

What's in a Word?

Jennifer A. Henderson

MSc English Language
The University of Edinburgh
2007

Abstract

Words are all around us to the point that their complexity is lost in familiarity. The term “word” itself can ambiguously refer to different linguistic concepts: orthographic words, phonological words, grammatical words, word-forms, lexemes, and to an extent lexical items. While it is hard to come up with exception-less criteria for wordhood, some typical properties are that words are writeable and spellable, consist of morphemes, are syntactic units, carry meaning, and interrelate with other words. Moreover, words can be classified and categorized in a number of different ways depending on how they are used, by whom, and to what extent they are established within the lexicon. English has many ways of adding new words to its repertoire through both productive and creative means. “Knowing” a word need not entail knowing every facet of its history and usage, yet there is still more to a word than simply the symbol-to-meaning relation.

Table of Contents

| | | |
|-------|-------------------------------------|----|
| 1 | Introduction | 5 |
| 2 | What We Mean By <i>Word</i> | 6 |
| 2.1 | Orthographic and Phonological Words | 6 |
| 2.2 | Lexemes and Word-forms | 7 |
| 2.3 | Grammatical Words | 8 |
| 2.4 | Lexical Items | 9 |
| 2.5 | Complicating Factors | 9 |
| 2.5.1 | Suppletion | 9 |
| 2.5.2 | Syncretism | 10 |
| 2.5.3 | Homonymy and Polysemy | 11 |
| 2.5.4 | Clitics | 11 |
| 2.5.5 | Periphrasis and Phrasal Verbs | 12 |
| 2.5.6 | Spelling | 13 |
| 2.6 | Preliminary Conclusions | 14 |
| 3 | Properties of Words | 15 |
| 3.1 | Orthography | 15 |
| 3.2 | Phonology | 16 |
| 3.3 | Morphology | 17 |
| 3.4 | Syntax | 17 |
| 3.5 | Semantics | 19 |
| 3.5.1 | Meaning and Meaning Change | 19 |
| 3.5.2 | Lexical Items and Lexicons | 20 |
| 3.5.3 | Predictability and Productivity | 21 |
| 3.6 | Words with Other Words | 22 |
| 3.6.1 | Sense Relations | 22 |
| 3.6.2 | Collocations | 23 |
| 3.7 | Summary | 23 |
| 4 | Subdivisions of Words | 24 |
| 4.1 | Building Blocks of Words | 24 |
| 4.1.1 | Morphemes | 25 |
| 4.1.2 | Simplex vs. Complex Words | 25 |
| 4.1.3 | The Trouble with Morphemes | 26 |

| | | |
|-------|--|----|
| 4.2 | Classifications of Words | 26 |
| 4.3 | Pragmatic Classifications of Words | 27 |
| 4.3.1 | Dialectal, Regional, and Cultural Words | 27 |
| 4.3.2 | Jargon | 28 |
| 4.3.3 | Informal Words | 29 |
| 4.4 | Bestowal of Wordhood | 30 |
| 4.4.1 | Existing and Established Words | 31 |
| 4.4.2 | Nonce Words and Neologisms | 31 |
| 4.4.3 | Possible, Actual, and Probable Words | 32 |
| 4.4.4 | Institutionalized and Lexicalized Words | 32 |
| 4.5 | Summary | 33 |
| 5 | Where (New) Words Come From | 33 |
| 5.1 | Why New Words | 34 |
| 5.2 | Morphological Productivity | 34 |
| 5.2.1 | Derivational Affixation | 35 |
| 5.2.2 | Conversion | 36 |
| 5.2.3 | Compounding | 37 |
| 5.2.4 | Restrictions on Productivity | 38 |
| 5.3 | Creativity: Non-Morphological Innovation | 39 |
| 5.3.1 | Borrowing | 40 |
| 5.3.2 | Reanalysis | 41 |
| 5.3.3 | Onomatopoeia and Phonaesthemes | 42 |
| 5.3.4 | Eponymy and Antonomasia | 43 |
| 5.3.5 | Metaphoric Extension | 43 |
| 5.4 | Rules, Analogy, and Usefulness | 43 |
| 6 | Word Intuition Survey | 44 |
| 6.1 | Demographics | 44 |
| 6.2 | Survey Analysis | 45 |
| 7 | Conclusions | 48 |
| | References | 51 |
| | Appendix A: Word Intuition Survey | 53 |
| | Appendix B: Word Intuition Survey (Key) | 55 |

1 Introduction

To paraphrase Shakespeare's eponymous Juliet we ask "What's in a word?" The tongue-in-cheek answer, which is nonetheless truthful, might be "letters" or "sounds." Juliet goes on to reason "That which we call a rose/By any other name would smell as sweet" (qtd. in Quirk 1968:122). Her argument regarding Romeo's troubling status as a Montague has merit. A name is but a word, an arbitrary label which the language community has agreed represents something or other. Indeed, a rose would smell and look and *be* the same regardless of what one calls it. But there is more to it than that. Words are more than mere labels, they are microcosms of language. They have their own histories, characteristics, and associations. They have specified functions regarding the roles they play in communication.

Despite Juliet's egalitarian approach to appellations, Shakespeare's play in a sense revolves around the fact that Montague is *not* just a name. It is a name and all the meanings associated with it. It entails a history of bitterness, rivalry, and feuding. The denotations of words are simple when compared to their connotations. Even a monosyllabic word such as *rose* is rich with connotations. *Rose* represents more than just a type of flower. Wrapped up in the word are associations with love, beauty, innocence, and devotion. Roses have been used to represent everything from the sacred to the secular to sports teams. Even when we know that the use of a particular word is arbitrary, most people would cringe at calling a rose by some other, equally arbitrary name such as a *hunkle*.

As we take a closer look at words, and even the term *word* itself, we find that they are more complex than our native speaker intuitions would initially perceive. We will begin by outlining the various ways in which both linguists and speakers use the term *word*, while looking at some of the factors which complicate definition. We will then look at the various properties and characteristics which words typically have followed by discussion of how words get subdivided and sorted. From there we turn to how and why new words come into the language. Lastly, we will analyze a survey of how a few native speakers view words and what they think it means to be a word. For the purposes of this paper, we will deal strictly with what it means to be a word in the English language.

2 What We Mean By *Word*

Words are ubiquitous. In a literate society words are everywhere and unavoidable. Every day people read, write, speak, and hear words. Words can be readily found in books and magazines. They can also be found plastered on signs, engraved on buildings, scrawled on food, printed on clothing, tattooed onto people, and they often even reside on the tips of our tongues. But rarely do people stop to genuinely ponder what constitutes a word to begin with. They may occasionally stop to ponder if a certain form is one word or two, such as *alright* or *a lot*. They may also utter something, pause, and ask themselves or others if what they just said is a word. Speakers of different generations may argue the wordhood of certain slang terms or colloquialisms (e.g. *bling* or *ain't*). The point is, the complexity of words is either taken for granted or shrugged off. It does not help that *word* itself is not readily definable.

Even in linguistics the term *word* is often tossed about in an ambiguous way. Intuitively, speakers have a sense that words are composed of sounds, carry meaning, are the basic units of phrases and sentences, and are typically found in dictionaries. While such intuitions help to describe words, they do little to elucidate what is or is not a word. For example,

(1) “I’m going to go to the grocery store.”

could be counted as having as few as five or as many as nine “words.” The number varies based on what counts as a word as well as what *kind* of word is being counted. *Word* can refer to a number of different linguistic concepts, all of which are similar yet distinctive.

This ambiguity surrounding the term *word* is typically not a problem. Indeed, for everyday speech, and even in general linguistics, having it be “deliberately vague” (Bauer 1983:13) can be useful when the distinction between uses is unimportant. In a discussion of words, however, it is helpful to disambiguate the various uses of the term. The following section details the ways in which we use the term *word* as well as some of the issues which complicate the matter, such as syncretism, homonymy, and clitics.

2.1 Orthographic and Phonological Words

In the written language, the most basic sense of the term *word* is an orthographic word. This is how wordprocessors conduct a “word count.” Plag (2003:4) defines an

orthographic word as “an uninterrupted string of letters which is preceded by a blank space and followed either by a blank space or a punctuation mark.” By that definition then, there are eight orthographic words in (1). The term “uninterrupted” however can be misleading since orthographically we do allow some punctuation to intervene. In (1), *I’m* has an apostrophe in the middle of it, and although it is a contraction of two words it still counts as one orthographic word. Likewise hyphens may also intervene as in *word-formation*. These clitics and hyphenated compounds will be discussed further below. We briefly note here, though, that compounds can orthographically vary. The forms *word formation* and *wordformation* are as acceptable as *word-formation* (Katamba 1993:294; Plag 2003:5). That the word *word-formation* can be represented as either one or two orthographic words is a good example of how wordhood comes in layers. What is one word on one level may be two words at another level. It also demonstrates some of the slipperiness involved in labeling something a word.

The spoken equivalent of orthographic words are phonological words. In normal speech speakers do not pause before and after each word. Indeed the first three orthographic words in (1), when spoken, may come out more like one word: [aɪm nə] or I’m’na. Clearly pauses cannot delineate words like spaces can. Plag (2003:6) points out that even trying to define spoken words by “potential pauses” falls short, since speakers may pause even in the middle of words, perhaps for emphasis. Instead of pauses then, phonological words are demarcated by stress and rhythm. Knowing, even unconsciously, that words in English can bear only one main stress helps listeners judge what is or is not a word. Phonology and stress will be looked at again in §3.2.

2.2 Lexemes and Word-forms

Most of the time the term *word* is used to mean either lexemes or word-forms. Lexemes are abstractions whereas word-forms are the concrete “units which actually occur” either in speech or writing (Bauer 2003:9). In example (1) above we sense that the word-forms *going* and *go* are related to each other and in some sense the same “word.” Also related to these are the forms *gone*, *goes*, and the irregular *went*. What these word-forms have in common is they are all inflected forms of the same lexeme, namely the verb GO. Because lexemes are abstract, they require particular word-forms to realize them in any given context and the lexeme then encompasses all the word-forms

which realize that particular lexeme (Bauer 2003:9). It should be noted, though, that this correlation of lexeme to word-form is by no means 1:1. For instance, the word-form *stores* can represent different inflections of two lexemes, the noun STORE (*two grocery stores*) or the verb STORE (*he stores food*). More will be said about this overlap below.

Because lexemes are abstract and include all the related inflectional word-forms, lexemes are also the dictionary form. That is, dictionaries will not list *go*, *going*, *goes*, and *gone* as separate entries. Rather they will all be found under the single heading GO. So when people talk of looking up a “word” in the dictionary, they are really looking up lexemes. Since they are listed in dictionaries, lexemes can be considered one form of lexical item (§3.5.2).

It should also be noted that lexemes do not include derivational forms. Inflection “produces a new word-form of a lexeme” whereas derivation “produces a new lexeme” (Bauer 2003:14). Hence derivation is a major method of productively creating new words (§5.2.1). This means that while the word-forms *inspires* and *inspirations* clearly share a meaningful base, they represent two different lexemes: INSPIRE and INSPIRATION.

2.3 Grammatical Words

Word can be used to mean grammatical words. Such words are defined by “their place in the paradigm and named by descriptions” (Bauer 2003:10). These descriptions are the morpho-syntactic features of the word in context (Katamba 1993:19). Thus in (1) *going* and *go* are the grammatical realizations of, respectively, the perfect participle and infinitive forms of the verb GO. Once again there is not a perfect correlation between grammatical words and word-forms. Take for example the following:

- (2) (a) John talked to Sally.
(b) John has talked to Sally.

Both (2a) and (2b) contain the word-form *talked* and each instance realizes the lexeme TALK. Yet the two instances do not represent the same grammatical words. In (2a) *talked* represents TALK + past tense. In (2b) *talked* represents TALK + past participle. Such instances of homonymy, where one word-form represents multiple grammatical words or even lexemes, are common in English (Matthews 1991:28) and will be treated in slightly more depth momentarily.

2.4 Lexical Items

When people pause to wonder if something is or is not a word, they may use the dictionary as the deciding voice. People generally tend to think of words as things in dictionaries, that is as lexical items listed in a lexicon. This is a fair assumption since lexemes are listed in dictionaries and in any standard dictionary they will be the most prevalent type of lexical item. The category lexical items, however, is far more inclusive. By definition, a lexical item is a “linguistic item whose meaning is unpredictable and which therefore needs to be listed in . . . dictionaries” (Carstairs-McCarthy 2002:144). Any item which must be learned and memorized is a lexical item. This includes such items as phrasal verbs, particular collocations, idioms, lexical phrases, and even proverbs, all of which consist of multiple word-forms. Using the term *word* to refer to lexical items therefore is misleading. We will return to lexical items when discussing the properties of words (§3.5.2).

