Linguistic, cognitive, and social constraints on lexical entrenchment

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Paula Chesley

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bruce Downing, Adviser

August 2011

Acknowledgements

My committee members have been excellent mentors. I have very much enjoyed working with and learning from them.

The help of my family throughout graduate school has been invaluable. They are worthy of, and have received, much more eloquent prose than I presently have time in which to express my gratitude.

In addition to my committee members and family, a number of people greatly facilitated this research.

Conceptual advice: Jean-Pierre Koenig, T. Florian Jaeger.

Experimental expertise: Koji Miwa, Patrick Bolger.

Data collection: Muhammad Abdurrahman, Lindsay Griener, Tom Lindsay, Deniz Rudin.

Statistical analysis: Steven Nydick, Sanford Weisberg.

Draft reading: Grace Deason, Jeanette Gundel, Betsy Kerr, Peter Hendrix, Kaitlin Johnson, Jean-Philippe Marcotte, Morgan Sonderegger, and most tirelessly, Eden Kaiser.

Generally awesome and helpful people to have around: Tineke Baayen, Patrick Bolger, Jürgen Bohnemeyer, Jeanette Gundel, Eden Kaiser, Barbara London, Jean-Philippe Marcotte, Chad J. Marsolek, Steven Nydick.

This research was partially supported by a National Science Foundation Graduate Research Fellowship, for which I am truly grateful.

For reasons of space, I am forcing myself to limit these thanks to only those who helped on the specific research in this dissertation. Countless other graduate students, faculty, collaborators, staff, and participants (in Buffalo, France, Minnesota, Edmonton, and elsewhere) have contributed to my research and positive experiences throughout graduate school.

Abstract

How do new words become established in a speech community? This dissertation documents linguistic, cognitive, and social factors that are hypothesized to affect *lexical entrenchment*, the extent to which a new word becomes part of the lexicon of a speech community. First, in a longitudinal corpus study, I find that linguistic properties such as the range of a word's meaning and the donor language of a borrowing affect lexical entrenchment. Contextual factors such as frequency, dispersion, and a borrowing's cultural context also play a role in lexical entrenchment. Second, a psycholinguistic study examines the extent to which speakers remember previously unseen words. Through eye-tracking, lexical decision, and free recall tasks, I determine that, again, linguistic and contextual information plays a role in the memorability of a new word. Speakers are more likely to remember words used in particular contexts, and they are also more likely to remember certain word types than others. In a third study, I find that musical preferences, knowledge of popular culture, and social ties influence comprehension of African-American English vocabulary. Together, these studies suggest that lexical entrenchment is predictable to an extent previously undocumented. Results indicate that information relating to dynamical, non-linear systems could be profitable in further studies on lexical entrenchment.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview: what is lexical entrenchment?

LEXICAL ENTRENCHMENT refers to the process by which a new word becomes part of the lexicon, or vocabulary, of a language. In other words, given that a new word occurs, how likely is it that the word will become part of the lexical stock of the language, as opposed to a fleeting lexical item? One central question for those studying lexical entrenchment is how lexical innovations are diffused throughout a speech community. Another question is the role individual speakers' memories play in lexical entrenchment.

Lexical entrenchment differs from MORPHOLOGICAL PRODUCTIVITY, sometimes referred to simply as PRODUCTIVITY, because the study of the latter seeks to generate, or predict, possible words, either with rule-based approaches (see, for example, Aronoff 1976, Selkirk 1982, Halle & Marantz 1993, and Ussishkin 2005) or by assigning a probability of productivity to a lexical item or lexical process (Baayen 1992). For example, productivity deals with whether or not a form like *hateable* is possible, or how likely we are to see this form. Entrenchment deals with how likely it is that this word will recur, and how often it will recur, given that we have already seen the word *hateable*. In this sense, lexical entrenchment takes up where productivity leaves off in the study of new words. Researchers can ask if different word formations are more or less likely to undergo lexical entrenchment, but they are not limited to

studying a particular type of word formation.

To date, lexical entrenchment has been understudied. This is a result of insufficient data as opposed to the inherent interestingness of the phenomenon. The majority of research in American linguistics in the latter part of the 20th century has been in the generative tradition, which seeks to generate all the possible words of a language, and to reject the impossible words. Continuing with our example, a generative approach has as a goal the acceptance of a new word like *hateable*, but the rejection of new words like **strawberryable* (because *-able* attaches only to verbs) or **dieable* (because verbs taking *-able* must take a direct object). The study of productivity is intimately related to this tradition.

As opposed to productivity, lexical entrenchment is an inherently probabilistic phenomenon (perhaps productivity is too but lexical entrenchment cannot be examined any other way). Once we have seen a new word, there is no way of saying, "This word will definitely not become entrenched" or conversely, "This word will absolutely become entrenched". Furthermore, lexical entrenchment is caused by many factors. In order to study lexical entrenchment, then, a large amount of diachronic data is ideal. We can imagine an unfortunate linguist writing down all of the new words he or she hears, and then trying to keep track of whether he or she hears them again for the next five to 10 years. This method would be time-consuming and would give way to many gaps. The data necessary to study entrenchment, longitudinal corpus data sampled from the same speech community, were not available until recently.

Having answered what lexical entrenchment is, we can move on to other important questions, namely, *why* we should study it, and *how* we should go about doing so.

1.1.1 Why we should study lexical entrenchment

Why do linguists study language? The answer will vary according to who we ask the question of, but for many linguists, the answer lies in the expressive power of language. How can we communicate an infinite number of ideas with a finite set of

words and syntactic rules? This must surely be one of the defining characteristics of human cognition.

When we talk about a finite set of words, we are saying that the number of (stored) words in a person's mental lexicon is finite, yet the number of possible words, either in a language's lexicon or possibly that a speaker can create on the fly, is infinite. It suffices only to examine one possible method of lexical enrichment, lexical borrowings, to show that this is so for a language. In English, borrowings have been attested from many languages. Each of these languages continually has new words added to it, and all those words are also possible candidates for borrowings into English. We cannot put a limit on the number of new borrowings in English. Since the number of new words is infinite, we would expect an unknown but large number of new words to be uttered each day in a speech community.

Although the possible number of new words is infinite, the percentage of these words that actually become entrenched is quite small (Metcalf 2004). In other words, speakers have implicit constraints on the type of words that tend to stick in their lexicons. This is reflected in preferences for the type of lexical items we see in languages. Therefore, studying lexical entrenchment tells us about preferences for the type of lexical items that languages have, in addition to the organization of the lexicon and how quickly lexicons evolve.

The previous paragraph touches on a question that is central to this dissertation: what is the relationship between words that are new for individual speakers and words that are new in a language or speech community? A specific concern is the relationship between memory and lexical entrenchment: are words that are more memorable more likely to become entrenched in a speech community? Possibly: a related phenomenon concerns stock prices shortly after companies' initial public offerings. New York Stock Exchange stocks with more pronounceable acronyms as ticker symbols were initially more likely to have higher gains in share prices than stocks with acronyms rated as less pronounceable (Alter & Oppenheimer 2006). The effect was tested after a day, a week, six months, and a year. While a significant difference was detected only after

one day, the trend persisted throughout the first year.

In Chapter 3, I show that acronyms rated as pronounceable as words as opposed to a series of letters are recognized easier (albeit not recalled easier)[1]. Similarly, new words that speakers remember easier could be more likely to become entrenched – at least initially. That is, memorable new words could be more likely to become trendy memes. A new meme word is one type of new word that is within the purview of this dissertation. A meme word rapidly spreads throughout a speech community; this spread is often greatly facilitated by media such as the internet. Accompanying this rapid diffusion process is an equally fast death: memes often fade quicker than other new words. An example of a recent meme word is Charlie Sheen's new coining *bi-winning* when he said in a March 1, 2011 interview, "I'm bi-winning... I win here and I win there".

In this dissertation, I do not explicitly examine recognition or recall of new words as a predictor of lexical entrenchment. This would require either longitudinal data or a longitudinal laboratory component I did not have time for given the other studies in this dissertation. However, I return to this idea in the General Discussion. In the present work, I assume that a link between individual memory and lexical entrenchment in a speech community is possible and proceed to examine possible factors making new words more memorable to individual speakers (Chapter 3).

Applications of lexical entrenchment studies

We should also ask how we can apply our knowledge of lexical entrenchment outside of linguistics. Three fields that could make use of information about lexical entrenchment and the applications of the methods I propose in this dissertation to new words are natural language processing, social networking and sociology in general, and marketing.

[1]Some call the former "true acronyms" and the latter "initialisms". I choose not to make this distinction and call all such words *acronyms*. As will be seen later, this distinction is somewhat artificial as changes across time and across the words themselves indicate different pronunciations.

One hypothesis of mine is that new words are marked lexical items in a language. If this is the case, natural language processing applications would do well to take advantage of their distribution. The speaker can use new words to call attention to what she is saying. If the discourse is coherent, we should assume that new words are used at salient points in the discourse and that they refer to the speaker's central intent. That is, in general new words should not occur in digressions, and the topicality of new words should be the same as the overall discourse topic. To test this hypothesis, I examined the new borrowings in Chapter 2 of my dissertation. Of the 281 borrowings, 249 (88.6%) are related to the same topic as the main topic of the article. This has implications for natural language processing: the detection of new words can aid in topic detection, and perhaps in automatically determining salient points in the discourse as well.

Social networking reseachers should also take advantage of the fact that new words are marked linguistic elements: as such, they are more likely to indicate membership in a particular social group than unmarked linguistic elements. Thus when describing a social network (in terms of influential speakers, etc.) an appropriate variable to examine would be the new words used by speakers.

Another domain that could benefit from the study of lexical entrenchment is marketing (see section 1.2.3). While the primary goal of this dissertation is not to benefit commercial enterprises, the study of lexical entrenchment can benefit those outside of linguistics proper. Advertizers want to know what makes a good product name, which can include factors such as memorability.

1.1.2 How we should study lexical entrenchment

As stated above, lexical entrenchment is an inherently probabilistic phenomenon with multiple causes. The complexity of lexical entrenchment warrants the use of probabilistic models to study the phenomenon. Therefore, in this dissertation I seek out possible causes of lexical entrenchment. In this dissertation, I examine linguistic,

cognitive, and social factors as possible causes of lexical entrenchment. Next I quantify these linguistic variables into random variables, and use these random variables as predictor variables in regression analyses to determine possible causes of lexical entrenchment. Using this methodology, we can discover complex relationships that we could not discover by introspection. For example, Metcalf (2004) suggests that both Frequency and Diversity of users affect lexical entrenchment. In Chapter 2, I find this to be true, but the two variables are so highly correlated that such a statement is arguably unhelpful in determining the causes of lexical entrenchment. Once dispersion, our predictor variable that corresponds well to Metcalf's Diversity, is accounted for, I find that frequency is actually negatively correlated with entrenchment in a linear regression model. Subsequent work (Altmann *et al.* 2011) shows a similarly surprising trend.

1.1.3 What this dissertation does not examine

For reasons of time, this dissertation does not take into account a few aspects of lexical entrenchment that nevertheless merit study. Phonological aspects of lexical entrenchment other than length, discussed in Chapter 2, are not considered. In fact, the phonological aspects of lexical entrenchment are worthy of a dissertation unto themselves. To date, I have seen little work on this area of study (although see Pierrehumbert in press, in addition to my work presented here).

New words formed by compounding are not examined in this dissertation. This is unfortunate, as at least in Germanic languages including English it is any extremely productive method of forming new words. This study is limited to neologisms that are string types, lexical items with no white space between them. Compared with compounds, string type neologisms are easy to detect in corpora automatically. Of course, some compounds are string types, but many are not, and any conclusions on compounds that are only string types would be speculative[2].

[2]For the purposes of this dissertation, the distinction between string-type compounds and non-string-type compounds is not relevant. The first two studies look specifically at borrowings, blends,

1.1.4 Goals of this dissertation

One goal of this dissertation is to expand our understanding of lexical variation. Lexical variation has sociolinguistic dimensions such as regional distinctions as well as age- and profession-related differences. Additionally, we can examine the lexicon in looking at properties of the words themselves. We can examine, for example, new or moribund words. I have opted to study lexical entrenchment – how new words become established in language – because in so doing we can examine not only linguistic and cognitive but also sociolinguistic aspects of the lexicon. By studying lexical entrenchment, we gain insight into how words are established in the collective vocabulary. This will give us insight into the linguistic, cognitive, and social properties of the lexicon and of language in general.

A second goal is to further our knowledge of the diffusion of linguistic innovations. Most current systematic work on language change examines phonetic (sound) change. An example of this is Labov's (1972) work on the Northern Cities Chain Shift (NCCS) for vowels. Although Minnesota is on the border of the region in which the NCCS occurs, it is much more pronounced in cities like Buffalo and Rochester, New York. An example of a vowel change in the NCCS is the raising of the vowel in the word *trap* so that it sounds more like *tree-ap*.

Less work examines the diffusion of lexical, phonological, or syntactic change. Does the diffusion of linguistic innovations in all domains follow patterns similar to those established for sound change? New words could be more salient than slight shifts in vowel pronunciations: if so, this difference in salience could distinguish between the diffusion patterns of phonetic change and lexical entrenchment.

This dissertation is composed of three main components: a corpus study, a psycholinguistic experiment, and an online survey. Each experiment has a diachronic element to it. The first study looks at new words at two different time periods, when they first occur and again approximately 10 years later. The psycholinguistic study

acronyms, and derived words, and the latter study looks at new coinages irrespective of word type

includes a learning phase, in which participants learn new words, and a recall phase, which takes place either on the same day or at least one day later. The third study looks at comprehension vocabulary of African-American English (AAE), examining words that have not yet crossed over into Mainstream American English (MAE) but that are likely candidates for active use in MAE.

1.2 The lexicon, with a focus on new words

Although the lexicon has become an increasingly important component of models of language in both theoretical and experimental linguistics, we still know very little about what drives lexical entrenchment. From its fledgling role in linguistic theory as a "repository" in which all unpredictable information about words is stored (Chomsky 1965:87), the lexicon has evolved to play a central role in theoretical linguistics in formalisms such as Lexical-Functional Grammar (Kaplan & Bresnan 1982) and Head-Driven Phrase Structure Grammar (Pollard & Sag 1987; Pollard & Sag 1994).

While in current syntactic models of grammar, the lexicon is now seen to contain structure, generative rules, and templates, these theories still fail to account for the full range of lexical productivity observed, such as blending, clipping, or forming acronyms. Furthermore, most current theories of the lexicon do not account for lexical entrenchment, or how new words enter the lexicon. Ito & Mester (1995) touch on the idea of lexical entrenchment by positing that loanwords in the lexicon are at varying stages of integration in the native lexicon, but how they come to be at these stages is not discussed. One is left to assume that the stages must be determined on a case-by-case basis, outside the grammar.

The following sections deal with various research on new words. Morphological productivity is the theoretical linguistic area of research coming closest to lexical entrenchment. Methods of lexical enrichment that are not taken into account from a productivity perspective are next discussed. Research on new words in the mental lexicon is also summarized. Outside of academic linguistic research, works for the

general public discuss lexical entrenchment, and marketing researchers are interested in the phenomenon from the perspective of brand naming. Finally, I explore known and hypothesized properties of new words.

1.2.1 Productivity

Productivity-based research examines the degree to which newly coined lexical forms or processes combine to form new lexical items. The goal of productivity research is to examine characteristics of possible words. Once a word is uttered, whether or not it becomes entrenched in the lexicon is of little interest to those studying productivity, except to use it to study other possible new words with the same lexical process (e.g. the suffix *-ity*). Researchers on productivity (e.g. Bauer 1983a) have tended to relegate non-affixational, non-compounding methods of lexical enrichment to a single cluster of unpredictable processes. This is unfortunate: non-affixational lexical enrichment is vast. In addition to the "unpredictable" (i.e. non-morpheme-based) methods of word formation Bauer (1983a) lists, such as clipping, blends, acronyms, and mixed formations (a combination of the preceding strategies), contemporary English regularly uses analogy, metaphor, metonymy, proper names (*Ponzi scheme*), and coining of a totally new word, which can sometimes evoke a certain concept, either via phonaesthemes or with overlapping phonemic or graphemic components. The product name *kleenex* evokes *clean*. Borrowings and calques are not often employed in contemporary American English, but languages such as French and German regularly make use of these processes. Of these "unpredictable" processes, some research has been done on at least blends (Gries 2004, Gries 2006, Lehrer 1996). Gries finds that blends for which component words share phonological material are much more likely to occur than those that do not. For example, a blend like *beardo* (*beard* × *weirdo*, a weirdo with a beard, the March 18 Urban Dictionary word of the day `www.urbandictionary.com/define.php?term=beardo`) has overlapping phonological material in the *eard* part of *beard* (/ɛɹd/). Therefore, it is much more likely to

occur than the possible blend *musteirdo* (a weirdo with a mustache), although the two are similar conceptually. With lexical entrenchment, one studies the trajectory of words once they have occurred, including things like the social contexts in which they are used.

1.2.2 The mental lexicon

As mentioned above, we would like to know whether the ease with which a new word is remembered by individual speakers plays a role in lexical entrenchment. All other things being equal, are new words that are remembered better more likely to become entrenched throughout a speech community? Gaskell and Dumay (Gaskell & Dumay 2003, Dumay *et al.* 2004, and Dumay & Gaskell 2007) explore lexical representations of new words having phonological overlaps with existing words.

In Dumay *et al.* (2004), lexical competition with extant words is considered an indicator of lexical consolidation. Lexical items compete with other lexical items with similar-sounding inputs; and consolidation is indicated with longer pause detection times (see article for entire paradigm). Whether or not new words undergo these competition effects is a good measure of whether they have been consolidated into the learner's lexicon. Participants are given either semantic information along with the new word or only the new word: semantic information, in the form of a definition, does not aid in memory consolidation of the new word. In this study, lexical consolidation is found to occur between one and 24 hours after learning the new words.

In Dumay & Gaskell (2007), one night's sleep is found to be essential for lexical consolidation. The authors hypothesize that this delay could be essential so as not to prevent the representations of newly learned words from overwriting extant lexical representations.

1.2.3 Lexical entrenchment: broader research

Relative to other phenomena studied by linguists, lexical entrenchment has received a fair amount of attention in general-interest literature. Because of the relatively low amount of scholarly work on lexical entrenchment, much is to be gained from examining both books written for the general public as well as marketing research. First, Metcalf's (2004) *Predicting New Words* is the work most approximating the goals of the present work. I discovered the existence of this book relatively late (i.e. June 2009) in my research on lexical entrenchment, but in many ways it is similar to the first component in my dissertation, the work on lexical borrowings. This work advances the *FUDGE* hypothesis for lexical entrenchment (although it is not termed lexical entrenchment). This hypothesis is described in further detail in Chapter 2. As this work is geared toward the general public, it does not follow the methodological standards of psycholinguistic or computational work. For example, it ignores the inherently high correlation between the *Frequent* and *Diversity* factors: if a new word is used by a wide range of speakers, it must be fairly frequent.

In another general-audience work on new words and naming, Steven Pinker criticizes the *FUDGE* hypothesis as at least partially circular (Pinker 2007:308), because frequency, diversity of users, and generation of new forms of meaning are what someone who works on lexical entrenchment tries to predict. Yet I show in chapter 2 that these factors can take on a causal quality if they are used diachronically: the more frequent and well-dispersed a new word is, the more speakers will hear and eventually use it. For his part, Pinker does not advance a rival hypothesis for lexical entrenchment, saying only that lexical entrenchment is a social phenomenon, largely unpredictable (pp. 319-322).

Managers believe brand names influence sales more than packaging and that they are more important in fostering long-term sucess (Lerman & Garbarino 2002). This is one possible reason why, to date, marketing researchers such as Lerman & Garbarino (2002) and Lowrey *et al.* (2003) are so interested in lexical entrenchment.

These scholars have investigated how linguistic properties of brand names can influence brand-name memorability. Lerman finds that brand names that are already extant lexical items in the lexicon are *recalled* easier, but that brand names that are non-words are *recognized* easier. Here recognition is operationalized as simple discrimination between possible choices, whereas recall is considered to be constituted of both retrieval and recognition. Unfortunately, the methodological standards in marketing research have been insufficient for the psycholinguistic community (Lerman, p.c.). For example, in Lerman's study, there are no repeated measures; i.e. there is one brand name per category.

Lowrey *et al.* (2003) examine data on consumer memorability of brand names in commercials. These authors perform a linear regression using a combination of recall and recognition measures as a dependent variable and several linguistic factors as independent variables. For highly recognizable brand names, linguistic factors have less of an effect than for moderately recognizable brand names. For the former, only unusual spellings and the presence of a blend brand name have an effect on memorability. Unusual spellings aid in memory retention, while blends are detrimental to memory retention. The authors posit that linguistic variables would have an even greater effect on lesser-known brand names, although these were not examined in their research.

1.2.4 Properties of new words

How frequently do new words occur in speech and in the media? When they are encountered, do they require more processing effort than extant but infrequent words? What portion of the population has heard a particular new word? Empirical work done on actual new words tends to be limited to one or two words (e.g., Hohenhaus 2006; Renouf 2007), and very little psycholinguistic work has been done on the comprehension of new words. Theoretical work on lexical entrenchment is lacking, so it is difficult to get an idea of the properties and distributions of new words.

In the first component of my dissertation (Chapter 2), I show that new borrowings account for approximately 0.081% of all tokens occurring in the *Le Monde* corpus (Abeillé *et al.* 2003). This study also reveals a bimodal distribution of French borrowings originally occurring in the period 1989 - 1992 at the later interval of 1996 - 2006. Figure 2.1 suggests a difference between infrequent, non-entrenched borrowings that are part of the left bulge of the density distribution, and the well-entrenched borrowings on the right part. This result echoes the findings of Baayen & Lieber

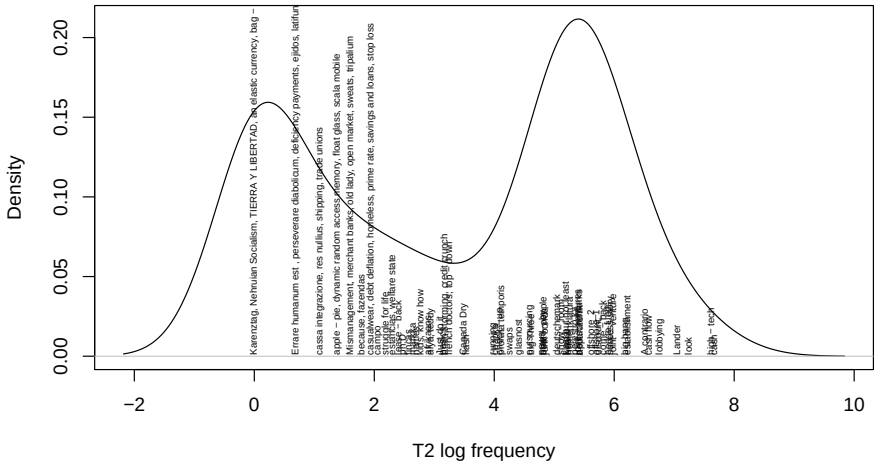

Figure 1.1: Estimated probability density function for logarithmic frequencies of borrowings at T2. Borrowed types are given at their T2 log frequencies.

(1997), who argue that the bimodal distribution of Dutch words with a particular prefix shows a difference between frequent, well-entrenched lexical items and infrequent nonce formations using the prefix. I would expect to see a similar distribution for other types of new word formations, as well as the agglomeration of all types of new word formations.

The findings in figure 2.1, coupled with those of Baayen & Lieber (1997), suggest that an emprical definition of lexical entrenchment is possible. In this figure, there is a relatively well-defined distinction between new words that are at the fringes of the lexicon, and new words that have become entrenched. One possibility for finding

the best place to make this distinction is to find the local minimum between the two bulges. In figure 2.1, there is even a small gap, just to the left of 4 on the x-axis, almost exactly where this local minimum occurs. In other words, lexical entrenchment does not occur randomly. If it did, we would see a normal distribution instead of the bimodal distribution in 2.1. This deviation from normality should lead us to suspect that various factors – linguistic, cognitive, and social in nature – are responsible for this distribution.

1.3 Dissertation components

The current state of research on lexical entrenchment leaves us with the following questions:

1. What are the linguistic and contextual properties affecting entrenchment?
2. What are the cognitive constraints leading toward the creation of a lexical representation of a new word in the mental lexicon?
3. What social and cultural factors could influence lexical entrenchment?

Chapter 2 addresses questions 1 and 3; Chapter 3 addresses question 2, and chapter 4 addresses question 3.

In what follows, I report on three studies, each looking at a different aspect of entrenchment. I hypothesize that linguistic and contextually determined properties affect entrenchment. To test to what extent this is the case, a study on lexical borrowings in French newspaper data is carried out. One could also imagine that cognitive factors such as memory retention of new words could also affect lexical entrenchment. In a psycholinguistic study, I examine factors influencing retention of new words in individual speakers' mental lexicons. Finally, I investigate the impact of social ties, knowledge of popular culture, and musical preferences on comprehension vocabulary of African-American English. The general discussion (Chapter 5) provides

synthesis of these experiments and gives ideas for future work on entrenchment and on the diffusion of linguistic innovations in general.

Chapter 2

Predicting new words from newer words: Lexical borrowings in French

This study addresses entrenchment into the lexicon of lexical borrowings. I search for all new lexical borrowings in a corpus of French newspaper texts and examine the frequency with which these borrowings reoccur in a second corpus of newspaper texts from about 10 years later. Lexical entrenchment emerges as depending on a variety of factors, including length in syllables, the original language of the borrowing, and also semantic and contextual factors. The dispersion of a word in the early corpus is found to be a better predictor of its frequency in the later corpus than its frequency, but both measures contribute to predicting the degree of entrenchment of a lexical item. The interaction between these two variables implies that borrowings are penalized for their burstiness.

2.1 Introduction

There are several ways speakers of a language can form new words. Affixation of pre-existing stems and affixes, as we see with the English suffix *-able*, is well studied. Speakers can form *hate* → *hateable* according to the same rules from which we have *love* → *loveable*; generative approaches such as those taken by Aronoff (1976), Selkirk

(1982), Halle & Marantz (1993), and Ussishkin (2005) treat affixation in great detail. Blends such as *gaybie* 'gay ∗ baby', which Urban Dictionary defines as 'the child of a gay couple'[1], offer a second possibility for lexical enrichment (see Gries 2004 for a discussion of blends in English). Further ways of creating new words, as pointed out by Bauer (1983a), include clipping, such as *convo* 'conversation', and acronym formation, such as *LDR* 'long-distance relationship'. Finally, the lexicon is enriched with borrowings like German *handy* 'mobile phone'. Because of their perceived lack of systematicity, borrowings have arguably received the least attention of all kinds of neologisms.

Borrowings can constitute a non-negligible portion of a language's lexicon. For instance, I estimate that new borrowings account for .082% of all tokens in the *Le Monde* treebank corpus of journalistic French (Abeillé *et al.* 2003) from 1989-1992. In other words, for every 1,000 words a reader of *Le Monde* encounters, it is very likely that she will run into a new borrowing.

Yet many of these borrowings could be nonce, or one-time, uses. To determine which borrowings are actually becoming entrenched into the French lexicon, we queried the online archives of *Le Figaro*, another full-coverage French newspaper, for the years 1996 - 2006. If we have the frequencies of the borrowings at two different time periods, can we predict whether or not a lexical borrowing will "survive" and become entrenched into a RECIPIENT LANGUAGE, the language into which the borrowings enter? Theoretical morphology is well developed for neologisms formed by affixation. Yet the internal structure of a borrowing in a recipient language is largely monomorphemic, which complicates matters for theories wishing to derive new words from extant lexical entries. Since borrowings do not conform well to these analyses, the tendency is to treat them on a case-by-case basis, if at all.

Although in principle an infinite number of borrowings are possible, the borrowings that do survive are a highly constrained subset of the possible borrowings. My working hypothesis is that borrowings do not occur indiscriminately, but are con-

[1] `http://www.urbandictionary.com/define.php?term=gaybie`

strained by a range of different factors. Firstly, the initial DONOR LANGUAGE of the borrowings could co-determine entrenchment. For example, borrowings from a prestigious language like English could be more likely to undergo lexical entrenchment than borrowings from a less prestigious language like Polish. Furthermore, the FREQUENCY of a borrowing at a given time could be an influential predictor about the borrowing's entrenchment in the language at a later date. A borrowing's DISPERSION – the number of text chunks a word occurs in if a text is divided into several sub-parts – might also be a relevant predictor. Since shorter borrowings require fewer processing resources, we hypothesize that the LENGTH of the borrowing will be inversely related to its degree of entrenchment. I also examine a semantic factor, the SENSE PATTERN (monosemy versus polysemy) of a borrowing in the recipient language, as well as the CULTURAL CONTEXT in which the borrowing is used: whether or not the borrowing in a particular context refers to the culture of the language of the borrowing (e.g. a Russian borrowing when describing Russia, as opposed to describing China).

I am not the first to argue that a new word's frequency and dispersion are important in predicting their entrenchment. Metcalf (2004) proposes the *FUDGE* hypothesis: new words that are FREQUENT, UNOBTRUSIVE, have a great DIVERSITY of users, that GENERATE new usages and meanings, and that refer to ENDURING concepts are more likely to become entrenched than those that are not. These are common-sense factors; in order to develop this hypothesis into a full-fledged theory, we need to test and quantify each factor against empirical data. For some of the *FUDGE* factors, this can be easily done. Metcalf's FREQUENCY and DIVERSITY correspond well to the current FREQUENCY and DISPERSION measures, respectively. It is difficult to know whether borrowings are truly unobtrusive in the recipient language's lexicon: in fact, because they stem from a different language, with a different phonology and a different morphology, they may well stand out. Borrowings constitute a robust method of lexical enrichment for many languages, including French. A concept's ENDURANCE is contingent on so many cultural and societal issues that it is impossible to include in the present study.

This study extracts initial new borrowings from the *Le Monde* corpus (Abeillé *et al.* 2003) from 1989-1992 (Section 2.3.1) and then queries the online archives of *Le Figaro* for the same borrowings during 1996-2006 (Section 2.3.2), taking frequency in this second corpus as an indicator of entrenchment into the French lexicon. Based on the predictor variables of FREQUENCY, DISPERSION, LENGTH, the donor LANGUAGE of the borrowing, and the borrowing's SENSE pattern and cultural CONTEXT, I model the frequency information of the *Le Figaro* corpus (Section 7.3.3) and find this model to explain a high proportion of the variance in the *Le Figaro* frequencies. In Section 2.4.2, I discuss model results, while Section 2.5 offers a general discussion of the findings. First, however, I go into more detail about relevant definitions and the scope of the study.

