The representation of frequent word combinations in lexical memory

Jason McDevitt
School of Communication Sciences and Disorders
Faculty of Medicine
McGill University
Montréal, Québec, Canada

Submitted December, 2006*

*A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Speech-language pathology.

©2006 – Jason McDevitt
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.
Abstract / Resumé

Many current psycholinguistic theories view the mental lexicon as a listing of (only) unpredictable sound-meaning correspondences (primarily words and morphemes). Under this view, regular complex word forms and syntactic structures are built with rules during language production, obviating the need for storage of complex but regular linguistic material. This type of model conflicts with recent experimental evidence that suggests that lexical memory may in fact consist of a more heterogeneous set of linguistic units, including complex word forms and multi-word expressions that in theory could be constructed via rules. Storage of such material seems to be driven largely by frequency. The present research consisted of two experiments designed to investigate whether semantically transparent noun and adjective phrases are stored as single lexical units when they are very frequent. Results from the two tasks (grammaticality judgment and speech production) supported the notion that frequent word combinations can come to be stored holistically in lexical memory. It was argued that usage-based models of lexical memory (vs. dominant generative theories) best account for such data. Data from language acquisition, aphasia, and corpus studies were offered as complementary evidence in support of the more general claim that a large component of linguistic competence is knowledge of lexical co-occurrence patterns. Finally, it was hypothesized that an exemplar-based model of lexical memory best captures the range of available data.

Plusieurs théories psycholinguistiques actuelles postulent que le lexique mental est constitué d'une liste de tout élément linguistique (principalement des morphèmes et des mots) dont la correspondance forme/sens n'est pas transparente. Selon ce type de théorie, les formes morphologiques et les structures syntaxiques qui sont complexes et régulières sont générées par des règles au moment de la production d'un énoncé, permettant ainsi d'éviter d'avoir à stocker en mémoire des formes complexes mais régulières. Or, cette théorie n'est pas compatible avec les résultats de plusieurs études récentes qui suggèrent que la mémoire lexicale est composée d'un ensemble d'éléments linguistiques plus hétérogène, dont des mots morphologiquement complexes et des expressions de plusieurs mots qui pourraient être générés au moyen de règles. Des données empiriques semblent indiquer que le stockage de ces éléments dépend en grande partie de la fréquence d'utilisation. Ce travail de recherche portait sur deux expériences visant à déterminer si des syntagmes nominaux et adjectivaux sémantiquement transparents sont stockés comme des unités lexicales unitaires lorsqu'ils sont très fréquents. Les résultats des deux tâches utilisées (juge ment de grammaticalité et production de la parole) appuient la notion selon laquelle des expressions composées de plusieurs mots peuvent avoir le statut d'unité lexicale, à condition qu'elles soient suffisamment fréquentes. L'interprétation avancée est que les théories de mémoire lexicale basées sur l'utilisation du langage (par opposition aux théories générativistes qui prévalent) expliquent mieux ces résultats. Des données d'autres domaines tels que l'acquisition du langage, l'aphasie et les études de corpus sont présentées comme des preuves additionnelles permettant d'affirmer de façon plus générale que la connaissance de la cooccurrence lexicale représente une composante importante de la compétence linguistique. Finalement, la théorie de mémoire du « modèle à exemplaires » serait la meilleure pour modéliser les connaissances lexicales.
Acknowledgements / Remerciements

As this adventure in learning draws to a close, I am struck by how much help I had along the way. First and foremost, I owe my greatest debt of gratitude to my advisor, Shari Baum, who always found time to help me out, even when she had lots of theoretically more important (but perhaps practically less interesting?) things to be doing. To my great delight and benefit, Shari is very good at a lot of things an advisor should be good at: giving encouragement but not cajoling, patiently explaining really basic things with expert knowledge but without condescension, and keeping a sense of humour in the face of deadlines, equipment failure, and other stressors. For these wonderful personal qualities, I am grateful, and this piece of work would never have come to fruition without her guidance. I also had the incredible fortune of having Debra Titone and Karsten Steinhauer on my thesis committee. They, too, are wonderful people whose expertise allowed them to give me the perfect combination of encouragement and constructive criticism. My research and analysis were both improved immeasurably thanks to their input.

It also gives me great pleasure to sing the praises of the research assistants who work in Shari Baum’s laboratory, specifically Erin Vensel and Meg Grant. They ran all the experiments for me and helped me manipulate raw data. Much of this took place while I was out of the country, but I knew I could have 100% confidence in their abilities and they never let me down. They also get a good deal of the credit (along with Lisa Coady, Shani Abada, Inbal Itzhak, and Effrat Pauker) for making the lab a fun work environment.

Finally, like so many graduate students before me (what can I say? truth wears grooves in language), I would like to thank family and friends for providing a network of social and emotional support that made three years of way more than full-time work bearable. Particularly, I would like to thank my speech-pathology classmates (what a great opportunity to have met some many wonderful new friends), Alexis Lenk (for providing distraction and domestic comfort at various stressful times) and my parents, brothers and sundry friends (for keeping tabs on me but learning quickly not to pester me with “When are you going to finish that thesis?”-style questions).
1. Introduction ........................................................................................................6
2. Multi-word lexical units ..............................................................................7
   2.1 - Some definitions ...............................................................................7
   2.2 - Types of multi-word lexical units .....................................................8
3. Theoretical background – Models, debates, considerations 12
   3.1 - Is there really a dictionary in my head? ........................................12
   3.2 - Different models of morphology in lexical memory ......................13
   3.3 - Frequency Effects ..........................................................................14
   3.4 - Updated models .............................................................................19
   3.5 - “Usage-based” approaches to the lexicon .......................................21
   3.6 - Models of lexical memory: Summary ...........................................25
4. Empirical investigations of frequent word combinations.................26
   4.1 - Psycholinguistic studies ................................................................26
   4.2 - Speech Production Studies ..............................................................29
5. Motivation for the Current Studies .........................................................32
6. Experiment One .........................................................................................33
   6.1 - Overview .........................................................................................33
   6.2 - Method .............................................................................................34
      6.2.1 - Design .......................................................................................34
      6.2.2 - Participants ...............................................................................35
      6.2.3 - Materials ...................................................................................35
   6.3 - Procedure ........................................................................................37
   6.4 - Results and Discussion ..................................................................38
7. Experiment Two .........................................................................................42
   7.1 - Overview .........................................................................................42
   7.2 - Method .............................................................................................44
      7.2.1 - Design .......................................................................................44
      7.2.2 - Participants ...............................................................................44
      7.2.3 - Materials ...................................................................................44
   7.3 - Procedure ........................................................................................46
   7.4 Results and Discussion ......................................................................47
8. General discussion ......................................................................................51
   8.1 - Psychological mechanisms .............................................................51
   8.2 - Speech and language development ................................................53
   8.3 - Language impairment in adults .......................................................54
   8.4 - Empirical studies of large corpora ..................................................56
   8.5 - The importance of constrained lexical co-occurrence in language competence ..................................................57
   8.6 - Wider implications for models of language and psycholinguistics ..................................................................................60
      8.6.1 - Increasing the role of lexical memory in linguistic competence – Impact on linguistic theory .................................................................60
      8.6.2 - Lexical memory – Exemplar-based? ...........................................61
8. Conclusions, future directions ...............................................................67
   8.1 - A quick recap ..................................................................................67
   8.2 - Some future lines of inquiry .............................................................71
1. Introduction

The research presented in this document examines how the human language faculty stores and uses frequent, multi-word chunks of language. The central hypothesis is that sufficient frequency of use can cause word combinations to be stored holistically in lexical memory, even when these word combinations are semantically transparent. This is not a common area of inquiry in psycholinguistics - a domain that has traditionally concentrated more on the syntactic extremes of word and sentence — nor is it the mainstream view of lexical representation (which views the mental lexicon as a list of words and morphemes with unpredictable sound-meaning correspondences, but typically not longer and/or straightforwardly compositional utterances). In recent years, however, there has been both theoretical speculation and empirical evidence suggesting that frequently encountered complex linguistic structures (including frequently collocated words) become “entrenched” in memory, behaving more like single units than sequences of individual lexical items. That being said, there are relatively few studies that have explicitly examined this claim in a rigorous fashion. Furthermore, among the existing studies that are relevant, the researchers’ methodologies are disparate, their experimental controls sometimes lacking, and their conclusions heterogeneous. Such observations motivate the two original experiments described below, which examine frequent word combinations in the context of visual language processing and speech production.

In the text that follows, I begin by introducing the notion of multi-word lexical units. This is a heterogeneous class of linguistic material and certain concepts, especially that of semantic transparency or compositionality, are relevant to the present research. Next, I present two earlier models of lexical representation that make very different claims about the nature of the linguistic material stored in memory. I then review some experimental evidence suggesting that frequency of use plays an important role in shaping psycholinguistic representation and processing, and explain how the first two models cannot accommodate these frequency effects. Finally, some alternate models of lexical representation are presented, including some approaches that explicitly incorporate usage factors, such as frequency, as central components shaping representation.

