. If combinations of real morphemes can be successfully used, results can be generalized to a greater extent than results from nonword data.

In a study of Finnish morphology, Järvikivi and Niemi (1999) studied the influence of word-category in suffixation using morphological decision tasks with real word stimuli. Participants were asked to decide whether a given stimulus was a noun in one experiment, and a verb in another. Stimuli were composed of noun, verb, or adjective stems and the suffix *–us*. This suffix can form real words with all three classes, but does so at differing levels of productivity. Reaction times were found to be longer when there was a stem-suffix mismatch, that is, where the suffix was less productive, which the authors suggest reflects processing difficulties during left-to-right processing and an unconscious awareness of stem class, in addition to the class assigned by the attached suffix. These results differ from those obtained by Burani, Dovetto, Spuntarelli and Thornton (1999), who did not find any effect of selectional restrictions in a lexical decision experiment, but who did find effects of semantic interpretability.

Derivational Suffixes

The psychological realization of roots and affixes, and their interactions with each other have yet to be fully explored. Giraudo and Grainger (2003) performed a series of masked priming experiments to determine whether and how derivational suffixes are represented in the mind. Using pseudowords, with both prefixes and suffixes, they found priming effects for target items sharing a prefix, but not for items sharing a suffix. They determined that differences in semantic transparency between the prefixed and suffixed items were not the cause of this discrepancy, and that this was not a frequency effect of the suffixed items. Then they looked at suffixes boundaries, finding that that when the suffix boundary matched the syllable boundary, RTs were faster relative to the control condition. They concluded that in French, they could find no effects of suffix priming because the suffixes were more difficult to identify within a word than prefixes which more often corresponded with syllable boundaries, and because suffixes frequently changed the form of the base. Both make suffixes less salient in the studied French items than prefixes.

Reid and Marslen-Wilson (2003) reported on three experiments investigating the processing of affixes in relation to the mental lexicon. Contrary to Giraudo and Grainger (2003), they found an effect of suffix priming, although it was in an inflectional paradigm (secondary imperfective forms), in addition to stem priming. They also found a suffix-suffix *interference effect* when different derivational suffixes were presented on the same stem in a prime-target pair. They take this as evidence that suffixes are represented at some level in the mental lexicon, separate from whole-word entries to which they might be attached. Further, they take this to mean that once the base has been revealed, there is competition between suffixed family members, and that the interference effect is caused by the activation of more than one possible suffixed word. Semantic transparency was also found to be an important factor regulating priming: when prime-target pairs were not transparently related (e.g., *department-depart*), no priming was observed. It is suggested that such items are stored as whole units.

There is some interest in the etymological roots of suffixes, with potential differences between native Germanic morphemes and borrowed Latinate morphemes being addressed. Vannest and Boland (1999) examined suffixes from the so-called Level 1 and level 2 types. In this system, Level 1 affixes generally change the structure of the base and are less productive than Level 2 affixes, which are considered to be neutral in terms of phonology and semantics, being transparent. Data from this study suggest that the more transparent suffixes more readily undergo decomposition than their less transparent, Level 1 counterparts. However, some of their results were conflicting, and they conclude that affixes differ from each other in the ways they are represented lexically and in the ways they behave with the roots to which they attach, although in generally, Level 2 affixes remain more likely to be decomposed.

Hay and Plag (2004) discussed possible suffix combinations and their restrictions. They examined the possibility of lexical strata influencing the processing of derivational affixes, versus the concept of affix-specific selectional restrictions, although it was in the combinatorial properties of suffixes, and not in the combination of suffix and base (e.g., the combination of the suffixes *–ize* and *–ation.*). In this work Hay and Plag (2004) test the assumptions of Complexity Based Ordering. In this framework, less decomposable suffixes will occur closer to the root than more decomposable suffixes. Relevant to this study is that Complexity Based Ordering relies on the concept of selectional restrictions in suffix combinations. While not quite the same as base+suffix selectional restrictions, the concept is similar. They found that while this is the case, the ease with which a suffix can be phonologically parsed is also a factor in potential suffixation.

Luzzatti, Mondini and Semenza (2001) examined how an agrammatic patient read morphologically complex words, finding that there was little difference between the ability to correctly read simple forms and high frequency complex words. Hay (2001) found that the relative frequency of the base of and the derived form of a word influences the way in which it is processed, and that the relative frequencies will influence whether a word is decomposed. This indicates that properties of the suffix interact with properties of the base, and more to the point, that suffixes *have* statistical properties that are worth consideration. We turn now to a discussion of two of those properties, frequency and family size.

Morphological Family Size and Frequency

While effects of word frequency are well attested in studies of linguistics, it is only relatively recently that family size has been revealed as a statistical factor. Schreuder and Baayen (1997) found that the processing of monomorphemic Dutch nouns was affected by the number of words that were morphologically related to the targets. Nouns with more morphological relatives were processed more quickly than those with fewer relatives in lexical decision experiments. However, the cumulative frequency of all the forms did not correlate with reaction times. Because the family size effect was not present in a progressive demasking task, which can be used as a tool to examine form identification at an earlier stage in processing, Schreuder and Baayen (1997) suggested that the family size effect must come into effect after an item has been identified. In line with earlier experiments, they found that base word forms were affected by the frequencies of their plural forms. For the terms of this thesis, we will borrow some terminology from Schreuder and Baayen (1997): surface frequency refers to the frequency of a form, e.g., the frequency of the word "colour." Stem frequency refers to the summed frequencies of singular and plural forms for nouns, or more generally, the inflectional forms of a word (including verbs). Morphological Family refers to all the words that are formed with the base word via derivation or compounding, but not inflection, and the Morphological Family Size is a count of all these words. The cumulative frequency of a word can be found by summing all the frequencies of members in the morphological family size, but excluding the base stem frequency. Schreuder and Baayen (1997) suggest that the family size effect is in actuality a semantic effect, caused by the activation of all morphologically related words of a target, which in turn activates the semantic representations of their morphological family members. Bertram, Baayen and Schreuder (2000) extended this research on the effects of family size to complex words. They found that the family size of a base of a complex word affects reaction times, so that the effect is not restricted to monomorphemic bases.

De Jong, Schreuder, and Baayen (2000) further explored the concept of family size with verbs in Dutch. Again, they found results indicating that participants were sensitive to the morphological family size of the target items, indicating that the family size effect is not restricted to nouns and nominal morphology. Furthermore, the authors contend that the effect is truly morphological and not just semantic, because the family size effect is present across verbs related morphosyntactically by suffix type rather than by semantics alone. In their study, the family size of verbs taking irregular forms also showed that this was not a function of the form of the suffix. Recently, the family size effect has been shown to be context sensitive (de Jong, Schreuder, & Baayen, 2003). The addition of a qualifier (e.g. *very, not*) in front of a word influenced its reaction time in lexical decision. While this should not affect the calculations made in the present study, as the family size effect has been reliably found for words in isolation, it does highlight the semantic nature of the measure.

Inflectional family size has been implicated in a number of studies as affecting processing. It stands to reason that derivational family size should also have an effect. Traficante and Burani (2003) found that the number of morphemes in an inflectional paradigm affected processing speed of verbal and adjectival base classes. In Italian, verbs have a much larger inflectional family size (more suffixes), which the authors claim makes the roots more salient (or 'active' under some models) than the noun and adjectival roots, which have a comparatively reduced inflectional family size. Overall, verbs required more time to process, which the authors suggest is because of the necessity of parsing.

Family size and frequency are related lexical properties that can be selected for use in studies of morphological processing, and previous studies have looked to various measures of frequency to study the processing of complex words. There are a number of ways to measure frequency, and it is not necessarily clear which ones best predict response latencies (Ford, Marslen-Wilson, & Davis, 2003). Burani and Thornton (2003) performed a detailed examination of the role of frequency in processing derived words in Italian. They were operating under the assumption that "frequency is the major determinant of the relative probability that lexical access is either whole-word based or morpheme-based" (Burani & Thornton, 2003: 158.) Similar to Hay (2001), they were concerned with the differing frequencies of the bases used and of the whole-word frequency, in comparison. In a previous study Burani & Thornton (2003) report on, pseudowords were created with real roots and were combined with suffixes that did not usually combine with the chosen roots. The suffixes chosen varied in terms of relative frequency of occurrence, with one set being highly frequent and the other being low frequency, while root frequency was kept constant. They found that pseudowords with highly frequent suffixes were responded to less accurately and more slowly than pseudoword controls, but that low frequency suffixes did not show any such effect.

Burani and Thornton (2003) examined whether the relative frequencies of the constituents of derived words in Italian affected the processing of these words. In their first experiment, they found that highly frequent suffixes generated longer reaction times and higher error rates on pseudowords in a lexical decision task, but did not find any such effect for suffixes of mid or low frequency. They decided to collapse the medium and low frequency categories into one category. In the second experiment, real words were used. These words were real combinations of high frequency roots and high frequency suffixes, high frequency roots and low frequency suffixes, low frequency roots and high frequency suffixes, and low frequency roots and low frequency suffixes. They found the frequency of the root influenced the reaction time and accuracy of responses positively, but that the frequency of the suffixes did not seem to make a difference. A third experiment attempted to find differences between the processing of combinations of low frequency root+high frequency suffix and high frequency root+low frequency suffix. They found that these categories of combination patterned like their roots, that is, the frequency of the suffix did not appear to make a difference.

Meunier and Segui (1999) used cross-modal priming experiments to investigate the connection between derived words and their bases and their representation in the mental lexicon, with frequency as a modulating factor. In their first experiment, derived words were used to prime their stems. High and low frequency stems were used. It was found that the highly frequent suffixed words did not appear to prime their stems, while low frequency suffixed words did. It was proposed that this resulted from the decomposition of the low frequency derived word, while derived words with high surface frequency were proposed to have their own lexical representations in the lexicon. In the second experiment, stems were used to prime derived words that contained those same stems, as were full-form affixed words. It was found primes identical to their targets primed more than bare stems of the same morphological family as the target. However, stems were found to prime related targets, although targets with a higher surface frequency benefited more from the prime. The authors suggest that this asymmetry between the two experiments, where in Experiment 1 the low frequency derived words primed stems more than the high frequency derived words, and in Experiment 2 it was the high frequency derived words that benefited more from priming, is caused by the presence of differing lexical representations. It is proposed that when high frequency derived words are encountered, two lexical representations are activated: one full-form and one decomposed. In the first experiment, the full form exposure would be less able to speed priming of a target than the low frequency derived word, which is presumably decomposed because of its lower frequency, and is therefore represented in its component parts in the mental lexicon. However, when the stem is given as a prime, the high frequency derived word will benefit from spreading activation. While this would also hold true for the lower frequency derived targets, since they are lower frequency, the activation would be slower.

Feldman and Pastizzo (2003) studied semantic transparency and the role of family size in morphological priming. In a lexical decision task, partially transparent primes sped reaction time but decreased accuracy in targets when compared to unrelated primes. Family size correlated with the semantic relatedness effects.

It appears that frequency measures, including suffix frequency and family size, are viable affix properties that can be used in studies of lexical processing. We now turn to a more controversial measure, that of morphological productivity.

Morphological Productivity

As this thesis deals with suffixes in English and their combinatorial properties, it must also include a discussion of suffix productivity, which refers to the ability of suffixes to combine with bases (simple or complex) to form new words.

In terms of pure data and stimuli creation, the most important measure of productivity used in this paper was described by Baayen (1992) and relies on the use of language corpora to produce a statistical measure of productivity, *P*, which is based upon the occurrence of novel forms in the corpus. These novels forms, called *hapax legomena* (henceforth "hapaxes"), are words that occur only once in a given corpus or database (the source of data being used to determine the *P* values; in this case, the CELEX lexical database). The size of the database, of course makes a difference. When a database is larger, there is a better chance of encountering novel words, although it should be noted that not all hapaxes are neologisms. The formula for productivity, *P*, is: ($P = V_n(1,C)/N_c$, where *P* is the productivity of an affix, $V_n(1,C)$ represents the number of new words in a category under investigation, and N_c represents the total number of words in that category. Although there have been some criticisms of this statistical model, chief among them the observation that the values this formula generates do not always coincide with speaker intuitions about productivity and that, given their presence in a corpus, the hapaxes are clearly already existing words and no longer potential words (van Marle, 1992), Baayen's (1992) productivity calculation is one of the few quantitative measures from which to base further analyses.

Baayen (1997) addresses some of van Marle's (1992) concerns regarding the output of the productivity formula in the form of a discussion on productivity and markedness, in particular regarding counterintuitive *P* values for less productive forms. In this study, the Dutch agentive suffixes *–er* and *–ster* are considered, where *–er* is unmarked, but has a lower *P* value than *–ster*. Since morphologically marked forms are expected to belong to a lower frequency class, the members of this class will have more hapaxes relative to the higher frequency *–er*. Upon completing an experiment in which participants were asked to generate new words under set conditions, it was found that more new words, i.e., words that did not already appear in the corpus data, were coined with *–ster* than *–er*, although *–er* words were higher in number overall. This being said, *P* remains a valid measure, as it reflects this productivity, and he suggests for markedness with less obvious productive forms as well. Baayen (1997) concludes with a caution that other factors are likely involved in the productivity of this and other suffixes, including the complexity of the base to which a suffix attaches and non-linguistic factors such as socio-cultural influences.

Baayen's (1992) approach to productivity reflects the probability of the occurrence of a form, and that is the focus of the measure. For others, it is not probability, but *possibility* of word formation that is of interest. Dressler and Ladanyi (2000) suggest a categorical approach to productivity which focuses on the latter. The basis of their argument lies with language specific word-formation rules, as evidenced through the assimilation of new words from other languages. For example, a highly productive affix will be able to attach to newly borrowed words, and will change the structure of the borrowed word to fit itself (through word-formation rules). Levels of productivity range from very high to very low, where low productivity items cannot alter the shape of a novel or foreign stem, and will not be applied to borrowed words. The possibility of a word-form under this system is similar to legality in a rule-based system, that is, what will be legal under a rule based system is also possible, if not probable. This viewpoint has its detractors as well, as there is no way to quantify possibility. This means that, unlike Baayen's measure of productivity, there is no formula that can be used to derive the likelihood of occurrence. This system is also a constructed one, and while it is very thoughtfully constructed, the categories suggested could, in theory, be extended and broken down into smaller categories indefinitely. Thus, there is no naturally occurring break-point for the categories.

Building on the work of both Hay (2001) and Baayen (1992, 1994, 1997), a relationship between productivity and parsing is being explored. At its most basic, it can be said that not all complex words are decomposed to the same degree, and that there are differences in processing by suffix (Hay & Baayen, 2002). As noted in several experiments above, it appears that complex words with high surface frequencies are less decomposable than lower frequency complex words, and may be more likely to encounter semantic drift and to be stored as full forms. Parsability is also linked to the relative frequencies of a complex word's components, such that if a base is more frequent than the complex word that it forms, the word may still be decomposable because the base can be more readily recognized than the whole form (Hay, 2001). Likewise, if a low frequency complex word is relatively more frequent than the base of the complex word, then it would be expected to be less decomposable and more opaque. Within the last few years, corpus-based statistical models have been used to support the proposal of a *parsing line*, which can be used to statistically determine the number of words containing an affix that can be parsed (Hay & Baayen, 2002). The more words an affix occurs in, the more productive that affix is expected to be.

It has also recently been found that phonotactics play a larger role than has been thought in the parsing, and therefore productivity, of affixes, such that there is a correlation between phonotactic transitional probabilities and the *P* value (Hay & Baayen, 2003). When a morpheme boundary is over an unlikely phonotactics sequence (e.g., /pf/ in *pipeful*), then it is more likely to be decomposed, and it is presumably easier for this to occur. They call this the *junctural phonotactic probability*, or the likelihood of a phonotactics transition between morphemes at a particular phonotactic sequence. It is suggested that affixes that can be easily parsed are more likely to be highly productive, as they can be more easily recognized and taken out of context by speakers.

The Mental Lexicon

As this thesis is a psycholinguistic study of suffix processing, a few words should be said on the topic of the Mental Lexicon, or the nebulous "dictionary" that exists in the minds of speakers where knowledge about words is stored. There are a number of different models seeking to describe how this storage is achieved, and what parts of words are available. Some models claim that every word a speaker knows is listed in the lexicon, like a dictionary, in its full and complete form; others claim that all morphemes are listed, but that morphologically complex words must be parsed into their components in order to be understood, such that no complex words have representations (Hall, 1992: 130). Many models fall between the two extremes. In dual race models, both are possible, but the way in which a word is processed will depend entirely on the properties of that word and its components (Baayen & Schreuder, 1999; Frauenfelder & Schreuder, 1992). Words will be influenced by their frequencies, phonology, and semantics in deciding whether to be stored in a whole-word representation or whether to be decomposed into constituents to be understood (Frauenfelder & Schreuder, 1992).