2.2 Lexical borrowings: definitions and classifications

This study aims to determine what factors promote entrenchment of new lexical borrowings from various donor languages into French, the recipient language in question. This use of the term BORROWING is consistent with that of Thomason & Kaufman (1988), who characterize a word as a borrowing only when fluent speakers of the recipient language adopt the lexical item from the donor language.

For the purposes of the present study, a LEXICAL BORROWING is defined as a lexical item (lemma) from a donor language satisfying the following criteria:

1. the (approximate) form and meaning are copied from donor to recipient language, without adaptation to French morphological and graphical conventions;

2. the borrowing's form-meaning correspondence is not yet found in a particular French dictionary.

The first criterion allows for the proper amount of imprecision with which to take into

account speakers' probable perceptions of lexical borrowings. For example, Thogmartin (1988) notes that his respondents offer *week-end* and *parking* as examples of English borrowings in French. Yet the form of *week-end* is not exactly the same in the donor language, since in English there is no hyphen. And the meaning of *parking* is not exactly the same in French and English: French *parking* is equivalent to English *parking lot*. A more significant divergence from meaning in the donor language concerns the German borrowing *handy* 'mobile phone': Germanophones perceive this word as an English borrowing (Erling 2004:132), but the meaning of *handy* in English is quite different from this. Since the goal of the current work is to capture speakers' knowledge about their language's lexicon, it seems appropriate to disregard these differences in form-meaning transfers and to still accept these examples as borrowings under the above definition.

Under the first criterion, the written form of the lemma must have been adopted more or less 'as is'. For example, *dérégulation* does not constitute an English borrowing according to this definition: the written form changed from English *deregulation* to *dérégulation*. Although the French form is approximately the same as the English, the French form follows French morphological and graphical conventions. That is, the affixes *dé-* 'de-' and *-tion* '-tion' are productive French affixes, and the base resembles the French *règle* 'rule'. Since French speakers most likely do not know that this word is a calque from English, I would want to exclude this word from any study on borrowings.

The second criterion justifies using a term other than LOANWORD, for which typically only the first criterion applies. It also follows the initial requirements for neologisms of Baayen & Renouf (1996). While a dictionary is not a perfect representation of speakers' lexical knowledge, often lagging behind the spoken language, verifying the presence of a word in a dictionary does represent a concrete way to measure lexical entrenchment. The dictionary used in the present study is the *Trésor de la langue française informatisé*, or *TLFi* (Dendien & Pierrel 2003), available online at `http://atilf.atilf.fr/tlf.htm`. This is one of the more exhaustive French

dictionaries. This dictionary is somewhat conservative in adding new words. Thus borrowings in the *TLFi* tend to be well entrenched.

Under this definition, transliterations into the Roman alphabet, where no apparent adaptation to French morphology occurs, also count as lexical borrowings. For example, Russian *glasnost* 'transparency' is obviously Romanized. Yet *glasnost* does not exist in a French dictionary, and it is a transliteration of a Russian term, as opposed to a translation (*transparence*) or a written Gallicization such as *glasnoste*. Lexical borrowings are not limited to words and may include idiomatic and multi-word expressions such as *last but not least* and *res nullius* 'unowned property' (Latin).

Proper nouns such as names of places or entities, titles of address, movie and book titles, and product names were excluded from the study. Also excluded from the study are any morphologically inflected or derived realizations of lexical items already existing in the *TLFi*. For example, I eliminated an attestation of *self-made men* from the *Le Monde* corpus, since it is the plural form of *self-made man*, which does exist in the *TLFi*. Finally, only exact matches were counted in the *Le Figaro* data, and near matches, such as *self government* when searching for simply *government*, were excluded.

2.3 Lexical borrowings, then and now

2.3.1 Then (*Le Monde*, 1989-1992)

The version of the *Le Monde*, or T1, corpus that I initially received from Anne Abeillé has 645,746 tokens[2]. Foreign words are labeled as such in the T1 corpus, but often, the part-of-speech tags in the corpus did not correspond to the definition of a lexical borrowing given above. The labelling of foreign words was non-systematic. For example, some proper nouns, such as *Aston-Martin*, were labeled as foreign words, and some words labeled as foreign words were in fact in the *TLFi*, such as *manu militari*

[2]I have the 2006 version of the corpus. The corpus is updated from time to time.

'by force' (Latin). These words are not unfamiliar to (at least some) Francophones, and it is probably best not to discuss them in terms of foreignness. To estimate the rate of lexical borrowings not labeled as foreign words in the corpus, I manually examined 431 random sentences (4% of the corpus). According to this definition, only three out of 14 lexical borrowings in this sample were labeled as foreign words in the corpus. A second pass through the corpus was necessary to find more borrowings.

This second pass to capture the borrowings not labeled as such was done in a series of three steps. The first and principal step of the procedure consisted in examining letter combinations of low frequency in 'native' French words and of relatively higher frequency in other languages. For example, I looked at all words with the letters <w> or <qi> in them. These letters or letter combinations are not common in French words, and they often tell of recent foreign origins. In addition, letters with certain diacritics, such as <ö>, do not exist in French, so any words with these written forms are certainly foreign. The derivational and inflectional morphological paradigms of other languages were also queried. For example, the superlative form of Spanish adjectives has the suffix *ísimo*; hence *ísimo* and *isimo* were queried. Second, the context around these borrowings was examined, on the assumption that borrowings might occur in clusters. Lastly, if step 2 led to any new borrowed words, I added the relevant written forms of these borrowed words to the list of infrequent written forms in French and began a new search for unusual letter cominations. In total, 94 letter combinations comprised of letters or letter combinations of low frequency in native French words were queried.

The combination of initial corpus labelings and this search method yielded 281 borrowed tokens and 138 borrowed string types in the T1 corpus. Borrowed types and their frequencies are given in appendix 7.1. Table 2.1 presents a breakdown of tokens and types by language. It illustrates the extent to which English dwarfs all other languages for borrowings. For statistical analysis, I therefore contrasted English (93 types) with the set of other languages (45 types).

Table 2.1: A breakdown of tokens and types according to languages.

Language	Types	Tokens
English	93	145
Spanish	12	35
German	7	60
Russian	7	18
Italian	6	10
Latin	5	5
Hebrew	4	4
Polish	1	1
Dutch	1	1
Finnish	1	1
Portuguese	1	1
Total	138	281

Two types, *deutschemark* 'Deutschmark'[3] and its plural *deutschemarks*, were extremely frequent, with 25 and 24 attestations, respectively. This means that one lemma comprises 17.5% of the borrowings found. Hence I perform analyses with and without this frequent lemma in the dataset to see if it is unduly biasing the results.

This method retrieved 281 borrowings from the T1 corpus. An estimate of the number of borrowings in the corpus would allow for evaluating the method's RECALL, i.e., the proportion of borrowings detected to the total number of borrowings in the corpus. By manually examining a subsection of the corpus for borrowings, we can estimate the recall of the method proposed as well as the number of borrowings in the corpus.

Above, I mentioned that I manually examined 431 sentences for borrowings. The method for detecting borrowings, in combination with the initial labelling of the corpus, found nine of the 14 borrowings in these sentences. Undetected tokens were *dirham*, *dirhams*, *nairas*, *majors*, and *eurobag*, where the first three of these are foreign currencies. We thus have a recall of $9/14 = 64.3\%$. We can construct a confidence interval for the method's recall, and from this we can estimate how many borrow-

[3]The term in the *TLFi* for the former German currency is *mark*.

ings are in the corpus. The following score confidence interval is taken from Agresti (2002:15-16):

(1) $(\hat{\pi} - \pi_0)/\sqrt{\pi_0(1-\pi_0)/n} = \pm z_{\alpha/2}$.

With $\hat{\pi} = .6429$ (9/14), $n = 14$, and with a critical value of $z_{\alpha/2} \approx 1.96$ for a 95% confidence interval, we can say that the recall of the method is between 38.76% and 83.66%. Dividing the number of borrowings obtained by both upper and lower bounds of the proportional recall gives bounds for the number of borrowings in the corpus. Hence the 95% confidence interval for the number of borrowed tokens in the corpus is between 336 and 724. The midpoint of these two figures is 530; dividing this number by the total number of words in the corpus, 645,746, gives an estimated percentage of words in the corpus that are borrowings: 0.082%.

In order to evaluate the viability of these lexical borrowings, we consulted a second corpus sampled 10 years later.

2.3.2 Ten years later (*Le Figaro*, 1996-2006)

I chose the online archives of the *Le Figaro* as the T2 corpus for several reasons. First, no other large corpus of a comparable genre – general journalistic French, with a national and international perspective – was available for the time period desired. Second, as opposed to *Le Monde*, the archives of *Le Figaro* systematically provide the queried word in context. The *Le Figaro* archives date from 1996 to the present, and, wanting to have the maximum spread possible in frequencies at T2, we queried the archives from November 1, 1996 through December 31, 2006. Ideally, we would examine the borrowings continuously from 1989 - 2006 as opposed to separating the T1 and T2 corpora by ten years. This would allow us to see much more fine-grained trends in lexical entrenchment. However, no such continuous data was available for French.

I manually examined up to 200 *Le Figaro* occurrences of each borrowing to control for its sense in context. For example, one of the borrowings was *bush*, in the sense of

the Australian outback, but the query results for this borrowing overwhelmingly refer to the former American presidents, so we do not want to count the latter occurrences at T2. If a borrowing had more than 200 occurrences, I estimated the total number of borrowings with the same sense by manually examining a subset of the occurrences[4], and then estimating the number of occurrences with the same sense from the inspected sample.

When I started querying the *Le Figaro* archives, the maximum number of responses returned was 1000. Ten borrowed types gave 1000 responses, which we must take as a lower bound on the total number of borrowings at T2. I also estimated the frequencies of these borrowings, albeit in a different way. This was done by manually examining up to 200 tokens of the most recent borrowings for sense, and then, assuming a uniform distribution across the 10-year period, I estimated the total number of occurrences for the entire 10 years from this sample. For example, if there were 1000 occurrences of a borrowing returned by the *Le Figaro* archives, I examined the most recent 200 of these. Let us say that, of these 200, 125 corresponded to the sense I was seeking, and let us also say that these 125 senses spanned a period of two years from 2004-2006. Since there are five two-year periods in the T2 corpus, we would then assume that this borrowing occurred with the same sense about as frequently in the other four two-year periods, which means that the estimated total number of times this borrowing with this sense occurs in the T2 corpus would be 125∗5, or 625 times. Unfortunately, over the course of querying the corpus, the maximum number of responses was limited to 300. This only affected results for three borrowings, namely, *a contrario* 'on the contrary' (Latin); *apparatchik* 'organization or party official' (Russian), and *board*. For these borrowings, the estimation technique was the same as for the borrowings for which the lower bound of 1000 responses was returned.

[4]The exact size of the subset was determined by dividing the total number of occurrences by a divisor between two and five, in order to get the closest number of occurrences to 200 as possible. The estimated number of total borrowings was then found by multiplying the number of borrowings found in the subset by the original divisor. For example, if there were 865 occurrences of a borrowing, 173 occurrences were examined, because 865/5 = 173. If 112 of these borrowings corresponded to the sense seen in the original corpus, then the estimated number of borrowings is 112∗5, or 560.

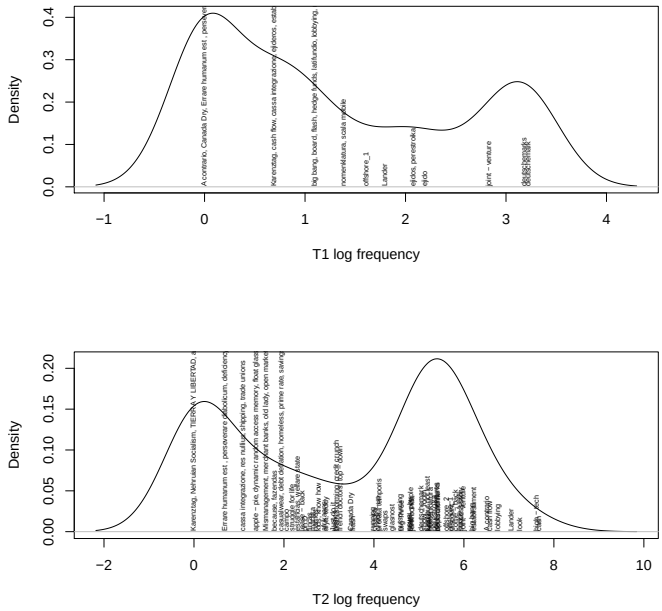

Figure 2.1: Estimated PDF for logarithmic frequencies of borrowings at T1 (above) and at T2 (below). Borrowed types are given at their respective log frequencies, and the T2 distribution is the response variable. Note that a density function will automatically smooth across a discrete distribution; hence frequencies smaller than 0 are given non-negative probability.

If the overwhelming majority of the occurrences were for a different sense, the borrowing was excluded from the T2 study. This was the case with *bush*. Unfortunately, the *Le Figaro* query interface does not differentiate between capital and lower-case letters, nor between accented and unaccented characters. Hence a query for *deregulation* also yields results for *dérégulation*. Seven borrowings were therefore discarded from the follow-up study.

The lower panel of Figure 2.1 shows the estimated probability density function (PDF) of the response variable, T2 frequency. The borrowings detected at T1 have a bimodal distribution at T2. This distribution suggests a difference between infrequent, non-entrenched borrowings that are part of the left bulge of the density distribution, and the well-entrenched borrowings on the right part. Examples of infrequent borrowings are *Karenztag* 'sick-day leave' (German) and *french mafia* [sic],

while examples of entrenched words are *high-tech* and *nomenklatura* 'high-level government officials [under Communism]' (Russian). This result echoes the findings of Baayen & Lieber (1997), who argue that the bimodal distribution of Dutch words with a particular prefix shows a difference between frequent, well-entrenched lexical items and infrequent nonce formations using the prefix.

In the present data, a difference of ten years between corpus dates is sufficient to show patterns of entrenchment. In other words, it is unlikely that a correlation between frequencies at T1 and T2 stems from sampling from the same distribution. The upper panel of Figure 2.1 gives the estimated PDF of the same borrowings at T1. Comparison of the upper and lower panels indicate that the T1 and T2 distributions are quantitatively different. As we move from T1 to T2, we see a pattern of entrenchment for many borrowings: much of the probability density at T1 is concentrated at the leftmost (low-frequency) bulge, while at T2 the rightmost, high-frequency bulge is more prevalent. For example, *cash flow* has moved from the leftmost bulge at T1 to the rightmost at T2. This visual inspection is confirmed via a two-sample Kolmogorov-Smirnov test ($D = 0.523$, $p < 0.001$): it is improbable that the T1 and T2 distributions of borrowings are the same.

2.4 Modelling the entrenchment of lexical borrowings

I studied the T2 frequencies with the help of a multiple regression model.

2.4.1 Predictor variables

Predictor variables fall into two categories: predictors representing properties that are intrinsic to a borrowing, and properties that are contextual in nature. The intrinsically defined variables examined in this study are LENGTH, SENSE PATTERN, and donor LANGUAGE of the borrowings. The locally defined variables are FREQUENCY at T1,

dispersion at T1, and cultural context.

Length

The borrowings in the T1 corpus vary with respect to length, with one of the shorter being *names*, while the longest is *Errare humanum est, perseverare diabolicum* 'To err is human, to persist [in so doing] is diabolical' (Latin). Some researchers (e.g. Pergnier 1989) hypothesize that short, "punchy" borrowings are more likely to become entrenched than longer borrowings. This hypothesis makes sense from a processing perspective: shorter words could be easier to process. New *et al.* (2006) found that for all but the very shortest words, processing times increase with length. If length has an effect on processing times, it could have an effect on lexical entrenchment. Longer borrowings might take longer to process, and over time, these words might die out due to their elevated processing costs, which are already high due to their novelty and foreignness.

There are several ways to quantify the length of a borrowing: one can count the number of letters, the number of syllables, or the number of morphemes. These different measures are highly inter-correlated: New *et al.* (2006) mention a correlation of $r = 0.81$ for the number of letters and the number of syllables for English words. The number of morphemes is a measure that is difficult to apply for borrowings, since borrowings do not fit into the morphology of the recipient language. I hence opted to approximate length by number of syllables. In New et al.'s multiple regression model for predicting reaction times of English words, the number of syllables has a higher standardized coefficient than the number of letters. The number of syllables is given in the variable length, which was log-transformed to reduce the effects of outliers.

Sense pattern

Consider the borrowing *news*, which is not attested in the *TLFi*. This word can indicate a *newsletter* (a sense attested at T1 and T2) or *new information*, a sense attested

only at T2. Another existing sense for *news* could help or hinder this borrowing to survive. From a utilitarian perspective, we could argue that the increased range of denotata of a polysemous word could make it be used more often. Since it can be used more often, it has greater benefit to speakers, and therefore is more likely to become entrenched.

From a processing perspective, the polysemy of a lexical item could facilitate quicker access, or it could confuse the reader, forcing him or her to disambiguate between senses. Results from the processing literature in this area are highly ambiguous. On one hand, for shallow processing such as lexical decision tasks, polysemy seems to be helping access (Piercey & Joordens 2000 and Rodd & Marslen-Wilson 2002; see Rodd *et al.* 2004 for a connectionist model), but on the other, for deeper processing, polysemy might be detrimental to lexical access (Klein & Murphy 2001; Piercey & Joordens 2000).

A lexical borrowing was determined to be polysemous if it has at least one other related sense elsewhere in the T1 or T2 data. It would have been ideal to examine the sense pattern at T1 only, but due to data sparsity at T1, this was not possible. Many of the borrowings only occur once in the T1 corpus, and other electronic resources from a similar genre from the same period are not available. I feel that the importance of investigating sense pattern as a factor influencing lexical entrenchment surpasses this methodological concern.

For some new polysemous borrowings, an existing sense was already present in the *TLFi*. For the purpose of this paper, a SENSE is defined as a first-level sub-entry of a lexical entry in the *TLFi* dictionary[5]. This dictionary-based definition is fairly intuitive: homonyms are considered different lemmata, but related senses are noted as different sub-entries of the same lemma. For example, a *pack* has one entry with

[5]A stipulation of this definition is that it includes items that are only in the dictionary as fixed expressions but that are used in a free context. The only example of this type of polysemy in the current findings concerns the word *cash*, where the *TLFi* lists this item only in the expression *payer cash*, yet it exists freely in my findings ('...30 % du prix devant être versé en *cash*', here with my italics. The data format of the corpus does not specify whether the borrowings were originally italicized.)

two immediate sub-entries: it is both a 'a number of individual units packaged as a unit' or 'a group of eight forwards in a rugby game', and not simply 'a group of teammates', which is the approximate definition it has in the T1 corpus. The variable SENSE is binary, and its values are MONO (monosemous) and POLY (polysemous).

Donor language

Another factor that could play a role in the entrenchment of a borrowing is the donor language from which it stems. Much literature discusses the privileged status of English borrowings among borrowings in French (see Etiemble 1964 and Hagège 2006 for examples of highly popular general-audience works on this topic). Yet for all of the claims about the reasons for and effects of English borrowings in French, most research (Picone 1996, Pergnier 1989, Rey-Debove 1987, etc.) examines only English borrowings and not lexical borrowings from other languages. Although the number of English borrowings detected at T1 is far greater than the number of borrowings from any other language, it is not obvious whether these borrowings will also occur more at T2. In other words, we would like to know whether English borrowings have the same probability of entrenchment as borrowings from other languages. Donor language is investigated in the variable LANGUAGE. As discussed in Section 2.3.1, this variable is binary (ENG, i.e. English, vs. NON-ENG, non-English) due to the distribution of the borrowings in Table 2.1.

Frequency and dispersion

The effect of frequency is extremely prominent in language and has been studied most recently with respect to processing cost (see Bybee & Hopper 2001 for a general overview). In the context of the present study, I ask whether a borrowing's frequency at T1 will predict the borrowing's frequency at T2. In this study, a borrowing's frequency is its overall log count in the T1 corpus, since word frequency is perceived on a logarithmic scale (see, for example, Howes & Solomon 1951 and Oldfield &

Wingfield 1965).

Frequency information is given in terms of counts in a corpus of texts, yet dispersion can also measure the degree of entrenchment of a lexical item into the lexicon. Which of these two measures is a better indicator of entrenchment? In other words, if we see a word with a frequency of 2 occurring in two different text chunks, and another word with a frequency of 6 which only occurs in one text chunk, can we say which word, if either, is more likely to be entrenched into the lexicon? Furthermore, perhaps it is possible that the presence of one of these measurements obviates the need for the other. I have no a priori opinion about which measurement, if either, will be more effective than the other. However, I suspect that these two measures will complement each other, and that knowledge of both will only improve model accuracy in predicting lexical entrenchment. This hypothesis is supported by the work of Gries (2008), who proposes that both frequency and dispersion measures should be given when discussing cognitive entrenchment of lexical items.

Ideally, we would measure dispersion using articles in the T1 corpus as different text chunks. Since an article is a cohesive text unit, this division reflects speakers' exposure to the borrowings well. However, this was not easily possible given the format of the T1 corpus.

The T1 corpus is divided into 44 sub-corpora. Since the literature on the corpus (Abeillé *et al.* 2003) does not specify to what a sub-corpus corresponds, I determined upon inspection that several articles comprise a sub-corpus, and that almost all of the time one article was contained in one sub-corpus. In the absence of clear boundaries between articles, then, a sub-corpus is an ideal unit for obtaining dispersion measurements. In using the pre-defined sub-corpora, I hope to approximate the essential of what a dispersion measurement aims to capture, i.e. how many times the borrowings are used in different articles, while minimizing the amount of manual inspection that would be necessary to divide the corpus into articles. The 44 sub-corpora range in size from 6288 to 18594 words.

Unsurprisingly, frequency and log dispersion are correlated in the data (r_s =

0.926). Predictor variables with such high correlations can lead to a spurious model, so we regressed log frequency on log dispersion and took the residuals of this model as the FREQUENCY variable. This variable gives all frequency information that is not already taken into account by the DISPERSION variable, and so these two variables are independent. The correlation between original frequency and the new variable, residual frequency, is $r = 0.272$, which is significant at $p < 0.001$. That is, a substantial portion of frequency information is already taken into account with the DISPERSION variable, but what is not is positively correlated with T2 frequency.

Cultural context

The context in which a borrowing occurs could provide a clue about its degree of entrenchment. A borrowing can occur in a context referring to a culture that is typically associated with use of the language from which the borrowing stems, or it can occur outside of that cultural context. I call the former RESTRICTED cultural contexts and the latter UNRESTRICTED cultural contexts. An example of a restricted cultural context is when using the Russian borrowing *perestroïka* 'economic restructuring' when referring to Russia or the former Soviet Union, while using it to refer to France represents an unrestricted cultural context. Using a borrowing in an unrestricted context could imply that the borrowing is less anchored to a particular culture, and that the borrowing could take on a more general meaning. This relative lack of restrictions of the use of the borrowing could correlate with lexical entrenchment.

Continuing with the example, in French, *perestroïka* occurs 210 times in the T2 corpus, in contexts referring not only to the former Soviet Union, but also to Iran, Syria, and even France and the European Union. There are four attestations in which this borrowing refers to an economic restructuring of a democratic nation; one occurrence even refers to a *perestroïka médicale*. On the other hand, in the *New York Times* for the same date range, *perestroika* occurs 174 times; this number is arrived at using the same techniques of estimation used in Section 2.3.2. In the *New York Times* data, all but one of the occurrences refer to economic restructuring of a non-

democratic regime. Perhaps in French then the meaning of this term is something closer to 'economic restructuring' than the term's meaning in English, where it still seems strongly anchored to the former Soviet Union, former Communist countries, and, to a lesser extent, other non-democratic regimes. The widened scope of reference of the French term, as opposed to the English, and its corresponding greater frequency at T2 is perhaps no accident.

Flaitz (1988:87) touches upon the notion of cultural context in her study on English borrowings in the French press when she notes that

> The words *success story*, for example, appear... in an article about the life and career of Lee Iococca, the American head of Chrysler Corporation. Had the article been focussed on the life and career of François Mitterrand, the anglophone phrase *success story* would have incited much well-deserved criticism.

Here *success story* is used in a restricted context when discussing the American Lee Iococca, but if it were used to describe François Mitterrand, it would be an unrestricted context. The above passage implies that the range of possible cultural contexts is a salient feature of a lexical borrowing. The current study translates this intuitive concept into a criterion for measuring the degree of a borrowing's lexical entrenchment in the recipient language: it is perhaps the entrenchment of the borrowing into the language, observable in the unrestricted use of a borrowing, that would incite the criticism. I predict that a lexical borrowing in an unrestricted context in the T1 corpus, such as *success story* when discussing François Mitterrand, will reflect a higher degree of lexical entrenchment in French and hence a higher frequency at T2.

Determining the cultural context of a borrowing was done manually, as no method for automatically determining cultural context was available. Use of a borrowing in direct connection with a culture in which that language is spoken was considered a restricted cultural context, while use of it outside of that context was considered an unrestricted context. The majority of cultural contexts were classified fairly easily.

For example, contexts discussing British or American companies' *junk bonds* would be restricted, while contexts discussing *junk bonds* from French or Italian companies would be unrestricted.

However, some classification difficulties arose when discussing international contexts. If the context concerned multi-national bodies such as the European Union, we classified these contexts as unrestricted. At present, it is an open question whether these contexts correspond to the definition of a culture typically associated with a particular language. When seeing borrowings in the context of international agreements between countries that could trigger a restricted cultural context, I classified the borrowings as restricted. For example, a *joint-venture* between British and Japanese companies would have a restricted cultural context, since Britain is associated with English. Any mention of a country typically corresponding to a particular language could elicit more borrowings, and so my guiding principle was to be conservative when classifying borrowings as unrestricted. The binary CONTEXT variable has two values, RESTRICTED and UNRESTRICTED.

2.4.2 Results

Ten of the 281 tokens were excluded from the study as it was impossible to gauge their frequencies in *Le Figaro*, or because their cultural contexts or sense patterns were unclear in the T1 corpus. On the remaining dataset (n = 271), I performed backward variable selection starting with main effects for all predictors and any two-way interactions, using the Akaike Information Criterion (Akaike 1974) to eliminate superfluous predictors. This yielded a model with 11 factors. I explored using a mixed-effects model (e.g. Baayen *et al.* 2008) on the data, but due to large differences in number of occurrences of the borrowings at T1, such a model could not be fit to the data.

The multiple regression model for predicting lexical borrowings, with the predictor variables given in Section 2.4.1 and log frequency at T2 as the response variable, is

given in Table 2.2. Intrinsic properties of the borrowings are listed first, followed by locally determined properties and then two-way interaction terms.

Table 2.2: Linear model for predicting lexical entrenchment of French borrowings.

	$\hat{\beta}$	S.E.	t value	Pr(>|t|)
(Intercept)	3.0098	0.7177	4.19	< 0.0001
LENGTH	−0.5868	0.3511	−1.67	0.0959
SENSE (POLY)	2.1125	0.5080	4.16	< 0.0001
LANGUAGE (ENG)	−0.6973	0.5247	−1.33	0.1850
FREQUENCY	−2.8443	0.8240	−3.45	0.0007
CONTEXT (RESTRICTED)	0.5845	0.8239	0.71	0.4787
DISPERSION	1.7503	0.0902	19.40	< 0.0001
FREQUENCY∗DISPERSION	−1.9479	0.4607	−4.23	< 0.0001
FREQUENCY∗CONTEXT (RESTRICTED)	2.3039	0.8553	2.69	0.0075
LENGTH∗CONTEXT (RESTRICTED)	−1.7410	0.4635	−3.76	0.0002
SENSE (POLY)∗CONTEXT (RESTRICTED)	−1.9864	0.7363	−2.70	0.0074
LANGUAGE (ENG)∗CONTEXT (RESTRICTED)	1.8067	0.5792	3.12	0.0020

The lower panel of Figure 2.1 above shows that even after a logarithmic transformation, the response variable is not normally distributed. This non-normality was further indicated in an S-shaped pattern in the model's residuals, which indicates that the model finds it difficult to predict the borrowings with very small or very high frequencies at T2. This same pattern emerged after I excluded outliers. After eliminating datapoints marked as potentially overly influential from the original model and refitting the model, the same S-shaped pattern was apparent. Fortunately, a plot of the observed vs. predicted values in the original model shows general homoscedasticity of the errors. At this point, I was fairly confident that the model in Table 2.2 was robust. Still, to ensure that violations of the assumptions of the linear model were not leading to spurious results, I fit a non-parametric model, a random forest, to the borrowings data. I used an implementation of a conditional permutation scheme for each predictor variable so as to avoid undue influence of correlated variables (Strobl *et al.* 2001). Information about the importance of each predictor variable in the random forest model is given in Table 2.3.

Table 2.3: Predictor variable importance for a random forest model of the borrowings data. Variable importance is given as a function of mean decrease in accuracy if the variable is held out of the model.

Predictor variable	Mean decrease in accuracy
SENSE	0.001
CONTEXT	0.033
FREQUENCY	0.492
LANGUAGE	0.505
LENGTH	1.344
DISPERSION	7.394

According to the random forest, DISPERSION is the most important predictor variable; it also explains the most variance in the linear model. The SENSE predictor variable is the least important, but is still helpful in predicting the response variable. The correlation between predicted and observed values for the random forest model is $r = 0.828$, while for the linear model we have $r = 0.832$ for an R^2 value of .6797. Since the parametric and non-parametric models are similar, we conclude that the violations of the linear model are not sufficient to warrant concern – they do not prevent the model in Table 2.2 from being robust. Hence all subsequent results are in reference to the linear model.

Two types, *deutschemark* 'Deutschmark'[6] and its plural *deutschemarks*, were extremely frequent, with 25 and 24 attestations, respectively. This single lemma comprises 17.5% of the borrowings found, and it is necessary to determine whether it is unduly influencing the model we propose. To answer this question, I excluded all occurrences of *deutschemark* and *deutschemarks* from the data and reran a model on the data without these types. The resulting model was very similar to the original model, and I conclude that this lemma alone is not significantly affecting the results.