2.5 Complicating Factors

Having outlined the sort of items *word* can refer to, we will now briefly turn to some of the idiosyncrasies of English which complicate the matter. In an ideal language there would be no overlaps, no possible confusion, and there would be a 1:1 ratio where every meaning had its own distinct form. English is not ideal. Nor are languages in general ideal. But while the overlaps and oddities of English may prove confusing at times, they also allow for much of the wordplay within the language. Numerous jokes and puns rely on the fact that one word-form can represent multiple lexemes or that a single phonological word can realize multiple word-forms. Below we look at some of these complications which add both color and confusion to the language and our sensibilities of what words are.

2.5.1 Suppletion

In discussing the verb GO above we merely glossed over the fact that the past tense is the irregular form *went*. The forms *go*, *goes*, *gone*, and *going* are clearly related to one another. *Went*, on the other hand, looks as though it ought to realize a completely different lexeme. Indeed *went* was once the past tense of the verb WEND before undergoing what is known as suppletion. According to Bauer (2003:342), suppletion is

“when two forms in a paradigm are not related to each other regularly.” This phenomenon of two morphologically unrelated word-forms realizing the same lexeme is relatively rare in English. Other examples are the grammatical paradigms *good, better, best* and *bad, worse, worst*. So while we might have little trouble thinking of *go* and *going* being the same “word” (that is the same lexeme), it is somewhat counterintuitive to think of *going* and *went* in the same way.

2.5.2 Syncretism

Suppletion is when two grammatical words of the same lexeme are morphologically dissimilar. Syncretism, on the other hand, is when “two grammatical words associated with the same lexeme are represented by the same word form” (Carstairs-McCarthy 2002:146). Whereas suppletion makes a paradigm more complex, syncretism makes it more economic. Syncretism is a common phenomenon. As seen in (2) above, verbs in English regularly syncretize the past tense (*talked*) and past participles (*has talked*). This is not true of all verbs though. Some irregular verbs such as SEE have distinct forms for all of their grammatical words. Compare the following:

(3) TALK and SEE

| | | |
|-------------------------------|----------------------|--------------------|
| basic form | <i>talk</i> | <i>see</i> |
| third person singular present | <i>talks</i> | <i>sees</i> |
| past tense | <u><i>talked</i></u> | <u><i>saw</i></u> |
| progressive participle | <i>talking</i> | <i>seeing</i> |
| past participle | <u><i>talked</i></u> | <u><i>seen</i></u> |

Such syncretism in verbs is typically economic and unproblematic. Examples can be found in many languages, including Russian and Latin (Bauer 2003). Non-grammarians likely never notice that there are “missing” forms.

One example of English syncretism, however, has resulted in a loss that has been felt and is frequently compensated for. Crystal (2003:71) notes that “by the time of Shakespeare, *you* had developed the number ambiguity it retains today.” Formerly *thee* and *thou* would have been the singular forms of the second person pronoun and *ye* and *you* the plurals. But as the language evolved, the forms underwent syncretism until *you* was used not only as both a subject and an object, but also as either singular or plural.

Since then speakers have come up with many compensating forms: *yous/youse, youall, y'all, you guys*, etc. The syncretism of the paradigm however seems to be fixed and none of these alternate forms can be considered standard. Thus the grammatical word for second person plural may be considered to have either no distinct word-form or many.

2.5.3 Homonymy and Polysemy

Homonymy is when two or more words (lexemes or grammatical words) share the same form (Crystal 2003:463). The form shared can be spelling (homography) as in *lead* 'metal' and *lead* 'present tense of the verb LEAD.' Or the form shared may be pronunciation (homophony) as in *lead* 'metal' and *led* 'past tense of the verb LEAD.' Words may also be both homophones and homographs as in *bark* 'dog noise' and *bark* 'part of a tree.' Often these words may not be related¹, as the previous examples show.

Polysemy is when two or more related words share the same form. The difference being that polysemous words are etymologically related whereas homonyms need not be. Polysemous words include *mouth* as in 'mouth of an animal' and 'mouth of a river.' Such examples are described as being the same "word." While they are the same word-forms, whether these are different senses of the same lexeme (MOUTH) or different but related lexemes (MOUTH₁, MOUTH₂) is more difficult to judge.

Both homonyms and polysemous words muddle the distinction of "what is a word" since the same word-form (orthographic or phonological) represents different grammatical words, lexemes, and senses. Generally though, these overlaps are fairly well tolerated and memorizing them is simply part of learning English.

2.5.4 Clitics

Another complicating factor are clitics, which are neither affixes nor words, but something intermediary. In (1) above we saw the word-form *I'm*. Here '*m*' is the clitic and *I* is a host. Clitics, like affixes, are obligatorily bound; they are "incapable of occurring in isolation" (Katamba 1993:246). Thus a clitic and its host always form one orthographic word (disregarding the apostrophe). However clitics and their hosts still behave separately.

¹Syncretism is a type of homonymy involving related grammatical words.

There are two types of clitics: simple and special. Simple clitics are “weakened forms of ordinary words” (Bauer 2003:132) where the clitic belongs to the same word class as the independent word which “could substitute for it in that syntactic position” (Katamba 1993:245). Examples in English include *have, is, will, and would* as, respectively, ‘*ve, ‘s, ‘ll, ‘d*. These clitics attach to whichever word would have preceded the independent word, regardless of word-class. They are semantic equivalents and there is no difference between *I’ve seen it* and *I have seen it*. Moreover syntactic operations never treat clitics and their hosts as single units (Katamba 1993:248). Both the clitic and the host still fill their respective syntactic roles within the sentence. So while an example like *would’ve* is orthographically one word-form, syntactically and semantically it functions as two separate grammatical words and two separate lexemes: WILL and HAVE.

Special clitics are not the reduced forms of independent words. The prime English example is the possessive ‘*s*. Unlike an affix, which can only attach to a word, special clitics may syntactically/semantically attach to full phrases. They show what Klavans calls “dual citizenship” (Klavans 1985:104 in Katamba 1993:248). For example:

- (4) a. The dog’s bowl.
 b. The director of the play’s hat.

In (4a) *bowl* belongs to *dog*. *Dog* is both the phonological host, to which the clitic attaches, and the syntactic/semantic host. But in (4b) *hat* does not belong to *play* it belongs to *director*. In this case the phonological host and the syntactic/semantic host differ and the clitic in a sense belongs to both as it attaches to the phrase. Special clitics, like simple ones, appear as single words orthographically, but to say they are simply a part of that word-form is an oversimplification of a bigger syntactic/semantic picture.

2.5.5 Periphrasis and Phrasal Verbs

Where clitics are two grammatical words represented as one orthographic word, periphrastics and phrasal verbs are just the opposite. They are grammatical circumlocutions when a single word will not suffice.

Periphrasis is “the use of separate words instead of inflections to express a grammatical relationship” (Crystal 2003:466). Two main instances of periphrasis in English are in verb paradigms and in the formation of comparatives and superlatives. In

a highly inflectional language like Latin, many of the verb tenses can be represented by a single word-form (e.g. *amabimus* ‘we will love’). English however frequently uses auxiliary verbs to express tense, as in *have gone*, *am going*, or *will go*. Some adjectives and adverbs also resort to periphrasis. While some form their comparatives and superlatives through the inflections *-er* and *-est*, others, namely most of the multi-syllabic ones, must use *more* or *most*:

- (5) *big* *lovely* *beautiful* *happily*
 bigger *lovelier* *more beautiful* *more happily*
 biggest *loveliest* *most beautiful* *most happily*

Phrasal verbs consist of a lexical verb and one or more adverbial or prepositional particles (Crystal 2003:466; Katamba 1993:307). Some phrasal verbs have straightforward meanings: *come in*, *turn off*, *take up*. Many though are idiomatic: *put up with*, *do in*, *blow over*. The meaning of such verbs is tied up in all the parts and thus they function as one unit. Phrasal verbs, along with periphrastic forms, use multiple orthographic words to represent single grammatical words and single lexemes.

Another very common category of multi-word words are compounds. These will be discussed later in §5.2.3. Additionally, English has a handful of conjoined words such as *nevertheless* or *insofar* which do not follow the general rules of compounds.

2.5.6 Spelling

How does spelling complicate what we mean when we talk of words? According to Burchfield:

An almost unqualified belief in a one-to-one relationship between most words in the language and the way they are spelt has been maintained since at least 1755 when Dr. Johnson’s dictionary was published.
 (1985:146)

Such a one-to-one relationship is only an ideal, and one that has not always existed. Prior to the invention of the printing press, and for a while after, most words could be spelled/spelt in a number of different ways. Bryson (1990:116) gives the example of *where* which “has been variously recorded as *wher*, *whair*, *wair*, *where*, *were*, *whear*, and so on.” Nor was it uncommon for two variations of the same word to occur in the same passage of writing.

Even today there are numerous words with multiple spellings. Above we mentioned the example of *wordformation*, *word-formation*, and *word formation*. There are many similar compounds that can vary orthographically. With *word-formation*, though, it is a matter of spacing and hyphenating. More complicating are those words which have two standardized spellings. A number of these are simply British vs. American spellings as in *programme/program*, *theatre/theater*, *learnt/learned*, *colour/color*, or *realise/realize*. Others, however, are non-regional variants such as *judgement/judgment*. These are each different word-forms, since a word-form is a concrete realization and includes spelling. But they also each represent the exact same grammatical word for the same lexeme. Which variant the lexeme is named after is merely a question of preference.

What then of misspellings? English, because of various phonetic changes and foreign influences, is notoriously difficult to spell. In fact Bryson (1990:111) describes English spelling as “so treacherous . . . that the authorities themselves sometimes stumble,” citing examples of dictionaries that have misspelled words in various editions. If even lexicographers misspell words it should be no wonder that the general populace struggles. We might then ask if *millenium* is a variant of *millennium* if enough people frequently misspell it that way. Though “wrong” we still understand which lexeme it is meant to represent and communication does not reach a gridlock.

While many misspellings are unintentional, there are also examples of deliberate orthographic tampering. Most of these are informal such as *nuff* for *enough*, *ya* for *you*, *gotcha* for *(I’ve) got you*, or *nite* for *night*. Such forms are usually shorter and reflect pronunciation. Though informal, some have become “standard deviants” or “accepted ways of writing colloquial form” (Crystal 2003:275). The example *gotcha*, like a clitic, combines multiple grammatical words into one word-form.

2.6 Preliminary Conclusions

We have now established that *word* can refer to orthographic words, phonological words, lexemes, word-forms, grammatical words, and to an extent to lexical items. Overlaps (e.g. orthographic words are word-forms) as well as other complicating factors have made the term more slippery than our native speaker intuitions might suspect. That said, *word* is vague because it can be, since the distinctions are not always

important. Indeed, for the remainder of this paper *word* will continue to be used in a vague manner, though typically it will mean lexemes and word-forms. For distinction, lexemes will continue to appear in small caps (e.g. GO), and word-forms in italics (e.g. *going*).

3 Properties of Words

Now that we have established what we mean when we talk about “words” we will look at those characteristics which determine whether something is or is not a word (using *word* in a vague way where the distinction between lexemes and word-forms is unnecessary). In the broadest sense, “a word is what native speakers think a word is” (Matthews 1972 in Bauer 1983:9). Yet even this definition is unsatisfactory since native speakers will not always agree about whether colloquialisms or slang terms (e.g. *ucky*, *spiffy*, *homie*, *fantabulous*) are words or whether compounds are one word or two (e.g. *anywhere*, *blackboard*, *apartment building*). Indeed no matter how wordhood is ascribed, there remain items which defy clear-cut definitions.

The following section, rather than coming up with a definitive definition for words, will analyze some of the properties words prototypically have. We will also look at some attributes we intuitively think they have and discuss problematic exceptions along the way. Included are features of orthography, phonology, morphology, syntax, and semantics. Additionally we will look at words as lexical items and words as they are interrelated with other words.

3.1 Orthography

One property of words is that they can be orthographic units. In a sense, words are spellable. This does not mean they are easy to spell, as shown above (§2.5.6), but that they should have some form of orthographic representation. This distinguishes words from mere noises. Some noises which become meaningful and established in a language community eventually find a way of being written: *um*, *er*, *pshaw*, *tut-tut*, *shh*, as well as a host of onomatopoeic words. People may not always agree on how to spell such units (e.g. *miaow*, *meow*, *me-yow*, *mew* for the sound of a cat). Nevertheless, such units can be spelled. They also meet the main orthographic criteria for words: they are units bounded by spaces. At least orthographically, they are words.