Models of the mental lexicon that allow for decomposition are in a better position to explain the processing of novel forms. In its strictest incarnation, a full-listing model would be theoretically unable to account for native speaker comprehension of novel forms, because a novel form will not have an entry in the mental lexicon. The current study makes use of pseudowords which contain real bases and real suffixes, but they are in combinations that should be novel to most, if not all, speakers. Therefore, these experiments assume that speakers have the ability to decompose words, which has been borne out in various studies. However, it cannot inform any discussion of full decomposition models and dual-route models.

To summarize, there is evidence that suggests that affixes can have representations in the mental lexicon that are independent of full-form representations of words and that affix-specific properties, such as frequency, can affect morphological processing. There is also some evidence to suggest that speakers of a language are sensitive to the morphological structure of words. The most reliable results pertaining to morphological structure have been obtained from studies using nonwords in combination with real affixes. When real roots have been used, results have been more difficult to obtain. Finally, there is evidence that suggests that speakers can be sensitive to the lexical category of a root during the processing of complex words, and that productivity as a quantitative measure can be used to help determine the likelihood of affixation and salience of affixation. These findings informed the construction of the following experiments and bring us into the first phase of testing.

CHAPTER 3

MOTIVATION OF THE STUDY AND PILOT INVESTIGATIONS

The Present Study

In the present study, pseudowords composed of real roots and real suffixes are presented to participants. Based on past research, this study asks whether or not native speakers of English are sensitive to selectional restrictions of affixes based on lexical category, whether they are sensitive to the family size of a suffix, and whether they are sensitive to the productivity of a suffix. The initial hypothesis is that pseudowords composed of noun roots and verb suffixes, or pseudowords that violate general selectional properties of an affix, will be easier to reject as non-words than bases that match the suffix type. It is also predicted that suffix family size will affect the speed of processing and difficulty of processing and that these will be reflected in accuracy and RT data. Pseudowords containing suffixes of a higher family size are predicted to take longer to reject in lexical decision experiments. Suffix productivity is predicted to be reflected in error rates and reaction times, with suffixes of higher productivity showing more errors and longer RTs. In sum, novel pseudowords that are composed of legal base+suffix combinations, where the suffix is both highly frequent and highly productive, are expected to show the most errors and longest RTs.

The combinatorial properties of suffixes, namely, the base categories to which a suffix is expected to attach, were extracted from Quirk & Greenbaum (1973: 436 – 441), Carstairs-McCarthy (2002: 45-57) and by independent analysis of existing words in the CELEX database. This was done to broadly determine whether and when base+suffix pairs violated the selectional restrictions of a given suffix, based on the root's lexical category. Lexical category violations were chosen as the selectional restriction because this is one of the most basic violations possible, meaning that it is applicable over a wide range of suffixes. Whether a combination is allowable because of a set of word formation rules or by analogy with other, existing, forms is a matter for further study. However, one might predict that a strictly rule-based approach would not be influenced by variations in suffix family size, given that all allowable combinations should be equally possible and should not vary according to any measure of frequency of occurrence (including cumulative frequency and family size). An assumption being made for the purposes of this study is that there exist in the English language the word categories of noun, verb, and adjective. They are used here as convenient labels for the word types under study.

Derivation is of greater interest in this study because morphological violations of selectional restrictions are being studied, and in the single-word presentation of a lexical decision experiment, it is easier to generate pseudowords composed of real bases and real suffixes that obviously violate these restrictions with derivational affixes. For example, while it would be structurally incorrect to use a nominal pluralizing suffix on a verb, in English it would be difficult to tell the difference between a written *–s* indicating a plural form and a written *–s* indicating the third person singular on a verb in the present tense (e.g., the "s" in "dogs" is theoretically distinct from the "s" in "he cooks," but it is phonetically and orthographically difficult to tell them apart in isolation). There are also a larger number of derivational affixes in English, so there are more items that can be manipulated.

The decision to use suffix family size as a measure of morphological activity is directly related Schreuder and Baayen's (1997) findings that the processing of nouns was affected by root family size and not cumulative frequency of all the forms (which was repeated with complex words). If a suffix is treated as the main unit of interest, and in doing so it is treated like a base, then one might propose the number of family members a suffix has should correlate with reaction times. In lexical decision experiments using root+suffix novel forms, one would expect longer RTs and higher error rates with suffixes of higher family sizes, because it should be more difficult for participants to reject pseudowords that contain more familiar forms, with familiarity being indicated by family size. Similarly, if participants are sensitive to Productivity as measured by Baayen (1992), then we might expect longer RTs and higher error rates when a suffix has a higher *P* value.

The following experiments investigate the role of suffix family size, suffix productivity, and selectional restrictions in novel root+suffix combinations in English. Lexical decision experiments have been used to maintain comparability with other experiments of lexical processing in the field of psycholinguistics. The task is well-known and easy to replicate with the right tools. The morphological decision task used by Järvikivi and Niemi (1999), which appears to be sensitive to word-internal morphological structure, is exploited in the final experiment of this set. It was felt necessary to use a task that would potentially be more sensitive to morphological structure. Two questionnaires have also been completed by participants, one of which addresses semantic interpretability and the other addressing overt knowledge of selectional restrictions.

Methodological Approach

The roles of selectional restrictions, suffix family size, and suffix productivity were investigated through the use of lexical decision experiments, offline questionnaires, and a morphological decision task. The main questions to be addressed through this research were: are lexical decision latencies affected by suffix family size or suffix productivity, are speakers sensitive to selectional restrictions, as evidenced through RT and error rates, and do these factors interact? The experiments were carried out at the Centre for Comparative Psycholinguistics at the University of Alberta. There were four experiments and two questionnaires in total.

The first step in investigating these questions was to run a pilot study, which consisted of two lexical decision tasks and one offline rating.

Pilot Study 1a: Lexical Decision

The first experiment in the pilot study was a simple lexical decision task. In this task, morphological legality, or morphological fit, was manipulated in a series of real morpheme base+suffix pairs. Suffix family size was also manipulated, with two categories (high suffix family size and low suffix family size) being defined through CELEX counts. In these experiments, suffix productivity was held low. This experiment set out to test whether participants were sensitive to selectional restrictions through the manipulation of base+suffix pairs, where the pairs were separated into "Match" (followed selectional restrictions) and "Mismatch" (violated selectional restrictions) conditions. In the pilot study, four suffixes were used, two of which had a high family size, and two that had a low family size. They were divided in this way so that potential effects of suffix family size could become apparent. If suffix family size is relevant to lexical processing, then we might expect novel base+suffix combinations containing a suffix with high family frequency to generate more errors because they are more likely to be incorrectly analyzed as real words. Likewise, for items that are correctly rejected as nonwords in a lexical decision experiment, we would expect items containing a suffix of high family size to take longer to reject. If speakers are sensitive to selectional restrictions specifying the base category to which a suffix should attach, then we would expect longer lexical decision latencies and more errors to morphologically possible (legal) combinations, specified in the "Match" condition. Together, the items that should be responded to least accurately and with the longest reaction times are those items composed of a lexically specified matching base and a high family size suffix.

Participants

Participants were recruited from the University of Alberta undergraduate population. Forty-eight native speakers of English, naïve to the purpose of the experiment were paid to participate. Their ages ranged from 18-65. Two participants were removed because of extremely slow response time (more than 2 standard deviations above the mean) and one other participant was removed because of a high number of overall errors (>30%).

Apparatus

The experiment was carried out on an Apple Macintosh computer using a program scripted in PsyScope 1.2.5.

Materials

Two-hundred and fifty-six novel pseudowords were created by combining existing words and suffixes. Forty-eight base nouns, 8 base verbs and 8 base adjectives were selected from the CELEX lexical database (Baayen et al., 1995) based upon length, family size, and frequency. Suffixes were selected from a list of 50 English suffixes for which *P* values have been calculated (Hay & Baayen, 2002).

Base Selection. Bases consisted of single syllable roots (nouns, verbs, and adjectives) with an orthographic CVCC structure where possible. The CVCC structure was used to control for vowel-induced phonological changes (e.g., resulting from the addition of a suffix beginning with a vowel). The rationale for having all bases combined with each of the four selected suffixes was to have each base serve as its own experimental control. Noun bases had a CELEX lemma lexical frequency ranging from 3 (*cusp*) to 1841 (*bird*), with an average lemma frequency of 281.67. Family size varied from 0 to 37. Verb bases ranged in lemma frequency from 36 to 883, with an average of 274.75, and adjective bases ranged from 11 to 1205, with an average lemma frequency of 269.75. See Table 1 for a summary of frequency and family size by base class. Choosing words from among relatively frequent tokens ensured that the bases would be recognized as corresponding to existing words of English. Each base word was monomorphemic and was rated in CELEX as an unlikely candidate for conversion (CELEX lists whether words are subject to conversion, and these items were not, according to their lists). See Appendix A for a list of all word bases and base+suffix combinations.

	Average Frequency	Family Size Lower limit	Family Size Upper Limit	Average Family Size
Noun	281.67	0	37	5.44
Verb	274.75	0	7	2.88
Adjective	269.75	1	9	4.88

Table 1: Frequency and Family Size by base class in Experiment 1

Suffix Selection. Suffixes were selected based upon family size, productivity (P), and word class attachment (i.e., the word class to which the suffix generally attaches). Suffix family size was calculated by counting the number of words in CELEX containing the suffix. Two of the suffixes used in Experiments 1 and 2 are of relatively low family size, while two have a relatively high family size. Productivity (P) was kept relatively low (ranging from P=0 to P= 0.002). The four suffixes selected were: *-ic* (Family Size = 726), *-ling* (Family Size = 24), *-ation* (Family Size = 543), and *–th* (Family Size = 169). Family size for *–th* may appear high, but it was the lowest Family Size available for suffixes that did not combine with nouns. See Table 2 for a summary of these statistics.

Suffix	Family Size	Family Size Category	Productivity (*P*)	Derivation (Base class → resulting class)	Example
-ation	543	High	.0001	V→N	colourize+ation→ colourization
-ic	726	High	0.002	N→A	cube+ic→cubic
-ling	24	Low	0	primarily N→N; some examples of N→A, A→N, V→N	duck+ling→duckling
-th	169	Low	0	Primarily A→N, some V→N	warm+th→warmth

Table 2: Summary of suffix properties in Experiment 1

Novel Stimulus Creation

Critical stimuli were created by combining each category base (e.g., noun) with each suffix. For example, the base *wasp* was suffixed to become **waspic*, **waspling*

**waspation,* and **waspth.* Two of the suffixes are legal according to word class restrictions while the other two are not (*-ic* and *–ling* suffix to nouns, so although **waspic* does not occur in common speech, it does not violate word class restriction like the combination of *wasp+ation*). There were a total of 256 target stimuli created (see Appendix A for word lists). Forty-eight were base nouns. Eight verb bases were suffixed with *–ation* and eight adjective bases were suffixed with *–th* to provide a comparison to the N-ation and N-th categories that did not violate selectional restrictions. Table 3 summarizes the combinations of lexical category and suffix family size that were present in the experiment.

		Suffix Family Size: High	Suffix Family Size: Low
Morphological Fit	Good	N+*ic* V+*ation*	N+*ling* A+*th*
	Bad	N+*ation*	N+*th*

Table 3: Suffix Family Size and Morphological Fit

Created stimuli were checked against the CELEX database and the Concise Oxford English Dictionary (Soanes, 2004) to determine whether or not the created items were commonly accepted words. *Porkling* and *fistic* both have entries in the OED, although they do not occur in CELEX, and results from early tests show that they are not widely considered to be acceptable words by speakers. The items were broken down into four experimental lists.

Procedure

Participants were seated at a comfortable viewing distance from an LCD iMac computer monitor. They were asked to answer, yes or no, as marked on the keyboard, whether or not a presented stimulus (word, nonword, or pseudoword target) was a real word of English. Stimuli were presented on the computer screen, one word at a time. An asterisk was presented in the middle of the screen at the beginning of each trial for 50 milliseconds, followed by a 50 millisecond blank screen, which was followed by the stimulus. Participants were asked to answer as quickly as possible while remaining accurate. When the subject responded, the screen cleared and the next trial began.

Results & Discussion

Prior to beginning data analysis for Experiment 1, all individual items with a response time faster than 350 ms or slower than 3000 ms were removed. Response times less than 350 ms were considered to be errors. Response times over 3000 ms may not necessarily have been errors, but given the timeframe, it is unlikely that they reflect the same type of automatic processing as indicated in faster decision latencies. Error rates were then calculated by subject and by item. Forty-five of the target pseudowords had error rates over 25%. Of the 45 of targets that had error rates greater than 25%, 15 took the suffix *–ation* (1 verb base; 14 noun bases); 8 were from the N-*ic* category; 20 took the suffix *–ling*, and one took *–th*. However, for the *-th* item, the construction of the target items resulted in a pseudoword that could have been mistaken for a real word (/tent+th/, which could have been mistaken for "tenth"). In the N-*ation* category, percent errors varied from 27% to 70%. In the N-*ic* category, error rates varied from 28-50%. In the N-*ling*, they varied from 28% to 70%. Although these made up a significant portion of the target items (approximately 21%), these items were left in the analysis because they spoke directly to the questions under study. In lexical decision, where the participant is asked to quickly decide whether an item is a word, more errors should be expected when targets are more word-like. While there will be genuine errors in the data, the large number of errors across stimuli speaks to another reason for participant error, and we can hypothesize that this source is the extent to which a target is like a word. Keeping these items in the analysis provides a more accurate description of participant behaviour than removing them.

Processing difficulty may be realized through increased reaction times or through increased error rates. The two main questions driving this pilot study were whether or not, when productivity is kept near-constant, native speakers of English are sensitive to violations of selectional restrictions and whether they are sensitive to suffix family size. One could propose that novel base+suffix pairs that do not violate selectional restrictions should be harder for participants to reject, since they could be considered more viable morphologically. Likewise, a higher suffix family size might be expected to increase the difficulty of the lexical decision task, making it harder for base+suffix pairs with high suffix family size to be rejected.

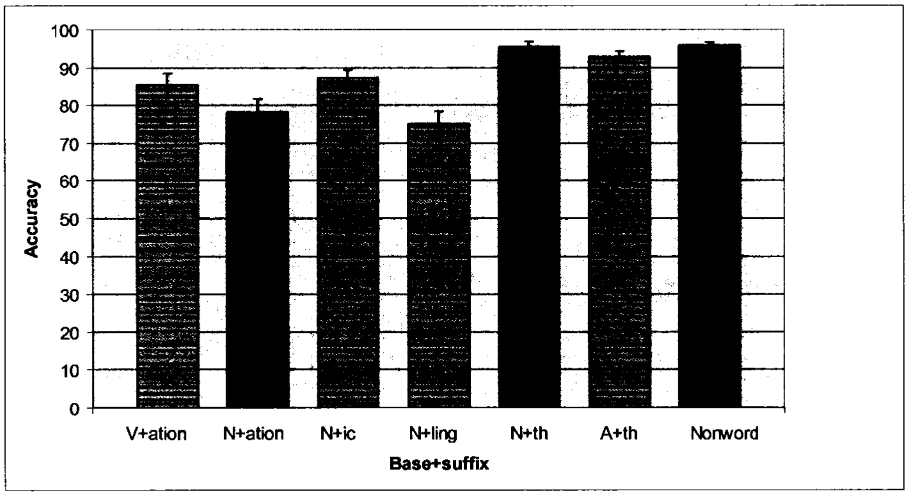

Figure 1: Errors by Base+Suffix Combination in Experiment 1 (Subject Analysis). The low family size suffix *–ling* showed the greatest number of errors, exceeding the high family size *–ic* and *–ation*. Low family size *–th* patterned like nonwords in Experiment 1.

Here, a lower score in a category means that participants answered incorrectly more often than for categories with higher scores. Both the Word (not pictured) and Nonword control categories are at near-ceiling levels of accuracy. Since a different response was required for each of these categories ("no" for "nonword" and "yes" for "word"), these results shows that participants were actively engaged in the task, and were not responding randomly.

If both selectional restrictions and suffix family size are relevant to morphological processing, then the most difficult items would be expected to be those items that do not violate selectional restrictions and that have a high suffix family size. There should be more errors found under these conditions. In this experiment, these criteria are met by the V+*ation* and N+*ic* categories. However, the N+*ic* category scores fairly highly in accuracy, above the morphologically ill-fitting N+*ation*. It also scores well above the N+*ling* category (morphologically legal, low suffix family size), which is the opposite of what would be predicted by suffix family size if it were to affect speaker judgments of wordness. The N+*ling* category shows the most errors of all items. While this might be predicted by the good morphological fit of nouns with *–ling*, it does not explain why *–ling* should show more errors than both the high family size conditions. That the *–th* categories do not also show higher error rates suggests that there is not an inverse relationship between error rates and suffix family size.