Since New *et al.* (2006) show a U-shaped curve for the effect of length on lexical decision times, I also added a quadratic term for syllable length to the model. This term was not significant: with respect to new borrowings, a linear effect of length, as

[6]The term in the *TLFi* for the former German currency is *mark*.

opposed to a polynomial effect, is sufficient for predicting lexical entrenchment. The lack of significance of the quadratic term in the model perhaps stems from the great majority of the borrowings in the corpus falling into the medium- and long-length categories of New *et al.* (2006). Also, I have measured length in terms of syllables, not letters. Therefore, a potential quadratic trend of length in the population is not likely to be significant in the model given the present dataset.

The final model given in Table 2.2 has an adjusted R^2 value of 0.6797 with a residual standard error of 1.402. To determine whether the model was overfitting the data, I performed bootstrap validation and compared the original goodness of fit statistics of the bootstrap samples to the full dataset. The R^2 decreases from 0.6851 on the training set to 0.6650 on the test set for an optimism of 0.0286. This optimism is no doubt due to the small dataset, but it is not large enough to call the model into question. Based on this information, I am reasonably confident that the model is predictive: it is capturing elements of lexical entrenchment of borrowings in French.

In this model, the significant main effects are SENSE, FREQUENCY, and DISPERSION, with the number of syllables being only marginally significant at $p < 0.10$. These main effects characterize the unrestricted contexts; all significant main effects except DISPERSION enter into significant interactions with CONTEXT. The DISPERSION predictor has the heaviest weighting and highest significance. All interaction terms are significant at the $p < 0.01$ level, with the interactions between DISPERSION and FREQUENCY as well as LENGTH and CONTEXT having the greatest influence in the model. As all significant main effects are modified by interactions, they will be discussed in conjunction with their interactions.

Figures 2.2 and 2.3 plot the partial effects for predictors retained in the model; the next subsections detail these results.

Length

As is seen in the upper left panel of Figure 2.2, length enters into a significant interaction with cultural context. When the cultural context is restricted, an increase

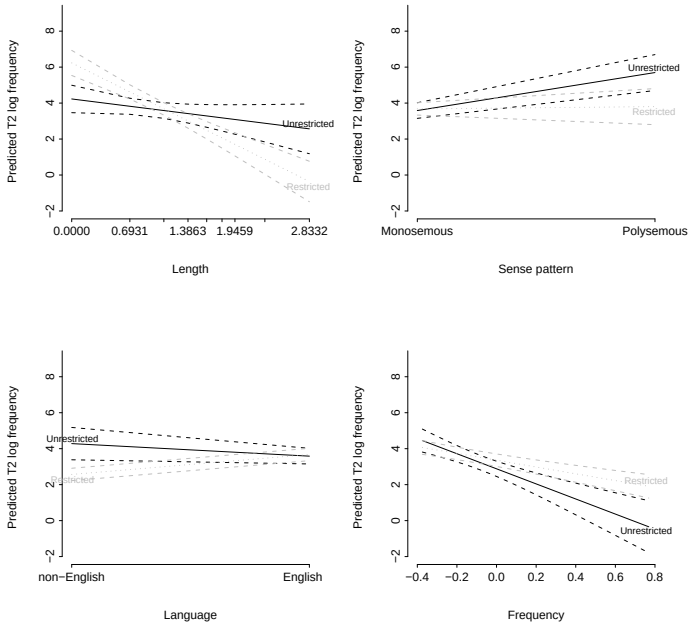

Figure 2.2: Effects for word length, sense pattern, donor language, and frequency by cultural context (restricted or unrestricted). Dashed lines indicate the 95% confidence intervals for each effect.

in length of a borrowing has an even more negative effect on the likelihood that the borrowing will become entrenched.

In examining the data, an explanation emerges as to why the effect of length is more pronounced when the cultural context is restricted. Of the 13 borrowings with the highest number of syllables, nine have restricted cultural contexts. Many of these borrowings refer to a specific policy in a foreign country, such as *an elastic currency*, *cassa integrazione* 'partial unemployment' (Italian), *classless society*, referring to the specific goal of John Major, the former Prime Minister of Great Britain, and *tierra y libertad* 'land and freedom' (the Spanish cry of Mexican revolutionaries). We are probably seeing this interaction because specific policies tend to have longer descriptors than more general concepts. In general, we can conclude that length probably does not bode well for a borrowing, and when length is used to denote a specific policy of a foreign country, the borrowing is even less likely to become entrenched.

Sense pattern

The next intrinsic property of a borrowing I examine is its sense pattern. The upper right panel of Figure 2.2 provides evidence for the hypothesis that in culturally unrestricted contexts, polysemous borrowings – i.e., new senses of borrowings with extant senses elsewhere in the language – are more likely to become entrenched than monosemous borrowings. While a polysemous borrowing has an extended range of denotata, borrowings in restricted cultural contexts have more limited ranges. The limited range of restricted cultural contexts could mitigate the positive effect polysemy has on a borrowing's probability of entrenchment. Monosemous borrowings, regardless of the context in which they occur, are more likely to be doomed to the margins of the lexicon.

Because of the interaction between polysemy and cultural context, we argue that the utilitarian explanation is better suited to explain the present findings on polysemy than the processing account. Polysemy is helpful in explaining the occurrences of the borrowings at T2, but only in culturally unrestricted contexts. Restricted contexts most likely narrow down the sense of a borrowing to such an extent that the advantage of polysemy can no longer emerge.

Recall that the current definition of polysemy requires us to include the English borrowing *cash*. Over the course of our research, the dictionary entry of the *TLFi* was changed from having a main entry of *payer cash* 'pay with cash', as is noted in footnote 1, to simply *cash*. Although *cash* is still only given as a part of the multiword expression *payer cash*, it could be seen as existing in the *TLFi*, and hence we might want to exclude it from the study. Since it is highly frequent at T2, including this word may unduly influence the influence of polysemy on lexical entrenchment. Upon excluding *cash* from the dataset, the effect of polysemy on lexical entrenchment is slightly diminished, but still significant ($\beta = 1.948$, $p < 0.001$).

The current findings contribute to general research on polysemy. With forms of language creation other than borrowings, we see that polysemy arises spontaneously

(Steels *et al.* 2002 look at simulations of language creation with artificial agents and finds polysemy of newly formed lexical items). Polysemy is highly prevalent in natural language: according to Fuchs (1996:29), 40% of French lexical items are polysemous, while Rodd *et al.* (2004) observe that 84% or more of relatively frequent English words in the Wordsmyth dictionary (Parks *et al.* 1998) are polysemous. The English figure is higher the French figure since more frequent words tend to be more polysemous, but we can still see that polysemy is rampant in these two languages. Because polysemy is common and is arises spontaneously, it might be advantageous in some way.

My hypothesis is that it is easier for speakers to make a transfer of meaning than to invent an entirely new word for a given concept. For example, on the French website Dictionnaire de la Zone (`http://www.dictionnairedelazone.fr/`), a website of neologisms in French similar to Urban Dictionary insofar as users propose content and give example usages, the word *fils* is proposed to mean 'friend, colleague, pal [masc.]'. This meaning no doubt is related to the standard definition of *fils*, i.e. 'son'. It seems less difficult to create this word with this meaning than to create an entirely new word with the same meaning.

Under this hypothesis, polysemy functions as a mnemonic device to a speech community when the neologism enters the language. On popular websites proposing or documenting neologisms in English such as WordSpy (`http://www.wordspy.com/`), Urban Dictionary (`http://www.urbandictionary.com/`), and Wiktionary (`http://en.wiktionary.org/wiki/Wiktionary:List_of_protologisms`), there are very few proposals for new words that do not have at least one element in common with an extant word; this element can be morphological, phonological, and/or semantic in nature (e.g., metonymy). This shared element most likely contributes to the retention of the new lexical item. Just as sharing a form with an existing lexical item makes learning a new word easier, so does sharing semantic content. Perhaps for these reasons borrowings relying on polysemy with extant lexical items are more likely to become entrenched than other borrowings. In this way, borrowings could also provide evidence toward a shared element with an extant lexical entry giving a

neologism an "evolutionary advantage" in the lexicon.

Donor language

Donor language did not yield a significant main effect, but did have a significant interaction with cultural context. As is seen in the lower left panel of Figure 2.2, for borrowings from English, the cultural context is irrelevant. However, for borrowings from other languages, context matters: non-English borrowings occurring in unrestricted contexts are more likely to become entrenched than restricted non-English borrowings. Hence we have Russian *nomenklatura* 'high-level government officials [under Communism]', which always occurs in culturally unrestricted contexts at T1, and which occurs 190 times at T2. In contrast, a restricted non-English borrowing is *scala mobile* 'method of adjusting salaries according to prices' (Italian), with a frequency and a dispersion of 4 in the T1 corpus but only 3 occurrences in the T2 corpus.

It is possible that markedness plays a role in this distinction. Context is irrelevant for English because it is the unmarked donor language. Francophones are so used to hearing English borrowings that the cultural context in which they occur does not matter for lexical entrenchment. For marked donor languages, borrowings in unrestricted cultural contexts are more likely to become entrenched than borrowings in restricted contexts because in unrestricted contexts, there are fewer limitations on what the borrowings can refer to. That is, the results for marked donor languages support my original hypothesis about cultural context.

Frequency

The lower right panel of Figure 2.2 indicates that, regardless of context, a higher frequency is a bad omen for a borrowing. This at first sight counter-intuitive finding can only be understood when dispersion is also taken into consideration (see Section 2.4.2). The negative effect of frequency is less pronounced when borrowings occur in restricted contexts. For example, the borrowing *board*, with the sense of *board of*

directors, occurs three times at T1, always in culturally restricted contexts. It has a dispersion of 1, yet it is still frequent at T2 with 184 occurrences. On the other hand, the borrowing *flash*, with the sense of *flash memory*, also occurs three times at T1 with a dispersion of 1. Two of these three occurrences are in culturally unrestricted contexts, and it only has 33 occurrences at T2.

I posit that frequency is less of a bad omen for borrowings in restricted cultural contexts because there are more cues to aid in integration into memory of these words. For example, a word such as *board*, occurring only in culturally restricted contexts at T1, is used to describe a particular company, and having a particular company tied to our memory of *board* will help us remember this lexical item. The more frequent this lexical item is, the more associations a speaker will have with it. Borrowings occurring in culturally unrestricted contexts have fewer salient associations with them, and they are hence less likely to benefit from increased frequency.

Dispersion

The DISPERSION main effect and the DISPERSION∗FREQUENCY interaction are solid predictors with the largest effect sizes in the model. Since the FREQUENCY variable is also significant, I can say that knowing only about a borrowing's dispersion and frequency in the T1 corpus will enable us to make a decent guess about its degree of lexical entrenchment at a later date. This finding is somewhat surprising: first, new words can be trendy, and hence frequent at a particular time, and all but forgotten some years later. For example, this can be the case with words to describe new technologies that are not adopted by society. A query of the *New York Times* archives for *laser disc*[7], a now-obsolete precursor to the DVD, shows that 18 of the 58 results are from 2000 or 2001, which is right around the time the production of laser discs stopped. This word is fairly frequent during these years, but is only seen three times from January 2007 to August 2008.

Second, the relationship between dispersion and lexical entrenchment is not straight-

[7]See `http://tinyurl.com/6dnxs8`, accessed on 27 August 2008.

forward due to the interaction between dispersion and frequency at T1. The main effect of DISPERSION shows a positive correlation with T2 frequencies, but the DISPERSION∗FREQUENC interaction has a negative coefficient value. As dispersion and frequency increase, the number of occurrences at T2 decreases.

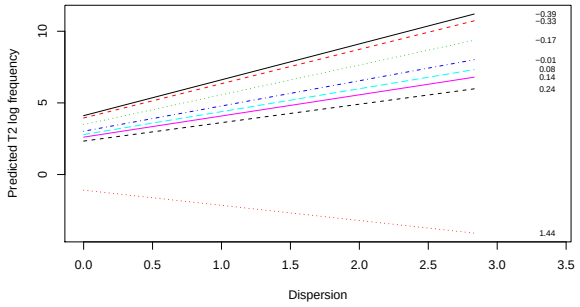

Figure 2.3: Interaction plot of log dispersion and residualized log frequency (in deciles given on right, with deciles 2 and 3 having equal values) on predicted T2 log frequency.

Figure 2.3 shows the dispersion-frequency interaction in quantiles of tenths. In this figure, residualized frequency deciles are in ascending order from top to bottom on the right side. For example, one tenth of borrowings have a residualized frequency of -0.39 or less, and 100% of borrowings have a residualized frequency of less than 1.44.

Like in Figure 2.2, in Figure 2.3 we see the same decrease in predicted T2 frequency as T1 frequency increases[8]. In this figure, the borrowings in the uppermost decile for FREQUENCY, the lowest dotted line, have the largest effect on this interaction. Borrowings in this decile have a residualized log frequency greater than 0.238. These borrowings are frequent, yet poorly dispersed in the T1 corpus. In fact, in Figure 2.3, they do not extend between the 1.0 hash mark on the dispersion axis. In general, these borrowings are infrequent in the T2 corpus. Examples of these borrowings include *ejido* 'communal agricultural land', Spanish, with 9 occurrences at T1 and 0

[8]The largest decile, shown in Figure 2.3 with a residual frequency of 1.44, is not given in Figure 2.2. This is a design principle of the `Design` package in `R` (Harrell, 2001). .

occurrences at T2, *ejidos* (*ejido*-pl), with 8 occurrences at T1 and 1 occurrence at T2, and *flash*, with 3 occurrences at T1 and 33 occurrences at T2.

This interaction suggests that borrowings with a low dispersion but high frequencies are getting penalized for their burstiness. A word that occurs, say, nine times in one article but nowhere else in the corpus is probably more indicative of that specific article than of language use in general. Unless there is another article about the socio-political aspects of agricultural practices in Latin America, for example, the borrowing *ejido* will probably not be used in French. In contrast, a word with a frequency of 9 and a dispersion of 3 is likely to be a frequent word at T2. The actual words are article-specific, but the penalty they incur is generalizable.

In the T1 corpus, many of the borrowings in the top decile for residualized frequency do come from one large article on the socio-political aspects of agricultural practices in Latin America. Because of this, a reviewer wondered whether a burstiness penalty is more appropriate for borrowings from this article or for borrowings in general. Upon exclusion of borrowings from this article in the multiple regression model, a positive DISPERSION main effect and a negative FREQUENCY∗DISPERSION interaction term were still significant. These results suggest that all frequent yet underdispersed borrowings in the dataset, and not just those in this specific article, incur a burstiness penalty.

The causal nature of frequency and dispersion

From a strictly synchronic perspective, the predictor variables of FREQUENCY and DISPERSION are only diagnostics, and not causal factors. It is essential to take diachronic corpora into account when examining lexical entrenchment, because otherwise the predictor variables of FREQUENCY and DISPERSION would only be diagnostics, and not causal factors, in determining lexical entrenchment. In fact, the *FUDGE* hypothesis has been criticized as at least partially circular (Pinker 2007:308), because frequency and diversity of users are, in a synchronic approach, what one would like to explain. But these factors can be causal predictors in a diachronic approach to

lexical entrenchment: the more frequent and well-dispersed a new word is, the more speakers will hear and eventually use it.

We can predict not only the frequency with which a word occurs at T2, but also whether it is still in use at T2. Crucially, a logistic regression model predicting whether the word will still be in use 10 years later is indeed significantly better than chance at predicting whether words fall out of use ($\chi^2(1) = 6.487$, $p = 0.011$). The model correctly predicts that borrowings are no longer in use in 39 cases and that they are still in use for 194 cases. My model can effectively predict lexical change across two time periods on the basis of FREQUENCY and DISPERSION, inter alia[9].

Cultural context

Of all the predictor variables, CONTEXT has the most complex relationship with the response variable. The main effect of cultural context is not significant, yet it interacts with four other predictor variables, LENGTH, SENSE, LANGUAGE, and FREQUENCY, in ways that are perhaps not intuitive.

A rule of thumb of linear regression is that there should be 15 datapoints for every factor (Harrell 2001:61). The present model has approximately 23.4, so I am not overfitting the data. Furthermore, in all of the models explored above in Section 2.4.2, i.e. the model without outliers, the model without influential datapoints, and the model excluding the frequent lemma *deutschemark*, the cultural context effects remained very similar to those of the original model. It is therefore unlikely that the cultural context interactions in the model are caused by the exceptional properties of the dataset.

My current hypothesis is that borrowings tend to first enter the language in restricted contexts and then expand their contexts to unrestricted borrowings. It follows straightforwardly that borrowings seen at T1 in unrestricted contexts are more likely

[9]The logistic regression model has main effects for FREQUENCY ($\beta = -6.394$) and DISPERSION ($\beta = 5.033$) and interaction terms LENGTH∗CONTEXT ($\beta = -2.732$) and LANGUAGE∗CONTEXT ($\beta = 3.188$), all significant at $p < 0.05$.

to become entrenched, since they are further along in the process of entrenchment than borrowings in restricted contexts. Future work can directly test this hypothesis by examining the contexts of the borrowings at T2 of borrowings occurring in restricted contexts at T1. If at T2 we see a positive correlation between frequency and unrestricted uses of these borrowings, we can conclude that some borrowings do initially have restricted cultural contexts before the set of contexts in which they can be used is expanded.

The cultural context can be conceived of as a coarse-grained content indicator. Is content otherwise relevant to a borrowing's adoption? Does topic matter? For example, one could hypothesize that technological borrowings are more likely to become entrenched than borrowings pertaining to fashion. Unfortunately, the lack of information about the breakdown of topics in the T1 corpus does not allow for us to answer this question in a meaningful way. We are not given a distribution of articles by topic and/or by genre for the T1 corpus. This means that if we see more financial borrowings, for example, we do not know if this is because a high proportion of the corpus relates to financial articles, or if borrowings relating to financial topics really are more likely to become entrenched than borrowings pertaining to other topics. This could be a feature of upcoming versions of the T1 corpus (Anne Abeillé, p.c.), so this may well be a promising avenue for future research. Another improvement to the T1 corpus concerns the division of the corpus into articles as opposed to sub-corpora. Divisions by articles will be a feature of future versions of the corpus (Abeillé, p.c.), and working with articles could allow for enhanced dispersion counts.

2.5 General discussion

Given that a new word occurs in a language, what makes it likely to "survive" and become part of the lexical stock of the language? To answer this question, I examine one method of lexical enrichment, lexical borrowings, in French, a language in which new lexical borrowings are not only common but also relatively easy to single out in

corpora. I look at two journalistic corpora in French from two different time periods to see if the new borrowings found at the first time period (T1; 1989-1992) are still in use at the second time period (T2; 1996-2006). Using frequency at T2 as a measure of lexical entrenchment, I find that several factors, such as a borrowing's dispersion and frequency, whether or not a borrowing is polysemous, the length and cultural context of a borrowing, and donor language of the borrowing help to determine a borrowing's degree of lexical entrenchment into French.

Of the factors examined, we would expect that dispersion, frequency, sense pattern, and length will also be relevant to a similar study on different types of neologisms, while donor language and cultural context are specific to borrowings. Given my results, all of these factors could be relevant when examining borrowings in other recipient languages, but we do not necessarily expect the factors to act the same way as they do in French. Non-English borrowings in culturally unrestricted contexts are more likely to be incorporated into the lexicon than culturally restricted non-English borrowings. Longer borrowings are generally less likely to become entrenched. Polysemous borrowings, i.e. borrowings with extant senses in the lexicon, are more likely to become entrenched into the lexicon than borrowings with no existing senses in the lexicon. Some factors, such as dispersion, are relatively straightforward to interpret. The interactions involving context are more difficult to understand, and the explanations I offer about them remain tentative.

It remains an open question as to whether the presence of pre-existing words or phrases in the recipient language at T1 describing the same concept as the borrowing can affect a borrowing's entrenchment. If extant native words are available for speakers, this might be reflected in a lack of lexical entrenchment for the borrowing. I opted to steer clear of this potential predictor, however, because it is impractical for at least three reasons. First, it sometimes is difficult to know at T1 if the equivalent word exists in the recipient language. For example, the English borrowing *short* is given in a stock context to refer to short sellers, and the French translation of *short seller* is *vendeur à découvert*, which is given in the T1 corpus directly preceding *short*.

But who knows if *vendeur à découvert* existed before this usage? Finding this out would require detailed sleuthing for each individual borrowing. This would be further complicated by the lack of electronic resources for pre-1989 dates. Furthermore, just how many times does a word have to be mentioned in the language in order for us to say it exists in the language? If we say just once, it could be quite difficult to say with certainty that an equivalent native term has never existed before in the language. Second, in my informal discussions on the meanings of borrowings with native speakers, it seems some speakers may detect a nuance of sense in the borrowing that others do not, so that the former will say there is no exact translation of a borrowing, while the latter will gladly given a native equivalent for the borrowing. Such subjectivity could complicate this avenue of research. Finally, perhaps a native phrase or multi-word expression exists to convey a similar idea, but it is considerably longer than the borrowing. All of these considerations combine to make the effect of extant native lexical items a challenging factor to examine.

For corpus and computational linguists, the principle finding of this research is likely to be the role of dispersion and frequency in modeling lexical entrenchment. These variables are both significant predictors, and the interaction between them is also predictive, with a negative effect on the response variable. This interaction penalizes borrowings in the dataset that are frequent but too bursty. Preliminary psycholinguistics findings (Gries 2008) support the idea that dispersion is a better predictor of reaction times than frequency. This is congruent with the present result and therefore could suggest that a burstiness effect could also be present in reaction time studies. Further studies and computational applications requiring word frequency distributions would do well to examine both frequency and dispersion for the construction of predictive language models.

For linguistics, the present findings indicate that lexical entrenchment of borrowings is predictable to a surprising extent. This is fortunate, since in rule-governed morphology, generative works give insight as to whether a form is possible or not. Productivity measures à la Baayen (1992) assign a probability to an affix or a lexical

process of occurring in new lexical items. Neither of these approaches provide insight as to whether the form will become entrenched in the language. I offer the present study as a first step in bridging the gap between the possible, the probability of the possible, and the probability of actual entrenchment into lexicon.

Chapter 3

Memory for new words

Recognition and free recall experiments addressed the retention in memory of meaningful new words (acronyms, blends, and derived words) collected from recent media sources. In a familiarization phase, participants were presented with the new words, were informed about their meanings, and answered a short questionnaire. In the recognition experiment, participants read through a list of words while their eye movements were recorded, and were asked to indicate whether they had seen these words before. Subsequently, they were asked to freely recall the words presented to them during familiarization. For half of the subjects, at least one night's sleep intervened between familiarization and testing, allowing us to examine potential consolidation of the new words into long-term memory. Hippocampal transfer during sleep enhanced recall for derived words, indicating that better consolidation in memory was restricted to rule-governed word formation. At every level of processing, acronyms emerged as more complex than blends and derived words. In recognition, word length and component frequency co-determined processing costs, whereas in free recall, the usefulness of the new words across different social networks emerged as the crucial predictor. Recall was superior for words suitable for informal, strong-tie networks, and impoverished for words appropriate for contexts relating to occupational, weak-tie networks. These results indicate a remarkable dissociation between memory for comprehension and memory for production, with comprehension memory for new words driven by lexical-distributional properties such as length and frequency, and memory for the

production of new words driven by their social valence.

3.1 Introduction

Humans have an amazing capacity to learn and remember words. Estimates of single-word vocabulary size alone for English speakers range from 60,000–150,000 words (Miller 1991, Kuiper 2006), with no clear upper limit. Furthermore, the acquisition of new words continues throughout most of a speaker's lifetime, so the size of an individual speaker's vocabulary size increases until old-old age (about 75 years old; McGinnis *et al.* 2008). Even more impressively, memory for linguistic units might not be limited to single words: frequency information for frequent, regularly derived phrasal units such as "do you know how" seems to be predictive of processing complexity independently of frequencies of the individual words (Arnon & Snider 2010, Tremblay 2009, Tremblay & Baayen 2010). Jackendoff & Pinker (2005) estimate the number of idioms speakers know to be of comparable magnitude to the number of known words. Jackendoff concludes that a substantial part of brain capacity is allocated to storing linguistic units (Jackendoff 1997).

Even though our lexical knowledge is vast, there seems to be no specialized mechanism for acquiring new words (Markson & Bloom 1997), and not all words to which we are exposed become part of our mental lexicons. Adult recognition for words encountered once range from approximately 85 to 95% after one week (Markson & Bloom 1997, Gaskell & Dumay 2003), with deeper retrieval mechanisms, such as free recall, at much lower rates, ranging from 5 to 30% (Dabrowska 2009, Swanborn & de Glopper 1999, Dumay & Gaskell 2007).

Previous studies of lexical learning have made use of simplex, monomorphemic words. Typical examples of novel words presented to subjects are *koba*, defined as a particular object (Markson & Bloom 1997), and *cathedruke* (Gaskell & Dumay 2003, Dumay *et al.* 2004, Dumay & Gaskell 2007), the latter modeled on the existing word *cathedral* but defined as a variety of vegetable. However, new words encountered by

adult speakers are rarely monomorphemic. Only 14% of the lemmata in the CELEX lexical database (Baayen *et al.* 1995) are monomorphemic; furthermore, in the list of word forms, which includes inflected variants, only 8% of the distinct forms are monomorphemic. One of the design features of language is that word formation rules and rules of inflection reduce the demands on memory by creating signs that are at least partially motivated. Recent evidence suggests that complex words may be easier to understand than monomorphemic words once differences in length and frequency are factored out (Balling & Baayen 2008). Given that most new words we encounter are morphologically complex, and given that the comprehension of such words depends on our knowledge of morphological rules, LEXICALIZATION, the efficient integration of new words into memory, may also depend on the availability of such rules.

The present study complements previous work on the integration of new words in lexical memory by focusing on new morphologically complex words. One recurrent problem in the experimental study of the lexicalization of new words is that it is difficult to find plausible meanings for new words: most plausible and useful concepts already have well-established words. This problem is somewhat alleviated when complex words are used instead of simple words, but the risk of unnatural form-meaning couplings remain when the researcher invents new words. I avoided this problem by collecting newly coined words from online media outlets to use as stimuli.

I explore two hypotheses about the lexicalization of new complex words. First, various STRUCTURAL FACTORS could favor entrenchment in memory. Second, a new word's SOCIO-CONTEXTUAL AFFORDANCES may also determine its subsequent accessibility in memory.

One structural factor likely to affect lexicalization is the nature and the productivity of the word formation process that gave rise to a new word. Morphological theory (Bauer 1983b, Bauer 2001) distinguishes between *regular* word formation processes, such as affixation and compounding, and the *creative* word formation processes leading to blends (*brunch*) and acronyms (*radar*). Often, regular word formation proceeds without specific attention or awareness on the part of the speaker that a new form

is being created (Schultink 1962). For example, while addressing a University Senate committee recently, a provost at the University of Minnesota spoke of *negativeness* instead of *negativity*. Since he did not hesitate or correct himself, and an existing word, *negativity*, serves the same function (not to mention that the gravitas of the situation would argue for conservative language use), I suspect this word was derived unintentionally as a result of the overwhelming productivity of *-ness* compared to *-ity* (Baayen & Renouf 1996). By contrast, creative word formation often involves a conscious, volitional act on the part of the speaker to craft a new form. When confronted with a creative new form, listeners often require at least a partial explanation of the intended meaning, even though a posteriori the semantic motivation for the new word may be fairly transparent (e.g., *brunch* as a blend of *breakfast* and *lunch*).

To clarify the consequences of creative versus regular word formation for lexical entrenchment, I included three kinds of stimuli in this study. As representatives of creative coining, I included blends and acronyms. Derived words were included to represent rule-based complex words. Since new regular derived words are supported by rules and/or a solid instance base of similar exemplars in memory, I predict the integration of new regular derived words in lexical memory to be easier. New derived words should have an "evolutionary" advantage compared to blends and acronyms, as there is a well-established ecological niche into which new derived words can be accomodated.

Other structural properties of new complex words may also affect how well they enter lexical memory. Consider blends such as *trumor* (from *true* × *rumor*) and *snowmageddon* (from *snow* × *armageddon*). In the former, both phonological forms are completely present, while in *snowmageddon* only *snow* is fully present. The meaning of *armageddon* has to be reconstructed from partial information (*-mageddon*). Acronyms are even less transparently related to their contributing component words or morphemes than blends. I define an ACRONYM as any reduction of a series of words to a series of their initial letters, regardless of whether or not the resulting series is a phonotactically legal sequence in English. Most American speakers are familiar

with acronyms such as *FBI* and *CFO*. Thus, derived words, blends, and acronyms differ not only with respect to word formation type, they also differ on a gradient of phonological transparency. Entrenchment in memory for new blends and especially acronyms might therefore be less effective compared to new derived words, due to reduced phonological support for linking forms to meanings.

For the comprehension of existing complex words, effects of word length, whole-word frequency, constituent frequency and constituent family size are well established (Baayen *et al.* 2007, Kuperman *et al.* 2010, De Jong *et al.* 2000; Moscoso del Prado Martín *et al.* 2004). For novel complex words, I expect these facets of constituent structure to co-determine how a word is stored. As a consequence, I predict that the comprehension of a novel complex word will be predictable, at least in part, from the lexical distributional properties of its component words, such as their frequency of occurrence.

Whether this prediction generalizes from novel derived words to novel blends and acronyms is an empirical question. If such words are understood decompositionally, in terms of their constituent components, a component frequency effect is a real possibility. Lehrer (2003) did not find evidence for decomposition of blends in a masked priming study, but in that study the blends were not explained, nor were they inferable from context.

Especially for acronyms, a component frequency effect would indicate a semantic locus for the frequency effect, as argued for comprehension in general by Baayen *et al.* (2006). To illustrate this, consider one of the acronym stimuli I heard "in the wild" some time after this study, CLM. On Urban Dictionary, (`http://www.urbandictionary.com`), CLM is defined as a *Career Limiting Move*, i.e, "Something you do at work that will get you fired, or end your career very soon." Without knowing its meaning, none of the meanings "career", "limiting" or "move" is activated when reading CLM. Once its meaning is known, it remains unlikely that CLM is understood by activating orthographic representations for *career, limiting*, and *move*. A much more parsimonious explanation is that the letters C, L, M provide access codes to the

relevant meanings. If so, a component frequency effect for acronyms must arise at the level of word meanings. For frequent, well-entrenched acronyms, the link to the original motivating components can be lost. For many speakers, *radar* has become a normal monomorphemic word, and its original components (*RAdio Detection And Ranging*) are no longer functional. For such acronyms, no component frequency effect is to be expected.