To introduce the two original experiments presented in this document, I first review a number of other studies that have examined the representation, processing and production of frequent word combinations. While I raise some points of criticism with some of these
studies, on the whole, the data are sufficiently suggestive of a special status for frequent expressions to justify further inquiry. The two experiments are then described. It is argued that the results of these experiments, somewhat parallel to the various studies of word-level morphology that will be reviewed (e.g. Alegre and Gordon, 1999; Hare, Ford and Marslen-Wilson, 2001) suggest that complex sequences of linguistic material (in this case, two-word phrases) can come to be stored in lexical memory, even if these sequences are regular and could thus, in theory, be constructed online via rules or rule-like processes. I argue that this finding supports the "usage-based" class of models of the lexicon (e.g. Langacker, 1987; see below), where the lexicon is viewed as a repository of conventionalized cognitive-linguistic routines. Finally, I speculate about some of the cognitive mechanisms that could underlie holistic storage of complex linguistic material and consider the implications of this phenomenon for psychologically realistic models of language.

2. Multi-word lexical units

2.1 - Some definitions

The present research considers how frequently encountered multi-word phrases are stored in lexical memory. Before going any further, it will be helpful to give some brief operational definitions of several key terms.

First, word is employed in the present work to refer to the type of linguistic unit typically listed in a traditional dictionary, namely a lexical unit with an identifiable syntactic category (noun, verb, etc.), that may have inflectional forms, that may be directly manipulated by linguistic rules or rule-like processes, etc. This contrasts with "multi-word phrases" or "word-combinations", which are discussed in detail below.

Another term employed throughout this work is lexical memory. This term is used where many other authors use "mental lexicon" and refers to a speaker's knowledge of lexical items and their linguistic properties. The term mental lexicon is avoided for several reasons. First, it evokes an image of a dictionary-style word list, a notion that is critically examined in this present work. Second, this special "mental dictionary" is often conceived as a component in a highly modular system (see section 3 below for further discussion of different theoretical accounts of lexical knowledge). While the present work does not directly address the question of linguistic modularity, it does speculate that there are important
interactions between language and general cognitive systems such as memory. As such, a term that highlights this connection is preferred.

2.2 – Types of multi-word lexical units

Terms such as multi-word phrase or word combination – terms I employ throughout this document – are intended to be merely descriptive, while avoiding claims about particular linguistic properties or endorsements of particular theoretical views. In the context of the experiments presented here, they refer to two-word noun and adjective phrases that are relatively semantically transparent. However, a range of multi-word lexical units have been studied and some important differences exist between them. Thus, a brief overview of some of the types of conventionalized expressions that have been identified in the linguistic and psycholinguistic literature is given here, with the goal of contextualizing both the specifics of the present studies as well as the more general discussion of constrained lexical co-occurrence and the lexical status of multi-word units that ensues.

Various authors have discussed the class of conventionalized word combinations alternately called fixed expressions (Moon, 1998), conventional expressions (Langacker, 1987), collocations (Manning and Schütz, 1999), or lexical phrases (Hunston and Francis, 2000) (see Van Lancker Sidtis (2004), Hunston and Francis (2000) and Moon (1998) for even more terminological variants). Such expressions represent a potpourri of linguistic material, consisting of (for example) idioms (spill the beans, under the weather), proverbs (Rome wasn't built in a day), frequently collocated terms (take care, shut up, salt and pepper), conventionalized speech formula (pleased to meet you, how's it going?), clichéd expressions (the pursuit of happiness, the American dream), memorized sequences (prayers, song lyrics) and proper names (George W. Bush). Needless to say, these expressions vary on a variety of structural, semantic and pragmatic properties, and a number of different theoretical characterizations of fixed expressions and idioms exist. Several aspects of this debate of relevant to the questions posed in the current research and are thus examined here.

Moon (1998) classifies “fixed expressions and idioms” into three main groups:

1) “Anomalous collocations” – These are fixed word combinations that are aberrant vis-à-vis typical lexical and/or grammatical patterns in the language. Consider for example the expression by and large; generally, prepositions and adjectives cannot be coordinated in English (other examples of this type include at all, stay put, etc.) Another type
of anomalous collocation consists of phrases containing lexical items that are not used elsewhere in free-combination in the language, such as the words *kith* in *kith and kin*, *shrift* in *short shrift*, *fro* in *to and fro*, etc.

2) **Formulae** – This label covers phrases that are routinely employed in a pragmatically-specified situation; they may be relatively semantically transparent but are ‘fixed’ in that they represent conventionalized ways of saying things (Mel’čuk (1995 *inter alia*) uses the term “pragmateme” to refer to a similar class of lexicalized expressions). Examples include such phrases as *can I help you?* [said by a store clerk], *I now pronounce you man and wife, I’ll get it* [meaning ‘I’ll answer the phone’], etc.. Moon (1998) also includes in this category phrases with special ‘discoursal function’ such as *not exactly*, sayings, quotations, catchphrases, and proverbs.

3) **Metaphors** – This category covers conventionalized figurative expressions with varying degrees of semantic opacity, ranging from somewhat conceptually transparent phrases such as *alarm bells ring, behind someone’s back, and pack one’s bags* to ‘pure’ idioms, such as *bite the bullet, kick the bucket, and shoot the breeze*.

As mentioned previously (and hopefully as illustrated by the typology and examples presented above), these multi-word lexical items are a heterogeneous class of linguistic items. In the now fairly extensive literature on idioms, collocations, and fixed expressions, a certain consensus has emerged among researchers in terms of criteria for the identification and classification of multi-word expressions. Manning and Schlüter (1999) suggest three identifying properties of such expressions: **non-substitutability**, **non-modifiability** and **non-compositionality**. Non-substitutability refers to the fact that certain word combinations are conventionalized, in the sense that certain lexical substitutions cannot be made, even if, in general, the words in question appear to have the same meaning. For example, one cannot use the phrase *yellow wine* to describe the beverage typically called *white wine*, even though ‘yellow’ is as good a (or arguably better) descriptor of the colour as ‘white’. Non-modifiability refers to the fact that certain multi-word units (and this is especially true for ‘pure’ idioms) are not easily transformable or modifiable in a way that one might expect based on the component lexical items. To use a rather classic example, the expression *to kick

---

1 Moon (1998) suggests very similar properties, referring to them as **non-compositionality**, **institutionalization**, and **lexicogrammatical fixedness**, respectively; I find Manning and Schlüter’s (1999) terminology more straightforward and thus use it here.
the bucket cannot be passivized (*the bucket was kicked by Grandma), even though kick is generally able to be put into the passive voice (the dog was kicked in the head by Grandpa). Finally, non-compositionality refers to the degree of semantic transparency of a multi-word expression. This point is perhaps the most relevant for the current research topic.

Manning and Schütz (1999) claim that “the meaning of a collocation is not a straightforward composition of the meanings of its parts” (p. 184). Moon (1998) makes a similar claim, stating that frequent co-occurrence alone is not sufficient evidence for the identification of fixed expressions, that they must also be some degree of non-compositionality, be it semantic or pragmatic. Nunberg, Sag and Wasow (1994), discussing what they refer to as ‘idioms’ (but what in reality encompasses various sorts of multi-word expressions, not just those in Moon’s (1998) “metaphor” category), claim that such expressions are “conventionalized”, in that “their meaning or use can’t be predicted, or at least entirely predicted, on the basis of a knowledge of the independent conventions that determine the use of their constituents when they appear in isolation from one another” (p. 492).

It would appear, then, that theoretical descriptions of multi-word lexical units are fairly consistent in identifying non-compositionality as a salient characteristic. To be sure, these expressions fall along a continuum of (non-)compositionality. At one end of this continuum would be pure idioms such as shoot the breeze, the typical use and meaning of the lexical building blocks of these expressions (e.g. shoot, breeze) has no semantic relationship to the sense instantiated by this particular word combination (‘talk informally to pass the time’). At the other end of the spectrum lie phrases that come very close to being the sum of their semantic parts, but have an added element of meaning (Manning and Schütz, 1999). For example, the meaning of the phrase weapons of mass destruction is close to being a straightforward summation of the meanings of its components. However, because this phrase is lexicalized (see section 8.5.1 for further discussion), it has accrued additional shades of denotation. It typically refers to chemical, biological or nuclear weapons and not, for example, conventional bombs, or heavy artillery, or cruise missiles, even though these are all weapons and do cause massive amounts of destruction.

2 This parallels common claims about morphologically complex (single) words, namely that only unpredictable or non-transparent forms are listed in the lexicon (see the next section for a more ample discussion).
As stated at the outset, the present research is a psycholinguistic investigation of
the lexical representation of common multi-word phrases. The question being asked is
whether the frequency of use of a word combination has an impact on whether the phrase
is stored as a whole in lexical memory. Frequency of use is independent of the semantic
and structural properties just discussed, in that frequent word combinations can be
entirely compositional, and non-compositional phrases may vary in frequency (e.g. shoot
the breeze is a more common idiom than trip the light fantastic). In fact, the non-
compositionality of many frequent word combinations represents a potential confound in
investigating usage effects. This is because it could be argued that non-compositionality
creates a logical need for speakers to memorize meaning-word combination pairings. To take
an obvious example, consider the case of ‘pure’ idioms, such as shoot the breeze. Because the
meaning of the combination of words is unpredictable (based on normal productive
linguistic mechanisms), the speaker has to store in memory the fact that (for example)
shoot+the+breeze maps to one meaning (unrelated to the component words). (The lexicon is
often claimed to be a list of all idiosyncratic sound-meaning correspondences; see section 3.1
below.)