The difference between the N-*ation* (ill-fit, high suffix family size) and V-*ation* (good fit, high suffix family size) categories was found to be significant, F_1 (1, 44) = 12.28, p = .0011, with more errors made in the N+*ation* condition. N+*ation* and V+*ation* show this pattern in the items analysis as well, but not to a statistically significant degree.

If selectional restrictions affected perceived wordness, then we would expect to see the opposite effect here, since V+ation is an allowable combination. Meanwhile, the difference between the N+*th* (morphologically illegal/bad fit, low suffix family size) and A+*th* (morphologically legal/good fit, low suffix family size) conditions showed almost no difference in accuracy. If the violation of selectional restrictions made target items easier to reject, then N+*th* would be expected to show a higher degree of accuracy than A+*th*. From Figure 1, it appears as if there might be a trend in this direction, but the difference between the categories is not statistically significant, $F_1(1, 44) = 1.20$, $p = .28$. Additionally, both *–th* categories fall just below the control error rates. The suffix *–th* shows some behaviour that is more consistent with nonword interpretation. There is no statistical difference in the error rates of the *–th* suffixed items and nonword controls, $F_1(1, 44) = 1.58$, $p = .21$. Results were equivalent in the items analysis. It does not appear that low family suffix size is involved in this behaviour, given that *–ling* shows a high error rate. It may be that *–th* is more difficult for participants to recognize automatically due to other properties, such as its syllable structure and related parsability. Analyses using *–th* must be tempered by this possibility.

In order to run a repeated measures ANOVA across the morphological fit and family size conditions, morphological fit was collapsed across base type. This resulted in two new categories that were composed of different base+suffix combinations. The new categories were High Family Size/Good Fit (N+*ic*, V+*ation*) and Low Family Size/Good Fit (N+*ling*, A+*th*). These categories correspond to those given in Table 3. Bad morphological fit (morphologically illegal combinations of base+suffix) did not need to be averaged. This repeated measures ANOVA (Figure 2) revealed a significant effect of

Family Size $F_1(1, 44) = 10.86$, $p = .0019$ and a significant interaction effect between Family Size and Morphological Fit, $F_1(1, 44) = 36.71$, $p<.0001$. Morphological Fit alone was not found to be significant. An ANOVA run on the items supported these results. Results must be cautiously taken, however, as the *–th* suffix may not be acting as a real suffix in these data, and instead might be regarded as a nonsense sequence. In addition, there were fewer tokens of the V+*ation* and A+*th*, compared to the numbers of N+suffix categories.

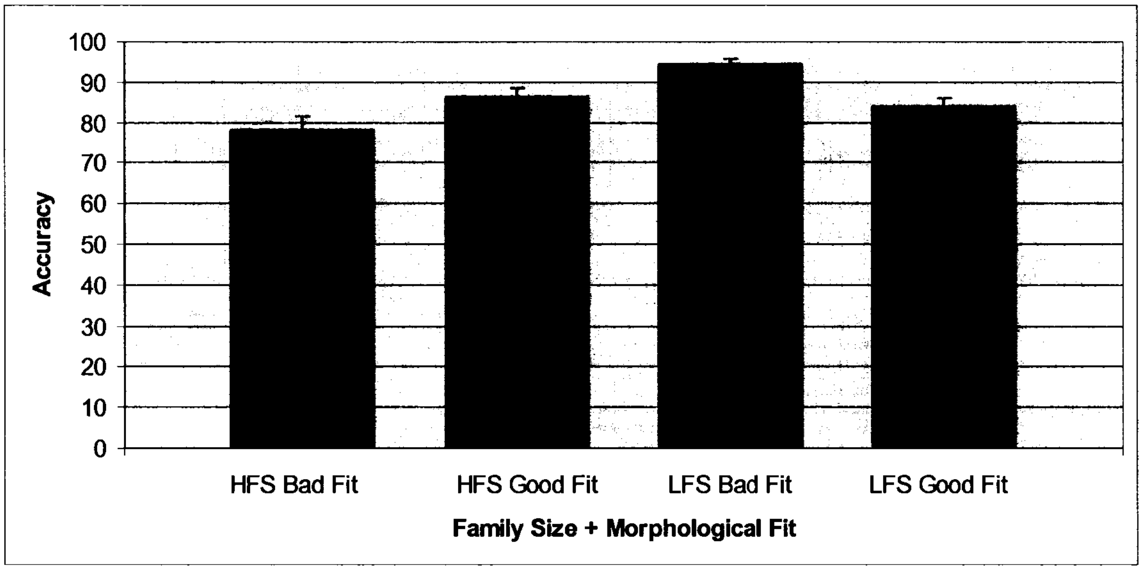

Figure 2: Repeated Measures ANOVA of Family Size and Morphological Fit in Experiment 2: There was a significant difference between the number of error generated in the high and low family size conditions, but violation of selectional restrictions did not appear to affect error rate across conditions.

Matters are complicated, however, since a similar Repeated Measures ANOVA run on the noun categories alone (N+*ation*, N+*ic*, N+*ling*, N+*th*) showed the opposite effect. Family Size was not found to be significant. Instead, Morphological Fit was found to be significant, $F_1(1, 44) = 9.75$, $p = .0032$. An interaction between Morphological Fit and Suffix Family Size remains, $F_1 (1, 44) = 38.04$, $p<.0001$. There are concerns about this data because *N-th* is not balanced by the inclusion of *A+th* in another group, and as already mentioned, items taking the *–th* suffix appear to behave more like nonwords. Thus, an effect of morphological fit based on the comparison of what could be a nonword and *–ling* might unfairly bias the results towards significance.

In summary, from the error data, it does not appear that morphological fit (good or bad) influenced these results consistently or in the predicted fashion. It is more difficult to explain the suffix family size data. The suffix *–ic* has a higher suffix family size than *–ling* (and *–ation*), but shows fewer errors, and the low family size *–th* shows fewer error still. Moving to an examination of RT may help to clear up the picture.

Reaction time was the secondary measure in this experiment because it is in the errors that participant interpretation of the word-status of an item is directly assessed. Reaction time data is still valuable, however, since items that are correctly perceived as nonwords, but that could be mistaken more easily for words, would be expected to take longer to reject. The main question in the analysis of reaction time is whether the lexical decision latency is affected by either suffix family size, morphological fit, or a combination thereof. Analysis of reaction time data was constrained by item length, so that the critical pairs of stimuli were V-*ation*/N-*ation*, A-*th*/N-*th*, and N-*ic*/A-*th* (Figure 3).

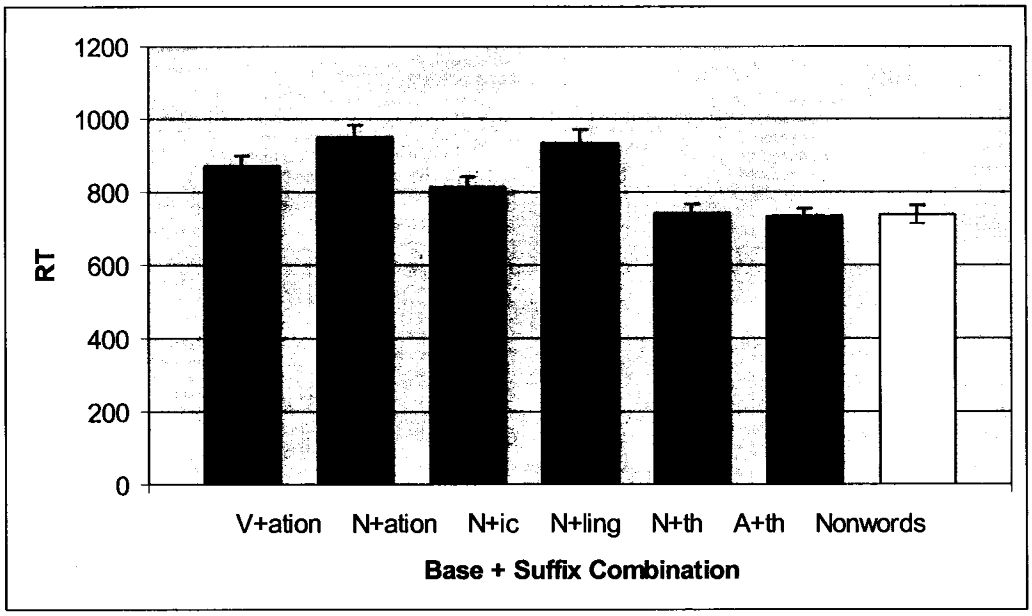

Figure 3: Reaction time by Base+Suffix category in Experiment 2 (Subject Analysis): Reaction times were longest in the *–ling* and *–ation* conditions. Participants responded more quickly to V+*ation* than N+*ation* targets. Items in the *–th* category patterned like nonwords in RT.

For the first comparison, between V-*ation* and N-*ation*, a planned comparison ANOVA found a significant difference between the N-*ation* (bad morphological fit) and V-*ation* (good morphological fit) categories, $F_1(1, 43) = 15.87$, $p = .0003$. Nouns took significantly longer to be responded to than verbs when combined with a morphologically illegal form (*-ation)*. Results in the items analysis neared significance ($p=.0505$). It was originally predicted that items violating selectional restrictions would be easier for participants to reject, and would therefore be expected to show faster lexical decision latencies. These results do not support this hypothesis, but further testing is necessary to ascertain whether this effect is reliable or a result of the unequal number of nouns and verbs.

A repeated measure ANOVA run on N-*th* (morphologically illegal/bad fit) and A-*th* (morphologically legal/good fit) revealed no significant differences in RT, $F_1(1, 44) =$

.58, p=.45. Similar results were found in an analysis of the items (p=.88). When reaction times for *–th* suffixed items were compared with reaction times for nonwords, there was no statistical difference found in either the subject analysis F_1 (2, 44) = .48, p=.62, or the items analysis, F_2 (2, 149) = .47, p=.63. This supports the idea that participants may treat items taking the *–th* suffix no differently than nonwords.

The maximally different *–ic* (attaches to nouns, high family size) and *–th* (does not attach to nouns, low family size) were compared in a repeated measures ANOVA. The pseudowords suffixed with *–th* were rejected significantly more quickly than pseudowords suffixed with *–ic* in both the subject and items analyses, $F_1(1, 44) = 14.66$, $p = .0004$, $F_2(1, 94) = 8.46$, $p = .0045$. However, since these two suffixes are very different, it is unclear which effect, or what combinations of effects, is responsible. No such effect was found for combinations of nouns with *–ation* (high family size, bad fit) and *–ling* (low family size, good fit).

A Repeated Measures ANOVA was also run to determine whether there was an interaction between morphological fit and family size. As in the error analysis, the high family size-good fit and low family size, good fit conditions were collapsed. Family size alone was found to be a significant factor influencing RT, $F_1(1, 43) = 67.56$, $p < .0001$. There was also an interaction effect found between suffix family size and morphological fit, $F_1(1, 43) = 69.516$, $p < .0001$. These results were supported in the items analysis. However, the effect of word length is a serious concern in these data. Direct comparison between N-*ation* (bad morphological fit/morphologically illegal, high suffix family size) and N-*ic* (good morphological fit/morphologically legal, high suffix family size), both high family size conditions, was not possible because of difference in letter length. The same is true for the low family size conditions, N-*ling* and N-*th*. Comparisons that combine the suffixes, as in the Repeated Measure ANOVA above, were possible where lengths could be averaged. In the nonword data, word lengths of 8 and 9 letters were not found to be significantly different in reaction times, $F_2(1, 38) = 2.80$, $p = .10$, but results relying on averaging should still be seen as preliminary.

Overall, it appears that participants were sensitive to suffix family size, but not in the way predicted. Participants made an unexpectedly large number of errors that did not appear to correlate with suffix family size or morphological fit. Items containing the low family size suffix *–ling* appeared to be the most difficult for participants to reject as nonwords, although *-ling* has a low family size suffix. Contrary to expectations, N+*ation* combinations generated more errors than V+*ation* combinations, even though the V+*ation* combinations are morphological legal combinations, where the N+*ation* combinations are not. N+*ation* combinations also took longer for participants to reject than V+*ation* combinations.

Reaction time correlated highly with length, but preliminary results suggest an effect of family size. Selectional restrictions (morphological legality/fit) were not consistently found to be significant. It is unclear what the role of selectional restrictions are and how they interact with suffix family size.

The suffixes used were carefully selected based upon family size, selectional restrictions, and productivity, but a number of questions remain unanswered in this pilot study. With only four suffixes, it is difficult to generalize results to other suffixes. Although there appears to be an interaction between suffix family size and morphological fit, this result must be tempered by the behaviour of *–th*, which appears to pattern with nonwords. Functionally, this leaves *–ling* as the only representative of low suffix family size. The suffix *–ling* showed the highest number of errors, and since there are no other items to directly compare *–ling* against, it is unclear why *–ling* was such a difficult suffix to process. There is no way to know from just this pilot whether this behaviour is restricted to *–ling*, or if other suffixes with a low family size will behave in the same way. There are also problems in the interpretation of the *–ation* data, where N+*ation* combinations appeared to be more difficult for participants, due to the small number of V+*ation* items.

What is clear from these data is that individual suffixes can be processed very differently. With the exception of *–th*, it appears that participants are sensitive to the presence of suffixes as viable morphological constituents, as can be seen through increased error rates and reaction times. Since errors and reaction times were not predictable solely based upon violations of selectional restrictions and suffix family size, it is likely that other factors influenced participant responses. One factor that might be involved in semantic interpretability. In the next experiment, an offline task, this possibility was addressed.

Experiment 1b: Acceptability Ratings

This offline questionnaire asked participants to rate how 'acceptable' they found novel base+suffix combinations as words of English. It was completed by participants following the online tasks of Experiments 1 and 2, and by students in a classroom setting. The main question here was whether or not violations of selectional restrictions influenced the acceptability ratings for the target items used in the first experiment. The goal was to assess the acceptability of possible complex words (words not violating word structure rules) and morphologically illegal words (those that do violate word structure rules) without referencing their identity as novel forms. Given that the same items were used as in Experiment 1a, suffix family size was also investigated. As participant response time was not recorded, effects relating to word length were not expected to influence results. Additionally, since the questionnaire was not asking for a bald judgment with respect to whether or not an item was a word, but rather asked for an opinion about a word-form, greater variation in acceptability was expected.

Participants

Participants were native English speakers both recruited from an Introductory Linguistics class at the University of Alberta and recruited to the Centre of Comparative Psycholinguistics. There were a total of 172 participants. 130 students participated during a classroom session, while 42 were tested at the CCP during an hour-long session. Participants at the CCP also took part in experiments 1 and 2 before answering this questionnaire. Laboratory participants were paid $10 for their time. In either case, participants were assured that their participation would in no way influence their grades or academic standing at the University of Alberta. The questionnaire consisted of 80 items, 20 of which were critical stimuli. There were four separate lists. Fillers consisted of real words of varying morphological complexity (e.g., simplex: wash, complex: colourize), nonwords without a morphological structure, and the complex pseudowords used in Experiments 1 (e.g., ?*waspling*).

Apparatus

The experiment was a pencil and paper task. No special equipment was necessary.

Materials

Target materials consisted of the same items used in Experiment 1 (see Appendix B for an example questionnaire).

Procedure

In the offline rating task, participants were asked to rate how acceptable the target items were as words of English, using the same list of words as in Experiment 1. No further instruction was given, and participants were not given any indication of base-suffix mismatches. No participants were excluded.

Results and Discussion

The highest-rated suffix was *–ling* and the lowest rated was *–th*. Between *–ation* and *–ic*, *-ation* was rated to be more acceptable in combination with noun bases than *–ic*. The complete order was, from highest to lowest: *-ling, -ation, -ic, th*. Figure 4 displays the acceptability of N+suffix combinations by suffix.

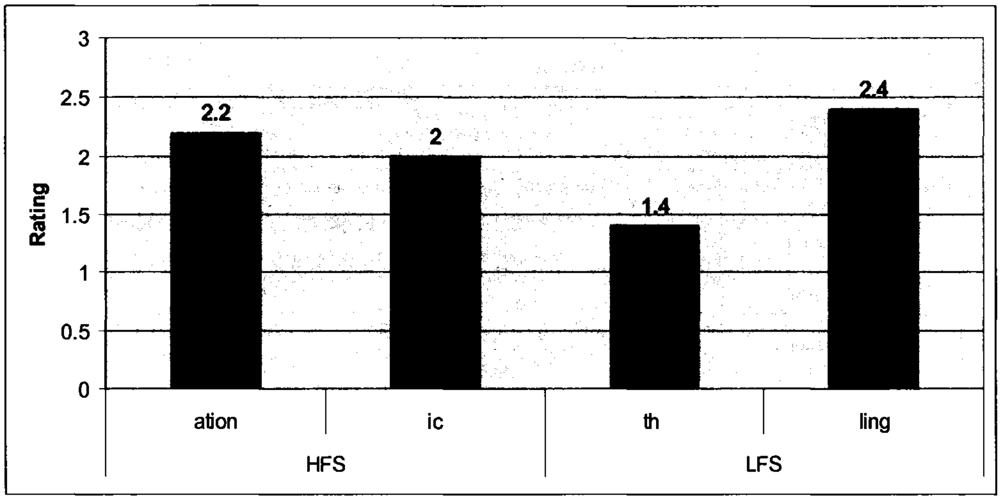

Figure 4: Acceptability ratings of N+suffix targets in Experiment 1b: Participants found N+*ling* combinations to be most acceptable, and N+*th* combinations to be the least acceptable. These results do not conform to a hypothesis where suffix family size alone influences the ability of a suffix to meaningfully combine with a base form.