Thus far, we have considered potential structural and distributional factors that I expect to co-determine lexical entrenchment. In this study I also consider a second set of potential predictors that we expect to affect encoding in memory, namely, what I will call a new word's SOCIO-CONTEXTUAL AFFORDANCES: new words that I encounter differ in the extent that they are useful for us to use with others in particular social and situational settings.

One dimension that is relevant here is the emotional valence of a new word. It is known for existing words that emotional words are more likely to be remembered than existing neutral words (Kensinger & Corkin 2003). Furthermore, existing words are remembered better when presented in humorous sentences (Schmidt 1994). This suggests that words that are not emotionally neutral are remembered better. I therefore expect humorous new words such as *snowmageddon* to have greater chances of lexical entrenchment, even though they run the risk of being perceived as obnoxious (Metcalf 2004).

Another second relevant dimension is social in nature. Listeners are attuned to a speaker's education level and social background (Labov 1966, Hay *et al.* 2006). For existing words, participants are known to be able to better remember words they use to describe close social ties than words with which they describe an unknown person (Bower & Gilligan 1979). Given that for an existing word its social affordances play a role in lexical processing, I expect that words that are of potential use to the subjects in communication with other speakers with which they have close social ties, will be better consolidated in memory.

There is an intriguing social correlate to the cognitive distinction between strong

and weak social ties. Not only do we remember words better if they are given in the context of close social ties, research on the diffusion of non-linguistic innovations (Rogers 2003, p. 340, Friedkin 1980) indicates that information spreads faster through close social ties. Thus, in anticipating whether new words will become entrenched into a particular speech community, we can posit a "virtuous cycle" for words that speakers would use with close social ties. Specifically, speakers are more likely to remember words they associate with close social ties. Since a word must be recalled in order for a speaker to employ it, a new word associated with close social ties, all other things being equal, is more likely to be re-uttered. This allows for more speakers to associate this word with a member of their ingroup, specifically, the one from whom they heard the word, and their memory for this word would hence be improved. Such a cycle would lead to the entrenchment of this new word in the speech community. I note that the slang of a closely knit speech community also conforms to this hypothesis.

In what follows, I report a three-part experimental study on memory for novel derived words, blends, and acronyms, and the role of a word's structural properties and socio-contextual affordances. During the self-paced familiarization phase, participants read new words, their definitions, and examples of their use in context; they then answered a series of questions probing different aspects of the words' socio-contextual and emotional affordances. For half of the participants, a testing phase took place 15 minutes after the learning phase; the other half was tested after a night's sleep to test for consolidation due to hippocampal transfer into long-term memory (Rudoy *et al.* 2009, Gaskell & Dumay 2003, Dumay *et al.* 2004, Dumay & Gaskell 2007). To probe whether a word had become a part of a speaker's mental lexicon, I used three different techniques: eye-tracking, recognition, and free recall. Even though speakers might not recognize a word, memory traces for previous experience might still be detectable in how the eye moves through the printed word. Hence the testing phase consisted first of a recognition task in which participants decided whether they had seen the word in their previous visit, while their eyes were tracked. After the recognition experiment, participants completed a free recall experiment.

3.2 Method

3.2.1 Participants

Forty-five participants at the University of Alberta took part in the experiment (23 females, 6 left-handed, mean age: 28.8). Thirty-one were volunteers and 14 participated in the experiment for course credit.

3.2.2 Apparatus

Stimuli for the recognition experiment were presented on a 24-inch screen with a 60-Hz refresh rate. A Microsoft Sidewinder Plug & Play Pad game controller was used to record participants' decisions (left = unseen; right = seen). Eye movements were monitored using an EyeLink1000 chin-rest eye-tracker (SR Research, Ltd., Mississauga, Ontario, Canada). Only movements of the left eye were recorded; for this, the pupil-only mode was used. The recording sampling rate was 1000 Hz.

3.2.3 Stimuli

The target stimuli consisted of 75 new words: 25 acronyms, 25 blends, and 25 derived words, obtained from websites such as Urban Dictionary (`www.urbandictionary.com`),
Wikipedia (`http://en.wiktionary.org/wiki/Wiktionary:List_of_protologisms`), National Public Radio (`http://www.npr.org/`), and the TV show *The Colbert Report*. These 75 words were selected from a larger set of 162 which were presented in a separate pre-study to 15 speakers in Minnesota who were asked to indicate whether they had seen these words before. Of the 75 stimuli selected for the experiment, 69 were words that none of these speakers had encountered before; six were words that only one speaker had seen before. The words varied in length from two to 20 characters, with an average word length of 7.83 characters (SD = 3.90). Stimuli are given in Appendix 7.2.

Sixty-four of the remaining 87 acronyms, blends, and derived words were used as fillers, and a further fourteen words were used as practice items. For the familiarization phase, the experimental list was composed of the 75 target words, seven practice items, and three fillers that were used as familiar practice items for the recognition experiment. In the recognition experiment, the experimental list was composed of the 75 target items, 64 fillers, and again seven practice items, three of which were familiar. In addition, a final block of 30 words was presented at the end of the experiment, for the following reason.

Acronyms are normally printed in upper case. I therefore presented acronyms in upper case in the experiment as well, while presenting blends and derived words in the usual lower case. In order to distinguish between case and word type (acronym vs. non-acronym), I included a final block of 30 words, half of which were presented in upper case, and half in lower case. The 15 words in lower case were comprised of eight words previously seen and seven previously unseen words. The 15 words in upper case consisted of seven words seen previously and eight unseen. I use the unseen words in this block to test for a potential effect of case in the discrimination task. All 15 unseen words were sampled from the same sources as the original 162 words.

3.2.4 Procedure

The experiment consisted of three parts. In the familiarization phase, stimuli were presented to the participants with a definition and an example context. Definitions were necessary, especially for acronyms, because their meanings are nearly impossible to deduce solely from their form. Acronym definitions always included all component words (e.g., LIMH, "Laughing In My Head. Like LOL except for not literally Laughing Out Loud"). The definitions for blends and derived words included at most one component word in 26/50 of the stimuli (the definition of *destinesia* does not mention its components: "What occurs when you get to where you were intending to go,

and you forget why you were going there in the first place"; *hypertasking* is defined with one component: "The simultaneous execution of an exceedingly large number of tasks. Multi-tasking to the Nth degree."). Both components were never given in any definition.

The definition and example context followed the original source as closely as possible. Some modifications, however, were required: profanity was removed, very long examples were shortened, and no stimulus was presented sentence-initially or sentence-finally in the example sentence. After presentation of the word with definition and context, the five socio-contextual affordance questions in Table 3.1 were presented to the speaker with the word, definition, and example sentence remaining on the screen.

Label	Question	Scale endpoints
emotional	How emotional do you think this example text is?	[not emotional at all/highly emotional]
funny	How funny do you think this example text is?	[not funny at all/very funny]
educated	How educated does the writer of this example text sound?	[not educated at all/highly educated]
party	Would you use this word at a party with your friends?	[no, not at all/absolutely]
work	Would you use this word at school or at work?	[no, not at all/absolutely]

Table 3.1: Socio-contextual affordance questions participants responded to for each word.

Prior to these questions, a measure of familiarity, or subjective whole-word frequency, was elicited: the trial always began with the question, "Have you seen this word before?" All of these questions were presented in the same order for all words. Participants responded by clicking the mouse on a 7-point Likert scale (see Figure 3.1). This part of the experiment was self paced. The experimental items were preceded by seven practice items. Short breaks were included after the practice session and one-third and two-thirds into the experiment. The order of presentation was

randomized for each participant separately.

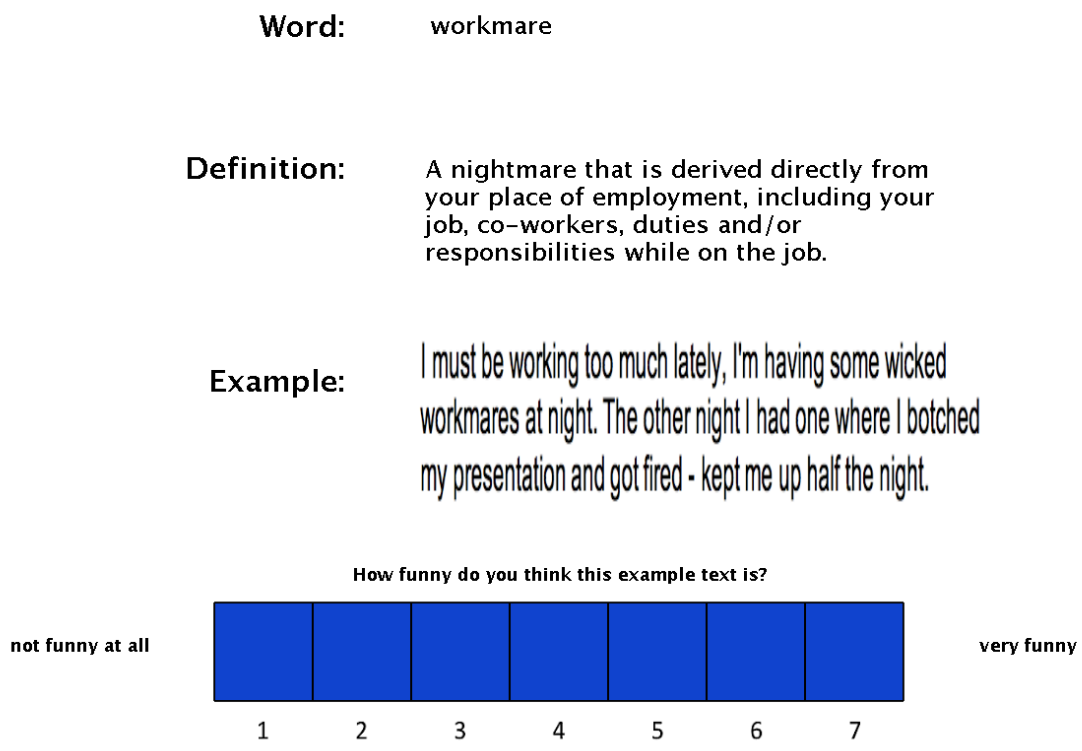

Figure 3.1: An example word-question pairing to which participants responded.

The familiarization phase was followed by a recognition experiment during which I registered participants' eye movements. Stimuli were presented in isolation on a monitor in white 44-point fixed-width font on a black background. Acronyms were presented in upper case; blends and derived words were presented in lower case. A trial consisted of the following sequence. First, a drift correct mark also served as a fixation point. Words were presented once the participant fixated on the drift correct mark. The left edge of the stimuli was 10.5 cm to the right of the left edge of the screen and 3.3 cm to the right of the fixation point. From the viewing distance of 70 cm of the screen, the visual angle was estimated to be 0.8 degrees for each letter. Because the optimal viewing position for a word is left of the center of the word (Farid & Grainger 1996, O'Regan & Jacobs 1992, Vitu *et al.* 1990), the fixation point was replaced by the stimulus such that its third letter coincided with the fixation point.

Stimuli remained on the screen until participants responded, with a timeout of 5000 ms. Immediately following a response, the next fixation point was displayed.

Here again, the order of presentation was randomized for each participant separately. The block of 30 control items testing for potential effects of case was randomized independently of the main list and was always presented as the last block. A short break followed the practice items and every set of 30 trials. During the breaks, participants were provided with feedback on their accuracy. At each break, the eye tracker was recalibrated using a three-point horizontal grid.

After a short break, the recognition experiment was followed by a free recall experiment. I asked participants to recall as many words as they could from the 75 words which they had seen during the familiarization phase. They were told not to worry about the precise spellings of these words. Participants were seated in front of a desktop computer and typed in their responses. They were given three minutes to complete this experiment.

To test the effects of memory consolidation during sleep, the experiment included a between-subjects sleep manipulation. Twenty-three participants completed the familiarization phase and after a short break proceeded with the recognition and recall experiments. The remaining 22 participants proceeded with the recognition and recall experiments with at least one full day after completing the familiarization phase[1].

3.2.5 Predictor variables

Participant was included as a random-effect factor. The between-subjects fixed-effect factor *sleep condition* (levels: immediate, sleep) was included to test for the effect

[1]Most sleep-condition participants (16/22) returned one day after their familiarization phase; the rest returned two or three days after the familiarization phase. To see if the additional nights of sleep were affecting the present results, I ran separate models for each response variable after excluding the six participants for whom the recognition and recall experiments occurred more than one day after familiarization. All resulting models (not shown) were very similar to those presented below which make use of all sleep-condition participants. These findings are in line with Dumay et al.'s findings that the essential ingredient to lexical consolidation is the binary condition of intervening sleep and not temporal distance from exposure.

of a night's sleep on recognition and recall. Given the results of Gaskell & Dumay (2003), Dumay *et al.* (2004), and Dumay & Gaskell (2007), we expected improved recognition and recall for participants in the sleep condition. Other participant properties that I examined were sex, handedness, age, and education level These other participant properties never reached significance in my analysis, and therefore will not be discussed below.

Word was included as a second random-effect factor, crossed with *participant*. The between-words factor *word type* distinguished between acronyms, blends, and derived words. As mentioned above, I expected acronyms to be the most difficult words to recognize and to recall, and derived words the easiest. This is because for acronyms, the meanings of its components are indexed by only a single letter of the corresponding orthographic forms. For blends, the component words are phonologically transparent to a greater extent. The components (constituents) of derived words are phonologically fully transparent.

In addition to *word type*, I considered the structural variables *word length*, *component frequency*, and *neighborhood density*. Word length was quantified as the word's number of letters. Component frequency was defined as the average of the frequencies of the words contributing to the acronym, the blend, or the derived word. For acronyms, function words were not included in frequency counts, and for derived words, only base-word frequencies were used. Before calculating the mean, frequencies were log-transformed. Frequency estimates were based on Google web queries (August 2009). Word length and frequency entered into a moderate correlation (r =-0.45), as predicted by Zipf (1949). I hence de-correlated length from frequency by taking residuals of a model regressing length on frequency. The correlation of the original length measure with the de-correlated length measure was 0.89. Finally, I scaled and centered all numeric variables.

I also took into account the possibility that a word's orthographic neighborhood density would affect recognition and recall. Neighborhood density was operationalized using the OLD20 metric of Yarkoni *et al.* (2009). To remove the high correlation

between a word's OLD20 score and its length ($r = 0.83$), we de-correlated the OLD20 variable from length. As neither the original nor the de-correlated measure ever reached significance, this predictor is not discussed further.

To investigate potential transfer effects from the familiarization phase, we included as predictors the definition length, the example context length, and the position of the new word in the example context. None of these exposure variables proved significant for any response variable and so will not be discussed further.

The final set of properties that I brought into the model was the word's socio-contextual affordances as inferred from the answers to the questions from the familiarization phase. As the ratings for the different questions were somewhat correlated, especially for the *funny*, *party*, and *work* questions, we orthogonalized the ratings using principal component analysis (PCA). Table 3.2 lists the loadings of the six questions on the principal components (PCs). PC 2 strongly represents *familiarity*, the measure of subjective frequency; PC 3, the *educated* question; PC 4 contrasts *emotional* vs. *educated*, PC 5 contrasts *funny* vs. *work*, and PC 6 represents *party* vs. *work*.

	PC1	PC2	PC3	PC4	PC5	PC6
familiarity	0.17	0.88	0.31	0.20	0.25	-0.06
emotional	0.34	-0.28	0.46	0.68	-0.36	0.04
funny	0.47	-0.25	-0.29	0.23	0.66	-0.39
educated	0.31	-0.22	0.67	-0.58	0.24	0.09
party	0.54	0.09	-0.35	-0.06	-0.04	0.76
work	0.50	0.17	-0.19	-0.33	-0.56	-0.52

Table 3.2: Loadings of the six familiarization phase questions on the principal components of the elicited ratings.

I brought temporal dependencies in the recognition experiment under control by means of two additional predictors, *trial index* and *previous RT*. *Trial index* specifies the position of an experimental item in the experimental list and often captures effects of habituation or fatigue, respectively shorter or longer durations over the course of the experiment. *Previous RT* is included to bring local temporal correlations in the

timeseries of responses into the model. Inclusion of these two predictors tends to be sufficient to remove autocorrelational structure from the model's residuals (Baayen & Milin 2010).

For the recall experiment, which was administered shortly after the recognition experiment, I included the trial index in the recognition experiment as a covariate to probe for potential recency effects. That is, I expect better recall for words appearing late in the recognition experiment, i.e. for words with a large trial index. Participants' responses during the recognition experiment were also included as a covariate since positive judgements can lead to improved recognition later on (Craik & Tulving 1975); the same could also be true for recall when administered after recognition.

3.3 Results and discussion

3.3.1 Recognition

I report information for five response variables from the recognition experiment: first fixation duration, gaze duration, fixation count, response latency (RT), and recognition accuracy. Together these variables inform about initial processing (first fixation duration), overall complexity (gaze duration, fixation count, RT), ocular fatigue (fixation count), participants' judgements about the words (recognition), and complete exposure time (RT). Six trials did not elicit a response prior to timeout; these trials were discarded for all analyses. Analyses for RTs and recognition were based on all remaining datapoints.

Gaussian linear models require response variables (RT, first fixation duration, gaze duration) to be approximately normal. I therefore inspected the quantile-quantile plots for these variables. As is seen in Figure 3.2.1, a straightforward negative inverse transformation sufficed for the RTs. Average RT was 1177.7 ms (SD = 545). The first fixation data is bi-modal with a long right tail (Figure 3.2.2) and did not respond well to transformations. I discarded 308 fixations (9.4%) that were shorter than 160

ms, thereby removing the first mode. Furthermore, I also removed from the tail of the distribution fixations longer than 350 ms. The resulting distribution was more symmetrical (3.2.3) and resisted further improvement from transformations. For the resulting first fixation data, the average fixation duration was 234.9 ms (SD = 41.1).

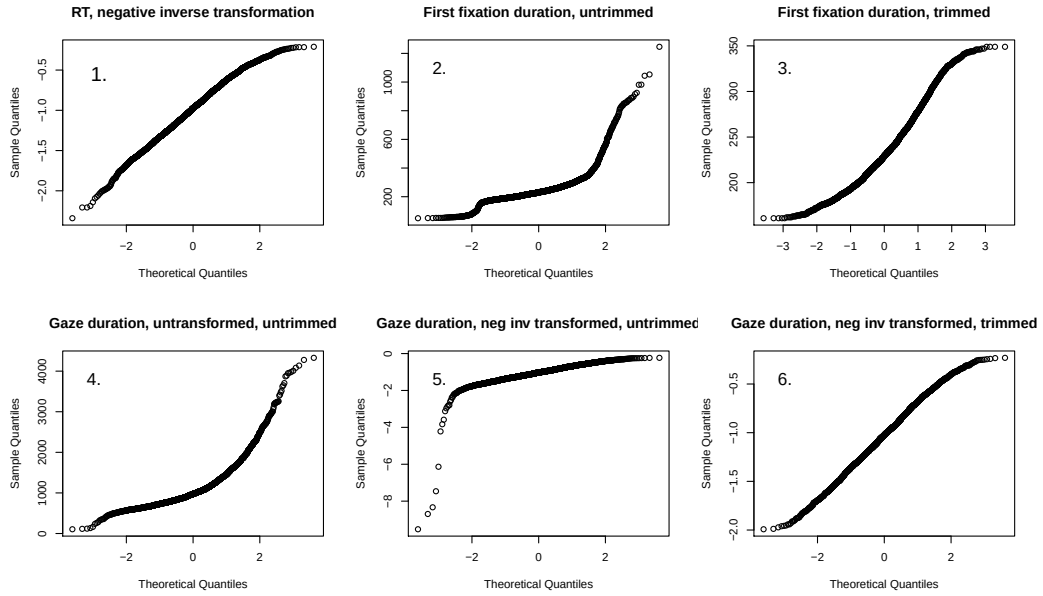

Figure 3.2: Quantile-quantile plots for numeric response variables RT, first fixation duration, and gaze duration.

Figure 3.2.4 shows that the gaze durations also did not follow a normal distribution. Upon applying the negative inverse transformation to the data (Figure 3.2.5), all but the left tail is normally distributed. This left tail is comprised of 32 trials (0.95%) with gaze durations shorter than 500 ms. Removal of these trials results in the approximately normal distribution seen in Figure 3.2.6. Failure to remove these atypical fixations from the dataset would result in distributions of the residual error that would closely resemble the quantile-quantile plots of the untransformed and untrimmed durations, indicating a systematic lack goodness of fit and violation of the fundamental assumption of normally distributed by-observation noise. The resulting gaze duration data has a mean of 1105 ms (SD = 478.4).

The fixation count is approximately Poisson-distributed (see Figure 3.3) and was

analyzed with a generalized linear model with a Poisson link function. As fixations shorter than 50 ms (3.3%) are probably misfixations that are not informative about the uptake of lexical information (Rayner 2009:p. 1460), these fixations were removed from the dataset. For the resulting dataset, the average number of fixations per word was 4.14 (SD = 1.89). The recognition and recall response variables were analyzed with generalized linear mixed models with binomial link functions.

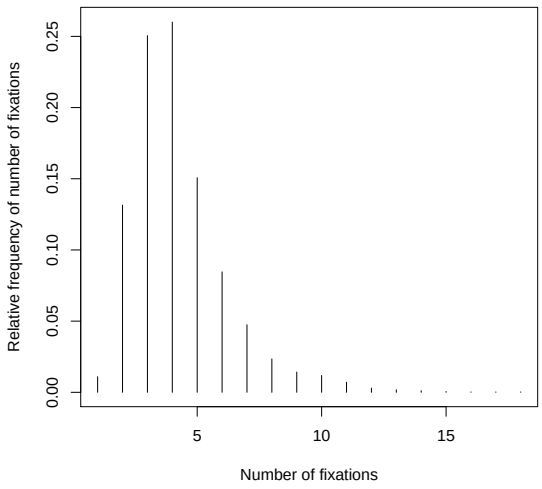

Figure 3.3: Distribution of fixation counts in the recognition experiment.

Fully 32% of fixations (4,459/13,933) were regressive, indicating that the target words were relatively difficult to read. By comparison, multi-morphemic Dutch compounds also shown in isolation (average length: 11.6 letters) studied by Kuperman *et al.* (2009) elicited a regression rate of 12.6%. The empirical probability density function for the number of fixations is given in Figure 3.3. The proportion of trials with a single fixation was small (0.011); the mode was at 4 (0.26), and the mean number of fixations was 4.14. Given an average length of 7.8 characters (SD = 3.90), the number of fixations for the present new words is large compared to the average of 2.6 reported for the above-mentioned Dutch compounds.

I used mixed-effects linear models to model the response variables (Baayen 2008)

using the predictor variables discussed in section 3.2.5 as fixed effects and covariates and with random intercepts for word and participant as crossed random-effect factors. I checked for random slopes and contrasts and report only those that reach significance. Random effects for each of the models are reported in Table 7.17 in Appendix 7.2. I used a stepwise variable elimination procedure with the aim of fitting minimally adequate, parsimonious models to the data. Outliers whose residual errors fell greater than 2.5 standard deviations from the mean were removed from the datasets and the models were refitted in order to ensure that effects are not due to a few atypical datapoints.

In what follows, I present findings for each significant predictor variable and covariate across all recognition response variables. Model summaries by response variable are reported in Appendix 7.2. Effects are visualized by response variable in Figures 3.4, 3.5, and 3.6. As the RT data patterned very similarly to the gaze duration data, the corresponding figure is not shown. Predictors in the model for recognition responses are a subset of those of the eye-tracking response variables, with similar effects, and also are not visualized separately.

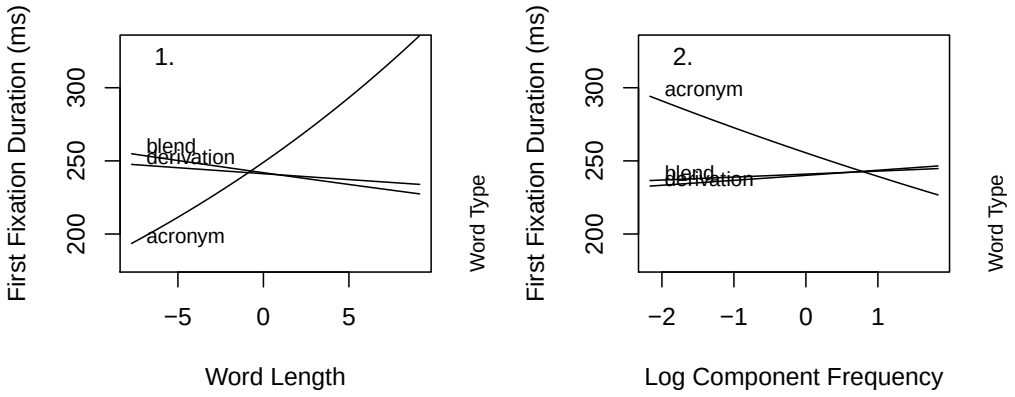

Figure 3.4: Partial effects of significant predictors for First Fixation Duration, with centered and scaled numeric predictors.

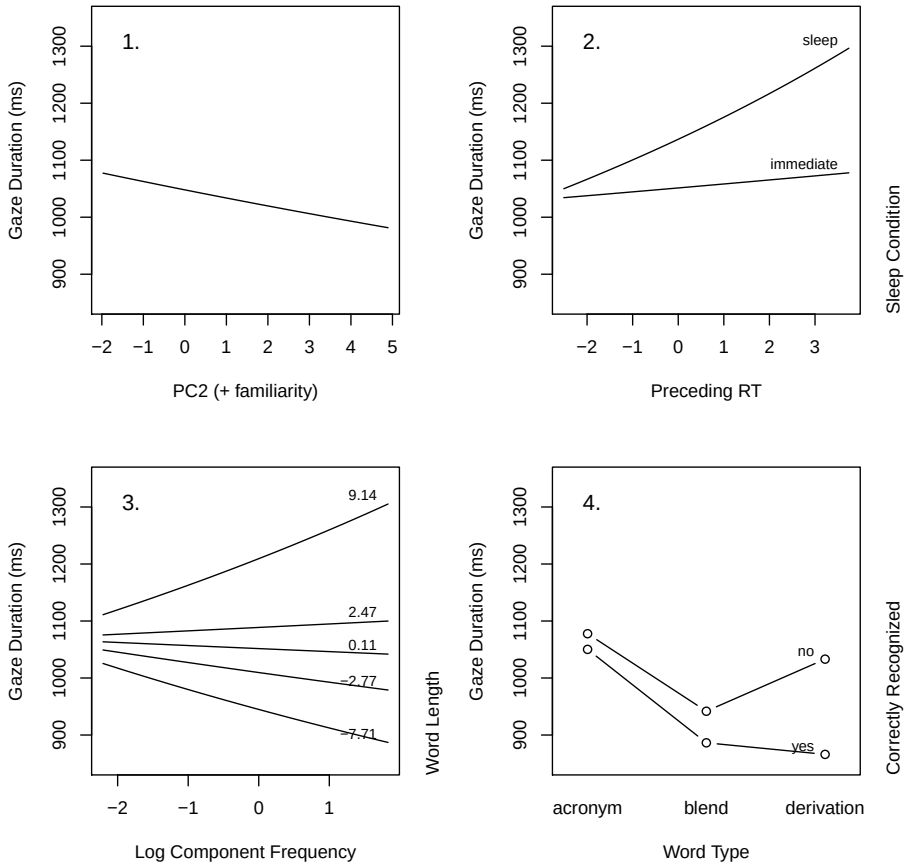

Figure 3.5: Partial effects of significant predictors for Gaze Duration, with numeric predictors centered and scaled.

Familiarity

The second principal component in Table 3.2 is highly loaded for the *familiarity* question, whether or not participants had seen the word previously. It emerges as significant in the analyses of the gaze durations and RTs. For both, greater familiarity predicts shorter durations (see Figure 3.5.1), as expected. This variable did not significantly predict participant recognition responses.

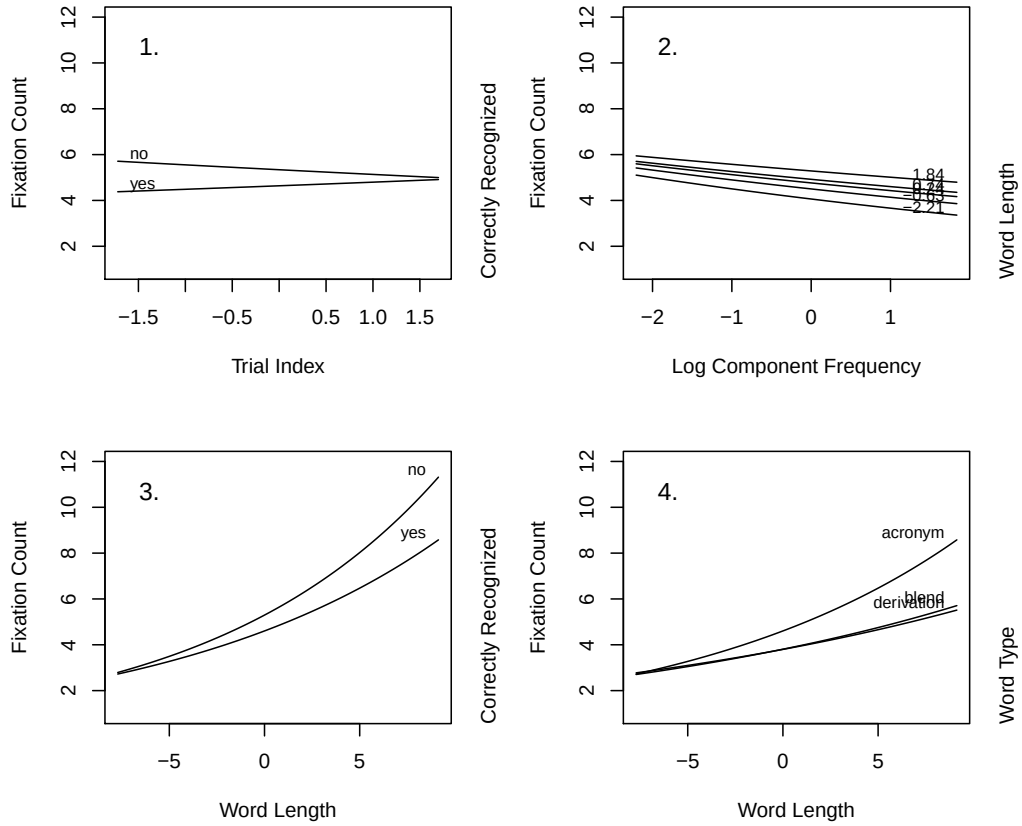

Figure 3.6: Partial effects of significant predictors for Fixation Count, with numeric predictors centered and scaled.