We have already covered that orthography can be an unreliable criterion when it comes to compound words which may take one—or more—of several forms (e.g. *bookshelf*, *writer-director*, *garbage can*). If orthography were the only standard, wordhood would be left to “the fancies of individual writers or the arbitrariness of the English spelling system” (Plag 2003:5). Additionally, this would mean that illiterate speakers and speakers of languages with no orthographic tradition would be unable to discern what is or is not a word. This is not the case. Thus while “bounded by spaces” is a property of prototypical words, it is not a foolproof linguistic criterion.

3.2 Phonology

Just as words are spellable, they must also be pronounceable. Like the orthographic criterion, this distinguishes words from mere noises. And if a noise is culturally meaningful enough, it will eventually receive a semi-equivalent pronunciation in the form of an onomatopoeic word like *boom* or *woof*.

Additionally, words must be pronounceable within their language. This incorporates the fact that words are divided into syllables. Syllables are “groupings of sounds for the purposes of articulation” (Katamba 1993:34). Without going into too much detail, syllables consist of one or more sounds and are divided into three parts: the onset, the nucleus, and the coda (Plag 2003:81). For a string of sounds to be considered a word within a given language only certain sound clusters can exist within the onset and coda. For example, in English *tr-*, *st-*, and *bl-* are common onsets. Additionally, English does not tolerate too many consonants or too many vowels clustered together. A word like **gpid* cannot exist in English because *gp-* is an “illegal syllable-initial combination” (Plag 2003:82) and therefore unpronounceable. Things become more complicated as foreign words enter the language with exotic sound clusters (e.g. *schmaltz*). Often, loan words become anglicized, or made to fit English sound patterns, as in *raccoon* from the Algonquin word *raugroughcun* (Bryson 1990:68).

Pronounceability, however, does not distinguish words from phrases or sentences. We have already noted that there are typically no pauses to delineate words in speech. Phonologic stress patterns, however, are helpful in delimiting words. In English, “every word can have only one main stress” (Plag 2003:6). Unlike orthography, this can account for compounds, where *gárbage can* is two orthographic words but one stressed

unit. Trouble with this criterion arises in that not all words typically bear stress. Function words like *the*, *an*, or *in* may only carry emphatic stress but are nonetheless some of the most basic words.

3.3 Morphology

There are three morphological features of words: they consist of morphemes and they are characterized by their “uninterruptability” and “internal integrity” (Plag 2003:5; Bauer 1983:105). Firstly, words are made up of one or more morphemes. Whereas syllables are units of sounds, morphemes “are the smallest units of meaning and grammatical function” (Katamba 1993:34). They are the building blocks of words. There are different types of morphemes (e.g. free, bound, roots, affixes) which will be discussed in detail in §4.1. Suffice it to say a word will always be at least monomorphemic (e.g. *he* or *cat*) and will frequently be polymorphemic (e.g. *un.do.able* or *nation.al.ist.ic*).

Unlike sentences, where words and phrases have some degree of mobility, the internal parts of words are fixed. Affixes are rule governed. Uninterruptability means modifications can only be added to the edges of words and in certain orders, thus *commonly* can only become *uncommonly*, never **commonunly*. There are some exceptions, however. The plural of *son-in-law*, for instance, is *sons-in-law*, a fact which could imply that the term is an idiomatic phrase rather than a word. There also exist slang terms like *abso-bloody-lutely*, though such exceptions are rare.

Similarly, internal integrity means that the internal components (i.e. the morphemes) “cannot be reordered within the word” (Bauer 1983:105). Thus *lawlessness* cannot become **lawnessless*. Historically, though, there are instances where metathesis has disrupted internal integrity so that *brid* became *bird*. Such examples occurred before the standardization of the language and metathesis now results only in misspellings and mispronunciations.

3.4 Syntax

Syntactically, words are the “fundamental unit out of which phrases and sentences are composed” (Carstairs-McCarthy 2002:146). In this sense they are building blocks. They are also the smallest unit of syntax with “positional mobility” (Bauer

1983:105). Words and groups of words can, to an extent, be repositioned within a sentence. However since words have internal integrity, units smaller than words cannot be repositioned. Compound words, even when two orthographic words, must always move as a single syntactic unit:

- (6) a. I love strawberries. > Strawberries I love.
b. The garbage can is full. > Is the garbage can full?

Because certain syntactic functions require the movement of complete phrases, this criterion cannot always distinguish between compound words and phrases (a distinction we will return to in §5.2.3).

Provisionally words can also be considered the smallest free-standing forms. This distinction, however, falls short with function words. It would take a very specific context to allow *the* or *my* to stand alone. Words are also the smallest omissible unit when given “appropriate discourse conditions” (Bauer 2003:66) as seen in (7):

- (7) A: James can swim. B: Mary can't [swim].
A: Is he coming in June? B: No, [in] July.

Affixes, however, cannot be omitted:

- (8) A: Is it repairable? B: *No, but it's [re]placeable.
A: Are the girls and boys going? B: *Just the girl[s].

According to Bauer (2003:67) a word is “the smallest unit which can be omitted when it would be identical with another element which occurred earlier in the discourse.” The repeated word is tacitly understood.

Additionally, words are classifiable. If something is a word it can be categorized into one of the word classes (Plag 2003:8) based on how it behaves syntactically. The word classes include lexical words—nouns, verbs, adjectives, and adverbs—as well as function words—articles, conjunctions, pronouns, etc. Thus if a unit can be classified into a syntactic category it can be considered a word. While this criterion manages not to exclude anything we would consider a word, some might consider it too inclusive.

Multi-word compounds and idioms may function as particular parts of speech:

- (9) a. A *devil-may-care* attitude (adjective)
b. Spoke *matter-of-factly* (adverb)
c. Three *jack-in-the-boxes* (noun)

(Carstairs-McCarthy 2002:67). At least in (9a), *devil-may-care* is syntactically a word

not a phrase. Furthermore, phrases are also classifiable according to their syntactic behavior. Words then may be defined not simply by their own classifications, but by the phrasal categories they head. That is, we can recognize a word as an adverb if it heads an adverbial phrase.

3.5 Semantics

One of the most intuitive ways to define words is that they carry meaning or represent a “unified semantic concept” (Plag 2003:7). Such meaning is arbitrary in that the “associations between most words and their meanings are purely conventional” (Carstairs-McCarthy 2002:7). There is nothing particularly canine about *dog*. If there were, it probably would be *dog* in all languages rather than *chien*, *Hund*, or *perro*. Because word meaning is arbitrary, words are also generally opaque and their meanings must be learned and memorized. This leads to another intuitive criterion, that words are things found in dictionaries. Yet as intuitive as these criteria are they are both highly problematic.

3.5.1 Meaning and Meaning Change

Firstly, it is difficult to pin down a meaning or definition for function words like *in*, *the*, or *that*. Such words are meaningful only in context. Additionally, while a “unified semantic concept” will include compounds like *harbor patrol* or *leadership training program*, there are ample concepts with no equivalent words in English, such as “the smell of fresh-baked bread.” It is also a stretch to say that the concept represented by a certain word is “unified.” The word *dog* represents everything from a Dachshund to a Labrador to a St. Bernard. As seen above in §2.5.3, individual word-forms may additionally convey multiple meanings, some of them completely unrelated (e.g. *bank* ‘financial institution’ and *bank* ‘mound of earth by a river’). Some perplexing words even manage to have dual meanings in opposition to each other. Bryson (1990:63) notes that “*cleave* can mean cut in half or stick together.” Additionally, one can argue that units smaller than words (i.e. morphemes) can also bear meaning. For example the prefix *re-* often means “do again” the action of whichever verb it attaches to.

Meaning is also not a fixed property of words. Words often change their

meanings over time². Or as Crystal (2003:138) puts it: “semantic change is a fact of life.” These changes have occurred throughout the history of the language and continue today. There are a number of ways in which a word may change its meaning. One is that words may broaden or extend their range of meaning. The word *butcher* for instance was once far more specialized and meant only a ‘slaughterer of goats’ (Wolfram & Schilling-Estes 1998:59). The opposite also occurs, where a word with a general meaning narrows or specializes. Classic examples of narrowing are *meat* (which meant food in general), *deer* (which could be any animal), and *girl* (which could refer to a child of either gender).

Meanings can also go through either amelioration or pejoration, as words gain or lose negative connotations. An example of amelioration would be *lean* which formerly implied being unhealthily thin but now means trim and athletic. Conversely, *lewd* has gone from meaning “of the laity” to “sexual impropriety” (Crystal 2003:138). Finally, words can undergo meaning shift, where a secondary meaning becomes a primary meaning. The word *bead* once meant “prayer,” but came to be associated with the rosary beads worn while praying. A specific type of meaning shift is metaphoric extension (§5.3.5) where new uses of a word are added “based on a common meaning feature” (Wolfram & Schilling-Estes 1998:60). This is one way in which polysemy occurs. For example *mouth* by extension can mean not only the oral opening of a human or animal, but also an opening for a cave or bottle.

3.5.2 Lexical Items and Lexicons

Defining words as things listed in dictionaries is also problematic. We have already seen that words—or more specifically, lexemes—are only one type of lexical item. Moreover, a word need only be listed if its meaning is thoroughly unpredictable. Though words are often opaque in their general sense (i.e. the arbitrary base form of a word), through derivational processes and compounding words can be predictable based on their parts. That is to say their analyzability may equate (though certainly not always) some transparency provided one knows the meanings and properties of the roots and affixes involved (Bauer 1983:19). For instance, *loveable*, *bathtub*, and

² Phrases too can shift in meaning as they or their components become idiomatic. For instance the phrase “a gay man” means something totally different now than fifty years ago.

gravedigger are relatively predictable when one knows the meanings of the various component morphemes. Terms like *blackmail* or *blackguard*, however, remain opaque. Highly predictable forms are not always listed in dictionaries.

Arguing wordhood based on the dictionary is thus like arguing fruithood based on the current selection at the local grocer. One grocer may prefer more exotic fruit than another just as one dictionary may be more discriminate about the words it includes. Quirk (1968:143) points out that people often refer to “the dictionary,” demonstrating a “tendency to think that there is only one dictionary.” In truth there are many and they do not always agree on the words to include. Some dictionaries may be more open to medical or scientific terms, others may lean toward American terms, still others toward World English terms.

More appropriate than saying “a word is something in a dictionary” might be “a word is something in a lexicon.” Granted, a dictionary is a type of lexicon, but it is not the only type. *Lexicon* can also refer to an individual’s vocabulary. This mental lexicon includes both the active vocabulary—words used frequently that can be recalled as needed—and the passive vocabulary—words used infrequently but nonetheless recognized and more or less understood (Crystal 2003:123). *Lexicon* may also mean all the words within a given language, which would in a sense be a compilation of all the words in all the dictionaries. Lastly, there is what we might term the On Call Lexicon. This would include those words which do not need to be either listed in the dictionary or memorized because they are derived from fully predictable means. One such word might be the verb *reswim*, as in “Next year he will reswim the race he lost.” It is an available word when needed. Otherwise it vanishes until it is needed again. Items in the On Call Lexicon are called *nonce words* (§4.4.2). To put such words in a dictionary would be “commercially unattractive” since their meanings are “immediately clear to anyone familiar with the basic meaning of productive affixes” like *re-* (Baayen & Renouf 1996:69).

3.5.3 Predictability and Productivity

Because of their occasional transparency and predictability, words (lexical, not functional) are productive and are frequently coined. That is, because we can understand or guess at the meaning of some unfamiliar words, new and unfamiliar words can be

added to the language. Affixes on the other hand, while themselves tools of this productivity, are not productively added to. Also, though new sentences are always being written and uttered, they are then forgotten. Rarely do sentences get coined, except possibly in famous quotes or advertising slogans. While some nonce words appear and vanish like sentences, neologisms (literally “new words”) are constantly being added to the lexicon. While productivity (§5) is a property of words, phrases, and sentences, only words are productively created with the intent of enduring.

3.6 Words with Other Words

A rather self-evident property of words is that they go together with other words. Words do not exist in isolation. They are interrelated in numerous different ways. Meaning itself does not exist entirely within a single word. According to Quirk (1968:139) “we cannot say what the meaning of a word is until it is put into an adequate context” because the meaning is spread over not only the word but the neighboring words as well. A word like *orange* assumes different meanings in the context of colors than in the context of fruits. Additionally, the term *uncle* takes much of its meaning from its relation to the words *brother*, *father*, *mother*, or *aunt* (Crystal 2003:156).