This result is opposite what might be expected if one only looked at family size, as *–ling* has significantly higher ratings for acceptability than *–ic*, but it is similar to the error data found in Experiment 1a.

These results indicate that one would not expect this to be solely an effect of length, where the longer a word is, the more acceptable it is found, as *–ic* and *–th* are the same length and there is a large difference between them. It is also worth noting that *-ling* still does have a higher rating and is the shorter suffix of the two. More interestingly, it does not appear that participants were influenced by the mismatched category of each suffix (Figure 5). It appears likely that semantic interpretability played a role in acceptability

decisions.

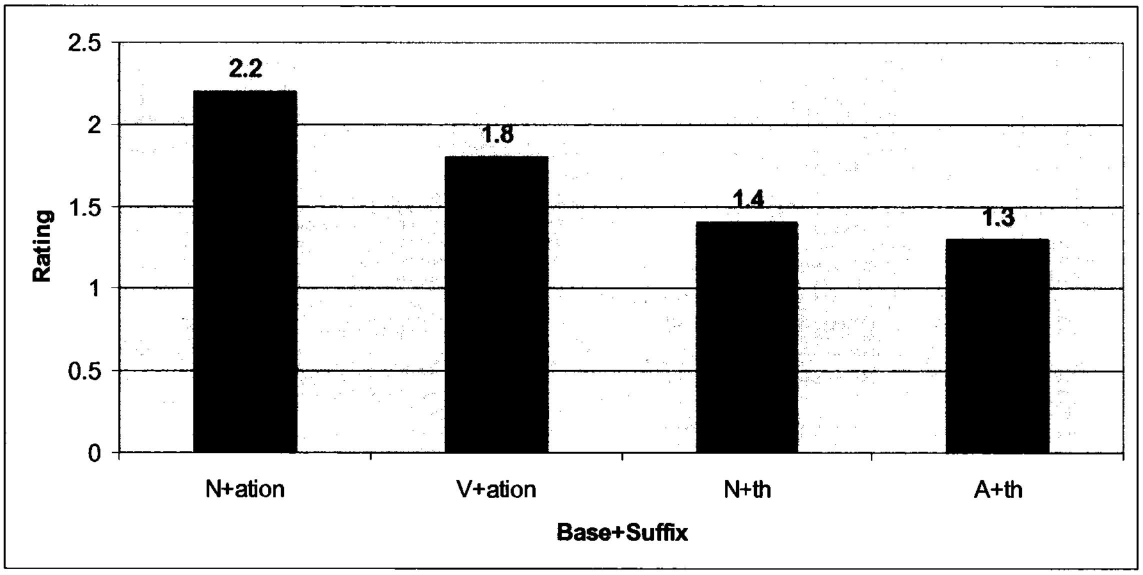

Figure 5: Acceptability Ratings by Category Match/Mismatch in Experiment 1b: Nouns were more acceptable in combination with suffixes than the matching category V and A, although there was not much difference between base types for *–th*.

Also included in the stimuli were 8 verb+*ation* and 8 A+*th* items. When base+suffix combinations are separated by base category, it is apparent that selectional restrictions are not strongly influencing acceptability. There is little difference between base types with the suffix *–th*, although N+*th* is rated slightly more acceptable. Nouns suffixed with *–ation*, which one would predict to be less acceptable if participants were being influenced by selectional restrictions, are actually rated more acceptable than V+*ation*. However, because of the small number of V and A bases used, it is possible that there were more semantically viable items included in the Noun base set, and therefore the results could be caused by the semantics of the bases, and not the affix properties.

Overall, the responses to the questionnaire reveal that, while some items were more acceptable than others, there was no clear difference based upon categorical information or base+suffix mismatches. The two suffixes *–ling* and *–ation* were rated the most highly. This result is similar to the error data in Experiment 1a. There was no clear difference based upon suffix family size or base category. It is interesting to note that pseudowords with noun bases were rated more highly than the expected, correctly classed items.

To get a sense of what effect acceptability could have had during the lexical decision experiment in Experiment 1, item acceptability ratings were plotted against error rates (Figure 6) and reaction times (Figure 7). When acceptability ratings were plotted against error rates, there was a ceiling effect where items neared 100% accuracy. There appears to be a relationship between acceptability and number of errors. When the relationship between accuracy and acceptability ratings was examined, the Spearman Rho Correlation was found to be highly significant ($p<.0001$) at -.447, where accuracy increases while acceptability decreases. The suffix *–th* clusters at the upper limit of accuracy, recalling that *-th* was the easiest of the suffixes to reject, and shows the lowest level of acceptability. The suffixes *–ic*, *-ation*, and *-ling* show greater variability, but overall have lower acceptability ratings at lower error rates.

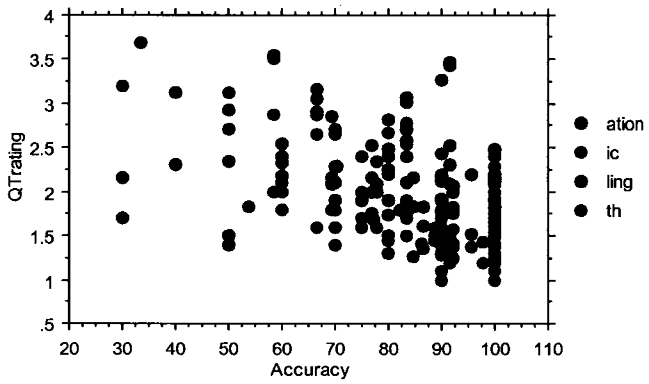

Figure 6: Questionnaire acceptability rating vs. accuracy (% correct) in Experiment 1: *-th* suffixed items show the highest accuracy and lowest acceptability ratings. The greatest spread in accuracy and ratings is seen in the *–ling* and *–ation* conditions. The Spearman Rank Correlation between rating and accuracy was highly significant.

Reaction time for correct responses was also compared against acceptability ratings. It appears that there is some relation between higher acceptability ratings and slower reaction times (Figure 7). This correlation was found to be highly significant ($p<.001$), with a Spearman Rank Correlation of .514. The graph below is separated by suffix, and it is apparent that *-ling* and *-ation* in general take have longer RTs than *–ic* and *–th*, as would be expected based on item length. Although there is not a significant difference in the reaction time between *-ation* and *-ling,* the graph shows a trend towards long RTs to *-ling* suffixed items.

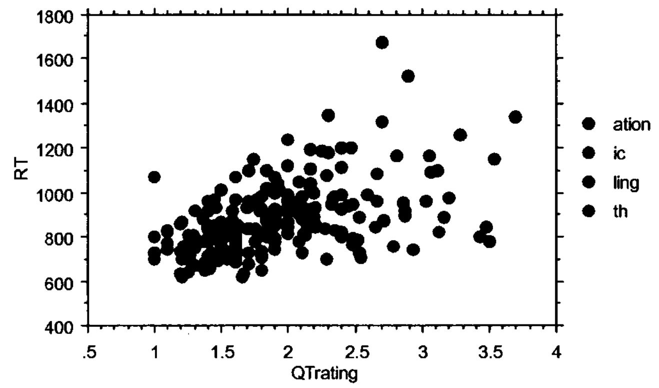

Figure 7: Reaction time and Acceptability ratings in Experiment 1: *-th* items cluster at the lowest RTs and acceptability ratings. Within the *–ic* and *–ation* conditions, there is a general trend towards longer RTs at higher acceptability ratings. Items suffixed with *–ling* show the greatest spread in RTs and ratings. Overall, there is a significant positive correlation between acceptability ratings and reaction time.

It is worth noting that the highest acceptability ratings were in the *-ation* and *-ling* suffixes, and it is these two suffixes that see the widest spread across RT and items with higher errors. Note that the *–th* suffix clusters at a relatively low RT and a relatively low acceptability rating, whereas *–ic* is spread further along the acceptability scale. This suggests that difficulty evident in the error rates and, to a lesser extent, reaction times, is not just a function of letter length, and that something else was making this task difficult. It is possible that unintended semantic factors related to base forms and the ability of bases to combine with suffixes were influencing reaction times, and more obviously, error rates in the first experiment.

If semantic interactions between suffix and base influenced results in Experiment 1a, then one should theoretically be able to remove the factor of base meaning by using nonword bases in the stead of real words. In Experiment 2, the same suffixes (*-ation, -ic, -ling, -th*) were combined with nonword bases to see whether there was an effect of the base on processing.

Experiment 2: Nonword Lexical Decision

As in Experiment 1, this was a lexical decision experiment. Unlike Experiment 1, suffixed nonwords were the items of interest. Selectional restrictions were momentarily put aside to see whether family size alone could affect processing. Phonotactically plausible nonword bases were paired with the four suffixes in Experiment 1 (*-ation*, *-ic*, *-ling*, *-th*). If native speakers of English are sensitive to suffix family size, then they might be expected to make more errors on items with a higher suffix family size (*-ation*, *-ic*), and to respond to them more slowly when answered correctly.

Participants

Forty-three participants took part in this experiment. One participant was removed for an excessive error rate (>30%). Three participants were removed for an overall performance (reaction time) being more than two standard deviations above the mean reaction time for the group.

Apparatus

Experiment 2 was run on the same equipment as Experiment 1.

Materials

The target nonwords in this experiment were created from the stimuli used in Experiment 1. In this experiment, each of the target items from Experiment 1 was changed by one phonetic feature. Each base word was altered by changing either the voicing of one segment or the place of articulation of one segment (e.g., ?*birdling* → **pirdling*). Suffixes were left intact. The experiments were carried out on an Apple computer using PsyScope (1.2.5) scripts.

Procedure

The experiment proceeded as in Experiment 1.

Results and Discussion

Overall, there were far fewer errors in Experiment 2 than in Experiment 1. No items were removed from the analysis. The suffix that showed the lowest accuracy rate was *–ling* (Figure 8). The suffix with the lowest error rate was *–th*. All suffixes were responded to with accuracy greater than 90%. When error rates were compared for high suffix family size (*-ation* and *–ic*) and low suffix family size (*-ling*, *-th*), no significant differences were found, $F_1(1, 38) = .85$, $p=.36$. In an analysis of error rate in the items, suffix family size approached, but did not reach, significance ($p=.07$), with fewer errors in the low suffix family size category. However, the considerations regarding the status of the *–th* suffix remain. It is unclear whether participants recognize *–th* as a suffix when it is presented in a lexical decision task. This might affect calculations regarding family size.

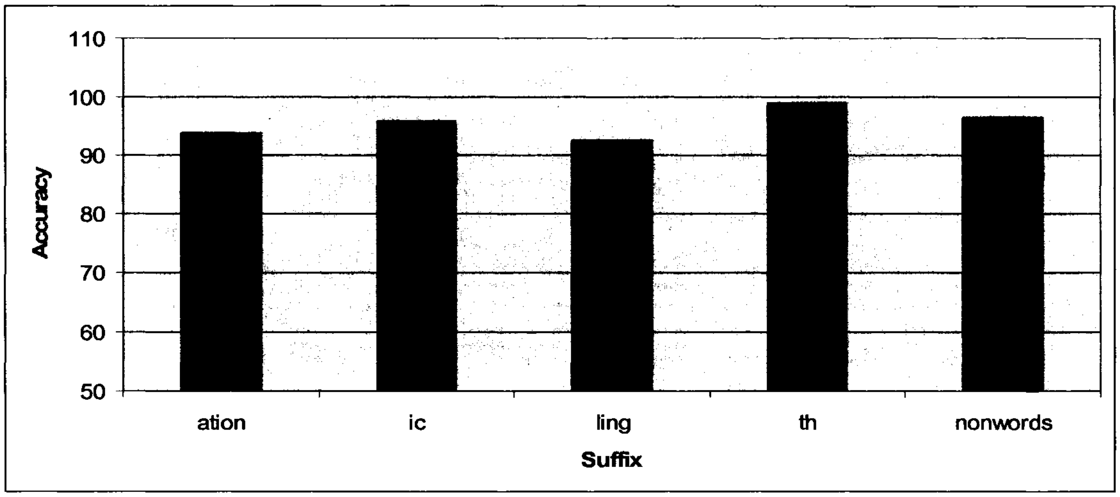

Figure 8: Accuracy by Suffix Type in Experiment 2 (Subject Analysis): All conditions showed over 90% accuracy. *-ling* suffixed items showed the highest number of errors, followed by *-ation*.

A repeated measures ANOVA of all four suffixes revealed that each of the suffixes was responded to differently, although this likely highly correlates with length, $F_1(3, 38) = 23.73$, p=<.0001. The only suffixes that can reasonably be compared to each other based on length are *-ic* and *-th*, although *-ation* and *-ling* are marginal. There was a significant difference found in an Repeated Measures ANOVA for *-ation* and *-ling*, $F_1(1, 38) = 9.8$, p=.003, although *-ation* showed higher RTs, which could be an effect of length. There was no significant difference found between them in the items analysis. In the case of *-ic* and *-th*, there was also a significant difference found between the reaction times to these suffixes, with participants taking longer to reject *–ic* suffixed items, $F_1(1, 38) = 13.89$, p=.0006. This is a more interesting result, as length remained constant. The suffix *–ic* has a higher family size than *–th*, and since there is no morphological fit category in this experiment, it is possible that family size might be affecting these results. Similar results were found in the items analysis. Alternatively, this could be more evidence that *–th* suffixed items are being treated like nonwords.

As in the analysis of Experiment 1, suffixes were collapsed into high and low family size categories. An ANOVA found a significant difference in the reaction times between high and low suffix family size, $F_2(1, 214) = 10.11$, p=.0017. However, it remains suspect due to length. There could be an effect of family size present, but since it took participants longer to respond to words in the high family size category, it is unclear whether the effects are due to family size or to the length of *-ation*. If they are due to family size, then the results are interpretable as a higher family size suffix requiring more effort to reject as a nonword than equivalent words with a low family size suffix.

There is one further type of analysis that can be made, based on reaction time. If it is not clear that differences in suffix family size affect processing, is it at least clear that the presence of suffixes as morphological units affect processing time? In order to examine this without length interfering, it is necessary to take suffixes on an individual basis, and compare them to nonwords of the same length. When ANOVAs were run on the items, only the Nonword-*th* comparison failed to reach significance. Based on these data, it is likely that the *–th* suffix does is not processed as a suffix when it occurs with nonwords.

These results suggest that suffix family size might play a role in processing, with a higher suffix family size correlating with longer reaction times. Unfortunately, it is not completely clear from these data whether that is the case. As in Experiment 1, letter length was a concern in this analysis. The large difference in error rates between Experiments 1 and 2 is proposed to have been caused by semantic effects, possibly the semantic interpretability of the base+suffix combination.

One cannot easily use nonword bases to study selectional restrictions, however, so in order to overcome this potential semantic interference, a new strategy for combining base and suffix pairs was used in Experiments 3 and 4. Furthermore, because of concerns about the generalizability of results, the number of suffixes was dramatically increased. In Experiments 1 and 2, there were only four suffixes. These four suffixes were split into 'high' and 'low' suffix family size, so there were only two suffixes per group. Additonally, the status of one suffix, *-th,* is questionable. With the addition of more suffixes, it is difficult to make a strong distinction between "high" and "low" family size, as suffix family size increases incrementally on a continuum, and is not easily parsed into high and low categories. Even so, the benefits outweigh the costs. By increasing the number of suffixes, it is to overcome possible difficulties caused by the properties of the limited subset used in Experiments 1 and 2, such as the status of *–th* and the high interpretability of *–ling*, as seen in the acceptability ratings. In the first two experiments, an attempt was made to keep the *P* values of the suffixes low while separating them according to 'high suffix family size' and 'low suffix family size.' Even so, some of the suffixes were more productive than others. While this may not have been a problem, we are in a better position in the main experiments to interpret any potential role of productivity. In the main experiments, with a greater number of suffixes, there is more variation and more available data.

CHAPTER 3

THE MAIN EXPERIMENTS

Results from the pilot study suggested that selectional restrictions may not play as important a role as other factors in lexical processing, and were unclear on how family size might affect processing. Results also suggested that there was a semantic factor active in the stimuli that was not fully predictable from either of these two variables. In order to more fully understand the role of suffix family size, productivity, and selectional restrictions, it was necessary to increase the number of suffixes to build a more general view of suffix processing.