Temporal dependencies

There was a general effect of habituation throughout the experiment: participants' first fixations and RTs and became shorter as the experiment proceeded (negative correlation with *trial index*). However, in the gaze and RT data, by-subject random slopes for *trial index* were required (see Table 7.17). While most participants experienced facilitation to the task, resulting in a decrease in response and reading times throughout the experiment, some experienced fatigue or no temporal effect.

An interaction of trial index by recognition accuracy for the fixation count measure, shown in Figure 3.6.1, indicates that at the beginning of the experiment, incorrectly remembered words elicited more fixations than correctly remembered words.

By the end of the experiment, correct responses came with slightly more fixations, and incorrect responses with slightly fewer fixations, suggesting subjects reached a compromise between speed and accuracy for optimally meeting the task demands.

Along with habituation to the task came lower recognition accuracy: participants became less accurate as the experiment progressed (main effect of *trial index* for the recognition data; see Table 7.7).

Local temporal dependencies emerged in the RT data. For this response variable, a positive correlation was present between the RT at trial t and the RT at trial $t-1$. This correlation was especially strong when participants were deciding on a "no" response. In a separate analysis of both targets and fillers (model not shown), the smallest slope of RT at trial $t-1$ emerged for known words (targets), responded to by participants who completed the recognition experiment after the familiarization phase without an intervening night's sleep. This slope was significantly greater for unseen words (fillers) and also after sleep (both $p < 0.01$). The effects for familiarity and sleep were additive in this supplementary RT analysis. In the gaze duration model, it was only in the sleep condition that the previous RT reached significance (see Figure 3.5.2). Apparently temporal auto-correlations are stronger when retrieval from memory is difficult, as when a word had been encountered further back in the past (at least 12 hours prior to testing). Similarly, when retrieval is impossible, as for the filler words, temporal autocorrelations are stronger as well.

An effect of temporal dependence was also present in the recognition analysis. Trials that elicited a longer latency at $t-1$ gave rise to significantly more accurate responses at trial t (main effect of previous RT in the recognition model; see Table 7.7).

Word type

Figure 3.4 reveals that at the very first fixation, acronyms pattern differently from blends and derived words (see also Table 7.3). The left panel shows that longer acronyms required longer first fixations. Acronyms with more frequent components

elicited shorter fixations (right panel). No such effects in the first fixations were present for the blends or the derived words. This effect of component frequency indicates that processing of acronyms is co-determined by the words contributing to the acronyms. The third quantile of acronym length is four letters, indicating that most acronyms are entirely visible with just one fixation. Hence it is not surprising to see effects of length and frequency at the first fixation for acronyms. As we shall see, for blends and derived words, these effects occur at later processing stages.

This difference between acronyms on one hand and blends and derived words on the other persists in the gaze durations (see Figure 3.5.4). The exception to this pattern is derived words that were not correctly recognized. This is the only significant difference between blends and derived words in the recognition experiment. Apparently, participants found it difficult to classify derived words as unfamiliar. Since derived words are fully phonologically transparent, with well-known bases and productive affixes, they are so easy to interpret that it becomes difficult to classify them as unknown. For a similar effect for novel interpretable compounds, see Coolen *et al.* (1991). The same pattern of results also holds for the RT data (see Table 7.6).

In the analysis of fixation counts, again acronyms distinguished themselves from blends and derived words (see Figure 3.6.4). Unsurprisingly, all three word types elicited more fixations as word length increased. Interestingly, the inhibitory effect of word length was strongest for the acronyms, even though acronyms were on average the shortest words. This must be due to retrieval and interpretation of multiple component words, compared to the two components contributing to blends and derived words.

Acronyms were also much more difficult to accurately recognize compared to blends and derived words (see Table 7.7). The more components have to be retrieved, the greater the probability of a retrieval failure leading to a "no" response.

Furthermore, there was significant by-participant heteroskedasticity in the first fixation, gaze, and RT models for different word types (see Table 7.17). Inspection of the by-participant coefficients for the three word types revealed greater variability

for the acronyms compared to the blends and derived words in the gaze durations and RTs. For the first fixation durations, the coefficients for the derivations revealed somewhat elevated variability.

A potential factor affecting the processing of acronyms is their pronounceability. Some acronyms are pronounced as a series of letters spelled out (e.g. *FBI*); others can be read out (e.g. *NASA*); a few allow both pronunciation strategies (e.g. *ASAP*); finally, both strategies can be employed simultaneously. For example, the Multicultural Summer Research Opportunities Program (MSROP) at the University of Minnesota is referred to by several as the [ɛm.ɛs.rap]). My hypothesis was that acronyms more likely to be pronounced as full words provide better chunks for recognition and recall. For a remarkable related effect of the pronounceability of stock-ticker acronyms on initial stock prices, see Alter & Oppenheimer (2006).

To determine if this was indeed the case, a survey on the pronunciation of acronyms from the target stimuli list was given to a separate group of 34 participants. Participants were given each acronym, along with its example sentence from the learning phase for context, and were asked to rate how they would pronounce it on a Likert scale from 1 to 5. An answer of 1 indicated a full-letter pronuncation, while 5 indicated a full-word pronunciation. A response of 3 specified that either response was acceptable or that a combination of both strategies could be taken (e.g. *ASAP*, *MSROP*). Mean pronounceability ratings were calculated for each word and added as an additional predictor for new models fitted to the acronym data.

There was no pronounceability effect at the first fixation duration. However, words with higher pronounceability ratings elicited fewer fixations ($p = 0.0565$, two-tailed test), shorter gaze durations and RTs ($p < 0.001$ for both), and fewer recognition errors ($p < 0.01$; models not shown). The effects of length and component frequency persisted when pronunciation ratings were added to the model specifications, indicating that pronounceability is an independent factor contributing to acronym processing.

To preserve ecological validity, acronyms were presented in upper case, while

blends and derived words were presented in lower case. As described in Section 3.2.3, I reserved a block of 30 trials, 15 of which were previously seen fillers, at the end of the recognition experiment to test whether any differences between acronyms on one hand and blends and derived words on the other were due to presentation in different case (upper/lower). Models fitted for the same response variables to the 15 previously unseen words, seven of which were in upper case and eight of which were in lower case, revealed a significant effect of word length but not of case. Hence, it is unlikely that greater processing costs associated with acronyms are due to the acronyms being presented in upper case.

The acronyms in the present dataset are fairly visually distinctive, with lower bigram transition probabilities than the blends and derived words. Since greater distinctiveness enhances memory for cues (Murdock 1960, Nairne 2002, Nelson 1979), I might expect to see the greater syntactic complexity of acronyms offset by their orthographic and phonological distinctiveness. However, it appears that retrieval of a larger number of components and lack of pronounceability as a word have much greater weight than their distinctiveness.

Component frequency

I already discussed some of the effects of component frequency in the context of the effects of word formation type. In summary of the above, facilitation emerged for acronyms (see Figure 3.4.2) but not for blends and derived words at the first fixation. We see this effect because speakers are accessing the words represented by the letters in the acronyms, which for most acronyms are all visible in the first fixation.

In the analysis of the gaze durations, component frequency again significantly predicted processing, as is shown in Figure 3.5.3. Component frequency interacted with length such that increased frequency led to shorter gaze durations for shorter words but to longer gaze durations for longer words. A possible interpretation of this interaction is that a short, newly learned word requires fewer fixations, in which case the two components provide a better joint cue for memory retrieval. Uptake for

longer newly learned words, on the other hand, is staggered over time, with multiple fixations. This may have lead to a weaker joint cue for retrieval and to components that are more likely to be processed invididually.

Increased component frequency also implies fewer fixations (see Figure 3.6.2). However, component frequency was not a significant predictor of recognition.

Component frequency has been shown to affect recognition of compounds (Kuperman *et al.* 2009) and derived words (Kuperman *et al.* 2010). In each of these cases a component is also a morpheme, or even an independent word. In my data, however, the orthographic components of blends and acronyms are sub-morphemic, yet frequency effects still occur. Apparently, in the familiarization phase, participants have learned to associate the new word with its components. When re-reading these words, these components are accessed, giving rise to the component frequency effect. The assumption made here is that the locus of frequency effects is at the semantic level, as argued on independent grounds by Baayen *et al.* (2006).

It is interesting that partial form information such as *desti-* and *-nesia* is sufficient to activate the component meanings *destination* and *amnesia*, respectively, when reading *destinesia*. This indicates that a full parse of the visual input into a string of morphemes is not necessary for comprehension. A similar conclusion was reached by Bergen (2004) in a priming study of phonaesthemes (e.g. *gl-* to indicate *light* in *glimmer*, *glow*, *glisten*, etc.), which, as for affixes, showed a priming advantage for phonaesthemes above and beyond phonological and semantic similarities. These results fit well with linguistic theories of construction grammar (Goldberg 2006, Dabrowska 2009, Jackendoff & Pinker 2005) and with the experimental results of Bowers *et al.* (2005).

Word length

Word length consistently entered into interactions with other predictors and was always inhibitory. As mentioned above, longer acronyms elicited significantly longer first fixations (see Figure 3.4.1), but no effect was present for blends or derived words.

Longer words also elicited longer gaze durations (see Figure 3.5.3). As discussed above, this effect is most pronounced for words with higher component frequencies. Likewise, longer words elicited more fixations, especially for acronyms (see Figure 3.6.4). Additionally, longer words that were incorrectly recognized had more fixations than longer words that were correctly remembered (3.6.3). Longer words also had significantly greater by-participant heteroskedasticity for the first fixation and gaze durations as well as for RTs (see Table 7.17). Finally, participants were more likely to accurately remember longer words than shorter words (see Table 7.7). Here we do see that more distinctive cues are remembered better.

Response

As expected, no effect of the response made was present in the first fixation durations. A participant's response did affect later measures, however. When deciding on a "no" response, participants had longer gaze durations than when deciding on a "yes" response. As mentioned above, this effect is most pronounced for derived words (see Figure 3.5.4), which have components that are fully present in the visual input. A very similar pattern was present in the RT data (figure not shown).

Sleep

An intervening night's sleep was a significant predictor only for gaze duration, interacting with *previous RT*. With a longer response latency at trial $t-1$, an intervening night's sleep lengthened participants' gaze durations at trial t (see Figure 3.5.2). We showed above that temporal auto-correlations throughout the recognition experiment are greatest when retrieval from memory is more difficult. Temporal auto-correlations are highest not only after a night's sleep, but also when making a "no" response to previously unseen fillers or previously seen targets. As a consequence, it is unlikely that the longer responses following a night's sleep would indicate deeper lexical processing made possible by lexical consolidation during sleep.

3.3.2 Free recall

Shortly after the recognition experiment, participants were asked to freely recall all the words they remembered from the familiarization phase. A word was considered correctly recalled only if its form provided an exact match to a word in the familiarization phase, modulo minor spelling errors. Under these criteria, *un-uptradeable* for *untrade-upable* was excluded, but *cannibista* for *cannabista* was retained.

Free recall was modeled with mixed-effects logistic regression. Significant partial effects are visualized in Figure 3.7. There was no effect of component frequency or word length, so these predictors are not shown or discussed further.

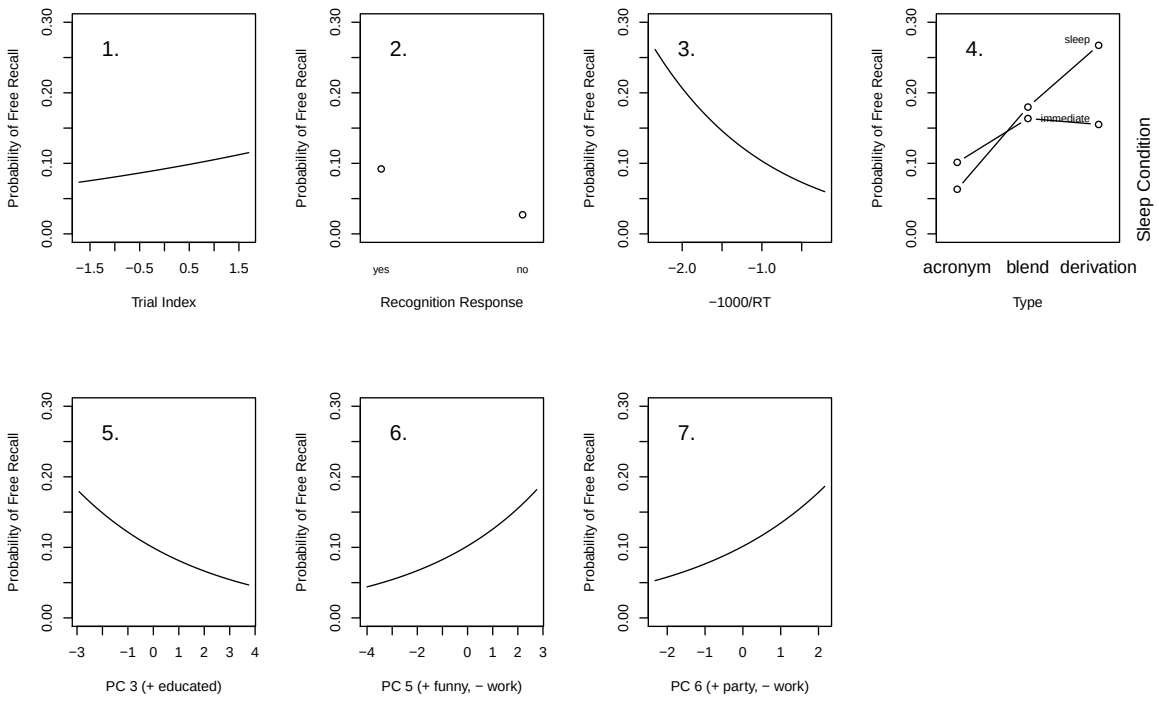

Figure 3.7: Partial effects of significant predictors for Probability of Free Recall, with numeric predictors centered and scaled.

Recency

Since the recall experiment was administered shortly after the recognition experiment, effects of recency are expected. Specifically, words that were presented closer in time to the recall experiment should be easier to recall. Since the variable *trial index* specifies a word's position in the experimental list during recognition, it can be used as a measure of recency for the recall experiment. Words with a higher trial index were seen more recently. Unsurprisingly, words with a higher trial index were more likely to be recalled, irrespective of word type (see Figure 3.7.1).

Previous positive decisions about an experimental item have been shown to afford improved recognition (Craik & Tulving 1975). I therefore included participants' decisions in the recognition experiment about whether or not they had seen this word in the familiarization phase as a predictor in the recall model. Indeed, positive responses were indicative of improved recall (see Figure 3.7.2), although there were still some instances in which words were freely recalled but not recognized.

Accessibility

A word's RT (or equivalently, a word's fixation count) from the recognition experiment was included as a predictor (see Figure 3.7.3). When either of these variables is included in the model, a word's length does not significantly predict recall. Time spent on a word in the recognition experiment was negatively correlated with free recall. The increased time spent making a judgement about an experimental item in recognition is an index of retrieval difficulty, and as such is consistent with increased problems in recall. Furthermore, we can conclude that the increased time spent making a judgement about an experimental item subsequently does not afford superior recall (see also Craik & Tulving (1975)).

Word type

Acronyms were recalled reliably less than blends and derived words. An intervening night's sleep aided recall, but only for derived words (see Figure 3.7.4).

For acronyms, I again included mean pronunciation ratings as a predictor for recall, but pronounceability did not reach significance.

Socio-contextual affordances

Unlike in the recognition experiment, in free recall the principal components of the socio-contextual affordance questions administered during the learning phase yielded significant main effects. In Figure 3.7.5, we see that PC 3 negatively impacts free recall – this principal component is highly correlated with the *educated* question. On the other hand, PC 5 positively impacts recall; this component is highly correlated with positive responses to the *funny* question and is negatively correlated with *work*. Finally, PC 6 was positively correlated with responses to *party* and again was negatively correlated with *work*; this predictor significantly improves recall.

These results indicate that socio-contextual affordances such as sounding educated or being appropriate for use in a work or school context render a word less likely to be entrenched in memory. On the other hand, well-recalled words are those usable in humorous contexts, with one's friends, at parties. In other words, words with affordances for use with other speakers in one's close social networks are more likely to be freely recalled. The more formal a social and/or topical context, the more free recall will suffer. A social or topical context that is less formal, however, will improve free recall.

3.4 General discussion

This research investigated how new, morphologically complex words are remembered and how lexical representations for these words are formed. Factors hypothesized to

affect memory included the word's internal structure (contrasting acronyms, blends, and derived words), consolidation of a memory trace due to hippocampal transfer during non-REM sleep (Rudoy *et al.* 2009), and a word's socio-contextual affordances. A recognition experiment with eye-tracking followed by a free recall experiment provided evidence that all these factors co-determine how a word is remembered, with a strikingly different pattern for recognition and recall.

Table 3.3 summarizes my findings, with columns 2–4 representing recognition findings and column 5 giving results for recall. The structural predictors were predictive mainly for recognition: longer words elicited more and longer fixations; words with more frequent components elicited more and longer fixations, but this effect is reversed for shorter words; words judged to be more familiar elicited shorter gaze durations; compared to blends and derived words, acronyms elicited more fixations and longer first fixation and gaze durations. By contrast, the only structural predictor that emerged for free recall was word type: acronyms had lower recall scores.

An intervening night's sleep led to stronger temporal auto-correlations in the recognition experiment, and afforded higher scores for derived words in the free recall experiment.

The predictors relating to a word's socio-contextual affordances are present only in the free recall experiment. Words that are judged as funny and appropriate for a party context were recalled better. By contrast, words evaluated as usable at work and characteristic of educated speakers were poorly recalled. This gradient of formality correlates with two factors, both of which may have contribued to better recall. First, a pragmatic affordance is involved: some of the stimuli are a bit humorous or slightly risqué, which makes these words more appropriate for use at a party than at work.

Second, a social affordance is involved, namely, usability as depending on the strength of social ties between speakers. Social ties at work tend to be weaker than the social ties between friends and family (Verbrugge 1979, Granovetter 1980, Gans 1962/1982). Interestingly, there is a cognitive correlate to the strong-tie/weak-tie distinction. For existing words, participants are able to better remember words they

	First Fixation Duration	Gaze Duration	Number of Fixations	Free Recall
		Structural variables		
Word Length	Longer for acronyms	Longer for frequent words	Increased for frequent words, acronyms, and unremembered words	
Component Frequency	Shorter for acronyms	Shorter for shorter words; longer for longer words	Shorter for shorter words; longer for longer words	
Familiar		Shorter		
Word Type	Longer for longer acronyms	Longer for acronyms, unremembered derived words	More for longer acronyms	Decreased for acronyms
		Sleep condition		
Sleep intervening		Stronger temporal auto-correlations		Increased for derived words
		Socio-contextual variables		
Educated				Decreased for educated words
Funny vs. Work				Increased for funny words
Party vs. Work				Increased for party words

Table 3.3: Summary of significant effects in recognition (columns 2 - 4) and free recall of new words.

use to describe themselves and their mothers than words with which they describe an unknown person (Bower & Gilligan 1979). Bower and Gilligan conclude that relating a word to a well-differentiated cognitive structure improves memory. A speaker's memory will have well-differentiated cognitive structures for her friends; these friends also comprise strong ties in her social network. Both research on social network theory (Rogers 2003, Granovetter 1973, Milroy & Milroy 1985, Milroy & Milroy 1992, Labov 2001) and cognition therefore predict that the words in the present study are more likely to gain currency in the strong-tie, interlocking social networks of friends than in the more formal, weak-tie social networks at work.

Comparing the recognition and free recall experiments clarifies that the former taps straightforwardly into language comprehension. Unsurprisingly, longer words take longer to comprehend, and words with more frequent components are more easily understood. These effects are typical of processing during reading of orthographic and possibly phonological information, which are initially encoded in episodic memory (Dumay & Gaskell 2007, Gaskell & Dumay 2003, Tulving 1972, Craik & Tulving 1975). In a levels-of-processing framework (Craik 2002, Craik & Lockhart 1972), the representations accessed in the recognition experiment are maintained on the lower levels of the hierarchy of cognitive representations. On a cline from general to specific (Bybee & McClelland 2005), this information is more specific.

Free recall, on the other hand, taps more into the general conceptualization processes that are also involved in active speech production. It is here that we see higher-level knowledge (in the sense of Craik (2002), Craik & Lockhart 1972) at work, such as intuitions about what kind of people would use a new word and in what social contexts this new word could be used appropriately. These higher-level factors do not play any role in the recognition experiment.

In summary, when we re-encounter a new word, what information in memory is of initial, primary importance depends crucially on whether we are reading or speaking. When reading, we access the memory traces of the components, and possibly the memory trace of the complex word itself. When speaking, the primary information

accessed is the word's socio-contextual affordances, which, in the speech production model of Levelt (1989), would be part of the conceptualization process.

Across recognition and recall, acronyms posed greater challenges to our participants. This disparity is due in part to the greater semantic complexity of acronyms, which denote phrases rather than single words: If you are a *BITGOD*, you reminisce about times Back In The Good Old Days. It is also partially due to the difficult phonotactic sequences in acronyms, as acronyms that are more likely to be pronounced as words are recognized better than those that are more often pronounced as a series of letters. Pronounceability played no role in predicting free recall, however, consistent with my hypothesis that in free recall, higher-order information instead of low-level information is of crucial importance.

The facilitatory effect of component word frequency indicates that speakers are consistently accessing the components of words, regardless of word type.

Compared with the acronyms, the blends pattern surprisingly similarly to the derived words. The two differences between blend and derived words that we were able to detect are that unremembered derived words, and not unremembered blends, have longer gaze durations, and that only derived words benefitted in free recall after a night's sleep.

Evidence of sleep-associated consolidation for new lexical items was modest compared to the studies of Gaskell & Dumay (2003), Dumay *et al.* (2004), and Dumay & Gaskell (2007). There are several possible reasons the present results are more restricted. First, my stimuli were meaningful, morphologically complex, and interpretable new words, some of which are actually coming into use, rather than invented monomorphemic word forms such as *cathedruke*. Second, my study probed for sleep effects targeting the learning of a word's meaning and use in social contexts. Participants were therefore presented with the new word, its definition, and an example in context; they were then asked several questions about the word's potential use. In contrast, the studies by Dumay and colleagues targeted the phonological learning of word forms which were presented repeatedly to participants, in the majority of their

experiments without being supplied with an interpretation.

The present study did reveal two sleep-related effects, though. First, sleep apparently strengthens the between-trial auto-correlational processes as the recognition experiment unfolds. This effect most likely reflects an increased difficulty in retrieving memory traces of the words' meanings established a day ago rather than consolidation in memory, as the auto-correlational processes were also stronger for previously unseen filler words.

Second, as mentioned above, an intervening night's sleep improved free recall, but only for derived words. The derived words in my study are orthographically and phonologically transparent and for the most part semantically compositional. If a reader knows the meaning of *innuendo* and the meaning of *-ish*, she can readily understand *innuendo-ish* as meaning "suggestive of an innuendo". Blends and acronyms, by contrast, are less phonologically transparent and less semantically compositional: without proper context, a speaker will not glean that the components of *tonorrow* are *tonight* and *tomorrow*. Even if the components are correctly inferred, the specific meaning ("tonight *or* tomorrow") requires further explanation. The present results suggest that semantic compositionality and perhaps phonological transparency are pre-requisites for effective sleep-associated consolidation.

Interestingly, the blends almost always patterned with the derived words in the recognition experiment. This finding harmonizes well with the possibility of parts of words, such as the *holic* in *alcoholic* to become a productive affix-like formative in words such as *workaholic* and *textaholic*. This suggests some fluidity between the word formation processes of blending and derivation. However, the advantage that the much more regular process of derivation affords for free recall may help explain, from an evolutionary perspective, why derived words are so much more common than blends and acronyms, and why this type of word formation is so well entrenched in the world's languages.

Chapter 4

The other meaning of Ph.D.: Learning words through hiphop music

Music listeners have difficulty correctly understanding and remembering song lyrics. However, results from the present study show that young adults can learn African-American English (AAE) vocabulary from listening to hiphop music. Non-African-American participants first gave free-response definitions to AAE vocabulary items, after which they answered demographic questions as well as questions addressing their social networks, their musical preferences, and their knowledge of popular culture. Results from the survey show a positive association between the number of hiphop artists listened to and AAE comprehension vocabulary scores. Additionally, participants were more likely to know an AAE vocabulary item if the hiphop artists they listen to used the word in their song lyrics. Together, these results suggest that young adults can acquire vocabulary through exposure to hiphop music, a finding relevant for vocabulary acquisition research as well as for models of adoption of linguistic innovations.

4.1 Introduction

In 1979, *Rapper's Delight* became the first hiphop song to receive national radio play in the U.S. For many outside of New York City, *Rapper's Delight* was a first exposure to hiphop. Couched in all the novelty was a non-mainstream vocabulary, much of which was African-American English (AAE). AAE words used in *Rapper's Delight*, such as *fly* ("cool/attractive") and *bad* ("cool"), subsequently enjoyed some prominence in Mainstream American English (MAE) throughout the 1980s. Could it be that non-African-American speakers learned these words through listening to hiphop songs such as *Rapper's Delight*?

Possibly. A speaker's vocabulary grows dramatically during adolescence and young adulthood (Anglin 1993); these are also times in which various media forms are used for socialization purposes (Arnett 1995). Perhaps it is not surprising that, given enough visual and linguistic context, speakers appear to acquire vocabulary from watching movies or television shows (see Chambers 1999).

It is less clear, however, that vocabulary acquisition can take place through listening to music, particularly hiphop. Listeners often misunderstand lyrics: they were seven times more likely to incorrectly transcribe sung words than spoken words (Collister & Huron 2008). Thus it is not suprising that music listeners have only a thematic understanding and memory of lyrics (Hansen & Hansen 1991, Greenfield *et al.* 1987). Furthermore, several barriers make it difficult to adequately understand hiphop lyrics in particular (Marc Martínez 2010):

- The presence of background music and samples;
- The fast pace of many rappers, often too fast for comprehension;
- The voice quality, which can be excited, shouting, or otherwise emotionally charged;
- Atypical language. Like other forms of verse, this includes atypical syntax and lexical items that better conform to verse structure. Hiphop is also rife

with double entendres and deliberately obscure language (Rickford & Rickford 2000:73). For speakers of MAE, the prevalence of specific African-American English (AAE) vocabulary in hiphop can make the lyrics even more difficult to understand.

These factors make for suboptimal conditions for comprehension of lyrics, not to mention vocabulary acquisition. A well-researched example of vocabulary acquisition under sub-optimal conditions concerns hearing-impaired populations, who have lower vocabulary acquisition rates than their non-hearing-impaired counterparts (Pittman *et al.* 2005). Similarly, it seems difficult, especially for non-African-Americans, to acquire vocabulary as a result of listening to hiphop.

On the other hand, adolescents and young adults listen to quantitatively more music than previous generations (Arnett 1995). Due to the increasing ubiquity of mp3 players, smartphones, and internet connectivity[1], these groups receive increased linguistic input from popular culture icons such as music artists. From its South Bronx roots and subculture origins, hiphop music has evolved to be fully mainstream (Chang 2005); many adolescents and young adults are regularly exposed to it. In listening to the same songs more, listeners benefit from repeated learning, enabling them to better process details. Furthermore, the availability of videos online enables immediate video viewing, which could facilitate vocabulary acquisition by offering visual context accompanying unclear lyrics whenever listeners want to watch. Finally, websites devoted to slang, such as Urban Dictionary (`http://www.urbandictionary.com/`), and to hiphop/AAE vocabulary in particular, such as Rap Dictionary (`http://www.rapdict.org/`), allow for explicit querying of words with unclear meanings. Widespread listening to particular artists using the same words could lead to large-scale vocabulary acquisition across social groups.

The context of AAE in the U.S. is ideal for testing vocabulary acquisition through music listening in young adults. AAE has regular linguistic features, including vo-

[1]For example, 79% of American teens and three-quarters of 18-24 year-olds have an mp3 player (Lenhart *et al.* 2010).

cabulary differences, that make it a legitimately distinct variety on par with Received Pronunciation (RP, the prestige variety in the UK) or MAE. Due to continued segregation patterns for African-Americans (Massey & Dalton 1989, Massey 2004), many MAE speakers might rarely interact with speakers of AAE. When they do, it is possible that via linguistic accommodation (Giles & Smith 1979, Pickering & Garrod 2004), AAE speakers do not use full-fledged AAE (Wolfram & Thomas 2002). On the other hand, as the prestige variety amongst hiphop artists (Alim 2004), AAE is often used in hiphop lyrics. Hiphop could thus represent a primary means of exposure to AAE vocabulary for many MAE speakers.

African-Americans have a long musical tradition in the U.S.; some prominent forms are spirituals, jazz, blues, soul, R&B, and hiphop. Each of these genres has highlighted AAE, but its use is particularly central to hiphop (Alim 2004). Table 4.1 gives example AAE words and phrases used by African-Americans in songs. Some of these lexical items are understood and have been adopted to varying extents by speakers of MAE at various points. Others are no longer used by AAE speakers. The following experiment provides evidence that comprehension vocabulary of items such as *player hater* and *Ph.D.* (see Table 4.1) stems in part from listening to hiphop.

Lexical item	Artist	Song title	Year
Straighten up and fly right "stop playing around"	Nat King Cole	"Straighten Up and Fly Right"	1944
Funky "soulful"	Sly and the Family Stone	"Music Lover"	1968
Player "promiscuous person"	Junior M.A.F.I.A. (H)	"Player's Anthem"	1995
Hater "jealous/resentful person"	Jay-Z (H)	"Izzo (H.O.V.A.)"	2001
Ph.D. "player hater degree", figurative degree held by a very jealous person	Kanye West (H)	"So Appalled"	2010

Table 4.1: African-American artists' songs often feature AAE lexical items, with (H) denoting hiphop artists. Words and etymologies taken from Smitherman (2000).