3.6.1 Sense Relations

Dictionaries list words in alphabetical order, regardless of meaning. Words, however, can be semantically grouped in a number of other ways. One common method is a thesaurus, which groups related words together according to categories such as food or business. Usually we think of using thesauri to look up synonyms, which are one of the most common types of sense relation. Typically people think of synonyms as being words that mean the same thing. This is rarely, if ever, true. Two words never mean *exactly* the same thing or are *perfectly* interchangeable. There are always nuances of meaning that distinguish a pair of synonyms or contexts that will allow one term but not the other. For instance *large* and *spacious* are synonyms when discussing room-size, but you would not ask for a “spacious slice of cake.” Moreover one would expect a psychiatrist to use the term *insane* or *mentally unstable* rather than *loony* or *nuts*.

Another type of sense relation is antonyms, which are more clear-cut. *Bad* is really the opposite of *good* just as *tall* is the opposite of *short*. Though while *light* seems

clearly the opposite of *dark*, *light* can also be the opposite of *heavy*. Antonyms exist in various forms: gradable, which can be made into comparatives/superlatives like *dry/wet*, *dryer/wetter*, *driest/wettest*; complementary, where the two terms are mutually exclusive such as *dead/alive*; and converse, where the two are mutually dependent as in *buy/sell* (Crystal 2003:165). Other sense relations are hyponyms (e.g. a *duck* is a type of *bird*), hierarchies (e.g. *second*, *minute*, *hour*), series (e.g. the days of the week), and part-whole relations (e.g. *finger/hand*). What all of these sense relations show is the interrelatedness of words and how the meaning of a word does not exist in a linguistic vacuum.

3.6.2 Collocations

Another way in which words interact and take meaning from each other is through collocation. The collocations of a word are essentially “the lexical company the word keeps” (Hoey 2003) or which other “words it goes with, likes, attracts” (Ter-Minasova 2005:450). Most words prefer certain words over others and the use of one word will often “call up” another (Crystal 2003:162). *Inconsolable* for instance is often paired with the word *grief*. Some collocations are stronger, or more fixed, than others. Carstairs-McCarthy (2002:11) gives the example of *white* in *white wine*, *white coffee*, *white noise*, and *white man*. These are not quite fixed enough to be compounds (each word still has its own stress). Nor are they quite idioms because they are half predictable (i.e. *white noise* is a type of *noise*). Other collocations are less fixed but frequently tend to crop up together. The verb *pay* often collocates with *attention*, *visit*, and *compliment* (Ter-Minasova 2005:450). Lastly, it should be noted that collocations are culturally or language based and by no means universal. Thus a primary difference between a native speaker and an advanced learner of a language is the former’s ability to appropriately use collocations (Hoey 2003).

3.7 Summary

We have now looked at some of the basic properties of prototypical words. These characteristic attributes are summarized in (10):

(10) Words:

- Are frequently bounded by spaces
- Are spellable

- Contain pronounceable syllables
- Can have only one main stress
- Consist of one or more morphemes
- Have internal integrity and are generally uninterruptible
- Are the building blocks of syntax with some degree of positional mobility
- Are the smallest free form and smallest omissible form
- Can be sorted into syntactic word classes
- Carry arbitrary, generally opaque meanings
- Are listed in lexicons, often including dictionaries
- May be predictable through derivation and are thus productive
- Interrelate with other words in various ways
- May prefer some words over others

4 Subdivisions of Words

So far we have discussed the types of units that *word* can refer to as well as the prototypical properties of words. For this next section we will look at some of the other ways in which words can be classified and subdivided, whether by grammarians, linguists, or the general populace. This includes the component units of words (i.e. morphemes), the various types of words based on both grammatical and pragmatic usage, as well as the various stages a word passes through as it becomes established within the language.

4.1 Building Blocks of Words

It has already been noted that words are not the minimal unit of language. Words may be the building blocks of syntax, but they themselves are built from morphemes. Because of the “buildability” of words, they are also productive and morphemes are like Lego bricks that can be “used again and again as building blocks to form different words” (Katamba 1993:20). Unlike Legos, where any combination is possible, morphemes are governed by rules. How new words are composed, or coined, is the topic of §5. Here we simply look at the types of morphemes used and the types of output possible.

4.1.1 Morphemes

There are essentially two types of morphemes: bound and free. Bound morphemes are those which can only occur in combinations, never alone. That is, a bound morpheme cannot be a word. Free morphemes can stand alone as well as occur in combinations. So far the distinctions are simple enough.

Morphemes can further be divided into the complementary categories bases and affixes. Affixes are always bound morphemes and must be attached to a base. Prefixes are those added before the base; suffixes are those added after the base. Suffixes can be further subdivided into inflectional and derivational (§5.2.1).

Bases are simply defined as “the part of a word which an affix is attached to” (Plag 2003:11) and may be either bound or free. The *press* in *impressionistic* is clearly a free morpheme. Whereas the *feas-* in *feasible* and the *loave-* in *loaves* are bound. Although the term *stem* is variably used in linguistics, for our purposes it will refer to such a bound base. Since multiple affixes are not only allowable but typical, a base may consist of more than one morpheme. For example, *impressionistic* can be seen as having four (free) bases: *press*, the base for *im-*; *impress*, the base for *-ion*; *impression*, the base for *-ist*; and *impressionist*, the base for *-ic*. The first of these bases (*press*) is known as the root, and is the minimal base that cannot be further analyzed.

One final type of bound morpheme is the cranberry morpheme. It is so named because *cranberry* is the prime example. The morpheme *cran-* is not only obligatorily bound, it is found in only that one combination. *Huckle-* and *-ric* are also cranberry morphemes found only in *huckleberry* and *bishopric* (Carstairs-McCarthy 2002:19).

4.1.2 Simplex vs. Complex Words

Morphologically speaking, then, there are two types of words: simplex and complex. Simplex words consist of one free morpheme. These are the root words of the language and many of the most basic: *love*, *dog*, *faith*, *happy*, *up*, *king*, etc. Since they are root forms and unanalyzable, simplex words must all be learned by memory. Complex words are those which have analyzable parts in various combinations. Free bases can be paired with affixes (e.g. *lovely*, *uncover*), bound bases combine with affixes (e.g. *feasible*, *aggression*), and even free bases can pair with other free bases to form compounds (e.g. *doghouse*, *proofread*).

4.1.3 The Trouble with Morphemes

Before moving on we should note that morphemes are more complicated than the Lego analogy might indicate. That they are building blocks is not the problem. The question is, are they building blocks defined by their meanings or their forms? Some morphemes do seem to have regular associative meanings. Others vary. Moreover the same form may serve different functions such as *-al* which can be nominalizing (e.g. *referral*) or adjectivalizing (e.g. *colonial*). Further complicating the notion of morphemes are things like conversion (§5.2.2) through the so-called zero-morph, words that are inflected through vowel change (e.g. *sing/sung*), and processes of reanalysis (§5.3.2) where monomorphemic words are treated as though polymorphemic (e.g. *workaholic* from *alcoholic*). As Plag (2003:26) puts it, the morpheme “is a useful unit in the analysis of complex words, but not without theoretical problems.”

4.2 Classifications of Words

We already noted that one of the properties of words is that they are syntactically classifiable (§3.4). We know these categories as word classes or parts of speech. It is not uncommon to think that meaning is the common denominator for a class. We often think of nouns as things or ideas, adjectives and adverbs as descriptors, and verbs as actions. Carstairs-McCarthy (2002:45), however, uses the example of *PERFORM* and *PERFORMANCE*, which both seem to denote the same activity. Additionally, a verb such as *RESEMBLE* (as in “John resembles a beaver”) can be interpreted as a type of description. Meaning is by no means a failsafe criterion for classification.

Instead, words are classified based on morpho-syntactic features. Verbs must follow the syntactic rules for verbs; prepositions must follow the syntactic rules for prepositions. For some classes, this includes how a word is inflected. Nouns may be singular or plural, which is regularly formed by adding *-s/-es*. There are those nouns which are irregular (e.g. *sheep, oxen, scissors*), but these still display other syntactic features of nouns. For a word to be a verb, it should be able to form (regularly or irregularly) all the necessary grammatical words associated with verbs. It should be noted that lexemes, not word-forms, are classified. There are plenty of instances of conversion (§5.2.2) where the same word-forms can be classified in two different word classes:

- (11) a. I will *dream* a *dream*. (verb, noun)
b. Tom ate a *chocolate*. Mary ate a *chocolate* strawberry. (noun, adjective)

One last thing about word classes. We mentioned that the classes fall into two larger categories: lexical and functional. These categories also line up with the larger categories of open and closed classes, where the open class can be added to and the closed class cannot. Lexical words which carry meaning belong to the open class. New nouns, adjectives, verbs, and adverbs are being added to the language all the time. Functional words, which only have “meaning” when in context, belong to the closed class. It is highly unlikely that English will coin a new determiner, article, or—despite some attempts—pronoun.

Words may also be classified according to how they came into the language. For instance they may be classed as borrowings, acronyms, abbreviations, blends, clippings, compounds, etc. These various word-formation processes will be the topic of §5.

4.3 Pragmatic Classifications of Words

A less formal way to classify words is based on the types of discourse settings they occur in as well as the ways people view them. The English lexicon is so large and varied that it is impossible to know all the words in it. Thus people specialize according to their needs and circumstances. People use different types of words depending on their job, their culture, their hobbies, where they are from, whom they are with, and even the mood they are in.

4.3.1 Dialectal, Regional, and Cultural Words

The English language can be thought of as consisting of a number of sub-languages. These of course include the major varieties of English—British, American, Australian, etc.—as well as the multitudinous regional dialects within each variety. Essentially, wherever English is spoken, its users will have their own needs and their own ways of making the language suit those needs. Wolfram and Schilling-Estes (1998:52) note that “one of the most noticeable differences among dialects are the different vocabulary words we find in different language varieties.” The *hood/bonnet* and *trunk/boot* distinctions between American and British Englishes are well known. In the United States, various regions refer to carbonated beverages as a *soda*, *pop*, or even

coke. These are matters of different words for the same thing. But words may also exist in one region which do not “exist” in another. A rural town off Chesapeake Bay will undoubtedly have a stock of different words when compared to those of a suburb in the deserts of Arizona, a village in the Highlands of Scotland, or a metropolis like London or Chicago. Each vocabulary is based on the local customs, food, history, climate, and geography. Put simply, speakers “need different words because they have to—or want to—talk about different things” (Wolfram & Shilling-Estes 1998:53) depending on their lifestyles.

While some words are regionally bound, others are culturally bound. People “belong to different social groups and perform different social roles” (Crystal 2003:364). Culture in this case can refer to a number of things and a person is not limited to one culture. Ethnicity, socio-economic status, gender, religion, and education can all play a part in the words a person knows and uses. The Jewish culture, for instance, uses many words from Yiddish (e.g. *schlep*, *schmuk*, *schlemiel*, *chutzpah*), some of which have become mainstream. Additionally, the more education one receives the more one is expected to use “big words” like *erudite*, *audacious*, or *aesthetic*.

4.3.2 Jargon

The occupation a person has also influences the words they use as well as the ways they must use them. Some words just seem suited to some professions. We expect doctors to use words like *hemorrhage* or *defibrillating*, lawyers to use *ad hoc* or *aforementioned*, scientists to use *quantum* or *bioluminescence*, sportscasters to use *overtime* or *scrimmage*, and advertisers to use *innovative* or *revolutionary*. Not only do the words differ, but so do the ways in which they are used. Occupations will vary in their level of formality. For instance, in the field of religion, not only will you get such words as *resurrection* and *atonement* but you may also find grammatical variation and older words such as *thee* or *giveth*.

Such occupational terms are known as jargon. Unfortunately, this term has developed negative connotations. Jargon ideally means “specialized vocabularies” (Wolfram & Schilling-Estes 1998:62) used by the insiders of a field who need specialized words. However, since outsiders do not always understand the words used, misunderstandings arise and the jargon can seem purposefully impenetrable and

recondite. Indeed jargon can be taken to extremes when, for example, businesses purposefully cloud the truth in layers of meaningless words. This is known as gobbledegook. One example is the use of “currently undergoing personnel surplus reduction” for “layoffs.” The fact remains, though, that everyone uses jargon of some sort, not only in their occupations but in their hobbies. Jargons are useful and make in-group communication run more smoothly and efficiently. Whether it be role-playing games, sewing, fishing, computers, Renaissance fairs, or movie/television/book fan clubs, almost everyone participates in some jargon. Moreover, people enjoy their own jargon and the “in-jokes which shared linguistic experience permits” (Crystal 2003:174).