In Experiment 3 below, the focus shifted from specific noun-suffix combinations to random target generation, with each participant seeing a new list of target items. Results from Experiment 1 suggested that subjects were sensitive to the semantic potential of base+suffix combinations, such that more 'possible' forms were rated more acceptable in the offline rating task and generated higher error rates in the lexical decision experiment. By randomizing the base+suffix combinations in Experiment 3, the effects of individual base+suffix combinations should not be present to the same extent. When each item is seen only once, the unique semantic combination can only affect results once. While it is likely that some specific combinations will be more acceptable to speakers than others, the results from these combinations should be offset by other results. This is different from effects that may result from suffix properties, as these should remain throughout. The goal of randomization is to remove the effects of the base.

Experiment 3

The goal of Experiment 3 was to increase the number of suffixes and base+suffix combinations, while removing possible unintended semantic effects that might have caused the significant results in Experiment 1 between word class categories. To this end, in Experiment 3 the base+suffix combinations were randomized for each subject. In Experiment 1, there were a total of four conditions, and each conditional had its own, stable, list. Because the combinations were randomized for each subject in Experiment 3, semantic acceptability should have averaged out if one item was more comprehensible than another.

Otherwise, Experiment 3 was a normal lexical decision task. Participants were asked to decide, yes or no, whether a given line of text that appeared on a computer screen was a real word or not.

Participants

Participants were primarily recruited from the University of Alberta undergraduate student body. No participants had participated in the first phase of testing. There were thirty-two participants in total. Two participants were removed due to slow reaction times (over 2 SD above the mean).

Apparatus

Experiment 3 used the same equipment as the previous two experiments.

Materials

Stimuli were created by randomly pairing bases and suffixes prior to experimentation. The number of suffixes was increased from 4 to 37. Many of the same base forms were used as in Experiments 1 and 2, but adjective roots were removed and more verb bases were added. Due to the restrictions on base choice, some of the verbs were capable of undergoing conversion to nouns, as listed in CELEX. Attempts were made to ensure that the frequency of the verb form was higher than the frequency of the noun.

Noun and verb bases were combined with suffixes randomly, such that each suffix appeared twice per list, once with a nominal base and once with a verbal base. Each base was seen only once per participant. The same bases were used as in Experiment 1, with the addition of 17 noun bases and 29 verb bases. The additional bases conformed to the structural constrained outlined in Experiment 1, following a CVCC format, and being equivalent in terms of family size and frequency. The number of suffixes increased to 37 from 4, and included all suffixes with reported *P* values (Hay & Baayen, 2002) that could legally combine with either nouns or verbs (as calculated from CELEX). Family size was recorded, but due to the number of suffixes, was not specifically controlled for. Suffixes that combine exclusively with adjectival bases were excluded (e.g., *-th*), as were suffixes that could be analyzed as full words (creating compounds, such as *–ful and –like*) or that could be analyzed as containing more than one suffix (e.g., *-ery*). Allomorphic suffixes were considered separately because of the method in which their *P* values are reported.

The suffix *–ly* was removed from the analysis. The reasons for this were twofold. First, while between verbs and nouns, there is a preference for *–ly* to combine with nouns, it is far more common for it to occur with adjective to form adverbs, and with neither of the base classes used here. Second, it combines relatively regularly with adjectives, and there have been some comments to the effect that *–ly* may be similarly processed to inflectional affixes, due to this regularity (Bybee, 1985: 84). Most of the suffixes under study cannot be viewed in this way, and while it is possible that inflection and derivation form a continuum (e.g., see Bybee, 1985), there is enough evidence at present to remove that suffix for the time being.

Procedure

Each participant saw a different list. In each list, there were two instances of each affix, one corresponding to the appropriate base class, and one which did not. There was no overlap between lists. The same timing of presentation was used as in Experiments 1 and 2. Errors were removed for reaction time analysis.

Results & Discussion

The three main questions of this line of study are 1) do selectional restrictions, in the form of morphological fit, make a difference in processing, 2) does suffix family size play a role in processing, and 3) does suffix productivity play a role in processing. In order to answer these questions, both error rates and RTs were considered. The same reasoning holds here as in the previous experiments: if an item is more word-like because of its morphological composition, then it should be harder for participants to reject. Are items more word-like when they do not violate selectional restrictions? Are they more word-like when a suffix is more recognizable or salient? Error data was analyzed first to address these questions.

First, error rates across violation of selectional restrictions and base class were analyzed. This was to see, for the first part, whether there was any difference in error rates visible based on whether or not a base violated the selectional restrictions of a suffix and for the second part, whether there were any difference in error rates based on base class itself. It would be useful to know whether participants treated noun and verb bases differently on the whole, as such a result would need to be taken into account for further analyses.

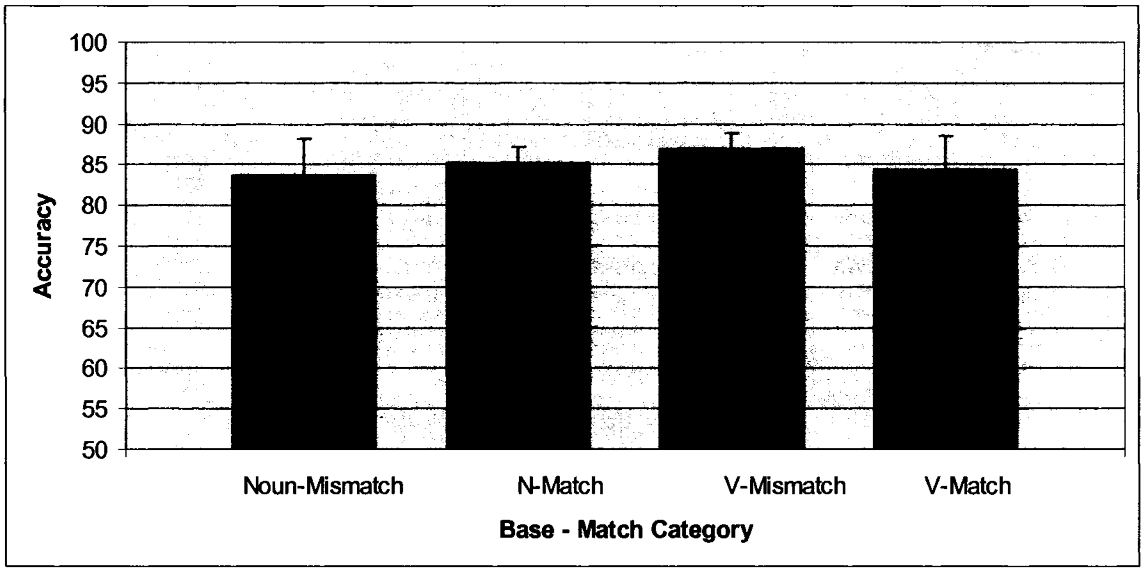

Figure 9: Accuracy for Base –Match/Mismatch Categories in Experiment 3. There is little difference between accuracy scores between the base+match/mismatch categories.

Error rates were calculated by suffix rather than by item, as items were unique to each participant. On average, the V+mismatch category showed the fewest number of errors (Figure 9, above), but there was variation within each category and all categories had average results over 80% accuracy. On the whole, neither base class nor selectional restriction violations were found to significantly affect error rates. There was also no significant interaction between them, $F_2(3, 68) = .56$, $p = .46$. Similar results were found in the subject analysis. The suffixes used in this experiment were categorized into high and low family size classes for preliminary testing. An unpaired T-test found no significant difference between High and Low suffix family size, p=.33.

However, the breakdown between family size categories was arbitrary. Items were split into either the high or low family size category based upon an even numerical split of the items, with half of the 36 suffixes being high family size, and half being low family size. This distribution was problematic, because there was no clear line between high and low categories for the moderately sized suffix families. For example, *-ling* (family size = 22) and *–y* (family size 613) are fairly unambiguous in their classification, as were the suffixes used in Experiments 1 and 2. However, different classifications of suffixes like *–en* (family size = 128, classified as a low family size) and *–ish* (family size = 131, classified as high family size) are harder to justify. Family size is more easily described as a continuous variable.

To reflect that, accuracy was plotted against suffix family size (Figure 10). Morphological fit and base class are collapsed because there was no significant difference found to be caused by either. Most suffixes were responded to at near ceiling levels of accuracy, which makes it difficult to discern possible trends. Near the upper end of family size, it looks as if there might be a trend towards more errors, but the variability throughout the data makes this a tenuous suggestion at best. Although we might expect a greater error rate for *–ness* based upon its higher family size, that higher family size also means the suffix should be treated with caution, as it could be interpreted as an outlier. The suffix *–hood*, which has a low suffix family size (25), might also be considered an outlier, due to the number of errors it generates. It is possible that the orthographic form "hood" was also being interpreted as a word.

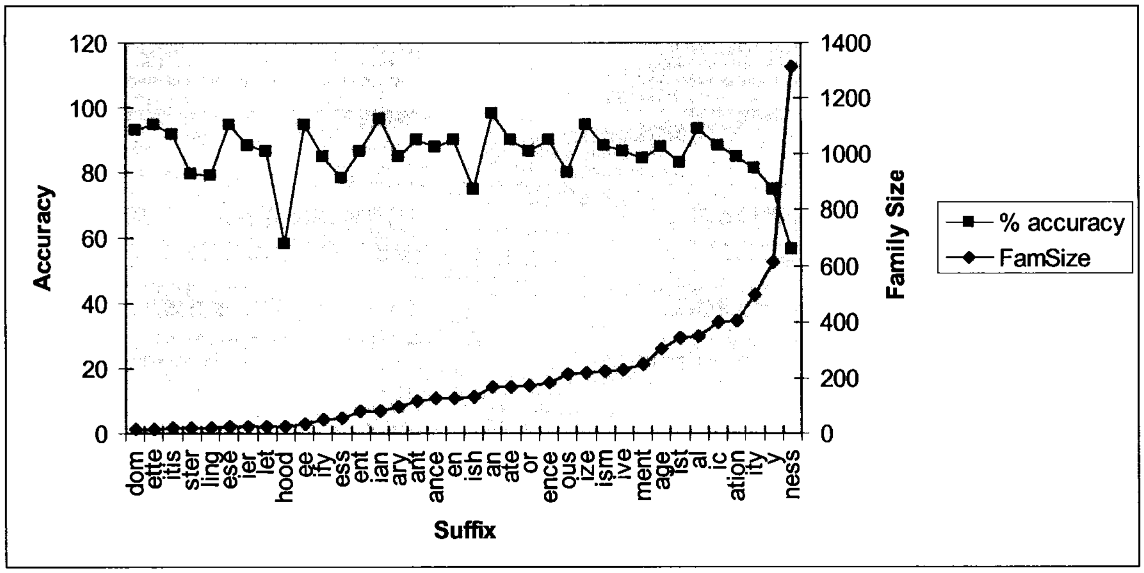

Figure 10: Accuracy by Suffix Family Size in Experiment 3: This graph shows near ceiling levels of accuracy for most suffixes, and may indicate a small trend towards more errors at higher suffix family size.

A similar procedure was followed to analyze the RT data. In comparing RTs based on selectional restrictions and base class, there were no significant differences found between any of the categories, and no interaction was found. $F_2(3, 982) = .97$, $p = .32$. Similarly, no effect for family size was found when suffixes were arbitrarily divided into high and low categories. Suffix family size was also compared, by suffix, with reaction time (Figure 11). Here, there is a descriptive difference between the match and mismatch conditions, where there appears to be a greater effect of suffix family size in the base+suffix match condition such that as family size increases, reaction times also increase to a greater extent than in the mismatch condition. There is a greater correlation between the match RT and suffix family size (cor=.36, $p<.05$) than in the mismatch condition (cor=.09, $p>.05$), such that there is a significant correlation between RT and suffix family size only when selectional restrictions are not violated. This difference is curious, for while the violation of selectional restrictions did not appear to affect RT on its own in a simple ANOVA, it appears as if it can have a descriptive effect when coupled with another factor, in this case, family suffix size. While the lines of best fit on this graph (Figure 11) are influenced by the possible outlier *–ness,* the difference warrants further consideration.

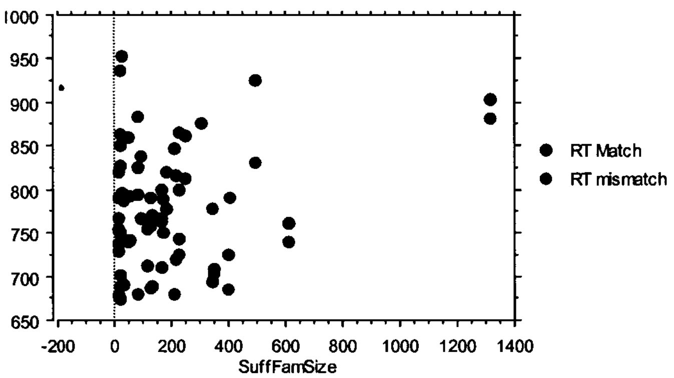

Figure 11: Reaction time (*y*-axis) by Suffix Family Size (*x*-axis) for individual suffixes in Experiment 3. This graph plots the average RT for suffixes by suffix family size, showing the general difference between the match and mismatch conditions. Base+suffix items following selectional restrictions see a higher correlation between RT and SFS, showing a trend towards longer RTs as suffix family size increases.

In consideration of *–ness*, Figures 12a and 12b show that the line of best fit generally predicts the position of the suffix *–ness* when *–ness* is removed from the analysis in the matching RT condition. There is a difference of about 50 ms between the actual placement of *–ness* and its predicted placement in the second graph.

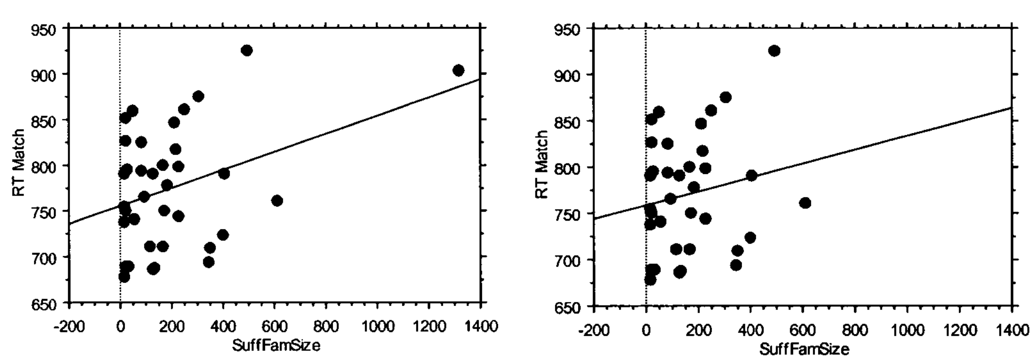

Figure 12a-b: Line of best fit for Reaction time & Suffix Family Size, before and after the removal of *–ness*. The line of best fit serves as a reasonable predictor, even when the possible outlier *–ness* is removed.

In this experiment, there was no effect of selectional restrictions found in either the analysis of the errors or reaction time, although base+suffix combinations in the mismatch condition showed less responsiveness to suffix family size. Likewise, there was no apparent difference based on the lexical category of the root. In comparison to Experiment 1, where a difference was found between reaction times of V+*ation* and N+*ation*, in this experiment, no such difference was found, $F_2(1,26)=.258$, $p=.6159$.

For the initial analysis along the productivity measure, the variable was split into two categories, high productivity and low productivity. While there were slightly fewer errors in the low productivity condition, this difference was not found to be significant (Figure 13).

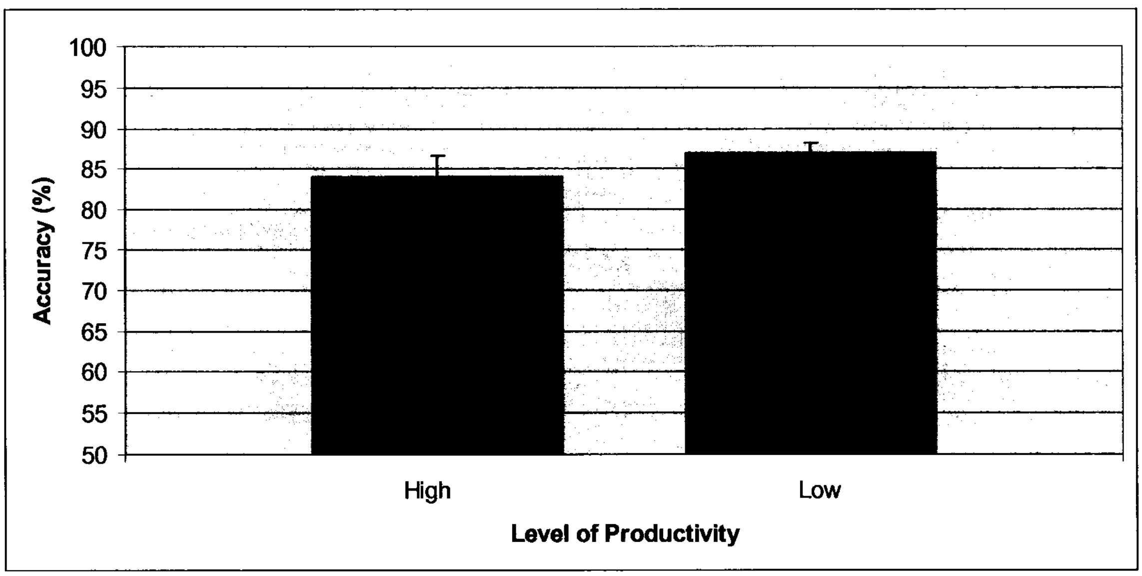

Figure 13: Accuracy by High and Low Productivity in Experiment 3: Participants made fewer errors when suffix productivity was lower, but this difference was not found to be significant.