To see if speakers might be learning AAE vocabulary from hiphop, I studied speak-

ers' comprehension vocabulary of AAE by asking them to give definitions for AAE lexical items. Because these words are not (yet) well-known by speakers of MAE, it is likely that familiarity of these words will have some link back to the AAE speech community, either from social ties or from exposure to popular culture including hiphop music. I hypothesized that a preference for hiphop music would be positively associated with participants' AAE comprehension vocabulary. If however the communication channel is too noisy for vocabulary acquisition, as previous research would suggest, then we would expect no effect of a preference for hiphop music on AAE comprehension vocabulary.

Several control factors were also assessed to determine whether a preference for hiphop was not masking any underlying variables that were actually responsible for AAE comprehension vocabulary. A series of demographic variables such as age, sex, ethnicity, and geographic area are important for other types of language use by adolescents and young adults (see, inter alia, Labov 2001, Eckert 2000, Trudgill 1983, Trudgill 1974). Second, social network variables such as strong and weak ties (roughly, close friends/family vs. acquaintances) are also important factors in the diffusion of linguistic innovations (Milroy & Milroy 1985, Milroy & Milroy 1992). Finally, it is possible that general popular-culture knowledge is responsible for AAE vocabulary knowledge. This possibility was assessed by asking participants five questions about popular culture that college-aged students could know.

4.2 Method

4.2.1 Participants, materials, design

One hundred sixy-eight undergraduate students in introductory-level undergraduate courses in music, linguistics, and sociology at the University of Minnesota participated in this online survey for extra credit. The stimuli, available in Appendix 7.3.4, consisted of sixty-four vocabulary items obtained from at least one of three sources:

a native speaker of AAE; a dictionary of AAE (Smitherman 2000); or Rap Dictionary (`http://rapdict.org/`), an online source of AAE and other words used by hiphop artists.

4.2.2 Procedure

Participants were first given the vocabulary items, presented one at a time, to freely define; they were told beforehand that they were participating in a survey on AAE and then were asked, "If you heard this word in a slang context, what would it mean?" The order of presentation was randomized separately for each participant. Next, participants were asked their sex and age as well as free-response questions addressing their ethnicity, hometown (city/state or foreign country), musical tastes, social networks, and popular culture knowledge in that order. The two questions about musical preferences were (question labels in brackets):

1. What are your favorite genres of music? [Genres]

2. Name some of your favorite groups/artists in each genre (up to 5 per genre given above). [Artists]

The two questions about participants' social networks elicited ethnic information about participants' strong and weak ties:

1. Please list the first names of all the people you know that you would ask to help you move, along with their ethnicity (up to 15). [Move]

2. Please list the first names of any African-American friends/acquaintances you might have that you interact with on a weekly basis, either online or in person (up to 15). [Weekly]

Asking for first names of participants' strong and weak ties helps participants clearly enumerate friends, as opposed to asking for an imprecise estimate (Brewer 1997, Marsden 2005). The Move question elicits participants' strong ties, assuming that

most people would not feel comfortable asking acquaintances to help them move. The Weekly question relaxes the criteria for association, thereby eliciting weaker ties.

Finally, participants responded to five popular culture questions. These questions were selected because of their diverse domains (given below in parentheses) as well as for the differing degrees of prominence of their subject matter in popular culture. All questions also incorporate varying degrees of AAE:

1. (music) Is Jay-Z married, and if so, to whom? [Jay-Z]
2. (sports, TV) What network is Charles Barkley a commentator for? [Barkley]
3. (TV) Please name a character from *The Boondocks*. [Boondocks]
4. (TV, movies) Who is Mo'nique? [Mo'nique]
5. (music) Please list as many song titles by Justin Bieber as you can (up to 5). [Bieber]

4.2.3 Data processing

The response variable, AAE comprehension vocabulary, was transformed to a scale of 1 to 5 from free-response form (see also Marinellie & Johnson 2004 for a similar coding scheme). A value of 1 indicated no knowledge of the word; 76% of responses were given this value. A value of 5 was assigned if the definition corresponded to any attested AAE meaning, but not a correct MAE meaning. Another 15.8% of responses were given values of 5. Responses indicating partial knowledge (8.2%) were given values of 2, 3, or 4. With 64 items, a participant's summed vocabulary score could range from 64 to 320.

Two researchers independently coded the definitions given for all 64 words; the Krippendorff's alpha coefficient (Hayes & Krippendorff 2007) for inter-rater reliability was $\alpha = 0.96$. The corresponding figure for the responses that were not simply "not sure" (53.7% of responses) was $\alpha = 0.944$. The researchers then met to examine the

definitions they did not classify identically. Genuine disagreements were resolved such that the response variable reflects rater agreement.

The 2009 estimates for a participant's City, County, and county's African-American (CountyAA) populations were recorded using U.S. census data[2]. Together, these population variables serve as rough estimates of the number of unknown speakers participants could potentially have interacted with on a regular basis, as well as casual exposure to AAE.

Responses to the first four popular culture questions were transformed from free-response form to a scale of 1 to 5. Again, 1s indicated no knowledge of the answer, while 5s indicated complete knowledge. Answers indicating partial knowledge were given values of 2, 3, or 4. A binary predictor variable of knowledge/no knowledge was determined to better fit the Jay-Z and Boondocks variables. For the Bieber question, responses were 1 plus the number of songs correctly identified. Two raters independently evaluated responses to each of these questions; all had reliability values of $\alpha > 0.9$.

For the social network questions, if a social tie listed had two ethnicities, that person was counted as both ethnicities. I grouped social ties listed for the Move question into six ethnicities: African-American, Asian/Asian-American, Caucasian, Hispanic, Native American, and South Asian/Middle-Eastern/African. This last ethnicity is not standard but was used due to data sparsity. Mean values for each variable were used when no ethnicity data was given (n = 11 or 12 for Move questions and n = 3 for Weekly).

All artists listed were coded into one of nine possible genres, which were a subset of genres used in Rentfrow & Gosling (2003): alternative, classical, country, folk, international, pop, rock, vocal/jazz/showtunes/oldies and hiphop, the primary genre of interest. Participants were assigned numeric values for each genre corresponding to the number of artists they listed in that genre. Hence there were nine music-related

[2]Census data available at `http://www.census.gov/popest/cities/files/SUB-EST2009-ALL.csv`.

predictor variables. Responses to the Genres question were used to disambiguate between two or more artists with the same name but different genres.

In total, 860 unique artists were listed, with The Beatles being the most popular (listed by 25/168 participants). Further information about artists in the survey, including the genre classifications for all artists, additional specifications concerning classifications, and the number of times each artist was listed will be available online shortly.

4.3 Results

One African-American and 166 non-African Americans participated in the survey. One participant gave his or her ethnicity as "bi-racial" without specifying the two ethnicies. Therefore, data from this participant and the African-American participant data were excluded because these participants could be familiar with the stimulus items from their everyday lives.

Several models were fit to the data to ensure that the findings were robust to potential departures from normality. The model reported here, a linear regression model with the summed vocabulary score of individual participants as the response variable, is consistent with others reported in the Supplementary Materials.

The Hiphop, Country, Weekly, Barkley, Mo'nique, and Boondocks variables were significant in predicting a participant's AAE vocabulary score. All predictors were positively associated with AAE vocabulary knowledge except Country, the number of country music artists that a participant listed as listening to. The model in Table 7.14 has a multiple R^2 of 0.39. These results are visualized in the added-variable plots of Figure 4.1, which plots the partial correlations between predictor variables on the x-axes and the response variable on the y-axes after partialing out all other predictor effects. The slopes of these lines correspond to the estimates in the linear model.

Variables such as Sex, Age, and hometown population information (City, County, and CountyAA populations) were not significant predictors. No strong-tie variables

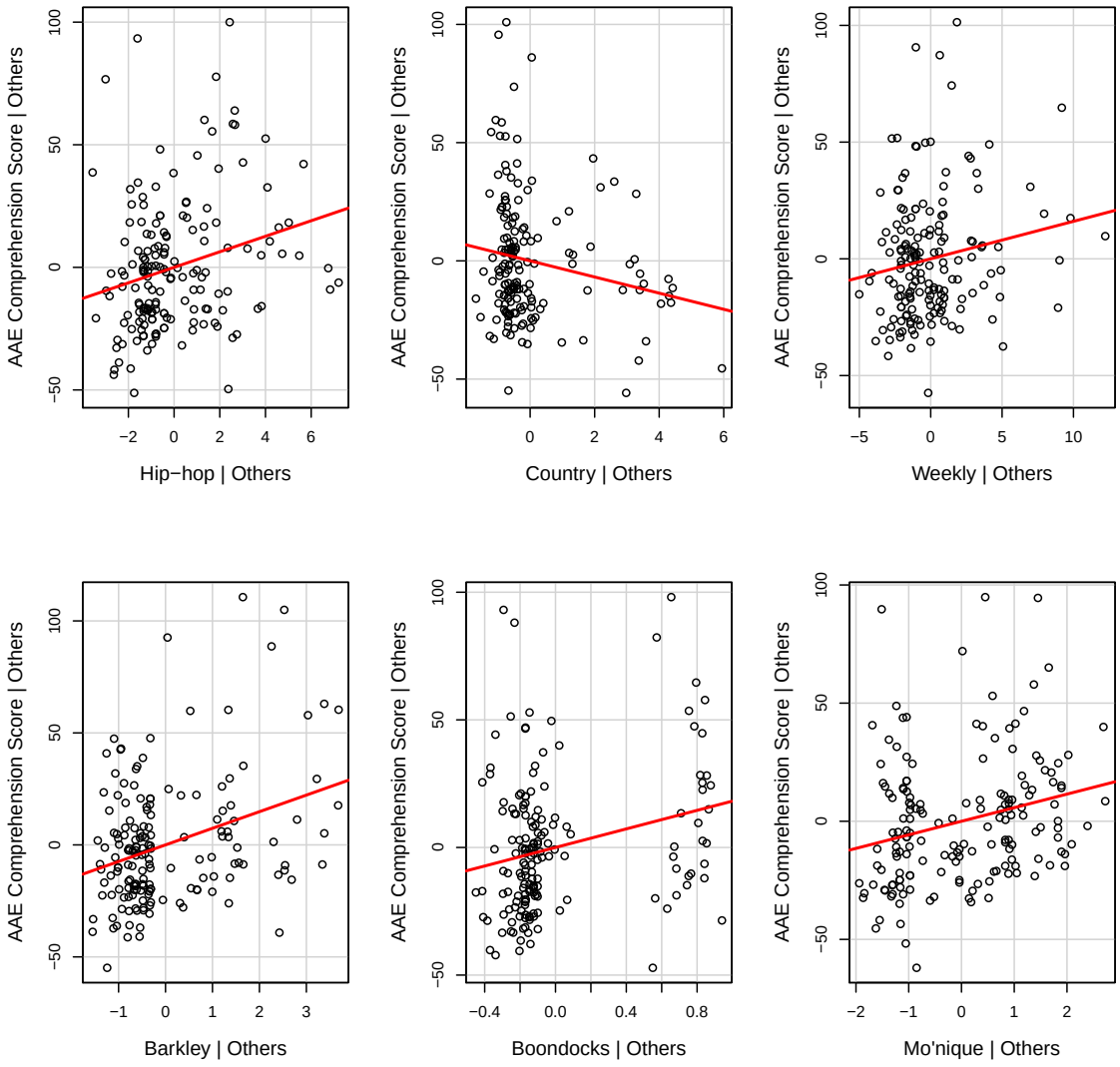

Figure 4.1: Added-variable plots of significant predictors for AAE comprehension vocabulary.

to any ethnicity were retained in the model.

To ensure that the effects were not due to multicollinearity, I inspected pairwise correlations between predictors; correlations were modest (Cohen 1992:138). For example, the Hiphop and Country variables were not signifcantly correlated ($r = 0.13, p = 0.09$), correlations between the popular culture variables retained in the

model, Barkley, Boondocks, and Bieber, were not above 0.16. Furthermore the highest correlation between predictors in the model (0.23, between the Hiphop and Weekly variables), was similarly low. The variables in the model also had low correlations with those not in the model, such that the former are not masking the explanatory power of the latter. For example, the Hiphop variable had correlations of 0.22, 0.23, and 0.17 with the hometown variables of City, County, and CountyAA populations, respectively.

To summarize, the number of hiphop artists a participant listens to was predictive of his or her AAE comprehension vocabulary. This effect was still significant when other factors, such as participant demographics, knowledge of popular culture, and overall musical preferences were taken into account.

4.3.1 The effect of preferred artists

To establish a firmer connection between speakers' actual linguistic input from hiphop music and their AAE comprehension vocabulary, it is necessary to examine the lyrics from specific artists participants listen to. To do this, the use of these words by specific artists participants listed was used as an additional variable to predict participants' knowledge of the words.

Eighty-one participants listed listening to at least one hiphop artist (48%; see Figure 4.2). I then did an automatic Google query for each hiphop artist listed together with each word on the Urban Lyrics website (`www.urbanlyrics.com`). Of the 92 total hiphop artists listed, 86 were present in the Urban Lyrics database. The lyrics each of these queries returned were manually examined to see if the usage of the word corresponded to the specific AAE meaning elicited.

The results for each word use by individual artists were then matched to participants' preferred artists such that for each word an estimate was made for the exposure of each participant to the word through hiphop music. Specifically, for a given word, the number of artists a participant listens to that use that word was used to predict

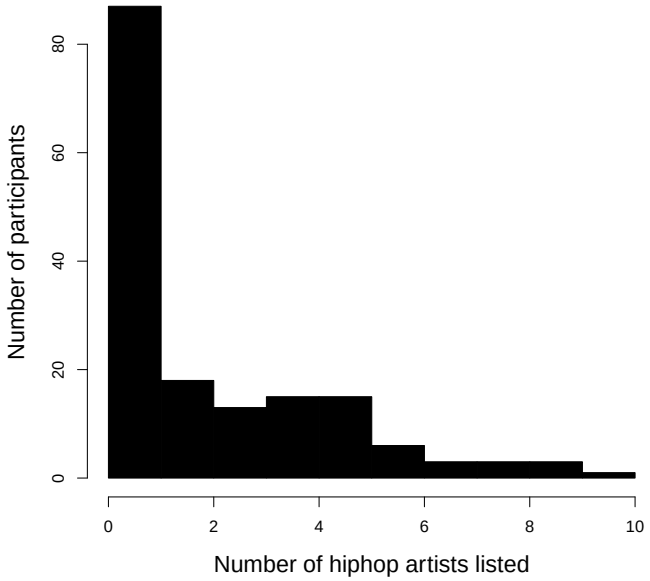

Figure 4.2: Summary of the number of participants who listed listening to hiphop artists.

the participant's familiarity with that word. The number of different contexts a word occurs in appears to be a better predictor of lexical processing and later entrenchment in a speech community than overall frequency counts (Adelman *et al.* 2006, Chesley & Baayen 2010, **?**); the same is likely true of comprehension vocabulary. Hence the number of artists was used as a predictor instead of word frequency counts.

Because this analysis required a separate treatment for individual participant-item pairings, a mixed-effects model, which allows for a distinct fit for each such pairing, was used. This new variable was added to a mixed-effects model that included as fixed effects all significant predictors from the previous study. Random effects of Participant and Word were included as crossed random-effect factors.

A possible confound in this study is the overall number of hiphop artists using a word. Because the number of contexts in which a word appears is predictive of both lexical processing and later entrenchment in a speech community, the number

of hiphop artists using a word could also be predictive of comprehension vocabulary. For example, a participant might not listen to Lil Wayne, who uses *fetti* ("money") more often than other artists in the present data. However, she could learn *fetti* in talking to friends who do. Additionally, the word's use in hiphop could be indicative of its general prominence in MAE. Finally, the number of preferred artists using a word could be highly correlated with the number of overall artists using the word, so any effect from the former could be due to the latter. Hence I also examined the number of times a word was used by all hiphop artists listed by all participants and included in the database (86).

A total of 42 of the 64 vocabulary items were detected in at least one hiphop artist's lyrics with the intended meaning. The predictor variables Barkley and Weekly did not reach significance. The predictors Hiphop, Mo'nique, Boondocks, and Country were still significant, with similar estimates. The number of overall artists using the word had a robust, positive association with vocabulary scores ($t_{(1)} = 6.72, p < 0.01$). In addition, the predictor variable of number of preferred artists using the word was positively associated with increased African-American English comprehension vocabulary scores ($t_{(1)} = 2.45, p = 0.014$). These two variables were moderately correlated ($r = 0.51, p < 0.01$). When a new, independent predictor for preferred artists was created by using the residuals of a model in which preferred artists was regressed on overall artists as the preferred artists variable, the number of preferred artists using a word was still significant. The effects of overall and preferred artists on AAE vocabulary score are visualized in Figure 4.3.

This analysis revealed that use of a particular AAE word by a participant's preferred hiphop artists was positively associated with the participant's knowledge of that word. This effect remained significant when the number of hiphop artists listed by participants and the overall number of hiphop artists using a given word were included as predictors.

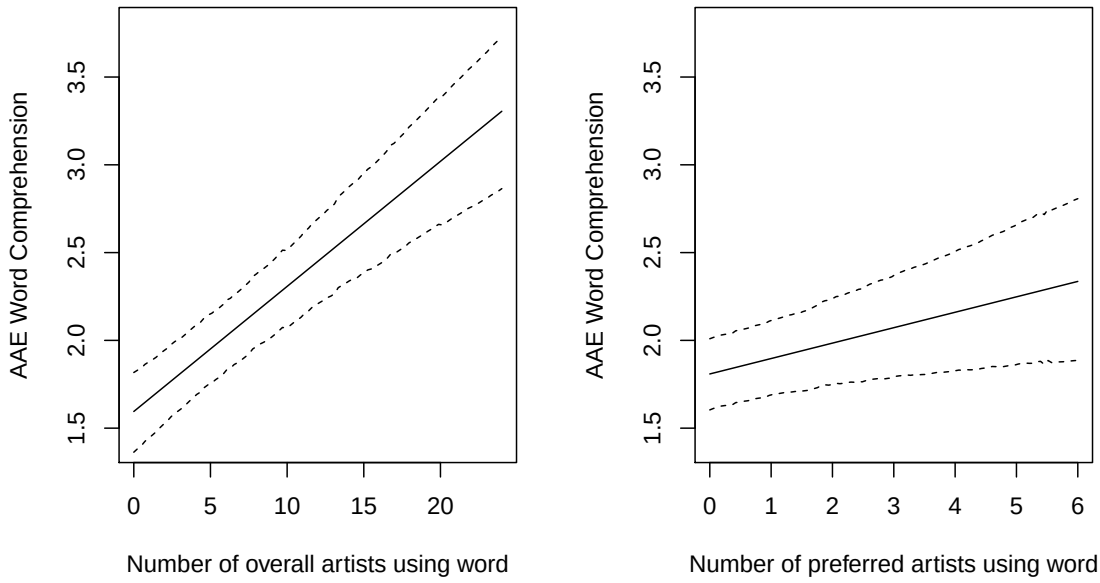

Figure 4.3: Partial effects of numbers of overall (left) and preferred (right) artists using a word on AAE vocabulary score. The dashed lines represent the 95% highest posterior density credible intervals.

4.4 Discussion

These results support the hypothesis that non-African-American young adults learn African-American English (AAE) vocabulary through listening to hiphop. A positive association was observed between AAE comprehension vocabulary and the number of hiphop artists participants listened to. Additionally, use of a particular word by a participant's preferred artists was predictive of the participant's knowledge of that word. This vocabulary acquisition takes place despite a suboptimal communication channel: in previous studies, music listeners made far more transcription errors on sung words than their spoken counterparts (Collister & Huron 2008), and they have only a broad, thematic understanding of song lyrics (Hansen & Hansen 1991, Greenfield *et al.* 1987) Hiphop listeners are learning vocabulary from linguistic input that is poetic, emotionally charged, accompanied by music, and from another variety of

English.

This hypothesis holds when other possible means of vocabulary acquisition, such as social ties, general popular-culture knowledge, and overall musical preferences are taken into account. Knowledge of African-American popular culture also appears to be associated with increased AAE comprehension vocabulary. In addition, the number of weak ties to African-Americans is a better predictor of AAE vocabulary knowledge than the number of strong ties, a finding which replicates the work on the strength of weak ties in diffusion of information starting with Granovetter (1973). Finally, a preference for country music was negatively correlated with AAE comprehension vocabulary, yet this predictor was not negatively correlated with a preference for hiphop.

Most work on vocabulary acquisition from the media examines vocabulary acquisition in infants or non-native speakers, often in a pedagogical setting. If explicit attention is directed to the media, L2 learners are able acquire vocabulary from that source (Koolstra & Beentjies 1999, Medina 1990). On the other hand, infants do not learn much vocabulary from baby videos (DeLoache *et al.* 2010). This study presents related results: young adult speakers of a different variety of the same language do appear to have enough context, linguistic and perhaps also cultural, to learn from musical input.

Another type of work on vocabulary acquisition from the media makes claims that appeal to common sense. Chambers (1999) observes that buzzwords are clearly propagated by the media and are integrated into the passive and active vocabularies of at least some speakers. Examples of this phenomenon in the U.S. include trendy memes such as *bi-winning*, Charlie Sheen's recent coining, as well as more enduring terms from the George W. Bush administration including *death tax* and *climate change*. Yet without an empirical study also examining social networks, such as the one presented here, it is impossible to tell whether this vocabulary is learned directly from the media for all speakers who know them, or if they pass first from the media to influential speakers and then to other speakers (see Labov 2001, 356-357). Sec-

ond, the words examined here are not buzzwords or memes; rather, they stem from a non-mainstream variety of English. Whereas buzzwords or memes generally have a clear instance of coining, the same is not necessarily true for AAE vocabulary. In sum, further theoretical and empirical studies are needed on vocabulary acquisition through the media.

Recent work shows an association between Scottish speakers who watch the London-based TV series *East Enders* and the adoption of phonetic features of Southen English (Stuart-Smith 2006). Similarly, since speakers' comprehension vocabularies are influenced by hiphop, it is possible that speakers will subsequently *use* AAE vocabulary they hear in hiphop songs. AAE is not the prestige variety throughout the U.S. and is perceived as indicative of lower socio-economic status than MAE (Green 2002:223). However, its *covert prestige* (Labov 2006) in the AAE and MAE speech communities also means its non-standard nature is perceived as "cool" (Bucholtz 2001). This could make MAE speakers more motivated not only to learn, but also to use, AAE vocabulary they hear when listening to hiphop.

Previous linguistic theories posit that the media only influences a small number of speakers, who then pass on linguistic innovations to the majority of speakers in a community (Labov 2001:356-357). In contrast, in this study, 48% of non-African-American participants listened to hiphop. Coupling this information with the present finding that young adults learn AAE vocabulary from listening to hiphop, as well as previous findings that adolescents use media as a form a self-socialization (Arnett 1995), it is possible that a larger percentage of the population than previously thought could be directly influenced from linguistic input they receive through the media.

Chapter 5

General discussion

This chapter focuses on a synthesis of the results in Chapters 2, 3, and 4. Chapter 2 showed that linguistic factors and contextual information can influence lexical entrenchment. Chapter 3 presented results supporting the hypothesis that contextual information can affect memory for previously unseen words. Chapter 4 showed that comprehension vocabulary is influenced by factors like social ties, popular-culture knowledge, and musical preferences. Together, these findings suggest that lexical entrenchment is predictable to a large extent, a finding previously not documented.

Implicit in Chapters 3 and 4 is the hypothesis that individual speakers' memories for and comprehension vocabularies of previously unseen words will affect lexical entrenchment in speech communities. In other words, I posit that individual cognitive processes, such as memory for new words, influence a system-level dynamic, specifically, whether or not these words become entrenched in a speech community. This assumption is discussed in further detail in section 5.1. Next, I explore alternative hypotheses to the central hypothesis of this dissertation, that various linguistic, cognitive, and social factors matter for lexical entrenchment. I then detail what this dissertation can add to findings on language change and the diffusion of innovations, linguistic or other. Finally, I outline future directions of research given the findings presented in this dissertation.

5.1 The mental lexicon and lexical entrenchment

In order for a new word (or any linguistic innovation) to become adopted in a speech community, it must be uttered by individual speakers. The studies on memory for new words in Chapter 3 and comprehension vocabulary in Chapter 4 do not explicitly address individual speakers' output; rather, they examine various components of individual speakers' mental lexicons that could likely lead to changed linguistic output. This section details the relevance of studying speakers' mental lexicon as it relates to lexical entrenchment, a phenomenon at the level of the speech community.

First, in a recent submission on the findings from Chapter 3 (Chesley & Baayen submitted), I posit that free recall is a necessary precursor to use in a speech community. I then use free recall as a measure that is easily testable in a lab setting in order to study potential use in a speech community. My hypothesis in this work is that, all other things equal, words that are more likely to be freely recalled by individual speakers should be more likely to achieve a greater degree of lexical entrenchment. Since the transition of a word from passive to active vocabulary is extremely hard to observe in natural contexts, some headway can be made by a laboratory study focusing on the ability to call up a newly encountered word at will. Of course, this study does not look at what happens after the initial exposure to the word; that is examined in Chapters 2 and 4. For example, a key finding from Chapter 2 is that the number of contexts a new word occurs in is highly predictive of its later entrenchment. Altmann *et al.* (2011) find a similar result: the number of speakers using a new word in a speech community is positively associated with degree of entrenchment.

Second, while each speaker has his or her own mental lexicon, speakers studied in this dissertation (young adults and adults) generally tend to increase their vocabulary by using input from other sources (speakers, books) or by manipulating existing words in their mental lexicons as opposed to completely new coinages. It follows that many words that become a part of these speakers' expressive vocabularies will first be encountered as the output of other speakers as opposed to their own coinages. That

is, a previously unseen word will most likely be a part of a speaker's comprehension vocabulary before becoming part of their expressive vocabulary, and a speaker's expressive vocabulary is a subset of his or her comprehension vocabulary.

In turn, the expressive vocabularies of many speakers is a necessary precondition to lexical entrenchment across a speech community. It follows that studying the comprehension vocabularies of speakers can inform us about lexical entrenchment. In fact, examining comprehension vocabulary before entrenchment actually occurs in a speech community could offer a first glimpse into the processes driving lexical entrenchment.

Concerning the study in Chapter 4, work in cultural studies and linguistics suggests that non-African-Americans are keen to adopt African-American trends (Bucholtz 2001, Bucholtz 2011, Tate 2003) and that hiphop does have an effect on language change (Jay-Z 2010:84). In this respect, this study examines the *possible words* non-African-American English speakers could adopt. There is no reason to believe these vocabulary items, given in Appendix 7.3.4, are not representative of the types of words that non-African-American speakers will eventually adopt.

An example of a non-African-American speaker's admiration for "new" African-American English (AAE that has not yet become entrenched in Mainstream American English, or MAE) is seen in a *Lopez Tonight* interview of Antoine Dodson (`http://www.youtube.com/watch?v=dObplOihwOI`, 4:47-5:14). Antoine Dodson is an African-American who, until recently, lived in the housing projects of Huntsville, Alabama. He became famous after a video of him being interviewed on the local news, in which he speaks AAE, went viral on YouTube during the summer of 2010. George Lopez, a successful Hispanic-American talk-show host, unabashedly envies and adopts Dodson's lexicon. At the end of the interview, the two have this exchange:

(2) GL: I'm always looking for some new phases... phrases. You said *I live*? I like that one.

AD: I live.

GL: And what did you say my shoes were, what?

AD: You is breakin' up [looking good], you breakin' it up...

GL: I'm breakin' it up! I'm breakin' it up! [does Michael Jackson-type walk and leg kick]

While I do not know if Lopez has repeated "breaking it up" again since this interaction, he is probably at least still aware of this term. Furthermore, by highlighting Dodson's language, Lopez' non-African-American viewers heard more AAE that they perhaps had not heard either. Other non-African-American speakers might not be as obvious about wanting to appropriate AAE lexical items as Lopez is above, but I believe this trend of MAE speakers adopting AAE vocabulary will continue.

5.2 Alternative hypotheses

Having established that studying the mental lexicon of individual speakers is relevant for understanding lexical entrenchment, we can turn to alternative hypotheses about lexical entrenchment. This dissertation examines the hypothesis that linguistic, cognitive, and social factors can influence lexical entrenchment. This sounds like a broad, perhaps obvious claim, so it is worth examining alternative hypotheses. The three I address here are:

1. Lexical entrenchment is essentially unpredictable;
2. Lexical entrenchment depends primarily on the influence of the speaker uttering the word;
3. Lexical entrenchment and the diffusion of lexical innovations is essentially the same process as the diffusion of non-linguistic innovations.

The first of these is simply the null hypothesis: lexical entrenchment is essentially unpredictable. Something similar to this hypothesis has been advanced by Pinker (2007:319-322):

> The surging and cycling of fashion is not the only internal dynamic that decides the fate of words. Even if tastes remained stable, the success of a new word would depend on lexical epidemiology – the way a neologism spreads from its coiner to a new speaker, who then may infect others, and so on. Eventually the neologism either dies out or becomes endemic, depending on how many people a word-knower talks to in a day, and on how readily they notice and remember it. As with real epidemics, it's hard to predict what will happen. Depending on small differences in how catchy the word is, and on how well-connected, trusted, or charismatic its first adopters are, it may or may not reach the tipping point that would lead it to become entrenched in the community and perpetuated down the generations... The naming... of things in general, is another example in which a large-scale social phenomenon – the composition of language – emerges unpredictably out of many individual choices that impinge on one another.

Despite Pinker's comments about linguistic factors ("how catchy the word is") and social factors ("how well-connected, trusted, or charismatic its first adopters are") potentially influencing lexical entrenchment, his final verdict is that the phenomenon is unpredictable. In this dissertation I show that the phenomenon can be predicted and modeled, but that the problem is complex and multi-factorial, perhaps operating in a non-linear dynamical system.

In this dissertation I show that factors like the dispersion of a word at an earlier time is positively associated with the word's entrenchment at a later time. This finding has subsequently been corroborated by work from other researchers (Altmann *et al.* 2011). These factors are only a subset of those at play in lexical entrenchment. There will of course be some randomness, the unpredictability to which Pinker alludes. Still, we can model the entrenchment of new words (at least new borrowings in French) to a surprising degree, if we are to believe Chapter 2. These findings represent a definite

advance in the state of the art.