4.3.3 Informal Words

There are a number of different types of informal words, some widely used, others highly stigmatized. One of the major categories is slang. The popular quote by Carl Sandburg describes slang as “language which takes off its coat, spits on its hands – and goes to work” (qtd. in Crystal 2003:182). Such a description seems to place slang as language used by the working class. Indeed, there are many who view slang as solely spoken by either lower/peripheral social classes or by youth. The fact is, like jargon, we all use slang, nor is it “forbidden in any social class” (Burchfield 1985:130). People just use different types of slang depending on who they are and who they “hang out” with.

The real difficulty is the “rather loose, imprecise way the term *slang* is often popularly used” (Wolfram & Schilling-Estes 1998:62). Wolfram and Schilling-Estes go on to describe slang as existing on a continuum, where a set of characteristics determines the “gradient nature of ‘slanginess’” (1998:63). For starters, slang is always informal and is often used to indicate in-group membership. Slang can also be a “special kind of synonym” which deliberately flouts “the conventional more neutral term” (Wolfram & Schilling-Estes 1998:64). An example would be using *bonkers* or *loony* for *insane*. The idea is the same, but the connotations are different. Slang terms are also usually thought of as having a short life span. For some, this means they simply fade out of the language and are forgotten (e.g. *squiffy* for “tipsy”). Others may become mainstream terms (e.g. *mob* or *slum*) while still others remain slang indefinitely (e.g. *flunk* or *cram*).

A second group of informal words are colloquialisms. Unlike slang, these words are “not closely associated with in-group identity or with flouted synonymy” (Wolfram

& Schilling-Estes 1998:65). Included in this set would be the simple clitics discussed earlier (§2.5.4). Words like *they'll*, *he's*, or *couldn't* are all in general usage, but are typically reserved for more informal forms of discourse.

Another class of words are taboo words. Words themselves are harmless. But over the years certain words have developed negative connotations. Thus to avoid embarrassment or giving offense, there are many words deemed unfit for polite society. Taboo words typically involve filth, sexuality, the sacred, or “physical, mental, and social abnormality” (Crystal 2003:172). Such words are also subjective, where what is taboo to one person may be fine to another. Taboo words are not necessarily swear words, though there is certainly an overlap. One thing both taboo words and swearing have in common is that they give rise to euphemisms such as *gosh*, *fetch*, *darn*, or *little girl's/boy's room* (for the slightly taboo *toilet*). According to Crystal (2006:132) everyone swears because it is a “natural response to an emotional state.” The difference is whether a person uses a taboo expletive or a mild euphemism.

Archaisms are another type of informal word class. These are typically “old-fashioned” or “dated” words used to evoke a former era. Medieval stories are rife with *damsel*, *quoth*, *yonder*, or *smite*. Whereas phrases like “capital idea” or “beastly weather” call forth more Victorian times. Archaisms can be found throughout poetry, literature, and films. Additionally, the jargons of religion and law often still use archaic forms. Similar are fossilized words, words which are essentially dead but are preserved in a phrase or expression (Bryson 1990:73) such as *hem and haw*, *raring to go*, or *out of kilter*.

4.4 Bestowal of Wordhood

We now return to additional linguistic classifications of words, this time the terminology involved in becoming a word. It is not easy to judge when a lexical item has become a full-fledged word. There is no elaborate ceremony where a lexical item kneels before a monarch who bestows wordhood upon it. We might consider induction into one of the main dictionaries as being close. But inclusion in a dictionary does not grant wordhood, it merely acknowledges that sometime since the dictionary's last edition, a particular word has become part of the speech community.

4.4.1 Existing and Established Words

In one sense, a word exists if somebody coins it. But that does not necessarily mean it is a word in the language. One difficulty with determining the existence of a word is deciding for whom or what it exists (Bauer 2001:34). No individual speaker can be expected to know every word in their native language. Various factors such as memory, education, or personal experience all limit a person's vocabulary. We have already covered that for multiple reasons dictionaries will not, and cannot, contain all of the existing words in a language. Bauer (2001:36) concludes that the best way to judge if a lexical unit is a "word" is to see if it is used within all or part of a speech community. Thus an existing word is one that has been coined and may be familiar to some speakers, but is not generally well-known. A word becomes established when it is known by "a large enough sub-set of the speech community" (Bauer 2001:36) and it becomes viable for inclusion in a dictionary.

4.4.2 Nonce Words and Neologisms

When new words are coined they are either nonce words or neologisms. Both of these are words which exist but are not established. Nonce words are "coined on the spur of the moment" (Bauer 1983:42) to fill an immediate need. They are temporary by nature and are not meant to become established. People use nonce words all the time either because no word exists for a situation/item or because the person cannot remember the correct term. Because nonce words are meant to fill immediate needs, they must be immediately understandable and are thus coined using productive means such as derivation or blending. Crystal (2003:132) gives the example of a "fluddle," which was used to describe something smaller than a flood but bigger than a puddle. If a nonce word is genuinely useful it may be coined on separate occasions by different individuals. In this way many of the words in the On Call Lexicon (§3.5.2) are nonce formations.

Neologisms on the other hand are simply new words. Some neologisms may become established whereas others will be vogue for a while then disappear again. There is really no way of predicting if a coinage will be a nonce word or a neologism, or if a neologism will become established. Factors involved in a neologism's staying power include who coins it (e.g. a celebrity or other well-known person), whether it fills a need or gap in the language, and who is willing to jump on the word's bandwagon. Regarding

the latter, a neologism (like *bling* for example) may be picked up by a certain group, become a slang term for a while (§4.3.3), and then eventually become established among the general speech community.

4.4.3 Possible, Actual, and Probable Words

According to Plag (2003:46), a possible word is one “whose semantic, morphological or phonological structure is in accordance with the rules and regularities of the language.” Possible words are those which *could* exist within a language and in some cases actually do exist. Along with being regularly formed, possible words must be predictable in meaning (just like the nonce words and neologisms they can potentially become). For example, *delouse* and *deaccessorize* are both regularly formed and predictable. They are both possible words. The difference is that *delouse* is also an actual, or established, word. Not all actual words are possible words though. Actual words need not be predictable and indeed are often idiosyncratic. Plag (2003:47) uses the suffix *-able* as an example. The words *affordable* (‘can be afforded’) and *manageable* (‘can be managed’) are both possible and actual words whereas *knowledgeable* (*‘can be knowledged’) is idiosyncratic in meaning. Thus *knowledgeable* is an actual word that is no longer possible, exemplifying what is known as lexicalization which will be discussed momentarily.

Many possible words do not actually exist. Moreover, some possible words are more probable than others. For a possible word to be coined it must fill “what is perceived as a lexical gap” in the language (Bauer 2001:41). Possible words are defined by linguistic factors. Probable words are “determined by extra-systemic factors” (Bauer 2001:42). Words may be less probable if they are blocked by other words with synonymous meanings (see §5.2.4) or because there is nothing for them to denote. Words may also be improbable for aesthetic reasons such as length or awkwardness (e.g. **sillily*). Bauer (2001:43) also notes that some words do not exist for no apparent reason: *neglect* can be either a verb or a noun, but the synonymous verb *ignore* has no equivalent noun form.

4.4.4 Institutionalized and Lexicalized Words

Possible words, including nonce words and neologisms, are often potentially

ambiguous. When coined, though, they are coined with a specific, contextually driven meaning. As a word becomes more established the other potential senses become less likely and the word becomes institutionalized. At this stage “potential ambiguity is ignored” (Bauer 1983:48). Such a word is still transparent and predictable. For example, *shoe box* could additionally refer to “a box worn as a shoe” or “a box shaped like a shoe.” Instead, the word has been institutionalized as meaning “a box for keeping shoes in.”

A word is considered lexicalized when it could no longer be formed by productive means. It is no longer a possible word. This can occur in different ways. Semantic lexicalization is when the meaning of a word has become opaque or idiosyncratic, as in *knowledgeable* or *blackmail*. Morphological lexicalization occurs when a formerly productive word-formation process ceases to be productive. Words formed with *-th* (e.g. *warmth*, *length*, *depth*) are lexicalized because that suffix can no longer be used to create words. All established words are either institutionalized or lexicalized.

4.5 Summary

Words can be subdivided in a number of ways. Words themselves can be broken down into their parts or morphemes. They can also be classified based on morpho-syntactic features, how and why they are used, where they are used and by whom, and whether they are new or established within the language.

5 Where (New) Words Come From

So far we have looked at what we mean by *word*, the properties words typically have, and some of the ways in which they can be classified. Now we look at the various ways in which new words come into the language to begin with. Words are essentially coined in one of two ways. They are either coined productively, according to the rules of morphology, or they are coined creatively. The situation, however, is more complicated than that and there is “no valid way of drawing a clear distinction between what is creative and what is productive” (Bauer 2001:71). For our purposes, creativity “changes the rules” whereas productivity “exploits the rules” (Bauer 2001:71). And when it comes right down to it, people will coin words any which way they choose.

5.1 Why New Words

First, we briefly address why people feel the need to coin new words at all. Plag (2003) gives three reasons: labeling, syntactic recategorization, and to express an attitude. The function of labeling is both straightforward and well-attested. The world is ever changing. For every new concept, idea, or thing there needs to be a way to talk about it. With the invention of the television came not only that particular label, but also the equally necessary verb TELEVISION. Conversely, as Kastovsky (1986:595) points out, if there is no plausible referent, as in ?*radishade* (vs. *lemonade*), a word—though linguistically possible—will not be coined. Even unlikely coinings cannot be completely dismissed since labeling is not restricted to the real world. It also functions in imagined worlds. Thus the realms of magic and science fiction may well need to label concepts such as *unmurder*, *deflame* (as in a dragon), or *particalizing* (able to obliterate something into nothing but particles).

The second function for new words is syntactic recategorization. A function which “nominalizes, verbalizes, adjectivalizes, or adverbializes sentences, thus transforming them in into parts of sentences” (Kastovsky 1986:595). In other words, information is condensed as one complex word takes on the meaning of a phrase. For instance we might speak of *hammering* instead of *hitting with a hammer*, or *rescuee* instead of *the person who was rescued*. Condensation of information is only one motivation for recategorization, though. Recategorization can also be used simply to add stylistic variation or to maintain textual cohesion. For example:

- (12) a. The army *destroyed* the city. It was terrible to behold.
b. The army’s *destruction* of the city was terrible to behold.

Additionally, Plag (2003:60) says that new words are coined “to express an attitude” such as fondness or familiarity. These are typically informal in tone such as *poppers* for *pop* (i.e. father) or *spidey* for a pet spider.

All of these come back to one main criterion: usefulness. “Word formation is conceptually driven” (Baayen & Renouf 1996:90). If a word is useful it is likely to be coined. If it is not useful, it will not be coined or will not become established.

5.2 Morphological Productivity

Morphological productivity is the coining of new words according to the rules of

the language. That words are described as being “coined” is telling. Words come into being and are then circulated through the language at varying degrees of permanence. Phrases and sentences on the other hand are perhaps more like checks: drafted on one end, cashed on the other, then promptly forgotten. Regarding the processes by which words are coined, not all linguists agree which should be regarded as productive and which as creative. For our purposes productivity consists of affixation, conversion, and compounding.

5.2.1 Derivational Affixation

As mentioned in §2.2, there are two types of affixation in English: inflectional and derivational. Inflections are those suffixes that form the various grammatical words associated with a particular lexeme. Typical examples are *-ed* which forms the past tense of regular verbs or *-s/-es* which form the plurals of most nouns. Adding an inflection to a base changes the word-form but not the lexeme. In this sense, they do not actually create new words. Derivational affixes, however, do create new lexemes and derivation is one of the most prolific means of coining new words.

English has many derivational affixes at its disposal. These affixes fall into different categories. There are affixes which deal with number (e.g. *multi-*, *poly-*, *uni-*) or negation (e.g. *-less*, *non-*, *de-*). A good number of affixes facilitate syntactic recategorization: *-ness* nominalizes, *-ify* verbalizes, *-able/ible* creates adjectives, and *-ly* often forms adverbs. Affixes are not all equally productive though. For one thing, certain affixes may fall in and out of fashion. A few current popular affixes like *mega-*, *e-*, or *-aholic*³ can, in part, reflect cultural changes. For whatever reasons, some affixes have vanished completely in terms of productivity. We have already mentioned (§4.4.4) that the *-th* in *warmth* can no longer be used to coin new words.