There was not a great deal of variation in the accuracy by suffix (Figure 14). Observing the accuracy of individual suffixes, there does not appear to be any obvious trend towards greater or fewer errors at high levels of productivity. Instead, there are a few items that have higher error rates, the most obvious being *–hood* and *–ness*, but these items do not appear to favour any particular productivity level.

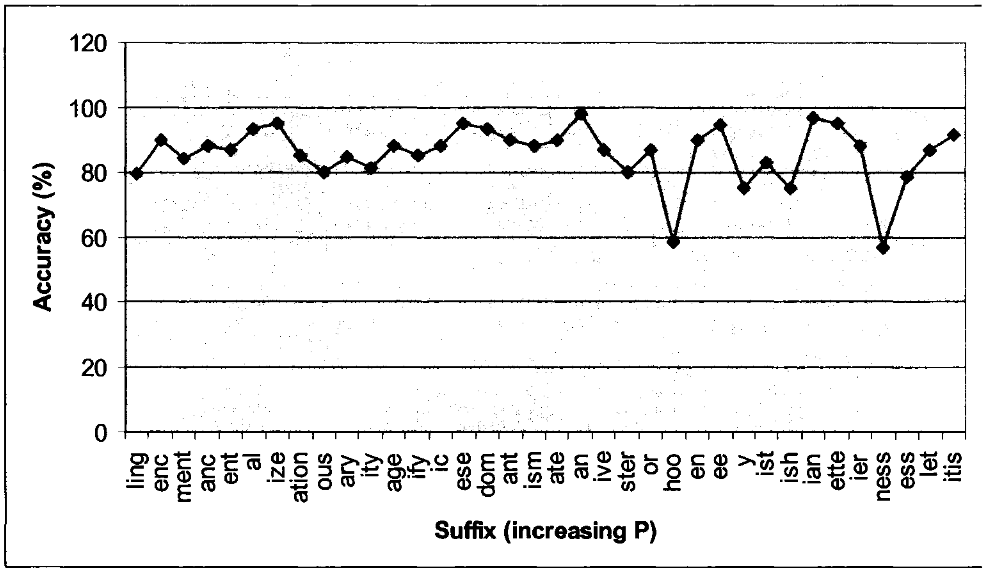

Figure 14: % Accuracy by Suffix in Experiment 3. Suffixes along the *x* axis are listed in order of increasing productivity. There does not appear to be a correlation between error rate and productivity.

In a comparison of reaction times in the Match condition (where suffixes matched selectional restrictions), highly productive suffixes were found to take *less* time for participants to respond to than low productivity suffixes (fig 15). This difference in reaction time was found to be significant, $F_2(1, 34) = 4.53$, $p=.04$. The reverse pattern was present in the mismatch condition, with highly productive suffixes taking longer to be rejected than low productivity suffixes, although this observation did not near significance.

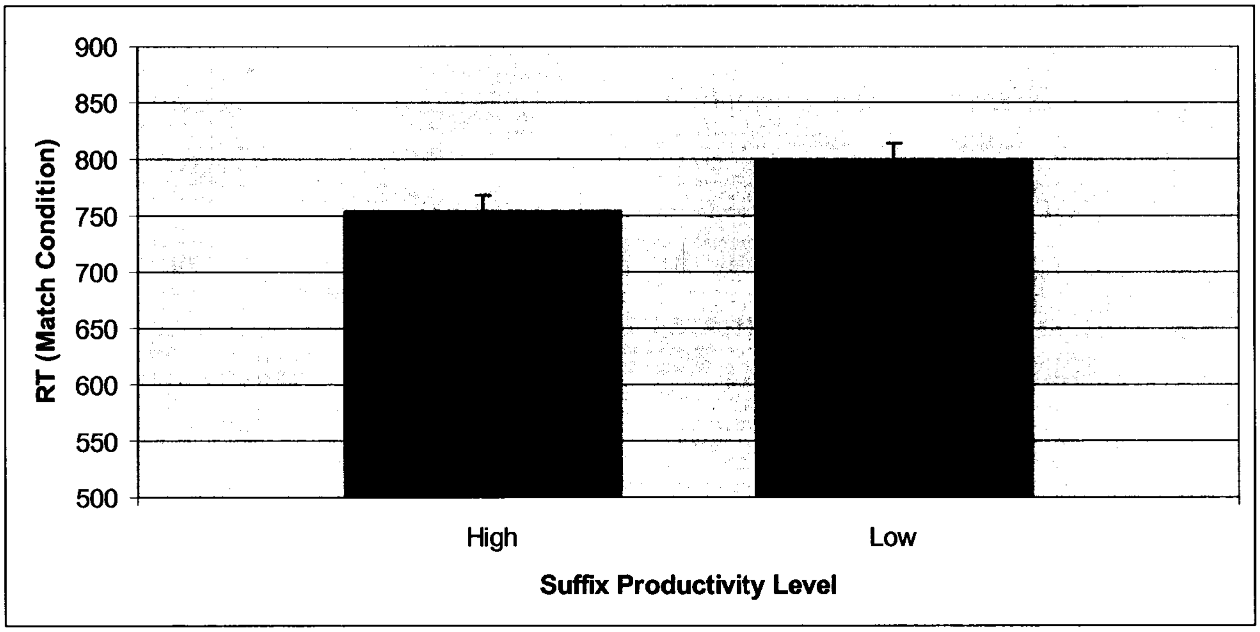

Figure 15: Reaction time to suffixes of high and low productivity in the Match condition in Experiment 3: possible base+suffix combinations were more rejected more quickly when suffix productivity was high.

Suffix productivity and suffix family size do not appear to correlate well. A high family size does not guarantee a high *P* value. Error rates and reaction times were both submitted to ANOVA tests to see whether family size and productivity interacted. There was no interaction in either case.

There appears to be a trend towards a correlation between higher suffix family size and longer reaction times. Most suffixes had fairly high rates of accuracy, making it difficult to determine whether higher family size was correlated with error rates. In cases where the base+suffix combination was possible, as determined by non-violation of selectional restrictions, suffixes of high productivity were rejected more quickly than those of low productivity, but this was the only place where productivity clearly appeared to affect results.

In this and the preceding experiments, the task set for participants was to decide whether or not an item was a word. While results from these experiments are suggestive, participants did not appear to be sensitive to the morphological structure of the novel pseudowords. A fourth experiment was run to capture possible differences that may have been too subtle for lexical decision to capture.

Experiment 4: Category Decision Task

Experiment 4 was a computerized decision task similarly constructed to Experiments 1-3. Experiment 4 made use of a category decision task, bypassing the standard word judgment made in a lexical decision experiment. Category decision tasks, where participants are asked to decide the lexical category of an item, have been found to be sensitive to morphological structure (Järvikivi & Niemi, 1999). If there is some effect of selectional restrictions (base+category match and mismatch) on accuracy or reaction time, then it might be more easily captured in this type of task.

Participants

Thirty-eight people, primarily recruited from the University of Alberta undergraduate population, participated in this experiment. One person was removed for excessive RTs.

Apparatus

Experiment 4 was run on the same equipment as Experiments 1-3.

Materials

The same list of suffixes and bases was used as in Experiment 3. As in Experiment 3, critical target items were random base+suffix pairs which were generated according to guidelines set out in that experiment.

Procedure

In Experiment 4, participants were asked to decide whether a given item was a noun, verb, or adjective. As in lexical decision, only one item appeared on the computer screen at a time. Participants were told that some items would look strange, and for those items, to please just decide what they thought the item looked like it could be. As in Experiment 3, the noun and verb bases were randomly paired with suffixes, and no base appeared more than once per participant. Each suffix appeared twice, once with a verb base and once with a noun base, but because there was extensive randomization in both Experiments 3 and 4, no participant saw the same target item more than once. The fillers remained the same.

Results and Discussion

Reaction times in this experiment were much longer than in a standard lexical decision experiment. For this analysis, RTs exceeding 10 s were removed, as were RTs less than 350 ms. Otherwise, a cut-off of 3 standard deviations above the average RT was used to exclude individual items on the basis of speed. This resulted in a loss of 1.9 % of the items.

The first question to be answered is whether or not violations of selectional restrictions in any way affected error rates, and whether or not family size played a role when split into high and low categories (Figure 16). Although there was a trend towards fewer errors in the low suffix family size category, it was not statistically significant. Likewise, there was no statistical difference between the match and mismatch categories. A comparison of accuracy by individual affixes and match-mismatch categories revealed what could be a slight trend towards more errors in the match condition, but this pattern is not clear (Figure 17).

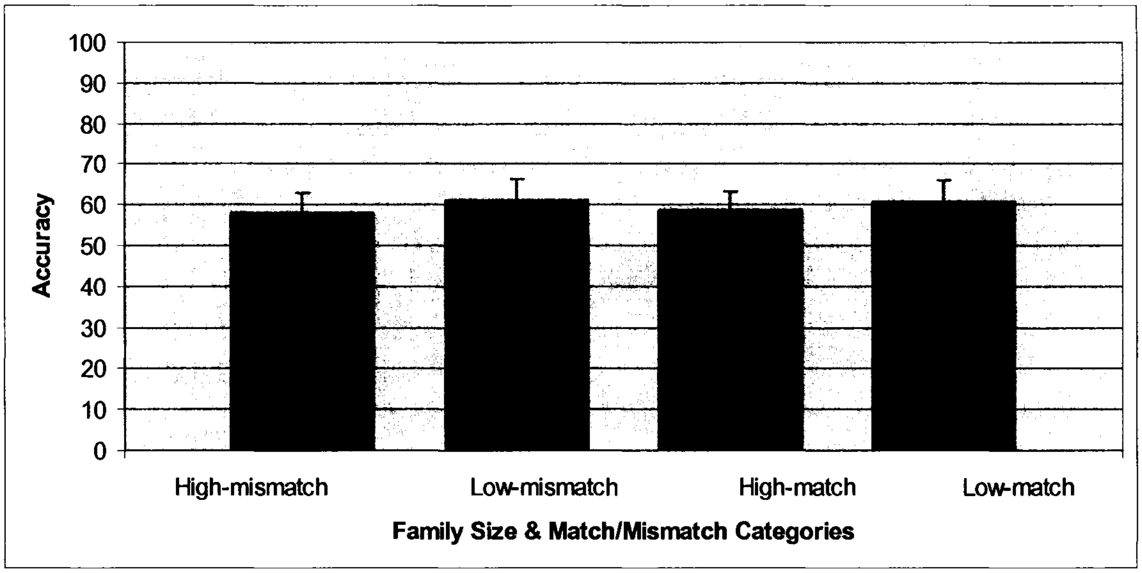

Figure 16: Accuracy by Family size (high and low) and Selectional Restriction match-mismatch in Experiment 4. There is a slight preference for more errors in the high family size condition, but this was not significant. It is not clear from Figure 16 whether error rates are in any way related to increasing family size.

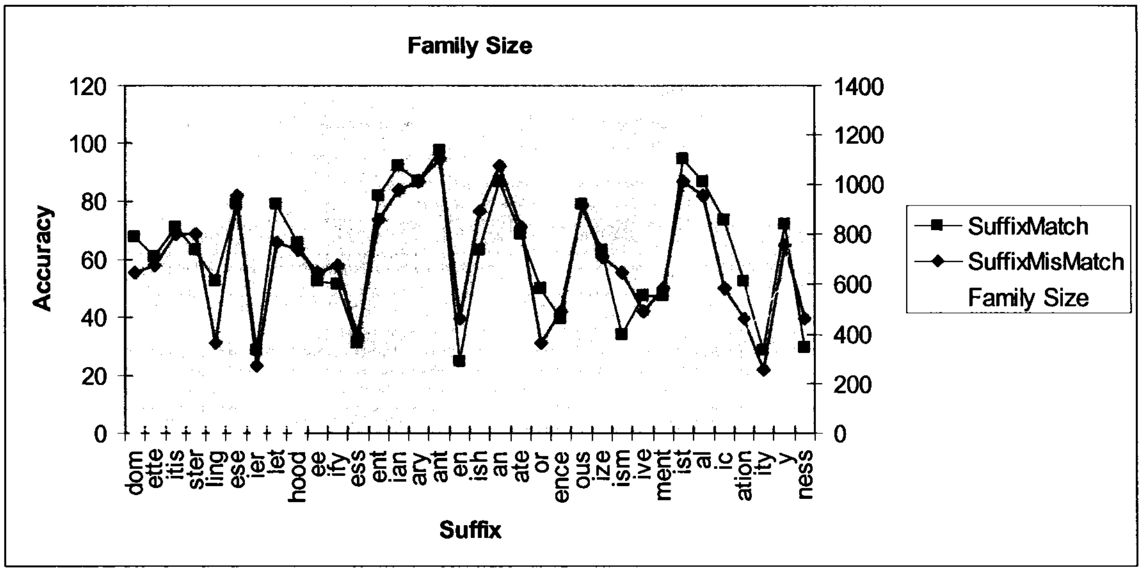

Figure 17: Graph of Accuracy by Suffix, Violation of Selectional Restrictions (Match/Mismatch) and Family Size in Experiment 4. There is a general trend for higher errors in the match condition, but this is not consistent across each suffix.

Error rates were also analyzed for any possible effects of productivity. Suffix productivity did not appear to correlate with the number of errors. Productivity and suffix family size were compared in a planned comparison ANOVA run on the error data by suffix (Figure 18). Error rates were highest when a suffix was in the low-productivity, high-family size group, and lowest when items were suffixed with low productivity-low family size suffixes. For high productivity suffixes, there were more errors made when a suffix had a high family size than when a suffix had a low family size. However, these results did not meet significance, $F_2(1,32) = 1.4$, $p=.24$.

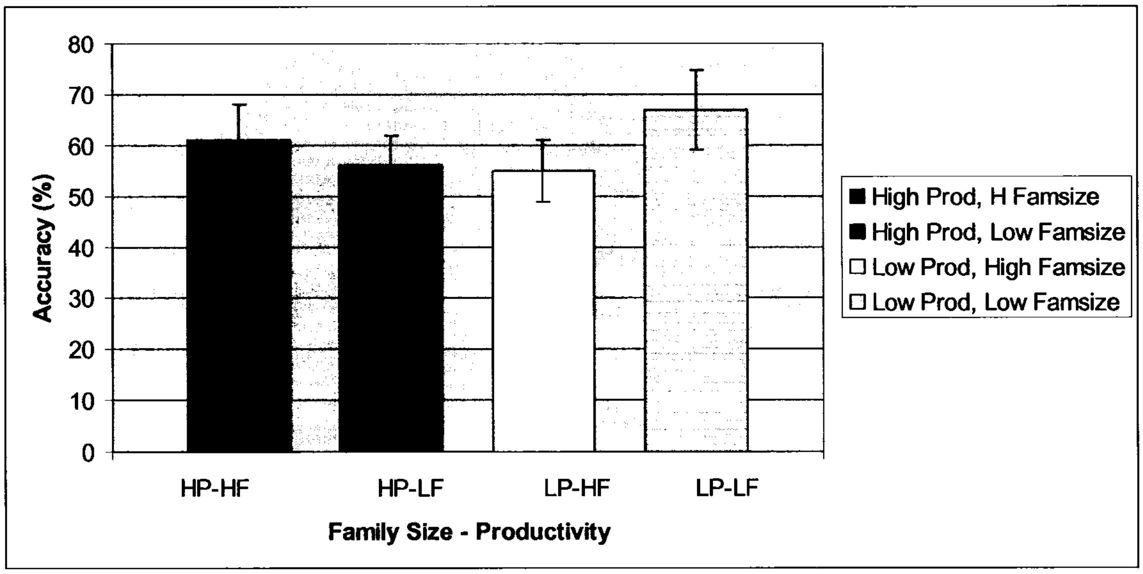

Figure 18: Accuracy rates by Family Size and Productivity in Experiment 4. High family size-high productivity and low-family size-low productivity showed the fewest errors. Results were not significant.