A second hypothesis is that the entrenchment of a word depends principally on the influence of the speaker uttering the word. This seems to be the position of Anatoly Liberman, a well-known etymologist at the University of Minnesota. He appears at least annually on Minnesota Public Radio's *Midmorning* show talking about new words. In his most recent appearance (Wurzer 31 December 2010) he states that the main criterion for a word to become entrenched is the prestige or influence of the speaker using it (specifically at 1:45 and 4:00). Yet this dissertation, and subsequent work (Altmann *et al.* 2011), finds that the overall number of contexts a word occurs in, and by different speakers, is a better predictor of lexical entrenchment. In other words, "influence" is a poorly defined term with respect to the diffusion of non-linguistic innovations.

I assume that lexical entrenchment, like the diffusion of other non-linguistic and perhaps linguistic innovations, roughly follows the Rogers innovation curve given in Figure 5.1. In this model for the diffusion of non-linguistic innovations, INNOVATORS, the first 2.5% of people to adopt an innovation, are generally percieved as socially marginal or deviant. This might not necessarily be true for the first 2.5% of people to purchase a new Apple product, but it definitely was the case in, say, a South-Asian community in which birth control is introduced (Lin & Hingson 1974). This procedure of diffusion seems to also hold for phonetic change (Milroy & Milroy 1985, Milroy & Milroy 1992, Labov 2001). In linguistics, one can think of innovators as being perceived as having affected or perhaps uneducated speech, or as talking like a foreigner. At this point in the adoption of innovations, their language is seen as abnormal and nonconformist. On one hand then, very few people adopt directly from innovators; in this respect we could say innovators have very *little* influence. My point here is that influence is not a detailed enough concept to describe the phenomenon.

On the other hand, the few people who do adopt the linguistic behavior of innovators are EARLY ADOPTERS (see Figure 5.1). These speakers have many social ties in a speech community. It must be at the early adopter phase in the innovation process

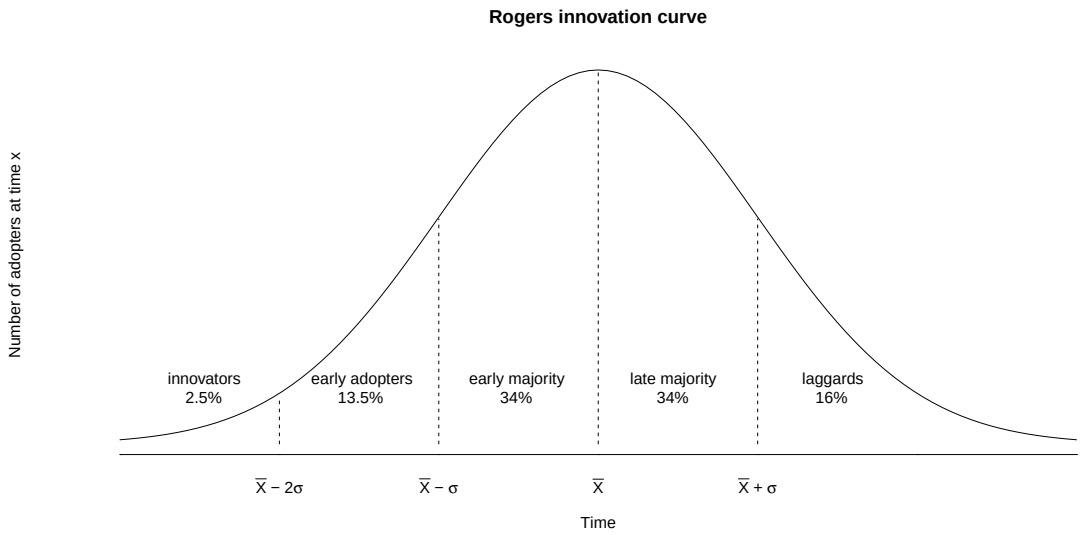

Figure 5.1: The Rogers innovation curve (see Rogers 2003, Figure 7-3, p. 281). The x-axis represents time; the y-axis, the number of adopters at time x.

that an innovation reaches a critical mass (Rogers 2003:343). Therefore, a division between innovators and early adopters, and not influential vs. uninfluential people, should be advocated in the discussion of the diffusion of linguistic innovations.

Relatedly, one could advance still another hypothesis about lexical entrenchment and linguistic innovations in general: that these processes are essentially the same as the diffusion of non-linguistic innovations. This interesting hypothesis merits an entirely separate dissertation. I have examined specific linguistic properties upon which it appears lexical entrenchment does depend, e.g. the word length of a lexical borrowing as well as the borrowing's sense pattern (see Chapter 2). But these are not unique to language in the same way that structural linguistic universals have been claimed to be unique to language. One could easily envisage similar non-linguistic phenomena. If word length is one example of processing complexity for language, then similar indices of visual processing complexity could be (and I'm sure have been) examined with respect to memory. Would an innovation that is easier to cognitively process, regardless of the mechanism, be more likely to become entrenched in a social group? I do not know.

There does seem, however, to be a distinction between the diffusion of linguistic and non-linguistic innovations in that the former are much more gradual and piecemeal. In deciding to purchase an iPad or to go on birth control, people generally fluctuate less in their decisions than in their vocabulary choices or the acoustics of their consonants and vowels for a given day. Once a person adopts a product, he or she either keeps it or returns it. She might decide to continue with that product line or eventually switch to another. A speaker's vocabulary, however, can vary by context, interlocutor, and previous linguistic experience, which is constantly changing. A speaker could utter a (new) word once, and then not again for several days, months, or even years. Or she could come to use that word relatively often, if for example it is part of a new job or lifestyle change. This gradual, uneven diffusion process would seem to indicate that the timecourse of the diffusion of linguistic innovations would be longer than that of non-linguistic innovations, but this hypothesis must be borne out by future empirical findings.

5.3 Language change and lexical entrenchment

In addition, the factors influencing linguistic diffusion could vary by the linguistic object of inquiry. To date, most diffusion studies in linguistics examine phonetic change, where the object of inquiry is the realization of phonological components. Yet syntax, morphology, lexical semantics, phonology, pragmatics, and (perhaps most noticeably) the lexicon also comprise language change. Phonological change, such as Grimm's law stating how Proto-Indo-European stops developed into fricatives in Germanic languages, appears to stem from phonetic change, and would appear to happen much more slowly than the latter. One could also ask whether syntax and morphology appear to change at different rates from phonetic change: these domains of language pertain to Saussure's LANGUE, the abstract system of a language. Phonetic change, however, occurs in PAROLE (see relevant discussion in Kiparsky 1965, 1-12). Perhaps a change pertaining to parole is more immediate, less abstract; as such it is able to

spread quicker. Perhaps not only timescales but also the underlying mechanisms by which change occurs vary by linguistic domain. Do the findings for phonetic change, the most well-documented type of change, hold true for other areas of linguistics? These are interesting questions, but I do not know of any literature addressing them.

I do think, however, that it is important to study all types of language change, including, of course, the diffusion of lexical innovations. Some who work on language change (e.g., Chambers 1999, Stuart-Smith 2006) note that vocabulary is not a "core" part of a language, and therefore, lexical change is less interesting to study, from a linguistic perspective, than say phonetic change. In my opinion, these statements stem from current research trends or perhaps ideological biases more than anything else. Interestingly, Saussure felt that phonetic change, nowadays what is most studied by those examining language change, was essentially an extra-linguistic phenomenon:

> An argument against separating phonation from language might be phonetic changes, the alterations of the sounds which occur in speaking and which exert such a profound influence on the future of language itself. Do we really have the right to pretend that language exists independently of phonetic changes? Yes, for they affect only the material substance of words. If they attack language as a system of signs, it is only indirectly, through subsequent changes of interpretation; there is nothing phonetic in the phenomenon. [Saussure 1966; cited in Kiparsky 1965, 1-13]

It seems that in the span of 70 or so years, phonetic change has become a "core" interest of linguistics; my hope is that the same will be true of lexical change in as many years from now. As evidenced by the positive reception of general-interest media on new words, such as Anatoly Liberman's annual radio appearances on Minnesota Public Radio, the general public thinks it is an interesting topic. Similar to my research, Armelle Boussidan is studying *semantic drift*, i.e. how existing words take on new meanings (Van den Bosch *et al.* 2011). An example of this phenomenon is the change in meaning of *factoid*: its first meaning was "a piece of information appearing to be

true but that isn't", while now it is often used to mean "a piece of information that is true, but trivial".

In addition, changes in the lexicon can lead to changes in other language domains. Lexical diffusion (see Bybee 2000), for example, is a phenomenon in which change occurs first in specific words with specific phonological environments. An example of possible lexical diffusion in Minnesota is the pronunciation of *pillow* as [pɛlow] ("pellow") as opposed to [pɪlow] ("pillow") and *milk* as [mɛlk] ("melk") as opposed to [mɪlk] ("milk"). In both of these words, the high lax vowel /ɪ/ (the "i" in *bit*) occurs before an /l/. Crucially then, this could be one instance of phonetic change that is phonologically conditioned and that spreads through certain words. If this type of change were to continue, it could lead to a merger of the vowels /ɪ/ and /ɛ/, which would affect speakers' phonological inventories.

5.4 Future directions of research

The results I present in this dissertation indicate that lexical entrenchment can be successfully modeled. Additionally, we see ways in which the linguistic, cognitive, and social factors interact to give rise to lexical entrenchment. Future directions of research include examining the theoretical issues behind the diffusion of lexical innovations (or any type of linguistic innovations really), as well as examining the cognitive and social interactions involved in language change.

From a theoretical standpoint, I have argued in the past (at conferences and workshops Quantitative Methods for Morphology and Morphological Development in San Diego) that change in the lexicon, and indeed in language in general, can be profitably modeled assuming that the lexicon/language is a DYNAMICAL SYSTEM, akin to those we see in physics and biology. Others have taken this perspective: for example, it was a central theme of the 2008 Language Learning conference (e.g., Beckner *et al.* 2009). Bill Kretzschmar (Kretzschmar 2010) is also very focused on this type of work. In addition, recent papers indicate others interested in the lexicon

and morphology have adopted this approach (Altmann *et al.* 2011 and Pierrehumbert in press). In his recent TED talk and Interspeech paper (Roy 2009), Deb Roy has stated that he also thinks of parts of language acquisition as a dynamical system. I believe this is an excellent framework for examining the dynamic nature of the lexicon and of language in general.

I received some friendly criticism of this conceptualization of language change, which I relay here informally. After listening to my ideas on this, Michael Ramscar (p.c.) said, "Saying something is a dynamical system is just another way of saying we have no clue what the hell is going on." It is a fair remark. What he is saying is that currently these explanations evoking dynamical systems, including my own, are only descriptive. More complex modeling tools are needed. Examples of such approaches are Niyogi (2006) and Nettle (1999). These works make use of modeling tools such as differential equations and computer simulations to model language change. I hope to be able to pursue this research in future endeavors.

Additional research I would like to pursue includes a laboratory experiment on the role of the media and high-profile individuals in the diffusion of linguistic innovations. This experiment follows the results of Chapter 4 and will examine the degree to which participants change their speech after listening to media clips of high-profile individuals who speak differently than them. For non-linguistic innovations, high-profile individuals are often innovators or early adopters (Rogers 2003), and such people have well-established behavior in the diffusion of innovations. Additionally, patterns of changed speech by participant characteristics (e.g. sex, ethnicity) will be noted. An initial intake questionnaire will examine degree of modification of speech as a function of affiliation or appreciation for the high-profile individual the participant listens to.

Chapter 6

Bibliography

Bibliography

Abeillé, A., L. Clément, & F. Toussenel. 2003. Building a treebank for French. In *Treebanks: Building and Using Parsed Corpora*, 165–188. Kluwer Academic Publishers.

Adelman, J.S., G.D.A. Brown, & J.F. Quesada. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17.814–823.

Agresti, A. 2002. *Categorical Data Analysis*. New York: Wiley, 2nd edition.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19.716–723.

Alim, H. S. 2004. Hip Hop Nation Language. In *Language in the USA: Themes for the Twenty-First Century*, ed. by E. Finegan & J.R. Rickford, 387–409. Cambridge UP.

Alter, A. L., & D. M. Oppenheimer. 2006. Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences* 103.9369–9372.

Altmann, E.G., J.B. Pierrehumbert, & A.E. Motter. 2011. Niche as a determinant of word fate in online groups. *PLoS 1* 6.

Anglin, J.M. 1993. *Vocabulary Development: A Morphological Analysis*. Monographs of the Society for Research in Child Development.

Arnett, J.J. 1995. Adolescents' uses of media for self-socialization. *Journal of Youth and Adolescence* 24.519–533.

Arnon, I., & N. Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62.67–82.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.

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

—— 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge, U.K.: Cambridge University Press.

——, D. J. Davidson, & D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412.

——, L. Feldman, & R. Schreuder. 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 53.496–512.

——, & R. Lieber. 1997. Word frequency distributions and lexical semantics. *Computers and the Humanities* 30.281–291.

——, & P. Milin. 2010. Analyzing Reaction Times. *International Journal of Psychological Research* to appear.

——, R. Piepenbrock, & L. Gulikers. 1995. *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

——, & A. Renouf. 1996. Chronicling the Times: Productive Lexical Innovations in an English Newspaper. *Language* 72.69–96.

——, L. H. Wurm, & J. Aycock. 2007. Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon* 2.419–463.

Balling, Laura, & R. Harald Baayen. 2008. Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes* 23.1159–1190.

Bauer, L. 1983a. *English Word Formation*. Cambridge: CUP.

—— 1983b. *English Word Formation*. Cambridge: CUP.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Beckner, C., R. Blythe, J. Bybee, M.H. Christiansen, W. Croft, N.C. Ellis, J. Holland, J. Ke, D. Larsen-Freeman, & T. Schoenemann. 2009. Language Is a Complex Adaptive System: Position Paper. *Language Learning* 59.1–26. supplement 1.

Bergen, B.K. 2004. The psychological reality of phonaesthemes. *Language* 80.290–311.

Bower, G.H., & S.G. Gilligan. 1979. Remembering Information Related to One's Self. *Journal of Research in Personality* 13.420–432.

Bowers, J.S., C.J. Davis, & D.A. Hanley. 2005. Automatic semantic activation of embedded words: Is there a "hat" in "that"? *Journal of Memory and Language* 52.131–143.

Bradley, A., & A. DuBois. 2010. *The Anthology of Rap*. New Haven, CT: Yale University Press.

Breiman, L. 2001. Random Forests. *Machine Learning* 45.5–32.

Brewer, D.D. 1997. No associative biases in the first name cued recall procedure for eliciting personal networks. *Social Networks* 19.345–353.

Bucholtz, M. 2001. The Whiteness of Nerds: Superstandard English and Racial Markedness. *Journal of Linguistic Anthropology* 11.84–100.

—— 2011. *White Kids: Language, Race, and Styles of Youth Identity*. Cambridge, Massachusetts: Cambridge UP.

Bybee, J. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In *Usage-based models of language*, ed. by M. Barlow & S. Kemmer, 65–85. Stanford: CSLI.

——, & P. Hopper (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Bybee, J.L., & J.L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22.381–410.

Chambers, J.K. 1999. Tv makes people sound the same. In *Language Myths*, ed. by L. Bauer & P. Trudgill, 123–131. Penguin.

Chang, J. 2005. *Can't Stop Won't Stop*. New York: Picador.

Chesley, P., & R. H. Baayen. 2010. Predicting new words from newer words: Lexical borrowings in French. *Linguistics* 48.1343–1374.

——, & R.H. Baayen. submitted. When listeners become speakers: An experimental study of factors leading to active entrenchment.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cohen, J. 1992. A Power Primer. *Psychological Bulletin* 112.155–159.

Collister, L. B., & D. Huron. 2008. Comparison of Word Intelligibility in Spoken and Sung Phrases. *Empirical Musicology Review* 3.109–125.

Cook, R.D. 2004. Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics* 32.1062–1092.

Coolen, R., H. J. Van Jaarsveld, & R. Schreuder. 1991. The interpretation of isolated novel nominal compouns. *Memory and Cognition* 19.341–352.

Craik, F.I.M. 2002. Levels of processing: Past, present... and future? *Memory* 10.305–318.

——, & R.S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11.671–684.

——, & E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104.268–294.

Cutler, C. 2007. The Co-Construction of Whiteness in an MC Battle. *Pragmatics* 17.9–22.

Dabrowska, E. 2009. Words as constructions. In *New Directions in Cognitive Linguistics*, ed. by V. Evans & S. Pourcel, 201–223. Amsterdam: John Benjamins.

De Jong, N. H., R. Schreuder, & R. H. Baayen. 2000. The morphological family size effect and morphology. *Language and Cognitive Processes* 15.329–365.

DeLoache, J.S., C. Chiong, K. Sherman, N. Islam, M. Vanderborght, G.L. Troseth, G.A. Strouse, & K. O'Doherty. 2010. Do Babies Learn from Baby Media? *Psychological Science* 21.1570–1574.

Dendien, J., & J.-M. Pierrel. 2003. Le trésor de la Langue Française informatisé: un exemple d'informatisation d'un dictionnaire de langue de référence. *Traitement automatique des langues* 44.11–37.

Dumay, N., & M. G. Gaskell. 2007. Sleep-associated changes in the mental representation of spoken words. *Psychological Science* 18.35–39.

——, ——, & X. Feng. 2004. A day in the life of a spoken word. In *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society*, ed. by K. Forbus, D. Gentner, & T. Regier, 339–344. Mahwah, NJ: Erlbaum.

Eckert, P. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.

Erling, E., 2004. *Globalization, English, and the German University Classroom*. University of Edinburgh dissertation.

ETIEMBLE, R. 1964. *Parlez-vous franglais ?*. Paris: Gallimard.

FARID, M., & J. GRAINGER. 1996. How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language* 53.352–368.

FLAITZ, J. 1988. *The Ideology of English: French Perceptions of English as a World Language*. Berlin: Mouton de Gruyter.

FRIEDKIN, N.E. 1980. Information Flow Through Strong and Weak Ties in Intraorganizational Social Networks. *Social Networks* 3.273–285.

FUCHS, C.. 1996. *Les ambiguïtés du français*. Paris: Editions Ophrys.

GANS, H. 1962/1982. *The Urban Villagers: Group and Class in the Life of Italian-Americans*. New York: The Free Press, 2nd edition.

GASKELL, M. G., & N. DUMAY. 2003. Lexical competition and the acquisition of novel words. *Cognition* 89.105–132.

GILES, H., & P. SMITH. 1979. Accommodation Theory: Optimal Levels of Convergence. In *Language and Social Psychology*, ed. by H. Giles & R. St. Clair. Oxford: Basil Blackwell.

GOLDBERG, ADELE E. 2006. *Constructions at work: the nature of generalization in grammar*. Oxford: Oxford University Press.

GRANOVETTER, M.S. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78.1360–1380.

—— 1980. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory* 1.201–233.

GREEN, L. 2002. *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press. (On AAE).

GREENFIELD, P. M., L. BRUZZONE, K. KOYAMATSU, W. SATULOFF, K. NIXON, M. BRODIE, & D. KINGSDALE. 1987. What is Rock Music Doing to the Minds of our Youth? A First Experimental Look at the Effects of Rock Music Lyrics and Music Videos. *Journal of Early Adolescence* 7.315–329.

GRIES, S.T. 2004. Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics* 42.639–667.

—— 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13.

Gries, Stefan Th. 2006. Cognitive determinants of subtractive word-formation processes: a corpus-based perspective. *Cognitive Linguistics* 17.4.in press.

Hagège, C. 2006. *Combat pour le français au nom de la diversité des langues*. Paris: Odile Jacob.

Halle, M., & A. Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. by K. Hale & S. J. Keyser, volume 24 of *Current Studies in Linguistics*, 111–176. Cambridge, Mass: MIT Press.

Hansen, C.H., & R.D. Hansen. 1991. Schematic Information Processing of Heavy Metal Lyrics. *Communication Research* 18.373–411.

Harrell, F.E. 2001. *Regression modeling strategies*. Berlin: Springer.

Hay, Jennifer, Warren Warren, & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34.458–484.

Hayes, A.F., & K. Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1.77–89.

Hohenhaus, P. 2006. Bouncebackability: A Web-as-corpus-based case study of a new formation, its interpretation, generalization, spread and subsequent decline. *Skase Journal of Theoretical Linguistics* 3.17–27.

Howes, D., & R. L. Solomon. 1951. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology* 41.401–410.

Ito, J., & A. Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *University of Massachusetts Occasional Papers in Linguistics [UMOP]*, ed. by J. Beckman, S. Urbanczyk, & L. Walsh, 181–209, Amherst, MA. GLSA.

Jackendoff, R. 1997. *The architecture of the language faculty*. Cambridge, MA: MIT Press.

——, & S. Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97.211–225.

Jay-Z. 2010. *Decoded*. New York: Spiegel and Grau.

Kaplan, R.M., & J. Bresnan. 1982. Lexical-functional grammar: a formal system for grammatical representation. In *The mental representation of grammatical relations*, 173–281. Center for the Study of Language.

KENSINGER, E., & S. CORKIN. 2003. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory and Cognition* 31.1169–1180.

KIPARSKY, P., 1965. *Phonological Change*. Massachusetts Institute of Technology dissertation.

KLEIN, D., & G. L. MURPHY. 2001. The Representation of Polysemous Words. *Journal of Memory and Language* 45.259–282.

KOOLSTRA, C.M., & J.W.J. BEENTJIES. 1999. Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development* 47.51–60.

KRETZSCHMAR, W.A. 2010. Language variation and complex systems. *American Speech* 85.263–286.

KUIPER, K. 2006. Knowledge of language and phrasal vocabulary acquisition. *Behavioral and Brain Sciences* 29.291–292.

KUPERMAN, V., R. BERTRAM, & R. H. BAAYEN. 2010. Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language* 62.83–97.

——, R. SCHREUDER, R. BERTRAM, & R. H. BAAYEN. 2009. Reading polymorphemic Dutch compounds: towards a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance* 35.876–895.

LABOV, W. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

—— 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

—— 2001. *Principles of Linguistic Change: Social Factors*. Malden, MA: Wiley-Blackwell.

—— 2006. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

LEHRER, A. 1996. Identifying and interpreting blends: An experimental approach. *Cognitive Linguistics* 7.359–390.

—— 2003. Understanding trendy neologisms. *Rivista di Linguistica* 7.371–384.

LENHART, A., K. PURCELL, A. SMITH, & K. ZICKURH. 2010. *Social Media and Mobile Internet Use Among Teens and Young Adults*. Washington DC: Pew Internet and American Life Project.

Lerman, D., & E. Garbarino. 2002. Recall and recognition of brand names: A comparison of word and nonword name types. *Psychology and Marketing* 19.621–639.

Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, Mass.: The MIT Press.

Li, K.-C. 1991. Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association* 86.316–327.

Lin, N., & R. Hingson. 1974. Diffusion of Family Planning Innovations: Theoretical and Practical Issues. *Studies in Family Planning* 5.189–194.

Lowrey, T., L.J. Shrum, & T.M. Dubitsky. 2003. The relation between brand-name linguistic characteristics and brand-name memory. *Journal of Advertising* 32.7–17.

Marc Martínez, I. 2010. Voix signifiantes : le cas du rap français. *Thélème: Revista Complutense de Estudios Franceses* 25.185–195.

Marinellie, S.A., & C.J. Johnson. 2004. Nouns and verbs: a comparison of definitional style. *Journal of Psycholinguistic Research* 33.217–235.

Markson, L., & P. Bloom. 1997. Evidence against a dedicated system for word learning in children. *Science* 151–163.

Marsden, P.V. 2005. Recent Developments in Network Measurement. In *Models and methods in social network analysis*, ed. by P.J. Carrington, J. Scott, & S. Wasserman, 8–30. Cambridge: Cambridge UP.

Massey, D.S. 2004. Segregation and Stratification: A Bisocial perspective. *Du Bois Review* 1.7–25.

——, & N.A. Dalton. 1989. Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation Along Five Dimensions. *Demography* 26.373–391.

McGinnis, D., R. J. Goss, C. Tessmer, & E.M. Zelinski. 2008. Inference generation in young, young-old and old-old adults: evidence for semantic architecture stability. *Applied Cognitive Psychology* 22.171–192.

Medina, S.L. 1990. The Effects of Music upon Second Language Vocabulary Acquisition. Paper presented at the Annual TESOL Conference.

Metcalf, A. 2004. *Predicting New Words: The Secrets of Their Success*. Boston: Houghton Mifflin Harcourt.

Miller, P., 1991. *Clitics and Constituents in Phrase Structure Grammar*. Free University of Brussels dissertation.

Milroy, J., & L. Milroy. 1985. Linguistic Change, Social Network and Speaker Innovation. *Journal of Linguistics* 339–384.

Milroy, L., & J. Milroy. 1992. Social Network and Social Class: Toward an Integrated Sociolinguistic Model. *Language in Society* 1–26.

Moscoso del Prado Martín, F., R. Bertram, T. Häikiö, R. Schreuder, & R. H. Baayen. 2004. Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30.1271–1278.

Murdock, B.B. Jr. 1960. The distinctiveness of stimuli. *Psychological Review* 67.16–31.

Nairne, J.S. 2002. The mytho of the encoding-retrieval match. *Memory* 10.389–395.

Nelson, D.L. 1979. Remembering pictures and words: appearance, significance and name. In *Levels of processing in human memory*, ed. by L.S. Cermak & F.I.M. Craik, 45–76. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nettle, D. 1999. Is the rate of linguistic change constant? *Lingua* 108.119–136.

New, B., L. Ferrand, C. Pallier, & M. Brysbaert. 2006. Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review* 13.45–52.

Niyogi, P. 2006. *The computational nature of language learning and language evolution*. Cambridge, Massachusetts: MIT UP.

Oldfield, R. C., & A. Wingfield. 1965. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17.273–281.

O'Regan, J.K., & A. Jacobs. 1992. Optimal viewing position effects in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance* 18.185–197.

Parks, R., J. Ray, & S. Bland, 1998. *Wordsmyth English dictionary-thesaurus*. University of Chicago. `http://www.wordsmyth.net/Chicago`.

Pergnier, M. 1989. *Les anglicismes: danger ou enrichissement pour la langue française ?*. Paris: Presses universitaires de France.

Pickering, M.J., & S. Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27.169–226.

Picone, M.D. 1996. *Anglicisms, Neologisms, and Dynamic French*. Amsterdam: John Benjamins.

Piercey, C.D., & S. Joordens. 2000. Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory and Cognition* 28.657–666.

Pierrehumbert, J. in press. The dynamic lexicon. In *Handbook of Laboratory Phonology*, ed. by A. Cohn, M. Huffman, & C. Fougeron. Oxford UP.

Pinker, S. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. New York: Viking Adult.

Pittman, A. L., D. E. Lewis, B. M. Hoover, & P. G. Stelmachowicz. 2005. Rapid word-learning in normal-hearing and hearing-impaired children: Effects of age, receptive vocabulary, and high-frequency amplification. *Ear Hear* 26.619–629.

Pollard, C., & I.A. Sag. 1987. *Information-based syntax and semantics: Vol. 1: fundamentals*. Center for the Study of Language and Information.

——, & ——. 1994. *Head-driven phrase structure grammar*. University of Chicago Press.

Rayner, K. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62.1457 – 1506.

Renouf, A. 2007. Tracing lexical productivity and creativity in the british media: The Chavs and the Chav- nots. In *Lexical Creativity, Texts and Contexts*, ed. by J. Munat. Amsterdam: John Benjamins.

Rentfrow, P.J., & S.D. Gosling. 2003. The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology* 84.1236–1256.

Rey-Debove, J. 1987. Effet des anglicismes lexicaux sur le système du français. *Cahiers de lexicologie* 51.257–265.

Rickford, J.R., & R.J. Rickford. 2000. *Spoken Soul*. New York: Wiley.

Rodd, J.M., M.G. Gaskell, & W.D. Marslen-Wilson. 2004. Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science* 28.89–104.

——, & W.D. Marslen-Wilson. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46.245–266.

Rogers, E.M. 2003. *Diffusion of Innovations*. New York: Free Press.

Roy, D. 2009. New horizons in the study of child language acquisition. In *Proceedings of Interspeech*, Brighton, UK.

Rudoy, J.D., J.L. Voss, C.E. Westerberg, & K.A. Paller. 2009. Strengthening Individual Memories by Reactivating Them During Sleep. *Science* 326.1079.

Sanneh, K. 2010. Word: Jay-Z's "Decoded" and the language of hip-hop. *The New Yorker* December 6.

Saussure, Ferdinand de. 1966. *Course in General Linguistics*. New York: McGraw.

Schmidt, S.R. 1994. Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.953–967.

Schultink, H. 1962. *De Morfologische Valentie van het Ongelede Adjectief in Modern Nederlands*. Den Haag: van Goor & Zonen.

Selkirk, E. 1982. *The Syntax of Words*. Cambridge: The MIT Press.

Smitherman, G. 2000. *Black talk: Words and phrases from the hood to the amen corner*. Mariner Books.

Steels, L., F. Kaplan, A. McIntyre, & J. Van Looveren. 2002. Crucial factors in the origins of word-meaning. In *The Transition to Language*, ed. by Alison Wray, chapter 12. Oxford: Oxford University Press.

Strobl, C., T. Hothorn, & A. Zeileis. 2001. Party On! a New, Conditional Variable Importance Measure for Random Forests Available in the `party` Package. Technical Report 50, Institut für Statistik, Ludwig-Maximilians-Universität München.

Stuart-Smith, J. 2006. The influence of the media. In *The Routledge companion to sociolinguistics*, ed. by C. Llamas, L. Mullany, & P. Stockwell, 140–148. New York: Routledge.

Swanborn, M.S.L., & K. de Glopper. 1999. Incidental Word Learning while Reading: A Meta-Analysis. *Review of Educational Research* 69.261–285.

Tate, G. 2003. *Everything but the burden: What White people are taking from Black culture*. New York: Broadway Books.

Thogmartin, C. 1988. The public impact of terminological planning in France. *The French Review* 64.1000–1006.

Thomason, S.G., & T. Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tremblay, A., 2009. *Processing Advantages of Lexical Bundles: Evidence from Self-paced Reading, Word and Sentence Recall, and Free Recall with Event-related Brain Potential Recordings*. University of Alberta dissertation.