The semantic unpredictability of non-compositional expressions of all types may
cause them to be stored in lexical memory. For this reason, in the research presented here, it
was decided that the best way to isolate the effects of frequency on lexical representation
was to focus on relatively semantically transparent frequent word combinations. While
expressions like shoot the breeze or trip the light fantastic may need to be memorized as wholes
because they cannot be analyzed as the sum of their parts, the same does not hold true for
last year or very good. If evidence is found for the holistic storage of these latter phrases, it is
more likely due to their sheer frequency, not to any element of idiosyncratic semantic
interpretation. It should also be noted that a claim for holistic storage of semantically
transparent phrases does not fit with some of the more theoretical and/or lexicographic
accounts (such as Moon, 1998 presented above) that argue that non-compositionality is a
necessary characteristic of multi-word lexical items. Finally, a caveat on this point: While an
attempt was made to use semantically transparent stimuli, it is very clear that semantic
transparency is a continuum, and that frequency of use often facilitates or perhaps even
drives the gradual acquisition of “additional elements of meaning” as discussed above. Thus,
even in the stimuli used in the present studies, there is probably some variation in degree of
semantic transparency (although presumably a fairly restricted amount; section 6.2.3 details the criteria used in stimulus selection).

3. Theoretical background – Models, debates, considerations

3.1 - Is there really a dictionary in my head?

The present research focuses on a claim about the nature of the linguistic material (specifically, word combinations) that the language system comes to store in long-term memory. It will be helpful, then, to situate this claim in the context of a brief discussion of some common views about the nature of lexical representation. When discussing how words are stored in the mind and brain, researchers typically use the term mental lexicon. The mental lexicon is often conceptualized as a speaker’s “mental dictionary”, namely a list of words of the traditional parts of speech (nouns, verbs, adjectives, etc.) paired with meanings and linguistic information, such as phonology (for example, Pinker (1999), a well-known work on the lexicon, employs the dictionary metaphor liberally). Psycholinguistic research has certainly expanded on this analogy, recognizing, for example, the necessity of including in the store of lexical knowledge syntactic information such as argument structure, as well as morphological information about derivationally and inflectionally related forms and the analysis and generation of new word forms. But the assumptions that 1) individual words are the privileged, organizing unit of storage in lexical memory (much as in a dictionary) and 2) words are typically retrieved individually from lexical memory and concatenated via syntactic rules remain fundamental for most researchers working in this domain (Jackendoff, 1995).

Given the common assumption that the lexicon is primarily a list of words, much of the theoretical debate around its content and architecture relates to morphology or, more specifically in the context of psycholinguistics, morphological processing in language production and perception. Morphological processing is not the focus of the current work and thus will not be exhaustively discussed here. However, some of the questions surrounding the type of morphological information represented in lexical memory and, more specifically, the effects of frequency of use on representation, are conceptually similar to those being asked about the frequent word combinations which are examined in the research.

3 In this document, I tend to avoid this term, as it is so tightly associated with the “dictionary” metaphor that is critically examined here. Ultimately, it will be argued that linguistic material (of various forms) that conventionally and/or frequently encodes a specific concept comes to be entrenched in long-term memory, and thus I prefer the term lexical memory when referring to the mental store of lexical units.
presented here. Thus, some of the different models of lexical access and representation that have been proposed to account for these phenomena are reviewed below.

3.2 - Different models of morphology in lexical memory

Butterworth (1983) and Pinker (1991) have proposed models that represent essentially opposite ends of the lexical representation spectrum. Butterworth proposes the Full Listing Hypothesis, namely that the lexicon explicitly lists all known word forms, including inflected forms that are entirely regular and predictable (e.g. *jump, jumps, jumped, jumping; dog, dogs,* etc.). In this model, novel words or word forms are handled (i.e. perceived or produced) via analogical induction from stored exemplars (although Butterworth also states that rules could exist as well, as a "fall-back procedure" (p. 263) for novel forms; he does not specify when analogy would be use vs. rules, or vice versa). This is an earlier model of morphology and the lexicon, and as such, it lacks some of the sophistication that later models have acquired as the amount of available experimental data has increased.

Butterworth motivates his Full Listing Hypothesis by comparing it to a hypothetical model where every morphologically complex word, including irregular inflectional forms and derivationally-related words, is rule-generated. He claims that, in such a system, as various lexical entries would necessitate the storage of idiosyncratic rules to produce their full range of morphologically related complex words (for example, the forms *be-is-was,* etc. are unique to the verb *be,* any computational advantage – in terms of reducing "memory load" – gained by not listing all possible forms would in fact be lost by requiring storage of the necessary idiosyncratic rules for each base form (p. 262) (note that most rule-based models of the lexicon – for example, Pinker’s (1991) model presented below – posit storage of irregular forms and not idiosyncratic lexical rules to produce e.g. *was* from the base *be*). Butterworth also reviews some evidence from speech production and reading experiments which may be considered consistent with his Full Listing Hypothesis (although by his own admission do not support it unequivocally). For example, he points to data from speech errors which indicate that transposition and stranding of word segments are not limited to morphemes (*contra* claims by Garrett, 1980, for example) and thus do not necessarily represent evidence for a morphological composition stage in speech production.
In stark contrast to Butterworth’s Full Listing Hypothesis, Pinker (1991) views the lexicon as a list of (only) unpredictable sound-meaning correspondences. Regular or predictable ways of forming words (for example, the regular verb inflections listed for *jump* above) are captured by rules and thus word forms like *jumps* do not need to be listed in the lexicon. Indeed, the only complex (i.e. multi-morphemic) words that need to be memorized are those that are unpredictable, such as irregular forms of verbs (e.g. *be*/*is*/*was*) and idioms (*kick the bucket*). As it turns out, this rule-based, maximally-economical view of the lexicon is widespread, especially among researchers influenced by generative linguistics, a school of thought with a commitment to structure-building rules. It can be traced back (at least) to seminal work by Chomsky, where he defines the lexicon as a listing of “the full set of irregularities of the language” (1965; p. 142). Models like Pinker’s (and indeed, several others discussed below) are modern-day psycholinguistic instantiations of this notion. Summarizing the current state of affairs (at least within the generative framework), Aronoff (1999) writes: “The prevailing model is that speakers each have a mental lexicon in which is stored every word of their language, inflected or derived, that has any unpredictable feature. Completely regular words are produced on the fly by productive patterns as they are needed and then discarded” (p. 563). In fact, this view has recently been taken even further; Stockall and Marantz (2006) claim that all inflectional forms are generated by rules (operating on morphological bases), including irregular verb forms.

### 3.3 - Frequency Effects

As it turns out, both of these ‘extreme’ views (Butterworth’s ‘list everything possible’ view vs. Pinker’s ‘list only what rules cannot generate’ view) fail to explain the full range of available experimental data. For example, Prasada and colleagues (Prasada, Pinker and Snyder, 1990; Prasada and Pinker, 1993) present data from a speech preparation latency experiment (a methodology used in the present work and discussed below) that suggest that irregular past tense forms of verbs are sensitive to frequency effects but regular past tense forms are not. In their experiment, the researchers showed subjects verb stems (e.g. *jump*, *sing*, *go*, etc.) on a computer screen and asked them to say the past tense form (*jumped*, *sang*, *went*, etc.). They found that for the irregular past forms, such as *sang*, the speed at which the

---

4 In fact, there is another dimension to Pinker’s model, namely that similarly sounding irregular forms (e.g. *sing*—*sang*, *sing*-rang*) are stored in “associative memory” and are available for processing of novel forms by analogy. This aspect is not relevant to the current topic and is thus not discussed further.
subjects produced the relevant inflected form was correlated with frequency (more frequent past tense forms were produced more quickly). No such effect was found for regular past tense forms like jumped. Prasada et al. argue that this result demonstrates that the irregular forms, by virtue of the fact that they are listed in the lexicon, can accrue frequency effects, whereas regular forms, which are handled on the fly by rules and thus not retained in long-term memory, cannot. Butterworth's Full Listing model, which hypothesizes storage in lexical memory for fully regular forms such as jumped along with irregular forms like sang, cannot account for the different behaviour of the regular and irregular verb forms.

The minimalist morpheme list and generative rules proposed by Pinker and his colleagues run into problems with certain experimental results as well. The experiment by Prasada et al. presented above notwithstanding, there is a significant amount of evidence to suggest that morphologically complex words with higher frequencies are processed differently than words with similar morphological structure but lower frequencies (see Ellis (2002) for a general review of frequency effects on language processing). Various researchers have interpreted these findings as indicating holistic storage of frequent complex regular forms. For example, Alegre and Gordon (1999) conducted a series of experiments in which participants were asked to perform a lexical decision task on visually-presented single words. The experimental stimuli were drawn from different parts of speech, (verbs, nouns and adjectives) and while word frequency differed, stem frequency was held constant across items, allowing the researchers to isolate effects due specifically to word frequency variations. They found that for nouns and verbs (the two parts of speech in their study that had inflectional variants), there was a certain frequency threshold above which complex forms were significantly faster in the lexical decision task. Alegre and Gordon concluded that "there are indeed whole-word representations available for regularly inflected forms" (p. 57), provided they are sufficiently frequent. Such a result poses a problem for both Butterworth's Full Listing Hypothesis as well as the generative rule-based model, as neither account predicts that some, but not all, regularly inflected forms are stored in memory, with storage conditioned by frequency.