Reaction times were also analyzed. There was no difference found between the match (meets selectional restrictions) and mismatch (violates selectional restrictions) conditions in either the subject or items analysis. When an ANOVA was run on the items to compare high suffix family size and low suffix family size, there were no significant differences. However, because the dividing line between high and low was purely numerical (half were high, half were low), reaction times were also compared against suffix family size (Figure 19). There is a difference in the degree to which RT correlates with suffix family size in the two conditions, where it appears to take longer for participants to reject items when they have a higher family size in the match condition (cor.=.55, $p<.05$) than in the mismatch condition (cor.=-.24, $p>.05$). This mismatch condition again was non significant, although the trend in reaction time appears to be opposite that of the match category. However, as it appears that it is the suffix *–ness* that is primarily responsible for the downward slope in the mismatch condition (bottom right).

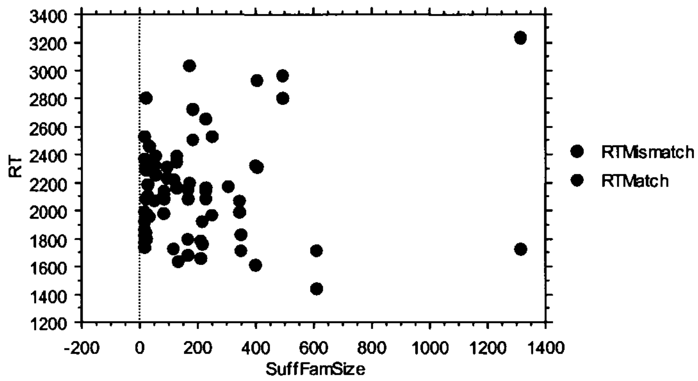

Figure 19: Graph of RT vs. Suffix Family Size, separated by the Match-Mismatch condition in Experiment 4. There is a significant correlation between reaction time and suffix family size, but only in the match condition.

When productivity levels were compared to RT, there appeared to be no correlation between them (figure 20). In most conditions, participants took longer to reject items that did not violate selectional restrictions. The exception was the low productivity-low suffix family size condition, where the mismatch (violating selectional restrictions) condition showed higher RTs. No effects were significant.

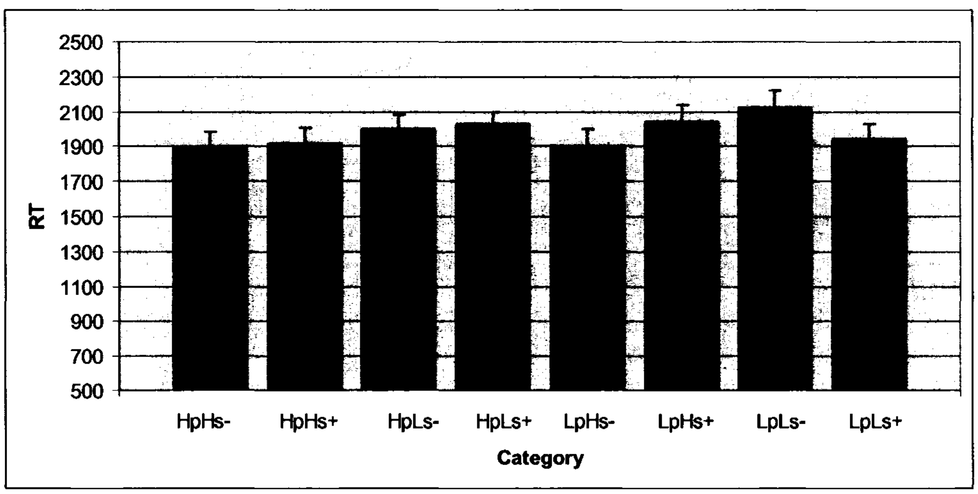

Figure 20: Productivity, Family Size and Selectional Restrictions by RT in Experiment 4: In most conditions, suffixes took longer for participants to reject when selectional restrictions were met. Legend: H refers to 'high,' L to 'low,' p to 'productivity,' s to 'suffix family size' and + or – to the match and mismatch conditions, respectively.

The results from Experiment 4 suggest that selectional restrictions at the level of lexical category do not pose great difficulty in processing. There was no difference between any of the conditions based on morphological fit in either the errors or reaction time, although there were hints of an effect in trends towards lower error rates and lower RTs. However, the story with suffix family size is more complex. Although the suffix *–ness* appears to be an outlier, an upward slope, towards higher reaction times at higher suffix family sizes, remains when it is removed from the analysis. Results are not conclusive.

With a possible difference observed between the match and mismatch RTs, and with participants otherwise not appearing to be sensitive to selectional restrictions, an overt test of participant knowledge was designed.

Experiment 5: Assessing Participant Knowledge of Selectional Restrictions

Returning now to the idea of selectional restrictions and morphological fit (match and mismatch), in particular to its apparent lack of influence (or very small amount of influence) on the results, we were curious as to whether participants had knowledge of the restrictions under study. Experiment 5 was an offline questionnaire that followed Experiments 3 and 4. The purpose of Experiment 5 was to determine whether native English speakers are explicitly aware of any selectional restrictions of affixes in English.

Participants

Participants were the same individuals who had participated in Experiment 4. There were 41 participants in total. None were excluded.

Apparatus

Experiment 5 was a pencil and paper task. No special equipment was required.

Materials

Items consisted of the 37 suffixes used to form novel pseudowords in Experiments 3 and 4. These were listed on the question sheet in alphabetical order. See Appendix C for a copy of the questionnaire.

Procedure

Participants were asked to indicate with which part of speech a given suffix combined. For example, given the suffix *–ness*, participants were asked to circle Noun, Verb, or Adjective, or any combination thereof, to indicate with which category *–ness* can combine. This question was asked for each suffix used in Experiments 3 and 4, for a total of 37. The participant ID was recorded so that results could be correlated with the results from online tasks where applicable.

Results and Discussion

There was a great deal of variability between subjects. No participant had explicit knowledge of every selectional restriction, although a few knew most of them. Others had very limited explicit knowledge of these selectional restrictions. Accuracy ranged from 30% to 95%. When match and mismatch conditions were plotted against explicit knowledge of selectional restrictions by subject (Figure 21), in the form of percent accuracy, it does not appear that knowledge of selectional restrictions reliably correlated with either increased or decreased reaction time latencies in either condition. The Spearman Rho correlations for both match and mismatch conditions did not support any correlation between RT and accuracy (match: cor.=.05, p>.05; mismatch: cor.=.09, p>.05).

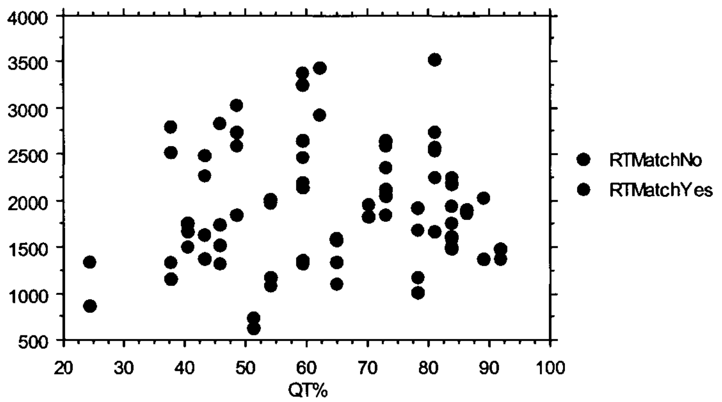

Figure 21: Reaction time (*y*-axis) for Match and Mismatch Categories by % correct on selectional restrictions, by subject. There does not appear to be any benefit to knowing "correct" selectional restrictions in either condition, by RT.

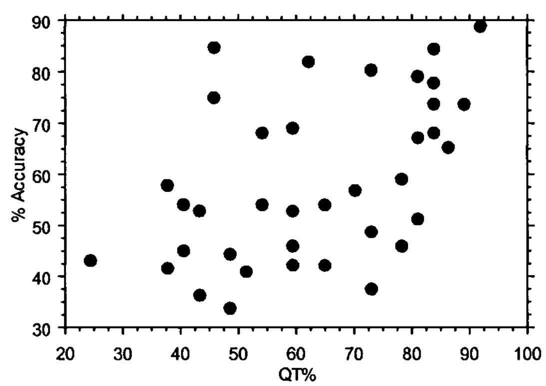

Figure 22: % Accuracy on the target items in Experiment 4 plotted against the offline selectional restriction knowledge test. There was a significant correlation between accuracy rates and explicit knowledge. Participants with greater explicit knowledge of selectional restrictions appear to have had an advantage in Experiment 4.

However, when results from the questionnaire are plotted against the number of errors participants made in the target items in Experiment 4, there does appear to be a trend towards higher accuracy during the experiment when participants scored more highly on the offline knowledge test (Figure 22, above). There was a significant correlation found between accuracy rates and explicit knowledge of selectional restrictions, which was revealed by a Spearman Rho Rank correlation, cor.=.49, p=.0036. This implies that some participants were being influenced by knowledge of selectional restrictions without explicitly being aware of them, as one would expect explicit decision making to be reflected in reaction time. This in turn implies that some speakers are sensitive to selectional restrictions, but that this sensitivity is very difficult to measure. It would appear that properties of suffixes and their combinations with bases can affect processing, but that these effects are likely to be quite small.

CHAPTER 5

GENERAL DISCUSSION

The experiments reported in this thesis investigated the roles of selectional restrictions, suffix family size, and suffix productivity in the acceptability and processing of novel pseudowords in English. Experiments 1 and 2 were part of a pilot study. In Experiment 1, participants were presented with novel pseudowords composed of lexical roots, primarily nouns, and four suffixes in a lexical decision task. These suffixes varied by suffix family size and selectional restrictions on base affixation, while productivity was kept relatively constant. Experiment 1b, a follow-up questionnaire, asked participants to rate how acceptable these base+suffix pseudowords were as words of English. Experiment 2 presented participants with nonword+suffix combinations that were structurally similar to target items in Experiment 1. In Experiment 3, the first of the main experiments, participants were presented with randomly generated base+suffix pairs in a lexical decision experiment. Verb and noun roots were combined with 37 English suffixes. Experiment 4 presented participants with randomly generated base+suffix pairs in a category decision task, where they were asked to decide the lexical category of the target items. A final offline questionnaire sought to determine whether native speakers of English have overt knowledge of selectional restrictions, and whether this knowledge impacted their responses in Experiment 4.

This thesis asked whether speakers of English are sensitive to selectional restrictions, suffix family size, and productivity in their judgments of novel pseudowords. These judgments, in both the online and offline experiments, reflect the degree to which participants consider the stimuli to be word-like. Another way to phrase the questions asked in this thesis is 'to what extent do suffix family size, productivity, and selectional restrictions affect the perception of novel pseudowords as real words?' Is a pseudoword that violates selectional restrictions of an affix less word-like than one does not? Are pseudowords containing suffixes with high family sizes more word-like than those containing suffixes with low family size? Do highly productive suffixes combine with roots to form more word-like items than low productivity suffixes? Do these factors combine? Each of these questions will be addressed in turn.

The first question is whether pseudowords violating selectional restrictions are less word-like than those that do not violate selectional restrictions. Selectional restrictions did not appear to have an effect on either the acceptability of pseudowords, as viewed through error rates and offline ratings, or on the processing of violations, as viewed through reaction time. With the exception of Experiment 1, where a difference was found between reaction times for V+*ation* and N+*ation*, and for which it is likely that the number of items in the V+*ation* condition influenced results, there were no significant differences in RT or error rates between conditions where there were and were not violations of selectional restrictions. However, in Experiments 3 and 4, there was a trend towards differences in RT, which were realized when the two conditions were plotted against Family Size.

The absence of a main effect of selectional restrictions could have resulted from a number of factors. It could be that the selectional restrictions needed to be more tightly defined. This series of experiments considered word class violations to be restrictions on suffixation. However, many suffixes occur in more specialized environments. It is possible that if one were to look at a smaller set of suffixes with more specifically defined restrictions that one might find results. Doing so in these experiments would have sacrificed the ability to examine the role of family size, however, since one cannot make generalizations about such lexical properties based on, for example, two suffixes that otherwise have many other differences (e.g., greater or lesser semantic interpretability, productivity, attachment properties). On the other hand, it may be that speakers are simply not sensitive to lexical restrictions in the formal descriptive sense. This would imply either that some roots are unspecified with respect to lexical category in the minds of speakers of English, and that specification occurs only with markers (e.g., affixes) or in sentence context, or that affixes are not truly constrained by selectional restrictions in the traditional sense.

Alternatively, it is possible that the absence of a main effect of selectional restrictions was related to specific properties of English words. In English, conversion from one lexical category to another is a relatively productive process. During the construction of the stimulus set, many potential bases were excluded on the grounds that their lexical category was completely ambiguous without supporting context. The rejected bases were not clearly nouns or verbs, but could serve as both. While the words used in these experiments were meant to be unambiguous, it is not impossible that by analogy, when suffixes were applied, conversion processes became active. It may have been the case that for some of the items, the addition of the suffix itself could have assigned lexical class to the root. This potential confound could be overcome in future experiments through design and language choice. If one were to continue using English, then it might be best to look to morphologically complex words as potential bases, so that lexical category is unambiguous. One could also look to monomorphemic English words, but even in longer words it is not necessarily clear that lexical category will be well defined, and if it is, if it will be possible to balance the stimuli. It might be more constructive to expand this sort of work to other languages where category is more clearly defined.

Another question regarding selectional restrictions remains. This question is whether or not native speakers of English aware of selectional restrictions. Results from Experiment 5 suggest that there is a great deal of variability among individuals with regards to their level of overt knowledge of selectional restrictions. Overt knowledge of selectional restrictions did not appear to correlate with reaction times to correctly categorized words in Experiment 4, but it did correlate with error rates, with a greater number of errors reported with less knowledge about selectional restrictions. Experiment 4 was a category decision task and on average the response latencies were much longer than those seen in lexical decision, so is unclear whether participants who had overt knowledge were explicitly accessing this knowledge during Experiment 4. However, the lack of correlation in reaction times suggests that explicit knowledge was not accessed, which implies that at least for some speakers, selectional restrictions may unconsciously affect their decision about what can and cannot be a word. In this respect, selectional restrictions appear to be active for some native speakers as influencing word-status judgments.

The next question asked is whether words that are affixed with suffixes of high family size are more word-like than those affixed with low family size suffixes, or in other words, does suffix family size affect speaker decisions about word-status. The short answer is that yes, family suffix size does appear to influence how word-like an item can be, at least at a very general level. When suffix family size was broken down into well-defined, discreet High and Low categories (Experiments 1 and 2) with productivity remaining constant, there was a trend towards longer reaction times and higher error rates for higher suffix family sizes. In Experiments 3 and 4, where there were many more suffixes, there was a general trend towards longer decision latencies at higher suffix family sizes when stimuli did not violate selectional restrictions. This could mean that when a suffix occurs frequently as a part of different words, it is more likely to be used to form a new word than a suffix that does not combine with as many items.

Since suffix family size appears to be a factor in the processing of real base+suffix pseudowords in English, it implies that suffixes are available as entities in the mental lexicon independent from the lexical forms in which they occur. One might expect such results under a morphological model which includes decomposition. In addition, although the meanings of suffixes are ill-defined compared to those of roots, the fact that there is a family size effect implies the availability of some semantic content. If we accept that there is a semantic component to the family size effect, then although it may not be intuitive, it suggests that individual suffixes are connected in admittedly abstract ways. One might then expect to be able to find priming effects to different words containing the same suffix. Giraudo and Grainger (2003) only found priming for prefixes in French, but it is possible that parsability and semantic transparency interfered with the suffix priming task. With regards to parsability, English could be employed in a similar priming task, using suffixes that are more easily identifiable because the morpheme and syllable boundaries are aligned and where they do not change the form of the base (e.g., *-ness*, *-dom*, *-ling*).

The next question asked was whether suffix productivity influenced native speaker judgments of word-status. In Experiment 3, there was some evidence that highly productive suffixes showed higher error rates than suffixes of lower productivity. This result might be expected, as the more productive a suffix is, the more items it can conceivably interact with to form new words. This result was not apparent in Experiment 4. However, because productivity and family size do not correlate, it is possible that effects from both were lost in competition. There was no statistical interaction between them, but if both factors have very small effects on lexical processing, then the overwhelming impact of semantic interpretation may have overridden these smaller effects. In addition, while Baayen's (1992) productivity formula may be clear, the factors underlying the phenomenon of morphological productivity are not fully understood, and more research needs to be done before we come any closer. This lack of specificity may have also influenced the results in these experiments, as it is difficult to control for factors that could confound results when those factors are not clear. In summary, results from Experiment 3 suggest that productivity might affect the acceptability of novel forms, but in order to be certain, more experiments are necessary.