Another limiting factor of affixes is that they cannot combine willy-nilly with any base. English’s affix repertoire consists not only of native prefixes and suffixes, but also Latinate ones (including French). While native affixes freely combine with Latinate roots (e.g. *regally*, *curiousness*), Latinate affixes are less likely to combine with native roots (e.g. *disbelieve*, but **smallity* vs. *smallness*). Other constraints include the fact

³*Workaholic* is modeled off *alcoholic* through the reanalysis of the morphemes.

that some affixes only combine with certain word classes, words with certain phonological properties, or even in certain orders. For instance *un-* attaches mainly to adjectives (e.g. *unforgettable*) while *-able* attaches mostly to verbs (e.g. *readable*). Additionally, *-en* can only combine with monosyllabic, obstruent-final bases as in *fatten* and *deaden* but never **candiden* or **equivalenten* (Plag 2003:62). And while *unhappiness* and *unhelpful* are both acceptable, the bases for those two are, respectively, *unhappy* and *helpful*⁴.

All of these constraints limit the productivity of affixes and the number of possible words they can coin. Some affixes have more constraints than others, putting affixation on a continuum. On one end are those affixes like *-th* which are no longer productive. On the other end are affixes such as *-ness* or *-ly* that are highly productive. These constraints and the productivity of certain affixes over others are “known intuitively by native speakers” (Burchfield 1985:107).

We should also note that affixes are not always straightforward. While some morphemes are easily discernible as affixes (e.g. *-ly* or *-ness*), others seem to blur the division with bound bases. For instance, if the *bio-* in *biochemistry* is a prefix, and if the *-logy* in *neurology* is a suffix, then *biology* would be a prefix and suffix with no base. We could say that *bio-* is sometimes an affix and sometimes a base. Alternatively, we can call *bio-* and *-logy* bound bases rather than affixes, making *biochemistry* and *neurology* compounds (§5.2.3).

5.2.2 Conversion

Conversion, sometimes known as “zero-derivation,” is the productive process “whereby a lexeme belonging to one class can simply be ‘converted’ to another, without any overt change in shape” (Carstairs-McCarthy 2002:48). In other words, conversion deals primarily with syntactic recategorization where the output lexeme looks identical to the input lexeme. Conversion typically occurs with nouns, verbs, and adjectives. It can also go in either direction (i.e. noun > verb or verb > noun), and it may be difficult to decide which direction it is going. Typical examples of conversion are:

(13) a. a bottle to bottle (noun-verb)

⁴**Unhelp* seems odd because *un-* combines more freely with adjectives than verbs.

- b. blind to blind (adjective-verb)
- c. poor the poor (adjective-noun)

Conversion, as defined above, is without a change in shape. There are, however, instances of noun-verb pairs where there is a change in stress:

- (14) a. a cónvert to convért (noun, verb)
- b. a pérmit to permít (noun, verb)

Conversion may also be used as proper nouns gain generalized meanings as in *to xerox* from *Xerox* machines or *to google* from *Google* search engine.

5.2.3 Compounding

Plag (2003) describes compounding as the most productive and most controversial type of word-formation. A simple definition of compound words is that they are two bases combined to create a new word. Compound nouns are the most prevalent, but there are also compound verbs and adjectives. Nouns, verbs, adjectives, along with prepositions can combine in a number of different ways:

| | | | |
|-------------|-------------------|------------------|-----------------------------|
| (15) | Noun | Verb | Adjective |
| Noun | <i>bookcase</i> | <i>handwash</i> | <i>knee-deep</i> |
| Verb | <i>turncoat</i> | <i>stir-fry</i> | <i>failsafe⁵</i> |
| Adjective | <i>greenhouse</i> | <i>dry-clean</i> | <i>red-hot</i> |
| Preposition | <i>underdog</i> | <i>outlive</i> | <i>overconfident</i> |

Most compounds have a head which defines the compound as a whole. A compound “inherits most of its semantic and syntactic information from its head” (Plag 2003:135). If the head is a noun, the compound will be a type of that noun (e.g. a *greenhouse* is a type of *house*). Often the head will be the right-hand element. Not all compounds have heads. Some are headless (or exocentric) and their “status . . . is not determined by either of [their] two components” (Carstairs-McCarthy 2002:64). For example, the noun *sit-in* is composed of a verb and a preposition and while *pickpocket* does contain a noun, it is not a type of pocket. Compounds may also be double-headed (or dvandvas) where neither component is more important: *writer-director* or *singer-songwriter*. Because of headedness, the reversal of a compound results in either

⁵This is the lone verb-adjective example.

nonsense (e.g. *greenhouse* vs. **housegreen*), a phrase (e.g. *red-hot poker* vs. *hot, red poker*), or a different word (e.g. *bookcase/casebook*).

For our definition, we used the ambiguous term *base* since not all compounding elements are words. For instance, we noted above (§5.2.1) that *bio-* and *neuro-* are not affixes but bound bases which form compounds. There are many such neoclassical elements that cannot stand alone but nonetheless contribute meaning to the compound (i.e. the meanings of *bio-* and *-logy* contribute to the meaning of *biology*). Based on this we might say compounds are formed from roots and stems, but there are also compounds formed from inflected words as in *road works* or *potter's wheel*. One element may even be a phrase, as in *over-the-fence gossip* (Plag 2003:134).

Further complications arise because while compounds are most often built of two elements, they are also naturally recursive and their structure can be repeated. Thus *office management training seminar video* is a single compound. When diagrammed though, such compounds can still be analyzed as being binary (Plag 2003:134). An *office management training seminar video* is, in fact, a type of video which can be broken down into [[[[*office management*] *training*] *seminar*] *video*].

Perhaps the chief complication with compounds is distinguishing them from phrases. The primary difference between compounds and phrases is stress. Compounds “tend to be stressed on the first element” (Plag 2003:137) whereas phrases often have final stress. Thus we have a *green hóuse* and a *gréenhouse*. Unfortunately, this distinction does not hold for all compounds and there are some which have final stress (e.g. *apple píe*). Another distinguishing feature is that while compounds *can* be idiosyncratic in their meaning (e.g. *blackguard*) phrases will not be unless they are idioms. Plag (2003:132) admits that solutions to the compound-phrase dilemma are hard to come by and “numerous issues remain unresolved.”

5.2.4 Restrictions on Productivity

There are a number of constraints that limit productivity. We already noted that affixes are constrained regarding the bases they can combine with. Phonological and morphological constraints tend to be of this sort. There are also semantic and pragmatic constraints. There is no use for a word that will not make semantic sense or a word for something that is unnameable. Even aesthetics can constrain productivity. A word like

adjectivalisation is possible, but in terms of deciphering a meaning, its length makes it more effort than it is worth.

Another type of constraint is blocking, or the “nonoccurrence of one form due to the simple existence of another” (Aronoff 1976:43 in Bauer 2001:136). Plag (2003) outlines two types of blocking: token, where a potential word is blocked by an existing one; and type, where one word-formation process blocks a rival one. Type blocking, Plag argues, ought to be abandoned since it is problematic and cannot account for doublets such as *curiousness/curiosity*. Such failures typically result in either one form ousting the other or the two words diverging in meaning, making (16) possible:

(16) The curiousness of the situation piqued my curiosity.

Token blocking, does work, but is dependent upon a number of factors. One such factor is the frequency of the blocking word. Outside language acquisition or Orwell’s Newspeak, **ungood* for *bad* or **goed* for *went* simply do not occur due to the high frequency of the blocking (albeit irregularly formed) words. Other examples might be *gloriously* being blocked by *glory* or *liver*⁶ ‘a person who lives’ being blocked by *liver* ‘an organ.’ When token blocking does fail it is usually due to either ignorance of the correct word or a temporary memory lapse (e.g. *bringed* for *brought*). There are examples where token blocking has completely failed, though: *inflammable* was unable to block *flammable* and if something is *raveling* it is also *unraveling*.

5.3 Creativity: Non-Morphological Innovation

Above we defined creative coinings as redefining the rules, rather than following them. Creativity, to a degree, can both make and break the rules. By definition, morphological productivity can only deal with complex words (taking conversion to be a form of zero-derivation which can turn *hammer* into *hammered* or *hammering*). The question then, is where do all the underived, uncompounded root words of a language come from? The most basic of these are the very foundation of English and are as old as the language itself. These include “almost all the most frequently used words in the language” (Crystal 2003:124) such as *love, see, in, have, be, hand, name, house, dog, white, and dark*. These words may be the core of English, but they are far from the bulk.

⁶More accurately, this *liver* may be a nonce word kept from becoming established.

Over the centuries, English has found a number of creative ways of adding to its lexicon.

5.3.1 Borrowing

A tremendous number, estimated at well over half (Adams 2001:11), of English words have actually come from other languages. Indeed one might call this “willingness to take in words from abroad” (Bryson 1990:66) a hallmark of the language. Over the centuries English has borrowed words from over 350 languages around the world (Crystal 2003:126). By far the most prominent sources, however, are Latin, French, and Greek. The history of borrowing is inextricably linked with the history of the language and the language’s speakers. With the Norman Conquest in 1066 came an influx of French words, most in particular spheres such as law, religion, or culture. Latin—the language of science, religion, and learning in general—has provided a fairly constant stream of loans throughout the centuries. Moreover, wherever English has traveled or colonized it has picked up words like souvenirs. From as close by as Gaelic and Norwegian to as far afield as Chinese, Tagalog, and Inuit, English has scoured the globe for its vocabulary.

English’s appetite for new words is so strong that it has been said that “we don’t just borrow words; on occasion, English has pursued other languages down alleyways to beat them unconscious and rifle their pockets for new vocabulary” (Nicoll 1990). In reality though, the term “borrowing” or even “loan-word” is misleading since the words are rarely ever returned to the donor language. Nor does the donor language have any reason for complaint since it retains the words in its lexicon as well. If anything is “beaten unconscious” it is the borrowed words themselves. Historically, as foreign words entered the language they were “made to conform to the vernacular patterns of [English] spelling and pronunciation” (Burchfield 1985:25). Many words were anglicized to the point that their foreignness is completely hidden such as *puny* from French *puisne* or *raccoon* from Algonquin *raugroughcan* (Bryson 1990:68). While anglicization does still occur, later borrowings underwent far less modification. Thus while *button* and *baron* show typical English fore-stress, later borrowings like *balloon* and *platoon* retain final stress. Additionally, *baggage* and *language* have the anglicized /dʒ/ whereas *camouflage* and *sabotage* retain the foreign /ɑ:ʒ/ (Burchfield 1985:18). Borrowings from the last century or so tend to fully retain their foreign look and sound

such as *tortilla* (with a /j/), *perestroika*, *sauerkraut*, or *fjord*.

Foreign borrowing continues to be a vast lexical source because of the global nature of English. Not only do English speakers travel all over the world, the language itself has settled all over the world as a primary or secondary language. All of these varieties of English add words to the lexicon which may or may not become mainstream.

5.3.2 Reanalysis

Another very common type of creative coining is reanalysis, which actually includes a number of methods for creating new words. The common denominator is that rather than adding on affixes, reanalysis involves breaking words apart, and not always into actual morphemes.

Backformation occurs when a “shorter word is derived from a longer one by deleting an imagined affix” (Crystal 2003:130). Some words we might not expect have come from this means, such as *edit* from *editor* or *reminisce* from *reminiscence*. In these cases it is easy to see how the *-or* and *-ence* could be analyzed as the same morphemes as in *conqueror* and *abhorrence*. Accidental backformations may even displace the original word. For instance *cherry* came from the already singular *cherise*.

Blending, like compounding, takes two words and makes them into one. The difference is that with blends (or portmanteaux) one or both words appear only partially. A prime example is *smog* which combines *smoke* and *fog*. Some words lend themselves frequently to blends. For instance, the tail end of *marathon* has practically become an affix and has been used to create *telethon*, *cyclethon*, *talkathon*, and many other formations denoting lengthy events. Blends can also be used for emphasis or to be eye-catching (e.g. *ginormous* or *fantabulous*).

A third type of reanalysis is abbreviations, which can be further subdivided into clippings, acronyms, and initialisms. Clipping is the deletion of some part of a word. Unlike backformation though, both the new and old form are synonymous and both typically remain in the lexicon. Clippings are usually short, either one or two syllables, and are often taken from the first part of a word, as in *ad*(vertisement) or *intro*(duction). There are, however, examples where the front is clipped (e.g. (heli)*copter*) or even where the front and back are clipped (e.g. (in)*flu*(enza)), though these are more rare. Acronyms involve deleting all but the first one or two sounds from a longer compound, combining

them, and pronouncing them as a word (e.g. *NASA* or *scuba*). Initialisms are like acronyms except that the separate letters are pronounced individually (e.g. FBI or BBC).