5 "Stem frequency" (or, as Alegre and Gordon (1999) refer to it, "cluster frequency") refers to the combined frequency of all the inflection variants of a stem. For example, the stem frequency for jumps would be the sum of the frequencies for jump, jumps, jumped, jumping. In contrast, "word frequency" simply refers to the frequency of (only) the word form in question.
Another set of relevant experiments is presented in Hare, Ford and Marslen-Wilson (2001). They used an innovative methodology exploiting homophony to examine frequency effects and past tense inflection in English, taking advantage of the fact that various past-tense forms are homophonous with unrelated, monomorphemic words (e.g. made/maid, fined/find, ducked/duct, etc.). The task employed in their first experiment was a form of writing to dictation. Various homophones of the type just described (i.e., a phonological sequence ambiguous between a past tense verb form and an unrelated monomorphemic word) were recorded and played for the participants, who were given a short amount of time (6 seconds) to write a sentence containing the word they heard (the sentence allowed the researchers to determine which sense of the homophone the participant had used, as some of the stimuli, such as ground and spoke, were also homographs). Crucially, Hare et al. controlled for the frequencies of the different homophone senses, categorizing the stimuli as either dominant for the past tense form or the other homophone sense (they also had a group of stimuli with approximately equal frequencies). Some examples of their stimuli are given in Table 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Past tense form</th>
<th>Other homophone sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past tense dominant</td>
<td>allowed</td>
<td>aloud</td>
</tr>
<tr>
<td></td>
<td>made</td>
<td>maid</td>
</tr>
<tr>
<td>Other homophone sense dominant</td>
<td>fined</td>
<td>find</td>
</tr>
<tr>
<td></td>
<td>ate</td>
<td>eight</td>
</tr>
<tr>
<td>Approximately equal</td>
<td>ducked</td>
<td>duck</td>
</tr>
<tr>
<td></td>
<td>heard</td>
<td>herd</td>
</tr>
</tbody>
</table>

Given previous research demonstrating that, in a semantically neutral context, the more frequent meaning of a homophone will be accessed first, they predicted that in this task, the participants would tend to write the dominant (more frequent) homophone sense. To address the issue at hand – namely, whether frequent, morphologically regular forms are stored holistically in lexical memory – Hare et al. included both regular and irregular homophonous verbs (examples of both are given for the different frequency conditions in Table 1). As explained above, only forms which are stored in lexical memory should be able

6 Based solely on the frequency of the past tense form, not the stem frequency of the word.
to accrue frequency effects. Thus, in the present work, if regular past tense homophones, in
addition to irregular past tense form (which are uncontroversially assumed to be stored
holistically) show the "dominant meaning first" effect, this would provide evidence for their
storage in the lexicon. On the other hand, if only the irregulars show the effect, this would
suggest that the regular forms are not stored in memory, but rather built by rules during
speech production.

Hare and colleagues (2001) analyzed the percentage of past tense responses provided
for the various frequency conditions. For both the irregular and the regular forms, the
percentage of past tense responses was significantly greater in the past tense dominant
condition relative to the other conditions (homophone dominant and approximately equal).
Furthermore, there was a significant correlation between the percentage of past tense
responses and the log of the past tense/homophone frequency difference, indicating that the
effect size varied with frequency; the more frequent the past tense reading was, the more
likely that sense was to be chosen. Both of these statistical effects were stronger for the
regular past tense forms. The authors concluded that such results can be explained by an
account allowing for storage for both regular and irregular past tense forms but are more
difficult to reconcile with a model that lists only unpredictable (i.e. irregular) forms in the
lexicon.

Nonetheless, Hare et al. (2001) identified some potential confounds in their
experiment: For one, this was an off-line task and the results could in theory reflect post-
lexical access strategies. More importantly, they point out that "although a rule-based
account does not predict an effect of past tense frequency per se in the regular verbs, it does
predict that speed of access will be affected by the frequency of the verb stem" (pp. 190-
191). Thus, it is possible that the effect observed for the regular past tense forms is an effect
of stem frequency, not past tense frequency. To address both of these concerns, the authors
conducted a second experiment, this time employing a cross-modal immediate repetition
priming task. This experimental methodology (an online lexical decision task) reduced the
possibility that a given effect may be due to post-access strategies; participants heard a
spoken prime, then immediately at the offset of the prime, were shown a visual target for
lexical decision. In addition, for this experiment, the authors chose stimuli in such a way as
to avoid the confound of stem frequency. Specifically, they chose groups of 54 regular and
54 irregular verbs. In each group, 2/3 of the verbs had a past tense form that was
homophonous with an unrelated word (as in the last experiment). The remaining 1/3 of the verbs had unambiguous, non-homophonous past tenses (e.g. jumped). The homophonous and non-homophonous verbs were matched in frequency. In the lexical decision task, the visually-presented target was always the stem form of the verb, while the prime was either the past tense or an unrelated control. Table 2, adapted from Hare et al. (2001), gives example stimuli for the different conditions:

<table>
<thead>
<tr>
<th></th>
<th>Prime (homophone)</th>
<th>Visual target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>homophone</td>
<td>paced (paste)</td>
</tr>
<tr>
<td></td>
<td>unambiguous</td>
<td>jumped</td>
</tr>
<tr>
<td>Irregular</td>
<td>homophone</td>
<td>blow (blue)</td>
</tr>
<tr>
<td></td>
<td>unambiguous</td>
<td>slept</td>
</tr>
</tbody>
</table>

As in the first experiment, the homophonous verbs were classified according to whether the past tense or the other homophone sense was the more frequent (dominant) reading. Finally, the control primes were inflected nouns and verbs (nouns in regular plural, verbs in regular past tense) that were matched in frequency and syllable length to the past tense primes, but which had no phonological, morphological or semantic relation to the target verb.

The results of this second experiment were consistent with the pattern suggested by the first. For homophonous regular past tense forms (e.g. paced vs. paste), these forms were found to facilitate lexical decision for the related verb target (pace) when the past tense form was the dominant sense, but not when the other homophone sense (paste) was dominant. (The regular past tense forms of non-homophonous verbs like jump also had a facilitating effect.) Importantly, in order to verify that the priming effect was specifically related to the frequency of the inflected past tense forms, and not simply to verb stem frequency, the authors computed correlations between priming effects and the frequency differences between the homophones (e.g. paste) and both stem (pace) and past tense (paced) forms of the verbs. They found that “the Regular verbs do show a reliable independent effect of past tense frequency...demonstrating that the regular PT [past tense dominant] effect cannot be solely an artifact of stem frequency” (195).
Hare et al.'s (2001) results thus offer more evidence that predictable, linguistically-regular material is sometimes stored in memory, especially if it is sufficiently frequent. According to the authors “…these results suggest that lexical knowledge is highly redundant,” and that “regular past tense verbs, particularly frequent past tense verbs, are lexically represented in some fashion and, indeed, a number of models of lexical organization are consistent with these assumptions” (p. 196) (several such models are reviewed below). Importantly, for the discussion at hand, these data reinforce the notion that models of lexical knowledge must be able to incorporate the pervasive effects of frequency, even for forms that can, in theory, be generated by rules or other productive mechanisms.

3.4 - Updated models

The first two models of lexical knowledge presented, namely Butterworth's Full Listing Hypothesis and Pinker's rule-based account, fail to account for the full range of experimental data available. Specifically, neither model considers that one contributing factor in lexical representation is usage frequency and that thus some but not all complex word forms may be stored in lexical memory. In fact, however, the studies reviewed above (amongst many others) clearly indicate that frequency shapes the way linguistic information is stored and accessed.

In recent years, some more sophisticated hypotheses about lexical representation have been advanced. Most of these models allow for the storage of some complex words (e.g. frequent ones) as well as mechanisms for producing novel complex forms. One example of such a model is the Augmented Address Morphology (AAM) model developed by Caramazza and colleagues (Caramazza, Laudanna and Romani, 1988; Burani and Laudanna, 1992; Chialant and Caramazza, 1995, inter alia). In this model, upon reading (or hearing) a complex word, both the whole word access representation of the word (if it exists) and the access representations of its component morphemes are activated. These two types of activation (whole word vs. decomposition into component morphemes) compete, and the winner is the route that activates a lexical entry first. It is assumed that if a whole word representation is available, it is this representation that will typically be accessed first. This could explain, for example, the experimental effects discussed above where frequent regular forms are found to behave differently than less frequent ones. The AAM model allows for holistic storage for morphologically complex words that are “known”, but Caramazza et al.
do not speculate about how much experience with a complex word is required before it comes to be stored holistically. Presumably, this model is not fundamentally at odds with the notion that frequency is a determining factor in the storage of complex regular forms.

Schreuder and Baayen (1995) present a similar model of the lexicon. While they qualify their model as reflecting "the central role of semantic computation" (p. 132), equipped with "a mechanism for symbolic computations on meanings" (p. 133) (i.e. rules), like Caramazza et al.'s AAM model, they also attempt to account for the numerous word-frequency effects found in the psycholinguistic literature by allowing for the possibility that morphologically complex words can be stored holistically. They write: "[F]requency effects are autonomous in the sense that any word leaves a memory trace, irrespective of its morphological properties. Thus, we do not exclude that, due to repeated exposure, fully regular complex words (including inflections) may develop their own lexical representations" (133). In contrast to the AAM model, however, Schreuder and Baayen argue that semantic transparency plays a crucial role in shaping lexical memory. Fully semantically transparent forms — at least relatively simple ones — may sometimes have access representations, but never have full lexical entries (i.e. they never map directly to semantic and syntactic information). They explicitly endorse Pinker's (1991) notion that production of regular, semantically transparent complex forms is rule-based.