——, & R. H. Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In *Perspectives on formulaic language: Acquisition and communication*, ed. by D. Wood. London: Continuum International.

Trudgill, P. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

—— 1983. *On Dialect*. New York: New York University Press.

Tulving, E. 1972. Episodic and Semantic Memory. In *Organization of memory*, 381–403. New York: Academic Press.

Ussishkin, A. 2005. A Fixed Prosodic Theory of Nonconcatenative Templatic Morphology. *Natural Language & Linguistic Theory* 23.169–218.

Van den Bosch, A., W. Daelemans, & T. Weijters. 2011. Using topic salience and connotational drifts to detect candidates to semantic change. In *Proceedings of the Ninth International Conference on Computational Semantics*, 315–319, Oxford, UK.

Verbrugge, L.M. 1979. Multiplexity in adult friendships. *Social Forces* 57.1286–1309.

Vitu, F., J.K. O'Regan, & M. Mittau. 1990. Optimal landing position in reading isolated words and continuous text. *Perception and Psychophysics* 47.583–600.

Weisberg, S. 2002. Dimension Reduction Regression in R. *Journal of Statistical Software* 7.

Wolfram, W., & E.R. Thomas. 2002. *The development of African American English*. Malden, MA: Wiley-Blackwell.

Wurzer, C., 31 December 2010. The Year in Words with Anatoly Liberman. Mid-morning radio broadcast.

Yarkoni, T., D. Balota, & M. Yap. 2009. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin and Review* 15.971–979.

Zipf, G. K. 1949. *Human Behavior and the Principle of the Least Effort: An Introduction to Human Ecology*. New York: Hafner.

Chapter 7

Appendix

7.1 Predicting new words: Lexical borrowings found in *Le Monde* (T1)

Borrowings are given as they appear in the T1 corpus. A dash indicates that the type was excluded from the *Le Figaro* (T2) study for reasons discussed in Section 2.3.2. The sense *offshore_1* pertains to drilling for oil at sea, while *offshore_2* qualifies business conducted abroad. One occurrence of *offshore* was excluded because its sense could not be determined.

Borrowing	Frequency		Borrowing	Frequency	
	T1	T2		T1	T2
A contrario	1	667	inquilinaje	2	0
Canada Dry	1	32	investment banks	1	4
Errare humanum est , perseverare diabolicum	1	1	joint - venture	17	400
Gross Up	1	1	joint - ventures	2	221
Just do it	1	21	junk bonds	2	127
Karenztag	2	0	khazjajstvo	1	0
Lander	6	1147	kids	2	15
Last but not least	1	178	kippa	1	183
Mismanagement	1	4	know how	1	15
Nehruian Socialism	1	0	kollektivnoe	1	0
TIERRA Y LIBERTAD	1	0	land	1	–
afrikaans	1	17	latifundio	3	1
alya	1	18	lean production	1	0
an elastic currency	1	0	lease - back	2	10
apparatchik	1	289	leitender Angestellter	1	0
apple - pie	1	3	lobbying	3	865

bag - ladies	1	0	look	2	1387
basics	1	22	lottizzazione	1	0
because	1	5	male oscuro	1	0
big Three	1	101	markka	1	13
big bang	3	489	medium - term notes	2	0
board	3	184	merchant banks	1	4
boat - people	1	124	minifundio	3	0
brain storming	1	23	money funds	1	0
brocca	1	0	names	1	14
bush	1	–	news	2	120
campo	1	7	next jump	1	0
cash	1	2120	next root	1	0
cash flow	2	711	next wave	1	0
cassa integrazione	2	2	nomenklatura	4	190
casualwear	1	6	offshore	1	–
chapka	1	55	offshore_1	5	300
check - up	1	59	offshore_2	1	274
chehita	1	0	old lady	1	4
citizen' s charter	1	0	open market	1	4
classless society	1	0	outsourcing	1	98
come - back	1	333	pack	2	–
credit crunch	1	23	parity cracking	2	0
cross borders	1	0	perestroika	8	210
debt deflation	1	6	ph.D	1	11
deficiency payments	1	1	popiwek	1	0
deregulation	1	–	prime rate	1	6
deutschemark	25	154	prorata temporis	1	61
deutschemarks	24	223	res nullius	1	2
discount	1	305	roadbook	1	1
downgrading	1	0	running	1	53
dynamic random access memory	1	3	savings and loans	1	6
ejideros	2	0	scala mobile	4	3
ejido	9	0	shipping	2	2
ejidos	8	1	short	1	–
establishment	2	500	show - room	1	165
estancias	1	9	sommerso	1	0
ex- joint - venture	2	0	stand - by	3	123
fazendas	1	5	stop loss	1	6
fincas	1	12	struggle for life	1	8
flash	3	33	success story	3	382
flint glass	1	0	sustainable	1	1
float glass	1	3	swaps	1	71
french doctors	1	25	sweats	1	4
french mafia	2	0	taref	1	0

french travel way of life	1	0	teddy	1	18
geschäftsfürhung	1	0	terratenientes	2	0
glasnost	2	82	the	1	0
government	1	0	to regulate	1	0
half baked	1	0	top - down	1	25
hedge funds	3	368	trade unions	1	2
high - tech	2	2056	training groups	1	0
homeless	1	6	tripalium	1	4
huasipongo	2	0	under - class	1	1
industrial design	1	0	welfare state	2	9

7.2 Memory for new words

7.2.1 Stimuli

Stimuli are given with short definitions.

Word	Word type	Definition
AATP	A	All About The Physical
ADF	A	All Day Funk
BGP	A	BackGround Props
BITGOD	A	Back In The Good Old Days
CLM	A	Career-Limiting Move
CVOC	A	Cooler Version Of Cool
DBI	A	Douche Bag Index
DLS	A	Dirty Little Secret
DTR	A	Define The Relationship
FD	A	Fashion Disaster
FSBO	A	For Sale By Owner
FSP	A	Filthy Street Person
FUSSDIRAG	A	Former U.S.-Supported Dictator Internationally Reviled for Acts of Genocide
Halophile	D	A person that likes and plays the video game Halo a lot
ICTYIAS	A	I Can't Tell You, It's A Secret
LIMH	A	Laughing In My Head. Like LOL except for not literally laughing out loud
MIRF	A	Mom I'd Run From
McDad	D	Divorced father who gives his children special privileges, like McDonalds
NCMO	A	Non Committal Make Out
NGL	A	Not Gonna Lie
OMS	A	Open Mouth Syndrome
PINO	A	Private In Name Only
SMV	A	Sexual Market Value
UDI	A	Unidentified Drinking Injury
URST	A	UnResolved Sexual Tension
VOCD	A	Volume Obsessive-Compulsive Disorder
WAM	A	Walking Around Money
anablog	B	Old-fashioned journal with crushed tree pulp, binding, and perhaps a lock
anti-anti-Semite	D	A person who is opposed to anti-Semites
bio-accessory	D	A living being used as a fashion accessory
cannabista	D	One who knows a good deal about cannabis
chairdrobe	D	A chair on which clothes can be piled in lieu of a closet
chickchismo	B	An exhilarating sense of female cunning
compunicate	B	Talking via Instant Messenger to people in the same room instead of speaking
de-eat	D	Euphemism for 'vomit'

Table 7.2 – continued from previous page

Word	Word type	Definition
dejabrew	B	Remembering things you did while drinking too much beer the night before
destinesia	B	When you forget why you were going to your current location
e-haircut	D	A change in a person's online identity
e-tact	D	Tact exercised in an electronic environments
e-void	D	To use technology for the sole purpose of avoiding (human) contact
econnoisseur	B	One who insists on the highest quality at the lowest price
elecoustic	B	An electric guitar played 'unplugged'
enterdrainment	B	Passive form of mind-numbing entertainment
errorist	D	Someone who repeatedly makes mistakes when speaking
fiscalize	D	To publicly control and supervise state revenues/budget assignations
gaybie	B	The child of a gay couple
grade digger	B	A girl (or boy) who only talks to you for help with her classes
hater tots	B	The figurative snack you consume when you're hating on someone
hypertasking	D	The simultaneous execution of an exceedingly large number of tasks
innuendo-ish	D	Suggestive of an innuendo
man stand	B	When a man stands outside a shop while the woman he's with shops inside
manther	B	Male counterpart of a cougar
multi-slacking	B	Doing multiple slacker-esque things concurrently
mysterectomy	B	Taking the suspense out of a movie by revealing its spoilers
obliviation	D	The state of total lack of awareness to the universe at large
oldiephile	D	One who has sexual relationships with the older generation
playlistism	D	Discrimination based on an iTunes music library
pornocchio	B	A person who embellishes their sexual activities to sound cooler
pre-eat	D	To eat a small snack or meal in anticipation of going out for food later on
pre-walk	D	To walk in preparation for a future walk (as on a subway platform)
premature evacuation	B	Stealthy exit after a one-night stand
re-cop	D	To acquire more of a particular drug
resolutionary	B	A person who joins a gym right after New Year's
sapiosexual	D	One who finds intelligence the most sexually attractive quality in a mate
showmance	B	A relationship contrived to look good outwardly
spim	B	Like spam, but sent via an Instant Messenger program
stoptional	B	Optional stop due to an unnecessary stop sign
testosterphone	B	To make a quick and to-the-point phone call in the male fashion
tonorrow	B	Either tonight or tomorrow
trumor	B	A rumor that turns out to be true
un-trade-upable	D	Something that is so good, it can't be improved by trading up
underdrunk	D	Inability to achieve one's desired level of intoxication
unfull	D	Euphemistic term for 'empty' or close to it
wordanista	D	A person who spends their life telling others what is or is not a word
workmare	B	A nightmare derived directly from your place of employment

Table 7.2: Stimuli and their definitions.

7.2.2 Models

	Estimate	Std. Error	t value
Intercept	5.5949	0.0316	176.8565
Number of Fixations	-0.0438	0.0082	-5.3485
Word Length	0.0327	0.0059	5.5862
Log Component Frequency	-0.0651	0.0149	-4.3595
Type = blend	-0.0479	0.0294	-1.6323
Type = derivation	-0.0523	0.0293	-1.7817
Trial Index	-0.0090	0.0025	-3.6605
Log Component Frequency : Type = blend	0.0735	0.0177	4.1609
Log Component Frequency : Type = derivation	0.0794	0.0167	4.7522
Word Length : Type = blend	-0.0394	0.0065	-6.0781
Word Length : Type = derivation	-0.0360	0.0064	-5.6192

Table 7.3: Model coefficients – First Fixation Durations.

	Estimate	Std. Error	t value
Intercept	-0.9537	0.0469	-20.3345
PC 2 (familiarity)	-0.0131	0.0047	-2.7765
Previous RT	0.0096	0.0071	1.3479
Correctly Recognized = no	0.0417	0.0149	2.8037
Type = blend	-0.1798	0.0350	-5.1304
Type = derivation	-0.2121	0.0343	-6.1882
Word Length (residualized)	0.0136	0.0034	4.0081
Sleep Condition = sleep	-0.0058	0.0081	-0.7240
Log Component Frequency (centered)	0.0809	0.0544	1.4877
Correctly Recognized = no : Type = blend	0.0221	0.0296	0.7444
Correctly Recognized = no : Type = derivation	0.1599	0.0301	5.3094
Word Length : Log Component Frequency	0.0046	0.0021	2.1911
Sleep Condition = sleep : Previous RT	0.0198	0.0100	1.9855

Table 7.4: Model coefficients – Gaze Durations.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	1.5458	0.0521	29.6926	0.0000
Correctly Recognized = no	0.1391	0.0248	5.5992	0.0000
Trial Index	0.0335	0.0094	3.5648	0.0004
Word Length	0.0651	0.0098	6.6404	0.0000
Log Component Frequency	-0.0761	0.0118	-6.4337	0.0000
Type = blend	-0.1902	0.0437	-4.3542	0.0000
Type = derivation	-0.1922	0.0443	-4.3400	0.0000
Correctly Recognized = no : Trial	-0.0726	0.0208	-3.4925	0.0005
Correctly Recognized = no : Word Length	0.0150	0.0067	2.2396	0.0251
Word Length : Type = blend	-0.0238	0.0114	-2.0856	0.0370
Word Length : Type = derivation	-0.0273	0.0109	-2.5054	0.0122
Log Component Frequency : Word Length	0.0124	0.0029	4.2566	0.0000

Table 7.5: Model coefficients — Number of Fixations.

	Estimate	Std. Error	t value
Intercept	-0.8537	0.0406	-21.0450
PC 2 (familiarity)	-0.0101	0.0049	-2.0594
Correctly Recognized = no	0.0156	0.0156	1.0042
Previous RT	0.0170	0.0056	3.0107
Type = blend	-0.1979	0.0367	-5.3870
Type = derivation	-0.2264	0.0357	-6.3374
Trial Index	-0.0212	0.0079	-2.6815
Word Length (residualized)	0.0209	0.0034	6.0745
Log Component Frequency (centered)	-0.0217	0.0084	-2.5851
Correctly Recognized = no : Previous RT	0.0241	0.0110	2.1975
Correctly Recognized = no : Type = blend	0.0482	0.0313	1.5370
Correctly Recognized = no : Type = derivation	0.1896	0.0319	5.9534
Word Length : Log Component Frequency	0.0044	0.0022	2.0373

Table 7.6: Model coefficients – RT.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-0.6796	0.4221	-1.6100	0.1074
Previous RT	0.1692	0.0623	2.7165	0.0066
Word Length	0.5836	0.2964	1.9692	0.0489
Type = blend	2.2124	0.3961	5.5852	0.0000
Type = derivation	2.2923	0.3779	6.0661	0.0000
Trial Index	-0.1558	0.0530	-2.9418	0.0033

Table 7.7: Model coefficients — Recognition. Reference level = not recognized.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-2.9740	0.3194	-9.3123	0.0000
Trial Index	0.1459	0.0528	2.7620	0.0057
Response = no	-1.2914	0.2293	-5.6309	0.0000
RT (neg inv transformation)	-0.6924	0.1950	-3.5506	0.0004
PC3	-0.2015	0.0685	-2.9423	0.0033
PC5	0.2100	0.0813	2.5822	0.0098
PC6	0.2810	0.1064	2.6418	0.0082
Sleep Condition = sleep	-0.4717	0.2914	-1.6187	0.1055
Type = blend	0.4782	0.3090	1.5476	0.1217
Type = derivation	0.4249	0.3104	1.3688	0.1711
Sleep Condition = sleep : Type = blend	0.5662	0.3047	1.8580	0.0632
Sleep Condition = sleep : Type = derivation	1.0160	0.3019	3.3651	0.0008

Table 7.8: Model coefficients — Recall. Reference level = not recalled.

Effect	SD	Corr. Parameters	LR Statistic	p-value
		First Fixation Duration		
Word	0.035		$\chi^2(1) = 83.02$	< 0.0001
Subj by Word Length	0.0081		$\chi^2(1) = 33.74$	< 0.0001
Subj by Word Type	0.094 (I)	-0.439 (I/B)	$\chi^2(5) = 70.23$	< 0.0001
	0.076 (B)	1 (B/D)		
	0.079 (D)	-0.422 (I/D)		
Subj – all	0.094		$\chi^2(7) = 982.9$	< 0.0001
Residual	0.13			
		Gaze Duration		
Word	0.053		$\chi^2(1) = 67.26$	< 0.0001
Subj	0.22		$\chi^2(7) = 1047.7$	< 0.0001
Subj by Word Length	0.011		$\chi^2(1) = 16.78$	< 0.0001
Subj by Trial Index	0.043		$\chi^2(1) = 51.30$	< 0.0001
Subj by Word Type	0.22 (I)	-0.58 (I/B)	$\chi^2(5) = 101.68$	< 0.0001
	0.15 (B)	0.99 (B/D)		
	0.14 (D)	-0.49 (I/D)		
Residual	0.23			
		Number of Fixations		
Word	0.049		$\chi^2(1) = 6.08$	0.013
Subj	0.25		$\chi^2(1) = 752.21$	< 0.0001
		RT		
Word	0.055		$\chi^2(1) = 77.42$	< 0.0001
Subj by Word Length	0.0092		$\chi^2(1) = 10.07$	0.0015
Subj by Trial Index	0.045		$\chi^2(1) = 58.40$	< 0.0001
Subj by Word Type	0.24 (I)	-0.60 (I/B)	$\chi^2(5) = 140.18$	< 0.0001
	0.17 (B)	0.99 (B/D)		
	0.16 (D)	-0.51 (I/D)		
Subj – all	0.24		$\chi^2(8) = 1259.7$	< 0.0001
Residual	0.23			
		Recognition		
Word	0.64		$\chi^2(1) = 80.1$	< 0.0001
Subj by Word Type	0.72 (I)	-0.73 (I/B)	$\chi^2(5) = 37.90$	< 0.0001
	1.04 (B)	1 (B/D)		
	0.90 (D)	-0.74 (I/D)		
Subj – all	0.72		$\chi^2(6) = 83.76$	< 0.0001
		Free Recall		
Word	0.78		$\chi^2(1) = 113.63$	< 0.0001
Subj	0.46		$\chi^2(1) = 28.95$	< 0.0001

Table 7.9: Random effects for First Fixation, Gaze Duration, Number of Fixations, RT, Recognition, and Free Recall. I = Intercept (for word type, always acronym), B = Blend, D = Derived word. The likelihood ratio (LR) statistics and corresponding p-values were obtained by comparing the models with and without the given random effect substructure. Subj – all tests a model without any random effect structure for subject against a model with the full random effect structure for subject.

7.3 Learning words through hiphop music

7.3.1 Stimuli

Word	Definition
5-0	police
A-town	Atlanta
CP time	colored people's time
The Nation	Nation of Islam
What it is?	How are you?
(to be) ghost	to leave/be out of here
ashy	dry skin for African Americans
ay yo trip	check this out
ballin'	to live the good life/play basketball
beezy	woman
bones	dollars
boo	boyfriend/girlfriend/someone you love
boughie	bourgeois
break someone out	to tell on
catch the vapors	to get caught up in somebody else's affairs
chedda	cash
cheese	money
chitlins	chitterlings – pig intestines
cop my steezy	copy someone's style
crossroads	heaven
cuddie	good friend
dap	greeting with hands like handshake
dead presidents	cash; dollar bills
dollar cab	subway
domino	hundred dollar bill
duckets	cash/money
dukey rope	fat gold chain
dun	PERFECTIVE ASPECT MARKER
face gator	friend
facheezie	for sure
feel me	understand me
fetti	money
finna	gonna
ghostride	dancing while driving
good hair	straight hair
gouda	money
grip	a lot
grip grain	to have wood grain on your steering wheel
guap	lots of money
hard	tough

Table 7.10 – continued from previous page

Word	Definition
heezy	variant of off the hook; cool, fun
hella	very
humming	bad odor/breath
krump	style of street dance
mail	money
make it rain	throw down money on people
off the hook	very good/new/crazy wild
player hater	a person who's resentful/jealous of promiscuous people
plex	beef, as in issue
road dog	friend
roll deep	in large numbers
rollie	Rolex
saditty	cocky
skrilla	money
straight cash	only cash
stupid	very; cool
sweatbox	small club
talking jazz	say bad things about
toe up	tore up; disheveled
trife	troublesome/trifling
trill	true/real
twurk	dance
what it do	How are you?
wile out	to flip out

Table 7.10: Stimuli and their targeted AAE definitions.

7.3.2 Further data specifications

Table 7.11 summarizes our participants' self-reported ethnicities. Data from the participant who gave her ethnicity as bi-racial without specifying ethnicities was excluded, as was data from the African-American participant. However, analyses that included these participants yielded the same results. Due to data sparsity, other participants who gave their ethnicites as bi-racial were given their non-Caucasian ethnicities as opposed to a new ethnicity.

Some participants did not provide American city and state information, either because they were from foreign countries, or because their response was underspecified. Foreign-born participants were given hometown values of 0 for City, County, and CountyAA; these values measure how many (African-)American speakers with which these participants come into contact in their hometown. Treatments for these and other anomalous responses are given in Table 7.12.

Ethnicity	Number
African-American	1
Asian/Asian-American	17
Bi-racial (Asian/Caucasian)	2
Bi-racial (South-Asian [Indian]/Caucasian)	1
Bi-racial (ethniticies not given)	1
Caucasian	135
East-African	5
Hispanic	3
Native American	1
South Asian (Indian)	1
Not reported	1
Total	168

Table 7.11: Summary of self-reported participant ethnicities.

Non-African-American rappers can use language differently than African-American rappers (Cutler 2007). To ensure that the artists classified as hiphop artists were actually likely to use AAE, all non-African-American hiphop artists such as Eminem were classified as pop artists. A few speakers of AAE that are more likely to be classified as R&B artists, such as Beyoncé, were classified as hiphop artists so that use of AAE could be gauged in a coherent music variable.

The effect of preferred artists

The method described of obtaining information about the number of artists using a particular word has two potential sources of noise. First, it is possible that lyrics on the Urban Lyrics information are incorrect. To minimize this possibility, known spelling variants of words were also queried. Second, the lyrics could be incomplete: users contribute to lyrics websites at their leisure. However, there is currently no gold standard of hiphop lyrics: even *The Anthology of Rap* (Bradley & DuBois 2010) appears to have many erroneous lyrics (Sanneh 2010). Still, no better method is available. Additionally, the predictor variable measures use by particular artists and not frequency counts. It is much less likely that an error would be made for a given artist than for overall frequency counts.

7.3.3 Statistical analyses

The following section details the process by which statistical models were selected. See sections 7.3.3 and 7.3.3 for summaries of models described in the main text.

Information given	# participants	Data treatment
Foreign country	5	City, County, and CountyAA populations given values of 0
"Puerto Rico"	1	City and County given mean population values of Puerto Rican *municipios*; CountyAA population given value of 0
"USA"	2	mean sample values for City, County, and CountyAA populations for non-foreign participants assigned
"Small town in the middle of Iowa"	1	City, County, and CountyAA values given mean values of City, County, and CountyAA populations for Iowa, respectively
"suburbs"	1	mean values for City, County, and CountyAA populations for the Minneapolis-St. Paul metro area assigned
"California"	1	Mean values for City, County, and CountyAA populations for California assigned

Table 7.12: Treatment for data with anomalous hometown information (either foreign country/underspecified American city/state information). US census data does not divide Puerto Rico into first-order administrative divisions such as cities; a *municipio* is equivalent to a county. Because of this, city and county information for this participant are the same.

Model selection

At the beginning of this research, the 26 possible predictor variables given in Table 7.13 were potentially predictive of the response variable. Of course, not all these variables would turn out to be significant predictors, so a principled selection method was needed. Principal component analysis and factor analysis of the predictors are one way to reduce this number, but each is problematic. For example, in both of these methods, there is no relationship to the response variable.

To ensure that the findings were robust, three different methods were used:

1. A non-parametric dimension reduction technique (sliced inverse regression; see Li 1991, Weisberg 2002) to determine the predictors necessary for a linear model,

2. A linear model with response variable as an individual's summed vocabulary

Demographic	Social network	Musical preferences	Pop-culture
Sex (C)	MoveAA	Classical	Jay-Z (C)
Age	MoveAsian	Country	Barkley
City	MoveCaucasian	Folk	Boondocks (C)
County	MoveHispanic	Alternative	Mo'nique
CountyAA	MoveNative	International	Bieber
	MoveSAAME	Pop	
	Weekly	Rock	
		Vocal	
		Hiphop	

Table 7.13: Possible predictor variables for predicting AAE vocabulary knowledge. (C) = categorical variable; all other variables are numeric. SAAME = South Asian/African/Middle Eastern.

score,

3. A non-parametric random forest model (Breiman 2001). This model makes no assumptions about the distributions of the response or predictor variables.

The model reported in the main text is the linear model in (2) above, with predictors selected from the dimension reduction technique in (1) as well as a stepwise backward elimination using the Akaike Information Criterion (Akaike 1974) to eliminate superfluous predictors. This model is given in Table 7.14.

A dimension reduction analysis such as sliced inverse regression can indicate the number of linear combinations of the predictor variables necessary to adequately describe the response. For the present data, one linear combination sufficed; a visual inspection determined that this dimension was linear. Therefore, interaction and quadratic terms of predictor variables were not needed. The number of predictor variables was then reduced with the use of a stepwise backward elimination method for dimension reduction objects (Cook 2004). Significant predictors retained in the model were Hiphop, Country, MoveAA, Barkley, Boondocks, Mo'nique, and Jay-Z.

Finally, results from the random forest model, with conditional permutations to eliminate effects due to collinearity of predictors (Strobl *et al.* 2001), revealed that the predictors Hiphop, Country, Weekly, Barkley, Boondocks, Jay-Z, and Mo'nique are contributing to improved predictions for AAE comprehension score.

Model

The model reported in the main text is presented in Table 7.14. This model was chosen for its interpretability as well as for its generalizability, agreeing with the random forest model and (for the most part) the stepwise dimension reduction model. In this model, estimates correspond to β values, with the response variable being

participants' overall AAE vocabulary scores (range: 64-320). All variables except for Boondocks are linear; Boondocks is a categorical predictor with treatment (dummy) coding. The reference level for the binary Boondocks predictor is no knowledge of a *Boondocks* character.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	78.1890	5.2616	14.8604	0.0000
Hiphop	3.1670	0.9856	3.2133	0.0016
Country	-3.4141	1.4677	-2.3261	0.0213
Weekly	1.6055	0.7449	2.1554	0.0326
Barkley	7.4085	1.6839	4.3997	0.0000
Boondocks	18.1671	5.5942	3.2475	0.0014
Mo'nique	5.7439	1.7387	3.3036	0.0012

Table 7.14: Model coefficients — overall vocabulary score. N = 166.

Residuals from the model given in Table 7.14 were normally distributed, barring a few outlier datapoints. To check for undue influence of outliers, datapoints with standardized residuals > 2.5 were removed and the model was re-fit to the data. All fixed and random effects were still significant predictors, with the same direction and similar magnitudes, and residuals were normally distributed.

Another way of examining the significance of an individual predictor is to examine the individual contribution that predictor has in improving model fit. We can compare the model given in Table 7.14 to a model excluding an individual predictor: if there is a significant decrease in the F statistic as a result of the predictor being left out of the model, we should conclude that the predictor is helping to describe the data. Table 7.15 details differences in model fit when a given predictor is removed from the full model.

Predictor	F statistic	DF	p-value
Hiphop	10.32	1	0.0016
Country	5.41	1	0.021
Weekly	4.65	1	0.033
Barkley	19.36	1	< 0.001
Boondocks	10.55	1	0.0014
Mo'nique	10.91	1	0.0012

Table 7.15: Differences in model fit when a given predictor is removed from the model presented in Table 7.14. The F staistic indicates the change in degrees of freedom for number of predictors in the model. A significant result indicates the predictor is helping to explain the response variable (AAE comprehension vocabulary score).

The effect of preferred artists

The linear mixed-effects regression model given in Table 7.16 was fit to the data in the subsequent analysis on the influence of lyrics by participants' preferred hiphop artists. Estimates in the model are β values, with the response variable as vocabulary score for individual words (range: 1-5). As above, the reference level for the binary Boondocks predictor is no knowledge of a *Boondocks* character.

	Estimate	Std. Error	t value
Intercept	0.9756	0.1789	5.4527
Hiphop	0.0541	0.0250	2.1668
Country	-0.0833	0.0353	-2.3603
Boondocks	0.0851	0.0314	2.7078
Mo'nique	0.1240	0.0411	3.0159
# Overall Artists	0.0713	0.0114	6.2698
# Preferred Artists	0.0879	0.0358	2.4523

Table 7.16: Model coefficients — individual word vocabulary score. N = 81.

This model included random effects of Participant and Word, seen in Table 7.17. Including these random effects in models allows for higher precision modeling of systematic variation and reduces the chance of spurious fixed-effects findings (Baayen *et al.* 2008). The main effect of Trial (position of the item in the experimental list) was not significant, nor did this variable vary by participant.

Effect	SD	LR Statistic	p-value
Word	0.430	336.78	< 0.001
Participant	0.630	868.1	< 0.001

Table 7.17: Random effects for Recognition and Free Recall. I = Intercept (for word type, always acronym), B = Blend, D = Derived word. The likelihood ratio (LR) statistics and corresponding p-values were obtained by comparing the models with and without the given random effect substructure.

Residuals in this model were fairly normally distributed. Again, to check for undue influence of outliers, datapoints with standardized residuals > 2.5 were removed and the model was re-fit to the data. All effects were still significant, with the same direction and similar magnitudes.

7.3.4 Rater guidelines

For the response variable of word knowledge, no specific criteria were given for each word. The general guidelines as described above in section 4.2.3 were given to raters.

Because of a lack of specific guidelines, raters met to discuss each definition given, and the response variable reflects an agreement of the raters.

For the five popular culture questions, more specific criteria were available to readers for responses for each question. These criteria are given below.

- **Jay-Z**. Participants get a 1 if they say no, to someone other than Beyonce, or NS (not sure). They get a 5 if they say anything like "Beyonce" or "yes, to Beyonce". They get a 2 if they say "yes" but don't say to whom, and they get a 4 if they correctly say Beyonce but express some doubt, like a question mark or "to Beyonce I think".

- **Barkley**. Participants get a 1 if they say NS or BET. They get a 2 if they say something vaguely basketball-related but not more, and they get a 3 if they say any sports-related TV station like ESPN or golf. They get a 2 for TBS as that too is a Turner network but not the right one. Finally they get a 5 if they say TNT.

- **Boondocks**. Participants get 1 if they can't name any character and 5 if they can correctly name a character. If you see something marginal, such as information about the show without naming a character, feel free to give it a 2,3, or 4.

- **Mo'nique**. Participants get a 1 for NS and a 2 for "singer" as she technically did do some singing. Mo'nique was a TV show host and a comedian before starring in *Precious* an winning an Oscar for her role. If they list one of the following they get a 3: tv show host, actress, comedian. If they are list one of these but are unsure, give them a 2. If they list two of these, give them a 4. And if they list all 3 or can give details about any of these, as in "she was in Precious and she won an Oscar" then give them a 5. The more detailed info they can give the better.

- **Bieber**. Participants get a score of 1 for NS and one additional point for each Bieber song they can name. Also, song titles participants list that are almost right but not exactly (not only partially right) can get half a point.