Though we have classed reanalysis as subsuming a number of creative processes, Plag (2003) shows that there is certainly an element of rule-governedness to them. Clippings are almost always mono- or disyllabic. Blends typically involve the first part of the first word and the last part of the second word (e.g. *brunch*). Reanalysis does involve a good deal of semantic, syntactic, and phonological regularity with rules that are separate from, though perhaps parallel to, those of productivity.

5.3.3 Onomatopoeia and Phonaesthemes

It is also possible to coin words based as much on sound as on meaning. Or rather such words take meaning from their representative sounds. We have already mentioned (§3.1 & 3.2) how onomatopoeia is used to verbalize or orthographically represent sounds. These range from the fairly standard (e.g. *bow-wow*) to the nonce formations found in comic books (e.g. *fwoomph*). Onomatopoeia are essentially transparent by nature and thus easily coined whenever needed.

Less straightforward are phonaesthemes. We have already stated (§4.1.1) that morphemes are the smallest unit of meaning. In some cases, however, a group of words which share the same sound(s) seem to have similar meanings. The connection between such phonaesthemes is often vague and only operates with certain words, but it is nonetheless intuitively there. For example *flash*, *dash*, *crash*, *bash*, *slash*, and *smash* all seem to denote abrupt movements. Likewise there is a sense of “smoothness or wetness” in the set *slip*, *slop*, *slurp*, *slide*, *slither*, *sleek*, *slick*, *slaver*, and *slug* though this sense is not found in *slow* or *slumber* (Carstairs-McCarthy 2002:7).

Phonaesthetics also covers the aesthetic judgements native speakers make regarding certain sounds and words. Meanings aside, some words are intuitively pleasant (e.g. *mellifluous* or *lullaby*) while others intuitively harsh (e.g. *spiky* or *vitriolic*). While sound symbolism may well be the result of linguistic coincidence, such patterns do affect the way new words are coined. Although they may not be aware of it, those who consciously coin words will tend to play off the intuitive nature of phonaesthemes, whether it be for a new product on the market, a nonsense word in a children’s book, or a species of alien for a television show.

5.3.4 Eponymy and Antonomasia

Proper names can also be a source of new words based on association—real or imagined—with an item or idea. This is known as eponymy for people and toponymy for places (Crystal 2003:155). People may lend their names to things as in *teddy bear* from Theodore Roosevelt or to ideas as in *volt* named for Alessandro Volta. Eponymous people need not even be real: *herculean* comes from the mythic hero Hercules and *mentor* from a character in Homer’s *Odyssey*. Examples of toponymy include *champagne* from Champagne, France and *gypsy* from Egypt. Name brands may also become generalized so that a *xerox* machine may refer to any brand of photocopier.

Related to eponymy is the rhetorical device antonomasia which is the “use of a proper name to express a general idea⁷” (OED 2007). Examples would be calling a traitor a *Benedict Arnold* or a highly intelligent person an *Einstein*. Essentially, a particular characteristic is singled out and the name becomes a synonym for that trait.

5.3.5 Metaphoric Extension

One final method of creatively coining “new” words is to simply use old words in new ways. We have already seen that words can carry multiple meanings (§2.5.3) as well as change their meanings (§3.5.1). The figurative or metaphoric extension of one word into a new domain is common for English. Bauer (2001:63) uses the example of a *bypass*, which once had to do with roads, but can now be used regarding blood vessels and a type of operation. Such extension may or may not also involve conversion. A *heart bypass* and a *road bypass* are both nouns. The principle behind metaphoric extension relies on creativity and cannot be produced by morphological rules.

5.4 Rules, Analogy, and Usefulness

When coining new words, or deriving or inflecting unfamiliar words, a speaker is presented with two means: rules and analogy. Often these two methods will coincide, but occasionally they result in conflicting possibilities. Such is the case when an analogy can be made from an irregular word such as *oxen* or *sing* (§6). Bauer (2001) gives a number of arguments and counter-arguments for whether innovation is driven by rules or

⁷Antonomasia also works in reverse so that “the Bard” can refer to Shakespeare.

analogy. For instance analogy would seemingly allow too much whereas rules cannot account for variation or coincidences like phonaesthemes. What it comes down to is a compromise, where both methods are viable and rules align mostly with productivity and analogy with creativity. Or it may simply come down to speakers formulating words whichever route is mentally fastest for them.

Ultimately, word-formation leads back to Bauer's (2001:142) "unformalisable" but "overriding" constraint: "words will not be formed unless they will be useful." If a word will be useful, it will be formed even if it must defy some of the rules of the language by creating new rules. When creating new words (consciously or not), speakers will likely question "does this word make sense in this context?" and "does this word feel right?" The rules of affixation may be broken in what Baayen and Renouf (1996:83) call "affix generalization." In their corpus study they found such unlicensed examples as *whyly*, *oftenly*, *itness*, *thereness*, *terrority*, and even the phrasal *next-to-nothingness*. It would seem that speakers follow rules intuitively, but break them whenever pragmatic needs override. Or as Burchfield (1985:113) puts it: the "formative rules [of English] are no more than general guidelines, observed only when it is convenient to do so, and broken—because of the needs of euphony, analogy, or some other competing principle—at will."

6 Word Intuition Survey

Before concluding, we look briefly at an informal survey I conducted which looks at some of the phenomena considered above. This survey was not meant to be extremely thorough or diverse, merely a way of getting a sense for how people view words. While not all responses were what I was expecting/hoping for, the results were interesting. After quickly giving the demographics, we will examine each of the twelve questions and see how they apply to the various properties of words. The survey and key can be found in the back as Appendices A and B.

6.1 Demographics

When giving the survey, I emphasized that it was about the person's *intuitions* and not about "right" or "wrong" answers. Twenty-seven people took the survey, nine in person, eighteen via email. Because I gave this survey mainly to friends and associates,

the majority of survey-takers were females, mid-twenties, and American. There were a handful of men and older women as well as five UK natives and two ESL speakers. All were either college students or graduates. Typical overall reactions to the survey were that it was “hard” yet thought-provoking and interesting.

6.2 Survey Analysis

The first question was simply “what is a word?” and asked for either a definition or characteristics. Some mentioned that words are groups of letters that are able to be written and pronounced. Others put that they are symbols used to identify or represent things or concepts in the world. The most common answer was that words are units of meaning used in communication. The second part of the question asked if they had ever said something and wondered if it was a word or not. All but one person admitted to second-guessing the wordhood of something they had said. Most said this phenomenon happened all the time, but they couldn’t recall specific examples. A few examples that were given were *Old Testamentish*, *dongle* (a computer thumb drive), *o’clockish*, and *recognizability*. It would seem that people do notice when they are creating nonce words, but since communication is not hindered they just move along in the conversation and the “word” in question is quickly forgotten.

Question two dealt with what we mean by word. Survey takers were given a sentence, asked how many words were in it, and then how they arrived at that number. The sentence is repeated in (17):

(17) Jack’s garage door won’t open, so he is going to have to get it fixed before he can go anywhere.

Two thirds said that there were 20 words, a number arrived at by simply counting. That is, they counted the orthographic words. Only two people caught/mentioned the duplicate *to* and *he*, and one listed *going/go* as being the same word. The remainder of the people counted one or both clitics (either just *won’t* or *won’t* and *Jack’s*) as being two words. No one counted the compound *garage door* as one word, though two people counted *anywhere* as two words. Interestingly, two people acknowledged the difficulty in counting and gave two numbers, one with or without clitics and the other with or without duplicates. If I were to redo the survey, I would incorporate a nonce word to see if people counted it. Overall though, it seems that people come up with wordcounts the

same way word processors do, by counting orthographically.

The next question dealt with people's intuitions about what is or is not a word. A list of eleven words was given and people were asked which were *not* words. From the list, only *decipherment* and *antidisestablishmentarianism* are listed in the OED, though *littler* is listed in Merriam-Webster. The other words included the "grammatically incorrect" *broughten* and other made up words which derived from known bases (e.g. *deaccessorize* and *uninflatable*). Three people said that none of the items were words whereas three other people said all were. Most people went through and selected certain forms. *Uninflatable* was most often listed as a word, more often than either of the established words. This seems to indicate its status in what we termed the On Call Lexicon (§3.5.2). The forms most often marked as not being words were *smallify* and *broughten*. The latter we already noted is grammatically incorrect. *Smallify* violates a restriction on productivity by combining a Latinate affix with a Germanic base.

For the fourth question, fourteen slang or colloquial terms (e.g. *wannabe* and *splendiferous*) were given and people were asked which ones they considered to be words. All of the words are listed in either the OED or Merriam-Webster. Everyone said that at least some of the terms were words; seven people said all of them were. The most recent term on the list, *bling*, was nearly always listed as being a word. *Spiffy* and *ain't*⁸ were also often chosen. The least picked were *schlep* and *doh*, but even these were considered words by a third of the survey-takers. A number of people noted that while they considered most of them to be words, they would never use them in written English, which shows something of the status of slang and other informal words (§4.3.3).

Question five was simply a long list of words and people were asked to circle those which they felt had started out as foreign borrowings. This question was intentionally tricky and incorporated a number of words that were highly familiar and thoroughly anglicized. There were also some that should have looked somewhat foreign (e.g. *bazaar*, *poncho*). In fact, only two (*house* and *king*) of the twenty-eight are native words. Only one person correctly identified all the words, the one person with a linguistics background. The words that were most often seen as native were *they* (Old Norse) and *fact* (Latin) which both received more votes than either *house* or *king*. In

⁸This despite the familiar playground rhyme that "*Ain't ain't a word and you ain't supposed to say it.*"

fact *shampoo* (Hindi) was considered just as “English” as *king*. Rather predictably, *bazaar* (Persian) and *poncho* (Spanish) were the least chosen. A helpful follow up question would be to ask people if they knew a foreign language, and if so which one(s). The point being, to most people the foreignness of words borrowed into English is often lost as they gain familiarity.

The next question is related. This time all of the words were blatantly foreign (yet recognizable) borrowings and the question was which the survey-takers felt were English. Essentially, the two questions combine to ask what it means to be a word in English. All of the words are listed in both the OED and Merriam-Webster. A couple of people said none of the words were English; a few said they all were. The words most often seen as English were *voodoo*, *loch*, and *kosher*. *Batik* and *perestroika* were least often considered English. Given the chance to do the survey over, I would ask people what the difference was between the words in five and those in six. For one thing, there seemed to be no pattern to which words were seen as English. Twice as many people felt *tortilla* was English than thought *hacienda* was, though both clearly come from Spanish.

Questions seven and eight simply asked about people’s familiarity with the plural or singular forms of some irregular nouns such as *seraph*, *cactus*, *octopus*, or *data*. The plural to singular question was unremarkable except that a number of people noted they used the plural forms as the singular. The question regarding irregular plurals was more interesting. Some forms nearly everyone knew (*cacti*, *syllabi*, *appendices*). Other forms were split between the “correct” form and a regular plural (e.g. *phenomena* vs. *phenomenons*). As was expected given its proximity to *cactus* and *syllabus*, *octopus* was most often pluralized as *octopi*⁹, showing that analogy is indeed a strong factor when recalling/coining words.

Continuing with analogy, the next two questions asked people to inflect two made up words: *glox*, a noun, and *kring*, a verb. A couple of people came up with *gloxen* (cf. *ox/oxen*). More though came up with either *gloxi* (perhaps influenced by *cacti/syllabi* above) or *glox* (cf. *deer/deer*). The majority of people, however, made *glox* plural through regular means creating *gloxes*. *Kring* seemed to be more difficult, possibly indicating that irregular verbs are more troublesome than irregular nouns.

⁹*Octopus* comes from Greek and thus grammatically shouldn’t take the Latin plural *-i*.

Irregular *krung* had twice as many votes as regular *kringed*. But there were also a few who went with *krang*, one instance of *krought* (cf. *bring/brought*), and one instance of the hypercorrective *kranged*.

Question eleven delivered the most unexpected results. Because of poor wording, only one third understood that they were to create a word for removing the imaginary substance *shlorp*. The rest described ways in which to remove it. While at first annoying, it soon became evident that people were expressing their intuitive sense of phonaesthetics (§5.3.3). Nearly everyone who answered the question in this way put that *shlorp* removal would require mopping or wiping. They intuitively lumped *shlorp* with words such as *slirp*, *slippery*, or *slime*. For those who did coin a new word examples were *deshlorp/deshlorping*, *deshlorpifying*, and *unshlorping*. Why the prefix *de-* was preferred six times whereas the typically more productive *un-* was used once is uncertain.