Finally, Ullman (2001, 2004) has elaborated what he calls the "declarative/procedural model" of the lexicon. This model was directly influenced by Pinker's (1991) work (Pinker collaborated on some early publications, e.g. Ullman et al. (1997)) and postulates a fairly strict (in this case, neurofunctional and to some extent neuroanatomical) separation of the "mental dictionary" of memorized words and morphemes and the "mental grammar" which contains rules for word and sentence formation. The declarative/procedural model fundamentally adheres to the classic Chomskyan view of the lexicon, also echoed in the earlier work by Pinker discussed above. Lexical memory "is a repository of all idiosyncratic word-specific information. Thus, it includes all words whose phonological forms and

---

7 Schreuder and Baayen (1995) do not specify the level of morphological complexity necessary to trigger full lexical representation. This is the extent of their discussion of this question: "Fully transparent words require minimal computation and do not acquire independent concept nodes. While we assume that high-frequency transparent complex words have their own lexical access representations, we share Pinker's (1991) intuition that rule-based composition applies to such transparent forms. Note, however, that as computational complexity increases, complex words tend to receive their own concept nodes even though they may be fully regular" (150).
meanings cannot be derived from each other... It also contains other irregular – i.e. not entirely derivable – word-specific information... and any unpredictable forms a word takes” (p. 233). Storage is also allowed for more complex linguistic material (including multi-word expressions) “whose meanings cannot be transparently derived from their parts (e.g. idiomatic phrases such as kick the bucket)” (p. 233).

In contrast to Pinker’s (1991) model, however, the declarative/procedural model allows for the storage of complex regular forms (for example, inflected verbs). Ullman (2001) states that “there is nothing preventing the memorization of any form” (p. 43). Storage of regular forms is significantly influenced by frequency, with high frequency items more likely to be stored in memory (Ullman, 2004, p. 245; Ullman, 2001, p. 43). However, it should be pointed out that Ullman views this process as exceptional. In a review of studies reporting frequency effects for regularly inflected words (Ullman, 2001), a finding that would suggest holistic storage of such forms, he is critical of the data, citing researchers’ failure to control for stem frequency in various experiments. However, he fails to review the work by Alegre and Gordon (1999) cited above where such frequency effects were found even when the controls he suggests were in place. Thus, it appears that the declarative/procedural model of the lexicon, while exhibiting a certain tolerance for memorization of frequent, compositional linguistic material, views this phenomenon as exceptional, significantly marginal so as not to call into question the traditional generative assumptions.

3.5 - “Usage-based” approaches to the lexicon

The discussion of the mental lexicon and the effect of frequency on word storage presented above serves to highlight two basic points. First, the nature of the various models explored illustrates the widely-held conception of the lexicon as fundamentally word- and morpheme-based (the three models presented in the preceding section are concerned almost exclusively with word forms). Second, there is an increasing body of experimental data that suggests that frequently used words may have some type of special status (for example, they may be stored and accessed as wholes from lexical memory). However, most of the more recent models of lexical representation, especially Ullman’s (2001, 2004) and Schreuder and Baayen’s (1995) models, have attempted to account for this without fundamentally altering the traditional framework of the lexicon as a maximally-economical wordlist.
In some respects, the models presented above are variations on a similar theme.

Some significantly different alternatives have been proposed, however. One model that is particularly relevant to the questions at hand is the view of the lexicon sketched by Langacker (1987) in the framework of “cognitive grammar”. This theory has as a central tenet the notion that structures larger than words and morphemes are stored in lexical memory. “Cognitive grammar posits a gradation uniting lexicon, morphology and syntax,” writes Langacker (1987, p. 36). Grammar is “a constantly evolving set of cognitive routines that are shaped, maintained, and modified by language use... It can be characterized as a structured inventory of conventional linguistic units” (p. 57). This is a significant departure from the models described above, in which the lexicon is conceptualized as a repository of items from the traditional parts of speech. In contrast, Langacker defines a ‘linguistic unit’ from the point of view of language use, not syntax, describing it as “a structure that a speaker has mastered quite thoroughly, to the extent that he can employ it in a largely automatic fashion. Despite its internal complexity, a unit constitutes for the speaker a ‘pre-packaged’ assembly” (p. 57). Most importantly, Langacker claims linguistic material of varying length and complexity can have unit status, including multi-word phrases. Langacker specifically discusses the storage of what he refers to as ‘conventional expressions’ as being a crucial part of a speaker’s linguistic competence. He points out that conventional expressions have traditionally “been excluded from the lexicon because they are larger than prototypical lexical items, and many are obviously compositional (even fully regular)” (p. 36).

Cognitive grammar, as outlined by Langacker, takes the view that any conventionalized linguistic structure can become entrenched in lexical memory. This would allow for the storage of not only irregular forms (necessarily stored irrespective of frequency because of their conventionalized irregularity), but of any linguistic material to which a speaker is exposed with sufficient frequency. This would include frequent, morphologically complex but regular forms (in conformity with the experimental findings presented above), even in the presence of general strategies (e.g. rules) for the production of such forms. In fact, cognitive grammar explicitly rejects what Langacker (1987) refers to as the “rule/list fallacy”, namely the notion that a linguistic unit is either listed in the lexicon or able to be produced by a rule, but not both. This theory would also allow for the holistic storage of the

---

8 Defined as “stock phrases, familiar collocations, formulaic expressions and standard usages” (35) such as great idea, tough competitor, mow the lawn, I don’t care, play fair (his examples).
frequent word combinations that are the focus of the current research. Langacker writes: “It is plausible to suppose that speakers master many regular plural forms as fixed units in addition to learning the general rule of plural formation. Similarly, it is clear that speakers learn as fixed units a large number of conventional expressions that are nevertheless fully analyzable and regular in formation” (42). In fact, it is an assumption of this theory that storage of particular exemplars is exactly what allows for abstraction of more general rules.

Many of the claims made about the lexicon by cognitive grammar are also present in a model of lexical representation outlined by Bybee (1985). This is primarily a model of morphological storage, based on evidence from a number of disparate areas, including cross-linguistic studies of descriptive morphology, diachronic language change, child language acquisition data and psycholinguistic studies. Although focused on word-level representations, the proposal has a number of features that relate at a general level to the discussion at hand. Like Langacker, Bybee characterizes her model of lexical representation as dynamic and susceptible to the influence of usage patterns. She writes: “Unlike a structuralist (including generative) lexicon, which is built on the metaphor of a dictionary, set down once and for all, unchanging, the lexicon I am describing here changes with use” (118). Similarly to Langacker’s rejection of the “rule/list fallacy”, Bybee explicitly rejects the notion of a binary classification of types of linguistic structures as “in the lexicon or not” (116). Rather, words and expressions with relatively high lexical strength (a concept comparable to activation level in psychological parlance) are likely to have autonomous lexical representation, even if they could be built by productive rules. Lexical strength is determined in part by frequency. Words in the lexicon (both basic and complex forms) are stored in paradigm-like clusters, based on phonological and semantic “lexical connections” (a morphological connection is defined as the convergence of phonological and semantic connections).

Bybee’s (1985) model can account for the psycholinguistic frequency effects presented above, as lexical strength encodes frequency (amongst other variables) and any form with sufficient lexical strength can be stored in the lexicon. She also argues that her model explains certain aspects of language change. For example, irregular verb forms with low token frequency have a tendency to regularize over time. For example, choose, shoot, and

---

9 Other determinants include semantics (semantically basic or unmarked forms are likely to have individual lexical entries) and irregularity (word forms that are so irregular that they cannot be derived through productive morphological rules must be stored as lexical items).
lose belong to a class of Old English strong (irregular) verbs that form the past tense through a certain type of vowel change. These verbs, which are relatively frequent, retain irregular past tense forms in modern English (i.e. chose, shot, lost). In contrast, less frequent verbs that belonged to the same class in Old English, such as see, smoke, and show, have come to form their past tenses in modern English with the regular -ed suffixation rule. Bybee illustrates the effect of frequency on diachronic change by comparing the Francis and Kucera (1982) frequencies of the two classes of verb (those that retained their irregular forms vs. those that became regularized). The irregular verbs had an average frequency of 140 while the regularized verbs’ average frequency was 22. Another example discussed by Bybee is the class of negative contractions in English (don’t, won’t, can’t, shouldn’t, etc.). She argues, following an analysis by Zwicky and Pullum (1983), that these items, despite being multimorphemic, have holistic lexical representation. She cites as evidence the absence of certain contracted forms (*may not -> mayn’t; *am not -> *am’), “phonological irregularities such as the unpredictable phonological shape of won’t and don’t” (122), and “semantic differences between contracted and non-contracted forms” (122) (not elaborated on). She writes:

“Undoubtedly, these contractions developed gradually from non-contracted combinations, and must have had autonomous lexical representations even when the semantic and phonological differences were minimal. In fact, it is possible that such frequent combinations may have their own lexical representation even when their derivation by combination is still thoroughly transparent. Otherwise, it is difficult to explain how and why contractions would develop. The changes that occur in them later must stem from the speakers’ acceptance of the sequences as autonomous units” (122-123).