Finally, we turn to interactions between effects. ANOVAs did not reveal interaction effects between any of the variables when the variables were split into categories, but there is some suggestion from the differences in the base+suffix matching (follows selectional restrictions) and mismatching (violates selectional restrictions) conditions that participants were more sensitive to suffix family size in the matching condition. That is, they were more sensitive to suffix family size when the target item was morphologically possible. No effects were found between suffix family size and productivity, but it is possible that if the effects were very small, they might not have been large enough to be captured through the experimental methods used.

If future experiments find that participants are more sensitive to morphological family size when a base+suffix combination is morphologically legal than when it is not, it could speak to a process by which potential combinations are unconsciously evaluated first for their morphological structure and then for other factors, including suffix properties. For example, once a word has been deemed permissible, then the family size of the suffix might become more active, such that a higher family size will result in a greater level of acceptability. However, it is possible that both of these factors are trumped by semantic interpretability.

A very large factor not manipulated in these experiments was semantic interpretability. It is likely that semantic interpretability was, at least in part, responsible for the very high error rates in the target items in Experiment 1. Experiments 3 and 4 were designed to limit semantic effects in the hopes that the effects of suffix properties, if any, would become clear, but this is ultimately an unnatural approach to assessing novel forms. In a naturalistic setting, one would have to expect that semantics would play a role in both the creation and ultimate acceptance or rejection of a novel form. For example, results from the first questionnaire clearly indicated that the suffix *–ling*, which has a low family size, was highly acceptable to native speakers when it occurred in N+*ling* combinations. Indeed, it was rated the most highly of the tested suffixes. It is possible that semantic interpretability is ultimately more important for speakers than lexical category, which would suggest that meaning comes before form and that a strict rule-based approach might not able to account for all morphological behaviour. A previous experiment run in Italian also failed to find significant differences in processing based upon lexical category mismatches in a lexical decision experiment (Burani et al, 1999). However, they did find that results varied based upon the semantic interpretability of the stem+suffix combinations. This is in line with the results of the first questionnaire and with the high item error rates in Experiment 1.

There are three main directions which can be explored immediately from this research. The first focuses on selectional restrictions of specific affixes, at the expense of investigating the role of large scale suffix properties (e.g., family size, productivity). Results from the manipulation of suffix family size are cautiously interpreted as being active in morphological processing. Productivity needs to be addressed more closely, especially given that this particular measure of productivity is relatively new. The second direction focuses on the roles of suffix properties on a larger scale, possibly returning to nonwords to avoid interference from base properties. The third direction focuses on the factors underlying acceptability of novel combinations, or neologisms, with a focus on semantic interpretability.

Of the factors studied, suffix family size appeared to be the most strongly associated with real word interpretations of pseudowords. While results did not match the hypotheses made exactly, elements remained. For example, the suffix *–ness* has a relatively high family size and a relatively high productivity value, and it shows a relatively high number of errors and long reaction times. However, the suffix *–ling* provides an excellent counter example, as it also showed relatively high error rates and long reaction times, and yet it has a very small suffix family size and has a low productivity value. At most, general trends can be described from these experiments, even when they are significant. Individual suffix behaviour cannot be perfectly predicted from suffix family size, productivity, or selectional restrictions, and so for future studies, the factors underlying their individuality should be addressed.

References

Alegere, M. & Gordon, P. (1999). Rule-Based versus associative processes in Derivational morphology. *Brain and Language, 68*, 347-354.

Baayen, H. (1992). Quantitative aspects of morphological productivity. In G.E. Booij & J. v.Marle (eds.), *Yearbook of Morphology 1991,* Kluwer Academic Publishers, Dordrecht, pp. 109-150.

Baayen, H. (1994). Productivity in language production. *Language and Cognitive Processes, 9*(3), 447-469.

Baayen, H. (1997). Markedness and productivity. In Dressler, W.U., Prinzhorn, M., & J.R. Rennison (eds.), *Trends in Linguistics: Studies and Monographs 97: Advances in Morphology*, Mouton de Gruyter, Berlin, Germany, pp. 189-200.

Berg, T. (2003). Right-branching in English derivational morphology. *English Language and Lingusitics, 7*(2), 279-307.

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

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

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

Burani, C. & Thornton, A.M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 157 – 207.

Carstairs-McCarthy, A. (2002). *An Introduction to English Morphology*. Edinburgh University Press Ltd., Edinburg.

Clahsen, H., Sonnenstuhl, I., Blevins, J.P. (2003). Derivational morphology in the German mental lexicon: a dual mechanisms account. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 125-155.

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

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

Dressler, W.U., & Ladanyi, M. (2000). Productivity in word formation (wf): a morphological approach. *Acta Linguistica Hungarica, 47*(1-4), 103-144.

Feldman, L.B. & Prostko, B. (2002). Graded aspects of morphological processing: Task and processing time. *Brain and Language, 81*, 12-27.

Feldman, L.B. & Pastizzo, M.J. (2003). Morphological facilitation: The role of semantic transparency and family size. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, Pp. 234-258.

Ford, M.A., Marslen-Wilson, W.D., & Davis, M.H. (2003). Morphology and Frequency: Contrasting methodologies. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 89-124.

Giraudo, H. & Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked priming. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 209 – 232.

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

Hay, J., & Baayen, H. (2002). Parsing and productivity. In Booij, G.E. & J. v. Marle (eds.), *Yearbook of Morphology 2001,* Kluwer Academic Publishers, Dordrecht, pp. 203-235

Hay, J. and Baayen, R. H. (2003). Phonotactics, Parsing and Productivity. *Italian Journal of Linguistics 1*, 99–130.

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

Järvikivi, J. & Niemi, J. (1999). Linearity and Morphological Structure in Derived words: Evidence from category decision. *Brain and Language, 68,* 340-346.

Libben, G. (1993). Are morphological structure computed during word recognition? *Journal of Psycholinguistic Research, 22*(5), 535-544.

Libben, G. (1994). The role of hierarchical morphological structure: A case study. *Journal of Neurolinguistics, 8*(1), 49-55.

Luzzatti, C., Mondini, S., & Semenza, C. (2001). Lexical representation and processing of morphologically complex words: Evidence from the reading performance of an Italian agrammatic patient. *Brain and Language, 79*, 345-359.

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

Plag, I., Dalton-Puffer, C., & Baayen, H. (1999). Morphological productivity across speech and writing. *English Language and Linguistics, 3*(2), 209-228.

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

Reid, A.A. & Marslen-Wilson, W.D. (2003). Lexical representation of morphologically complex words: Evidence from Polish. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 287-336.

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

Soanes, C & Stevenson, A. (Eds.). (2004). *The Concise Oxford English Dictionary.* Oxford Reference Online. Oxford University Press. University of Alberta. 10 Jan 2005.

Traficante, D. & Burani, C. (2003). Visual processing of Italian verbs and adjectives: The role of the inflectional family size. In Baayen, R.H. & R. Schreuder (eds.), *Trends in Linguistics: Morphological Structure in Language Processing*, Mouton de Gruyter, Berlin, Germany, pp. 45 – 64.

Van Marle, J. (1992). The relationship between morphological productivity and frequency: a comment on Baayen's performance oriented conception of morphological productivity. In G.E. Booij & J. v.Marle (eds.), *Yearbook of Morphology 1991,* Kluwer Academic Publishers, Dordrecht, pp. 151-163

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

Appendix A: Experiment 1 Stimuli: Nouns

Base	BaseFreq	Base FamSize	-ation	-ic	-ling	-th
barb	33	2	barbation	barbic	barbling	barbth
barn	226	7	barnation	barnic	barnling	barnth
bird	1841	37	birdation	birdic	birdling	birdth
bulb	207	3	bulbation	bulbic	bulbling	bulbth
cask	20	0	caskation	caskic	caskling	caskth
cusp	3	0	cuspation	cuspic	cuspling	cuspth
desk	1633	7	deskation	deskic	deskling	deskth
dirt	356	11	dirtation	dirtic	dirtling	dirtth
disc	194	4	discation	discic	discling	discth
duct	45	2	ductation	ductic	ductling	ductth
dusk	268	2	duskation	duskic	duskling	duskth
fern	90	2	fernation	fernic	fernling	fernth
fist	470	5	fistation	fistic	fistling	fistth
fort	453	0	fortation	fortic	fortling	fortth
gang	433	11	gangation	gangic	gangling	gangth
germ	158	4	germation	germic	germling	germth
gulf	276	2	gulfation	gulfic	gulfling	gulfth
helm	18	1	helmation	helmic	helmling	helmth
hemp	34	2	hempation	hempic	hempling	hempth
horn	323	21	hornation	hornic	hornling	hornth
hunk	42	0	hunkation	hunkic	hunkling	hunkth
kelp	17	0	kelpation	kelpic	kelpling	kelpth
lamp	629	18	lampation	lampic	lampling	lampth
lens	229	4	lensation	lensic	lensling	lensth
mast	60	10	mastation	mastic	mastling	mastth
mink	57	0	minkation	minkic	minkling	minkth
monk	164	1	monkation	monkic	monkling	monkth
musk	28	6	muskation	muskic	muskling	muskth
pact	337	0	pactation	pactic	pactling	pactth
pest	164	7	pestation	pestic	pestling	pestth
pomp	33	3	pompation	pompic	pompling	pompth
pond	339	3	pondation	pondic	pondling	pondth
pork	150	7	porkation	porkic	porkling	porkth
sect	85	4	sectation	sectic	sectling	sectth
silk	474	7	silkation	silkic	silkling	silkth
silt	60	1	siltation	siltic	siltling	siltth
tent	785	3	tentation	tentic	tentling	tentth
tern	29	0	ternation	ternic	ternling	ternth
text	687	5	textation	textic	textling	textth
tuft	63	2	tuftation	tuftic	tuftling	tuftth
tusk	33	0	tuskation	tuskic	tuskling	tuskth
verb	58	28	verbation	verbic	verbling	verbth
volt	72	2	voltation	voltic	voltling	voltth
wand	37	0	wandation	wandic	wandling	wandth
wasp	107	4	waspation	waspic	waspling	waspth
wisp	65	1	wispation	wispic	wispling	wispth
yard	1563	22	yardation	yardic	yardling	yardth
zinc	72	0	zincation	zincic	zincling	zincth

Verbs:

Base	BaseFreq	BaseFamSize	-ation
bawl	75	1	bawlation
hark	36	3	harkation
lend	489	4	lendation
lurk	154	1	lurkation
melt	436	6	meltation
sift	71	1	siftation
warn	883	7	warnation
wilt	54	0	wiltation

Adjectives:

Base	BaseFreq	BaseFamSize	-th
bald	152	9	baldth
bold	252	6	boldth
daft	56	2	daftth
dank	30	1	dankth
fond	416	3	fondth
lank	11	5	lankth
rapt	36	7	raptth
vast	1205	2	vastth

Appendix B: Offline Acceptability Questionnaire

INSTRUCTIONS

Age: __________ Gender:__________ First language: ______________

If English is not your first language, please indicate your knowledge of English on a scale of 1-7
If English is not your first language, how many years have you studied English? _____

EXPERIMENT

On the following pages, you will see a list of English words and a rating scale (1-5). The following is an example

WORD	Rating Scale (1 = completely unacceptable, 5 = completely Acceptable)				
Legalness	1	2	3	4	5

Please use the scale provided to rate the acceptability of each word, with 1 meaning 'completely unacceptable as an English word' and 5 meaning 'completely acceptable as an English word.' Circle the number you think best describes how you rate the word.

For example, if you think that the word *legalness* is completely acceptable, then circle "5." If you think that it is a completely unacceptable English word, then circle "1." Even if you have seen the word before, you do not need to rate it a '5.'

Thank you for your participation.

WORD	Rating Scale (1 = completely unacceptable, 5 = completely Acceptable)				
1) shrelte	1	2	3	4	5
2) pactth	1	2	3	4	5
3) issue	1	2	3	4	5
4) research	1	2	3	4	5
5) deskic	1	2	3	4	5
6) independence	1	2	3	4	5
7) horror	1	2	3	4	5
8) fernling	1	2	3	4	5
9) spirit	1	2	3	4	5
10) scrume	1	2	3	4	5
11) level	1	2	3	4	5
12) fortth	1	2	3	4	5
13) telescope	1	2	3	4	5
14) porkling	1	2	3	4	5
15) lisk	1	2	3	4	5
16) walc	1	2	3	4	5
17) pestic	1	2	3	4	5
18) stern	1	2	3	4	5
19) zist	1	2	3	4	5
20) screige	1	2	3	4	5
21) gulfth	1	2	3	4	5
22) important	1	2	3	4	5
23) gilm	1	2	3	4	5
24) splamn	1	2	3	4	5
25) silkic	1	2	3	4	5

26) gusp	1	2	3	4	5
27) pilm	1	2	3	4	5
28) basic	1	2	3	4	5
29) smiln	1	2	3	4	5
30) fistth	1	2	3	4	5
31) design	1	2	3	4	5
32) splouge	1	2	3	4	5
33) woman	1	2	3	4	5
34) germling	1	2	3	4	5
35) nens	1	2	3	4	5
36) scruke	1	2	3	4	5
37) command	1	2	3	4	5
38) golt	1	2	3	4	5
39) ralp	1	2	3	4	5
40) lensation	1	2	3	4	5
41) twilight	1	2	3	4	5
42) dwolt	1	2	3	4	5
43) phlunch	1	2	3	4	5
44) nelp	1	2	3	4	5
45) pondic	1	2	3	4	5
46) chicken	1	2	3	4	5
47) fulb	1	2	3	4	5
48) cucumber	1	2	3	4	5
49) discation	1	2	3	4	5
50) jasp	1	2	3	4	5
51) lonk	1	2	3	4	5
52) trapezoid	1	2	3	4	5
53) hesk	1	2	3	4	5
54) lemon	1	2	3	4	5
55) textic	1	2	3	4	5
56) splague	1	2	3	4	5
57) waspling	1	2	3	4	5
58) sploom	1	2	3	4	5
59) lively	1	2	3	4	5
60) monkation	1	2	3	4	5
61) considerable	1	2	3	4	5
62) runner	1	2	3	4	5
63) bolf	1	2	3	4	5
64) voltling	1	2	3	4	5
65) thwult	1	2	3	4	5
66) sprep	1	2	3	4	5
67) bulbation	1	2	3	4	5
68) spreek	1	2	3	4	5
69) avalanche	1	2	3	4	5
70) economy	1	2	3	4	5
71) duskth	1	2	3	4	5
72) ether	1	2	3	4	5
73) instruction	1	2	3	4	5
74) board	1	2	3	4	5
75) barnation	1	2	3	4	5
76) solidify	1	2	3	4	5
77) aviator	1	2	3	4	5
78) feeble	1	2	3	4	5
79) scroon	1	2	3	4	5
80) bluebell	1	2	3	4	5

Appendix C: Offline Selectional Restriction Questionnaire

Post-Computer Questionnaire
Computer ID________________ (have the RA fill out)

Below you will find a list of suffixes, followed by "Noun Verb Adjective." Please circle the grammatical part of speech (noun [person/place/thing], verb [action/state] or adjective [modifying/describing a noun]) that you think a given suffix can combine with to form a new word. You can circle more than one. Don't worry about the resulting word class.

For example, you might see:

–like Noun Verb Adjective

Since there are words like *childlike*, you would circle "noun." The suffix *–like* can combine with the noun *child* to form a new word.

Please answer as best you can, but try not to take too long. If you really don't know the answer, you can leave it blank. There are no wrong answers.

-al	Noun	Verb	Adjective
-ary	Noun	Verb	Adjective
-ate	Noun	Verb	Adjective
-dom	Noun	Verb	Adjective
-en	Noun	Verb	Adjective
-ent	Noun	Verb	Adjective
-ese	Noun	Verb	Adjective
-ess	Noun	Verb	Adjective
-ette	Noun	Verb	Adjective
-hood	Noun	Verb	Adjective
-ian	Noun	Verb	Adjective
-ic	Noun	Verb	Adjective
-ier	Noun	Verb	Adjective
-ify	Noun	Verb	Adjective

-ish	Noun	Verb	Adjective
-ism	Noun	Verb	Adjective
-ist	Noun	Verb	Adjective
-itis	Noun	Verb	Adjective
-ity	Noun	Verb	Adjective
-ive	Noun	Verb	Adjective
-ize	Noun	Verb	Adjective
-let	Noun	Verb	Adjective
-ling	Noun	Verb	Adjective
-ly	Noun	Verb	Adjective
-ous	Noun	Verb	Adjective
-y	Noun	Verb	Adjective
-age	Noun	Verb	Adjective
-ance	Noun	Verb	Adjective
-ant	Noun	Verb	Adjective
-ation	Noun	Verb	Adjective
-ee	Noun	Verb	Adjective
-ence	Noun	Verb	Adjective
-ment	Noun	Verb	Adjective
-ness	Noun	Verb	Adjective
-or	Noun	Verb	Adjective
-ster	Noun	Verb	Adjective
-an	Noun	Verb	Adjective