Lastly, question twelve gave people the made-up word *grick*. They were told it was a “sticky, goo-like” substance and were asked to define the derived form *grickalization*. That nearly all answers fell into one of two categories neatly shows the potential ambiguity of a word until it is institutionalized or lexicalized (§4.4.4). *Grickalization* was defined as either being the process whereby something takes on *grick*-like properties or when the *grick* hardens (cf. crystallization).

Though more informal than in-depth, this survey still offered some interesting glimpses into how some people view words. The eldest of the survey takers (68) was least willing to count unfamiliar forms or slang terms as words. Other than this one generational insight, there seemed to be no patterns based on gender or nationality.

7 Conclusions

As familiar and prevalent as words are, not only in our language but in our culture, it is easy to overlook their complexity. People learn new words and coin new words all the time. Once we have learned to speak and read, we often take for granted the idiosyncrasies of English and the eccentricities of its lexicon. To begin with we asked “What’s in a word?” In response we have looked at the various properties and characteristics entailed by words. Words are units of both phonology, morphology, and orthography, made up of sounds, morphemes, and letters. They can be concrete units and abstract concepts. Words have various properties such as grammatical functions and

syntactic classes. Within words are not only meanings that change and accumulate, but also meaningful relations with other words and the people who use them. Within words—sometimes visible, often not—are their histories as they have come into the language through various productive and creative means. Finally, we looked at the intuitions of a handful of English speakers.

All this leads to a concluding question: What does it mean to *know* a word? The most obvious answer is that to know a word means being able to either recall or recognize it and know its meaning. But even this is not straightforward. Does recollection/recognition necessarily entail knowing the proper spelling or pronunciation? Ideally the answer would be yes. But as perplexing as English spelling is, and with the convenience of relying on computer spellcheckers, knowing a word and knowing how to spell a word are two different things. Pronunciation is also problematic because of the different varieties and dialects of English. There is a certain level of pronunciation that is involved with knowing a word. Vowel quality may differ, but knowing a word seems to imply knowing the proper stress pattern and which sounds the letters represent (e.g. that *facade* rhymes with *broad* and not *brocade*).

Additionally, what does it mean to know the meaning of a word? We do not know all words equally, but rather on a continuum. There are words which are easily recalled, words that are instantly recognized, and words we must pause and search our mental lexicon for. Moreover, when asked what a word means, even when it is a word we can recall easily, it is still difficult to give a dictionary definition. It is far easier to give synonyms or to use the word in context. According to Quirk (1968:140), “knowing the meaning of a word is knowing how to *use* it.” This means being able to understand it or put it into context, not necessarily being able to define it. Thus meaning is not actually enough. “To know a word is to know a great deal more than its meaning” (Hoey 2003). For instance we may not always be able to classify a word as an adverb or a participle, but we still must know how it works within a sentence. Knowing how to *use* a word also implies having some understanding of its particular connotations: is it formal, informal, dialectal, jargon, slang, etc. Knowing a word also entails having at least a general understanding of the word’s collocations. Part of knowing the word *consequences* is intuitively knowing that it is more likely to pair with *serious* or *grave* than *boisterous*.

Knowing a word, however, does not necessarily involve knowing its etymology. Indeed most people have no idea whether a word is Germanic, Latin, or French. Nor does knowing a word require knowing all of a word's inflections. This may seem counter-intuitive since knowing a word's inflections is part of knowing how to use a word. But take for example the well-known words *cactus* and *octopus* (§6.2). We can understand exactly what is meant by both of these words without knowing that their "correct" plural forms are *cacti* and *octopuses*. People muddle through with *cactuses* and *octopi* all the time.

Juliet, in her star-stricken naivete, would have a word or name be no more than a label. And to an extent she is not far wrong. Words, like names, are strings of letters and sounds, syllables and morphemes. They are arbitrary labels filling syntactic roles and formed through rule-driven processes. But words are also much more. They are slippery in meaning and rich with connotations and associations. They can build both bridges and barriers. They are tools of creativity which may also be coined through creativity. Though bound by rules, words seem also bound to break rules, leaving a trail of homonymy, polysemy, and general eccentricity in their wake. The words of the English language are as diverse and idiosyncratic as the speakers who use them. Perhaps the simplest answer to "what's in a word?" is merely "a lot."

References

- Adams, Valerie (2001) *Complex Words in English*. Harlow: Longman.
- Baayen, R. Harald and Antoinette Renouf (1996) "Chronicling the Times: Productive Lexical Innovations in an English Newspaper," *Language*, Vol. 72, No. 1, pp. 69-96.
- Bauer, Laurie (1983) *English Word-formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie (2001) *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bauer, Laurie (2003) *Introducing Linguistic Morphology*, 2nd edn., Edinburgh: Edinburgh University Press.
- Bryson, Bill (1990) *Mother Tongue*. London: Penguin Books.
- Burchfield, Robert (1985) *The English Language*. Oxford: Oxford University Press.
- Carstairs-McCarthy, Andrew (2002) *An Introduction to English Morphology*. Edinburgh: Edinburgh University Press.
- Crystal, David (2003) *The Cambridge Encyclopedia of the English Language*, 2nd edn., Cambridge: Cambridge University Press.
- Crystal, David (2006) *Words Words Words*. Oxford: Oxford University Press.
- Hoey, Michael (2003) "What's in a word?" *MED Magazine*, Issue 10, August 2003. Modern English Publishing.
<http://www.macmillandictionary.com/MED-Magazine/August2003/10-Feature-Whats-in-a-word.htm>
- Katamba, Francis (1993) *Morphology*. Basingstoke: Macmillan.
- Kastovsky, Dieter (1986) "The problem of productivity in word formation," *Linguistics*, Vol. 24, pp. 585-600.
- Matthews, P. H. (1991) *Morphology*, 2nd edn., Cambridge: Cambridge University Press.
- Nicoll, James D. (1990) "The King's English," posted 15 May 1990, Usenet newsgroup rec.arts.sf-lovers.
http://groups.google.com/group/rec.arts.sf-lovers/msg/c961c46670ca97d6?q=g:t13676756607d&hl=en&lr=lang_en
- Oxford English Dictionary Online (2007). Oxford University Press.
<http://dictionary.oed.com/>

Plag, Ingo (2003) *Word-Formation in English*. Cambridge: Cambridge University Press.

Quirk, Randolph (1968) *The Use of English*, 2nd edn., London: Longman Group Ltd.

Ter-Minasova, Svetlana G. (2005) "Traditions and innovations: English language teaching in Russia," *World Englishes*, Vol. 24, No. 4, pp. 445-454.

Wolfram, Walt and Natalie Schilling-Estes (1998) *American English: Dialects and Variation*. Malden, Mass.: Blackwell Publishers Ltd.

APPENDIX A

WORD INTUITION SURVEY

1. a. What is a “word”? Briefly define or give two or more characteristics:
 b. Have you said something and wondered “is that a word?” Any examples?

2. How many “words” are in the following sentence? How did you arrive at that number?:
 “Jack’s garage door won’t open, so he is going to have to get it fixed before he can go anywhere.”

3. Which of the following, if any, are **NOT** words:

| | | | | | | |
|-------------------|----------|---------------|-----------|------------------------------|----------|--|
| decipherment | littler | deaccessorize | wordhood | antidisestablishmentarianism | | |
| photogenicishness | smallify | uninflatable | broughten | unasphyxiate | chunkily | |

4. Which of the following, if any, **ARE** words:

| | | | | | | |
|-------------|--------|-------|-------|---------------|---------|-------|
| schlep | spiffy | homie | ain’t | ginormous | wannabe | gonna |
| thingamajig | zilch | snafu | wonky | splendiferous | doh | bling |

5. Which of the following, if any, **started** as **FOREIGN BORROWINGS** into English:

| | | | | | | |
|----------|---------|---------|--------------|---------|----------|-----------|
| entrance | kayak | courage | taboo | amok | king | boomerang |
| coffee | caravan | waltz | resurrection | opera | hammock | knapsack |
| they | house | erudite | ascend | fact | potato | charity |
| sauna | bazaar | poncho | shaman | shampoo | assassin | climax |

6. Which of the following, if any, do **YOU** consider to be English words:

| | | | | | | |
|----------|---------|-------|-------------|-----------|----------|---------|
| wigwam | loch | hula | perestroika | apartheid | kosher | blarney |
| tortilla | kung fu | batik | fjord | voodoo | hacienda | geisha |

7. What are the **plural** forms of the following **singular** nouns:

| | | | | | |
|----------|--|---------|--|------------|--|
| cactus | | octopus | | appendix | |
| syllabus | | seraph | | phenomenon | |

8. What are the **singular** forms of the following **plural** nouns:

| | | | | | |
|------|--|----------|--|------|--|
| dice | | criteria | | data | |
|------|--|----------|--|------|--|

9. One *glox*, two _____

10. If today I *kring*, tomorrow I will have _____
11. Imagine that a substance known as *shlorp* is all over the floor. What might be a highly specific word used to describe cleaning up shlorp?
12. Assume *grick* is a sticky, goo-like substance. What might *grickalization* mean?

APPENDIX B

WORD INTUITION SURVEY (KEY)

1. a. What is a “word”? Briefly define or give two or more characteristics:
e.g. Things in dictionaries, Units of meaning, Building blocks of sentences
b. Have you said something and wondered “is that a word?” Any examples?
2. How many “words” are in the following sentence? How did you arrive at that number?:
“Jack’s garage door won’t open, so he is going to have to get it fixed before he can go anywhere.”
e.g. 23 (anywhere), 22 (Jack’s), 21 (won’t), 20 (orthographic), 18 (duplicate to/he), 17 (garage door), 16 (going/go)

3. Which of the following, if any, are **NOT** words: *(criteria: normal font are in the OED, underlined is in Merriam-Webster)

| | | | | |
|-------------------|----------------|---------------|-----------|------------------------------|
| decipherment | <u>littler</u> | deaccessorize | wordhood | antidisestablishmentarianism |
| photogenicishness | smallify | uninflatable | broughten | unasphyxiate chunkily |

4. Which of the following, if any, **ARE** words: *(criteria: in the OED or Merriam-Webster)

| | | | | | | |
|--------------------|--------|-------|-------|---------------|---------|-------|
| schlep | spiffy | homie | ain’t | ginormous | wannabe | gonna |
| <u>thingamajig</u> | zilch | snafu | wonky | splendiferous | doh | bling |

5. Which of the following, if any, **started** as **FOREIGN BORROWINGS** into English:

| | | | | | | |
|----------|---------|---------|--------------|---------|----------|-----------|
| entrance | kayak | courage | taboo | amok | king | boomerang |
| coffee | caravan | waltz | resurrection | opera | hammock | knapsack |
| they | house | erudite | ascend | fact | potato | charity |
| sauna | bazaar | poncho | shaman | shampoo | assassin | climax |

6. Which of the following, if any, do **YOU** consider to be English words: *(in the OED and Merriam-Webster)

| | | | | | | |
|----------|---------|-------|-------------|-----------|----------|---------|
| wigwam | loch | hula | perestroika | apartheid | kosher | blarney |
| tortilla | kung fu | batik | fjord | voodoo | hacienda | geisha |

7. What are the **plural** forms of the following **singular** nouns:

| | | | | | |
|----------|----------------|---------|------------------|------------|-------------------|
| cactus | <i>cacti</i> | octopus | <i>octopuses</i> | appendix | <i>appendices</i> |
| syllabus | <i>syllabi</i> | seraph | <i>seraphim</i> | phenomenon | <i>phenomena</i> |

8. What are the **singular** forms of the following **plural** nouns:

| | | | | | |
|------|------------|----------|------------------|------|--------------|
| dice | <i>die</i> | criteria | <i>criterion</i> | data | <i>datum</i> |
|------|------------|----------|------------------|------|--------------|

9. One *glox*, two _____
e.g. *gloxen* or *gloxes*

10. If today I *kring*, tomorrow I will have _____
e.g. *krang* or *kringed*

11. Imagine that a substance known as *shlorp* is all over the floor. What might be a highly specific word used to describe cleaning up *shlorp*?
e.g. *unshlorp*, *deshlorp*, *deshlorpify*, or *unshlorpen*

12. Assume *grick* is a sticky, goo-like substance. What might *grickalization* mean?
e.g. To make something have the properties of *grick*, when the *grick* hardens.