The language change data presented by Bybee (1985) in support of her model certainly require critical examination. For example, the frequency based analysis of the irregular and regularized verbs relies on frequency data from a corpus of English from the 1960s. However, the morphological evolution of Old English strong verbs took place over hundreds of years, and it is not clear that the Francis and Kucera (1982) frequencies can be assumed to hold for, say, the 16th century. Of course, even leaving aside such concerns, the
way grammar has evolved historically is far from being a direct indicator of the psychological reality of present-day speakers. However, this type of evidence has rarely, if ever, been used to inform psycholinguistic models and thus its use here represents a unique and potentially important contribution, especially since it is merely one component of an analysis that combines multiple data sources (language acquisition, synchronic cross-linguistic analysis, psycholinguistic studies and language change).

3.6 - Models of lexical memory: Summary

The discussion above contrasts models of lexical memory on two different axes. On the one hand, it was shown that models differ in their ability to account for experimental findings suggesting regular, morphologically complex forms are stored in memory. Earlier models, such as Butterworth (1983) and Pinker (1991), did not allow for this possibility. Various other models do provide mechanisms for the storage of complex forms, although they differ in how they view this phenomenon. For example, Ullman (2001) views this as an exceptional, essentially last resort phenomenon in a generally rule-driven system, whereas Langacker's (1987) model gives a central role to storage of complex linguistic material. The other relevant axis of comparison is how the various models view the composition of the lexicon at a more general level. Most of the models reviewed above embrace, either explicitly or implicitly, the "dictionary metaphor", assuming that the linguistic material stored in lexical memory consists principally of single words from the traditional parts of speech. The models outlined by Langacker and Bybee are the exception to this mainstream view. They have as a core assumption the notion that units larger than single words are often stored in lexical memory.

The following sections begin to examine the lexical representation of frequent word combinations. Subsequent to a review of previous studies and the presentation of the current research, these important questions about the content and structure of lexical memory (i.e. effects of frequency, nature of lexical units) will be re-examined in light of the data pertaining to multi-word expressions.
4. Empirical investigations of frequent word combinations

4.1 - Psycholinguistic studies

There are several studies that have directly examined the processing and production of frequent word combinations. Harris (1998) reported two relevant experiments, both of which used a letter detection task similar to the one used by Reicher (1969) to demonstrate the so-called word superiority effect. In the original task, the subject briefly sees a string of letters. These letters may either form a word or a random sequence. After the initial brief exposure, the letters are masked (i.e. the image is degraded so the letters are unrecognizable) and the subject is asked to indicate whether a given letter appeared in the string. The finding that letter recognition accuracy is higher when the letters form a word is robust and has been replicated and expanded with different tasks (see Carr, 1986, for a review). A common interpretation of this finding, based on McClelland and Rumelhart's (1981) influential "interactive activation" model of word recognition, is that information from the processing of letters immediately begins to provide input into activation of word-level representations, and "this will produce a significant top-down influence onto letter-level representations, that is, increase activation for consistent letters and decrease activation for the inconsistent letters" (Balota, 1994; p. 311). This top-down letter activation and deactivation is presumed to underlie performance on the letter recognition task.

Clearly, an important assumption in the above analysis is the presence of word-level representations, i.e. the notion that orthographic words, although composed of strings of discrete characters (letters), are stored as unitary wholes in the mind. Harris (1998) extrapolates from this observation about words and letters and postulates that a similar claim can be made about collocations and letters. She hypothesizes that in the same way that readers recognize the "cohesiveness" of certain letter combinations (i.e. read them as words), they may also treat frequent word combinations as "unitized structures" (p. 56) and that this phenomenon may be observable on psycholinguistic tasks. Specifically, in the context of her experiments, she hypothesized that if letters are identified more accurately in the context of frequent versus random word pairs, then perhaps the same holds true for letter recognition in the context of frequent versus random word pairs.

Harris (1998) used three types of stimuli: frequently collocated word pairs (e.g. focal point, free world, night club), "collocation neighbours" (where one letter of each collocation is
changed to form an infrequent word pair, e.g. *vocal point, tree world, night clue*) and word pairs where one word from each collocation is randomly paired with another word (e.g. *cargo point, open world, night wall*). In the experimental task, the subject briefly sees a word pair. This word pair is then masked and two letters appear directly above a (formerly recognizable) letter somewhere in the word pair. The subject must indicate which of the letters was actually present. Harris found that letter identification was better in the collocation condition than in the other two conditions. In a second, related experiment, she used the same task, but designed some stimuli so that by choosing the incorrect letter in the letter identification task, a collocation would be formed. Thus, for example, if the presented word pair were “focal paint”, the letter choice would occur at the “a” position and would consist of “a” and “o”. She found that accuracy was significantly impaired in this “trick” condition (65% correct vs. 90% correct in the true collocation condition). In other words, subjects were misreading the “focal paint” type stimuli as their more frequent collocation neighbours (in this case, *focal point*). Harris interprets the results of these experiments as evidence for “collocation level representation” of the frequent word pairs (67).

These experiments are interesting, in that they are among the few psycholinguistic studies to directly examine the “holistic storage” hypothesis. However, there are some weaknesses in the methodology and analysis. The methodological weaknesses relate mainly to the construction of stimulus items. Items for the collocation condition were “familiar word pairs” from the Brown corpus (Francis and Kucera, 1982), defined as words that frequently occurred adjacent to one another in the corpus and that were judged to form units by native speakers. However, no actual frequency information is reported by Harris. It is unlikely that the one million word Brown corpus, which is tiny by the standards of modern day corpus linguistics, would be large enough to produce meaningful statistics for collocation (in comparison, the corpus used to construct the stimuli for the present work has 1.7 billion words; see section 6.2.3). This fact, coupled with the unelaborated-upon assertion that the selected collocations were judged to “form units” by native speakers (without specifying who the native speakers were, how many of them participated and what criteria they used to make this judgment), jeopardizes the empirically-grounded objectivity this type of experiment could have. Furthermore, Harris does not provide the details of any statistical analysis of her results, merely asserting (in one sentence) that her results support her hypothesis of holistic storage for the frequent word pairs. Finally, in the second experiment,
which contrasted performance on the letter recognition task with frequent relative to infrequent word pairs, the frequency of the individual words comprising the phrases was not controlled for. It is possible that letter recognition was facilitated for more frequent component words, independent of any effects caused by the collocation frequency. Thus, while these results are intriguing, they are best viewed as motivation for further research, as opposed to a definitive case for a usage-driven model of lexical representation.

Sosa and MacFarlane (2002) also looked at processing of frequently-occurring word combinations. In contrast to Harris' (1998) letter-recognition task, they used a word-monitoring paradigm where stimuli were word pairs containing *of*, taken from the Switchboard corpus (Godfrey and Holliman, 1993), a large\(^\text{10}\) corpus of spontaneous English speech in the form of telephone conversations. The independent variable was the collocational frequency of the word pairs, with the pairs being classified into four frequency groups, based on occurrence in the Switchboard Corpus. Table 3 gives examples of the stimuli for the four frequency conditions.

<table>
<thead>
<tr>
<th>Examples</th>
<th>High Frequency</th>
<th>High – Mid Frequency</th>
<th>Low – Mid Frequency</th>
<th>Low Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kind of, lot of</td>
<td>All of, couple of</td>
<td>Care of, because of</td>
<td>Sums of, colleague of</td>
<td></td>
</tr>
</tbody>
</table>

Subjects heard segments of spontaneous speech containing *of* collocations and were told to press a button whenever they heard the word *of*. Reaction time for high frequency collocations was significantly slower than for the less-frequent conditions. Furthermore, error rate for the high frequency condition was much greater than the other conditions. The authors conclude that these data "indicat[e] holistic storage for the frequent collocations and therefore impeded access to *of* as an independent word" (234). In other words, subjects had difficulty perceiving *of* in phrases like *kind of* and *sort of*, presumably because these strings, by virtue of their high frequency, have come to be stored as single units.

\(^{10}\) The corpus is comprised of 2400 5-10 minute phone conversations, collected from 543 speakers. It contains approximately 2.4 million words.
Sosa and MacFarlane's research overcomes some of the weaknesses present in Harris's (1998) studies. For example, they clearly indicate the frequencies of their stimuli, and they perform adequate statistical analysis of their results. A potential limitation of this work, however, is its focus on a very specific type of word combination (phrases of the form \( \text{WORD} + \text{off} \)). It is not immediately clear if the results of this research can be generalized to any and all word combinations (including, for example, combinations of open-class words of various parts of speech). Also, the authors do not explain why holistic storage of the word combinations they studied would decrease an individual's ability to access component words. Indeed, work on idioms has shown that semantic and syntactic processing of idioms (which are comprised of multiple words) occurs, despite their status as lexical units (see footnote 11).

4.2 - Speech Production Studies

In addition to the psycholinguistic investigations of frequent word combinations presented above, there are some studies that have considered the production of these structures in the context of spontaneous speech. For instance, Bybee and Scheibman (1999) focused specifically on the phonetic realization of the word \( \text{don't} \) in a corpus of spoken conversation collected from six speakers. They examined the degree of reduction of \( \text{don't} \); in terms of flapping of the /d/, reduction of the vowel, and deletion of /a/ and /t/. Each \( \text{don't} \) token was classified on a 4-point scale, where one end of the scale represented no reduction (i.e. realization as /dont/) and the other end represented maximal reduction (realization as a nasalized schwa). Bybee and Scheibman found that frequently occurring contexts for \( \text{don't} \) (e.g. I \( \text{don't} \) + frequent verb, such as know) were associated with more reduction (i.e. production of \( \text{don't} \) with more of the aforementioned phonological processes). For example, all tokens classified as maximally reduced occurred in the I \( \text{don't} \) + frequent verb context (the verbs were know, think, like, mean, feel). In contrast, \( \text{don't} \) tokens occurring with less frequent verbs (e.g. inhale, code) were produced in canonical form. The authors interpret these data as supporting their claim that frequent repetition of linguistic material "conditions chunking" and subsequent "autonomy in storage" for lexicalized chunks (p. 575-576).

Bush (2001) also studied the effect of collocational frequency on phonological processes in a corpus of spoken conversation. Specifically, he focused on palatalization of word-final /t/ and /d/ when the following word began with /j/ (as evidenced by the two
phonemes being blended together in production and realized as an affricate). Bush examined all word pairs that provided the necessary environment for this phonological process and, in a finding similar to Bybee and Scheibman (1999), determined that the probability of palatalization depended on string frequency (i.e. how often the two words appear together) and transitional probability (for a word pair, how likely the second word is to appear given the first word); higher rates of palatalization were associated with more frequent word combinations. Put more concretely, collocations such as “would you”, “could you”, etc., are more likely than infrequent word combinations (e.g. “would youth”) to be produced with an affricate. Bush hypothesizes that such collocations are stored and processed as single units — essentially as single words — increasing the likelihood of the palatalization/affrication process, which virtually always applies at the word level (for example, the verb ‘graduate’ is always pronounced /grædzuet/, never /grædjuet/).

A third study of the impact of collocation frequency on phonetic reduction in spontaneous speech was carried out by Gregory, Raymond, Bell, Fosler-Lussier, and Jurafsky (1999). These authors examined the occurrence of phonetic reduction in the Switchboard corpus (the same corpus used by Sosa and MacFarlane in their experiment on of phrases described above). Specifically, they were interested in the influence of word frequency and word predictability on reduction, focusing on three reductionary processes: word-final/t/ and /d/ deletion; tapping (flapping) of /t, d/; and durational shortening (i.e. reduction in the overall duration of the word). They found that one specific measure of probability, mutual information, was significantly correlated with all three types of reduction. (The mutual information of a word pair is a measure of how likely the two words are to appear together, taking into account the individual frequency of both words.) The finding is similar to that reported in Bush (2001): essentially, the more likely two words are to appear together, the more likely reduction is to occur. In contrast to the other studies presented above, however, Raymond et al. point out that their results, while compatible with the view that high frequency word combinations are stored holistically in the mental lexicon, need not be interpreted this way: “A probabilistic model… does not require that structural changes are necessary in the lexicon for this shortening effect [i.e. the reductionary processes explained above]” (p. 14). Rather, they conclude, quite straightforwardly, that “These results support a probabilistic model of production in which knowledge of the likelihood of words in contexts is used by speakers and affects their pronunciation of words” (p. 15). However, the authors
do not elaborate on how speakers would encode probabilistic knowledge of word co-occurrence, if not in the lexicon.

A final study of spontaneous speech production that examines the representational status of frequent word combinations was conducted by Backus (2003). In this study, Backus looked at code-switching in a corpus of spontaneous speech from Turkish-Dutch bilinguals. His general finding is that the phrases embedded in a matrix language (i.e. the 'main' or structuring language of the utterance) are often “multimorphemic lexical units...which blur the distinction between syntax and lexicon” (83). These multimorphemic units include plural nouns, compounds and, most pertinent to the present topic, adjective-noun and verb-object collocations. Backus explicitly adopts the hypothesis that many of the collocations are lexical units “called up whole from the...lexicon” (107). This analysis rests on the observation that when several words from the non-matrix language are inserted at the same time, these groups of words tend to be fixed phrases. For example, among the instances of verb + object constructions in the corpus where at least one word was in Dutch (n = 73), 10 of these utterances were produced with both the verb and the object noun in Dutch. Backus states that the majority (6/10) of these Dutch embeddings were collocations (e.g. ‘master a language’). This contrasts with the cases where a Turkish verb was produced with a Dutch object (n = 51), most of which (35/51) did not obviously correspond to any fixed phrase in Dutch. Backus states that these data “suggest that most Dutch transitive verbs that appear in the data are used because they are part of a larger unit also containing a noun. It is that unit that was selected” (111). Backus uses this main conclusion - that multi-word insertions in code switching frequently occur because the embedded words are retrieved as a single unit from the lexicon - to make a more general claim about the mental lexicon, a claim that echoes the dynamic, usage-driven view explained above. He writes that the “CS [code-switching] data give evidence for the view that the lexicon is not neatly decomposable into discrete word classes, containing minimal building blocks of larger structures. Though prototypical nouns, verbs, and prepositions obviously have an important place in the lexicon, they exist side-by-side with larger units, which combine morphemes in all kinds of syntactically possible ways and have achieved a degree of entrenchment high enough to be lexical units in their own right” (124).

While Backus (2003) draws some strong, clear conclusions from his work, his data and analysis do not necessarily project the same level of strength and clarity. The size of his
corpus appears small (although he gives no exact figures). No statistical analysis is performed on the data, he merely describes apparent trends and asserts his intuition about the findings (for example, discussing embedding of compound nouns, he states that “…about one-fifth of inserted nouns is [sic] a compound, surely a number that is higher than if compounds stood an average chance of surfacing as en EL [embedded language] insertion” (101)). His linguistic analysis of his data is at times ad hoc and subjective. For example, he lists verb-object combinations produced in his corpus, ordering them by “decreasing specificity of the verb” (111). It is not clear, however, how he determined that pour and drink are more “specific” than store and eat, respectively (or what this distinction would even mean, in semantic or conceptual terms). Finally, when assessing if a word combination used in a code switch is a fixed expression – a crucial point in his analysis, given his claim about multimorphemic units having lexical status - he is admittedly just stipulating based on intuition and does not avail himself of any empirical (e.g. corpus-based) or lexicographic studies. Thus, in general, while this study is intriguing, further experimental validation of its conclusions is necessary.

5. Motivation for the Current Studies

The research presented in the preceding section represents some preliminary evidence suggesting that factors such as frequency of use play a role in shaping the type of linguistic material that comes to be consolidated in lexical memory. The general pattern that emerges from these studies, which were conducted by different groups of researchers using varying methods, is one where, across different tasks and modalities, frequently recurring bits of speech behave differently than would be expected if they were also built up on the fly by stringing together individual morphemes. Yet some important caveats must be presented along with these studies. First, the various researchers do not all agree on the interpretation of their results. While most argue that their data necessitate a model where multi-word expressions are stored in the lexicon, Raymond et al. (1999) state that their data do not necessarily imply any “structural changes” to traditional views of the mental lexicon. Second, as discussed above, several of these studies used less-than-optimal methodologies. For example, a frequent problem was the assertion that a certain group of utterances or stimuli represented “fixed expressions”, “collocations”, “frequent phrases”, etc., without any
objective criteria for such a classification. Stated somewhat more generally, although many of
the researchers made claims about the impact of frequency of use on lexical representation,
objective data on lexical frequency were often lacking.

The combination of intriguing data and room for methodological improvement is a
good motivation for further research, and the remainder of this document will be devoted to
presenting two experiments designed to investigate the representation of frequent word
combinations. The first experiment is based on previous research done with idiom
processing and requires the participants to make a grammaticality judgment about visually-
presented phrases of varying frequency. The second experiment draws from prior
investigation of the representation of compound words in the lexicon. It considers the effect
of phrase frequency on the amount of time it takes a speaker to prepare the production of an
utterance containing the target phrase.

6. Experiment One

6.1 - Overview

This experiment compared the on-line processing of frequent vs. infrequent two-
word combinations by measuring the response time in a grammaticality judgment task. The
task used is an adaptation of a task devised by Swinney and Cutler (1979) to test the lexical
representation of idioms. Swinney and Cutler were asking a question similar to the one under
investigation here, but with regard to idioms as opposed to semantically-transparent,
frequent phrases. Namely, they were investigating if idiomatic expressions (such as buy the
farm, kick the bucket, eat crow) are stored as single lexical units. Their experiment consisted of a
grammaticality judgment task where participants were shown word strings and had to
indicate, as quickly as possible, if the words formed “meaningful, natural phrases in
English”. There were three types of word strings: idiomatic phrases, non-idiomatic
grammatical control phrases and ungrammatical distracter strings. (Table 4 gives examples of
the different types of stimuli.) The authors reasoned that “the task requires that candidate
strings be analyzed for sensibleness as a unit” and thus if idioms are stored as single lexical
units, “decisions made to idiomatic strings should be faster than those made to literal word
string controls”, since in the other conditions, multiple lexical units (i.e. words) need to be
analyzed before a judgment can be made (526).
### Table 4 – Swinney and Cutler (1979): Stimuli examples

<table>
<thead>
<tr>
<th>Phrase Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiomatic</td>
<td>Buy the farm</td>
</tr>
<tr>
<td>Grammatical (control)</td>
<td>Buy the car</td>
</tr>
<tr>
<td>Ungrammatical (distracter)</td>
<td>The buy green</td>
</tr>
</tbody>
</table>

The result of the Swinney and Cutler (1979) experiment was that reaction times in the phrase judgment task were indeed faster for the idiom condition. The authors concluded that “the results support a model in which idioms are stored and accessed as lexical items” (528).¹

The current experiment made use of a similar task. Participants were shown strings of words and were asked to judge if they formed acceptable English phrases. As in the Swinney and Cutler experiment, some of the phrases represented units hypothesized to have holistic storage – in this case, frequent word combinations. Grammatically acceptable phrases that are not frequently collocated served as a control condition. Importantly, these phrases were constructed by replacing one word from each frequent collocation with a word of equal or greater frequency (see the “Materials” section below for examples). Finally, there was a distracter condition of ungrammatical phrases.

### 6.2 - Method

#### 6.2.1 - Design

This experiment used a within-subjects, repeated measures design, with collocation frequency (frequent vs. infrequent) as the independent variable and response time as the dependent variable.

---

¹ Subsequent research has shown that the component words of idioms do undergo syntactic processing and that literal meanings of component words may also be activated (see Titone & Connine, 1999, for a review). In other words, idioms are not stored as unanalyzed “long words”. However, my assumption is that semantically non-transparent idioms are, by definition, stored “holistically” (which is not the same as being stored as an unanalyzed chunk), in that a speaker has to remember that this specific combination of words has a non-predictable meaning. It is presumably this fact that underlies the effect discovered by Swinney and Cutler.
6.2.2 - Participants

Forty native English speakers (29 female, 11 male) between the ages of 18 and 30 participated. The participants were all university students with no history of speech, language or hearing disorders and who had normal or corrected-to-normal vision. All participants completed the entire experiment and were paid for their participation.

6.2.3 - Materials

Approximately 150 phrases were selected quasi-randomly from the *BBI Dictionary of English Word Combinations* (Benson, Benson and Ilson, 1997), a specialized dictionary of grammatical and lexical collocations. The collocations chosen had the following characteristics:

1) They were all two words long, forming either noun phrases (e.g. *next week, emergency landing*) or adjective phrases (e.g. *mentally ill, sparsely populated*).

2) They were all semantically compositional, i.e. the meaning of the collocation as a whole is the sum of the component words. This excluded obviously idiomatic expressions such as *blue blood* (which is not actually blue), *kangaroo court* (which does not actually prosecute kangaroos), etc.

Subsequent to this initial selection, several further filters were applied. First, any word combination that appeared in the dictionary (*American Heritage College*) - i.e. as a compound word - was eliminated. Collocations were also checked against idioms listed in the *Oxford Dictionary of Idioms* (Siefring, 2004), again with an eye to ensuring that idiomatic (non-semantically transparent) expressions were not included; any collocation appearing in this dictionary was excluded. Finally, any collocations that consisted of two words that were clearly semantically associated were eliminated. (This resulted in the removal of collocations such as *infectious disease*.)

Frequency statistics (bigram frequencies) were obtained from the English Gigaword Corpus (Graff, 2003) for the remaining collocations. This corpus is a huge collection of journalistic-style text, collected in the 1990s. It contains approximately 1.7 billion words. Both bigram frequency (i.e. the frequency of occurrence of the two-word sequence in question) as well as the frequency of the individual words comprising the collocations were obtained. Frequency information was taken for the surface forms of the bigrams and words.
only; frequencies of inflectional or other morphological variants were not factored in. This means, for example, that if a word appeared in the plural in a collocation (e.g. human rights), the frequency of (only) this string was obtained (not e.g. human right).

Once these statistics were obtained, the fifty most frequent word combinations were retained (the frequencies for these bigrams ranged from 181769 to 1029). For these fifty word combinations, grammatical control items were constructed. Control items were formed by replacing one word in the original collocation with another word of the same part of speech with equal or greater frequency in the corpus. These replacement words also had the same number of syllables as the original word. The resulting bigrams had to be infrequent (defined as occurring less than 100 times in the corpus\(^\text{12}\)). Table 5 gives some examples of collocation and control items, along with their respective frequencies in the Gigaword corpus; frequencies for all collocations and their control counterparts are available in Appendix A.

Table 5 – Example stimuli and their frequencies

<table>
<thead>
<tr>
<th>Frequent bigram</th>
<th>Frequency</th>
<th>Infrequent control bigram</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>next week</td>
<td>80644</td>
<td>our week</td>
<td>43</td>
</tr>
<tr>
<td>very good</td>
<td>217544</td>
<td>very even</td>
<td>84</td>
</tr>
</tbody>
</table>

Finally, normative judgments were elicited from twenty subjects regarding the frequency of the fifty collocations and the corresponding control phrases. Twenty native English speakers, aged 18 – 30 with the same demographic profile as the experimental participants (but different individuals), completed a survey in which they were asked to rate, on a scale of 1 to 7, how frequently the two-word combinations (frequent and infrequent) occurred in everyday speech. The scores for each two-word combination were averaged across participants. A cut-off of 5 was established for “frequent” status, i.e. any of the high frequency word combinations that did not also receive an average rating of at least 5 in the norming task were not retained (this resulted in the elimination of three word combinations).\(^\text{12}\)

\(^{12}\) This criterion could not be strictly adhered to for three grammatical control phrases, given the extremely high frequency of the component words. However, for every pair of frequent and infrequent phrases, the frequent member of the pair was vastly (i.e. several orders of magnitude) more frequent in the corpus. Frequencies for all stimuli are available in Appendix A.
Similarly, a cut-off of 3 was established for the control phrases: any phrases not rated below this cut-off were not retained (this threshold, in turn, resulted in the elimination of 12 word combinations).

From this final list, the thirty most frequent collocations, along with their infrequent grammatical controls, were retained as experimental stimuli. Thirty ungrammatical word combinations were then constructed to serve as a distracter condition. These ungrammatical combinations were formed by changing one word in the original collocation to a word of another part of speech. For example, a noun-noun combination could be changed to an adverb-noun combination (e.g. next week -> originally week). Table 6 gives some example stimuli items for the three conditions.

<table>
<thead>
<tr>
<th>Collocation</th>
<th>Grammatical Control</th>
<th>Ungrammatical Distracter</th>
</tr>
</thead>
<tbody>
<tr>
<td>next week</td>
<td>our week</td>
<td>originally week</td>
</tr>
<tr>
<td>very good</td>
<td>very even</td>
<td>very pizza</td>
</tr>
</tbody>
</table>

For purposes of counterbalancing, the 90 trial items (30 frequent phrases, 30 grammatical but infrequent phrases, 30 ungrammatical phrases) were separated into 3 blocks, each block containing 10 different items from each condition. Order of presentation of the blocks was varied across participants.

6.3 - Procedure

Participants were tested during one session that lasted approximately 45 minutes. (During this session, they participated in both of the experiments described in this document.) The experiments were conducted on a Dell PC running the Windows XP operating system, using the Brown Lab Interactive Speech System (BLISS) software (Mertus, 2005).

For this first experiment, the participants were informed that they would see a series of phrases on the computer screen. They were instructed to make a “yes” or “no” decision...
regarding whether the words on the screen formed an acceptable phrase in English (the complete text of the instructions is available in Appendix B). Participants indicated “yes” or “no” by pressing a labeled mouse button. Button order (i.e. YES-NO vs. NO-YES) was alternated between participants. As explained above, stimuli were blocked and counterbalanced across subjects. For each participant, the software assigned a different random order of presentation within each block at run time. The stimuli were presented in 24 point font on a single line in the centre of the computer screen.

Before commencing the actual experiment, participants completed a series of three randomly presented practice items (one from each condition) and were given an opportunity to ask questions regarding experimental procedure. Practice items were all different from experimental items. Reaction times were equal to the latency between the appearance on the screen of the phrase and the participant’s response. Stimuli remained on the screen until a response was recorded. After 3900 ms without a response, the trial was considered a “no response” and discarded during analysis. The inter-trial interval was 4000 ms.

6.4 - Results and Discussion

The dependent variable in this experiment was response time, i.e. the time it took the participants to make a decision about the acceptability of the visually-presented two-word sequences. The reaction times for the three frequency conditions (frequent, infrequent and ungrammatical word combinations) were averaged and are presented in Figure 1 below. These figures reflect mean reaction times after the exclusion of any trials for which the participant did not produce a response within the allowed time (3900 ms), where a participant’s judgment was incorrect, or where a participant’s response time was greater than two standard deviations from his/her mean for the relevant condition. In addition, a software error affected the recording of reaction times for a small percentage (1.7%) of trials; these trials had to be excluded. Finally, four participants were excluded entirely from the final analysis, three on the basis of very high error rates (over 50% of trials in each case) and one due to hardware malfunction.
The response times shown in Figure 1 show a clear pattern whereby the participants were faster in judging the acceptability of the frequent word combinations than for the infrequent combinations. Repeated measures analyses of variance (ANOVA) were performed on these data, with collocation frequency as the independent variable. This analysis confirmed a significant main effect of collocation frequency on reaction time both by subject ($F(2, 70) = 41.881, p < .001$) and by item ($F(2, 87) = 53.868, p < .001$). Post-hoc analysis using the Newman-Keuls procedure revealed significant ($p < .01$) differences between the frequent word combination condition and the infrequent and ungrammatical conditions. No significant difference was found between the infrequent and ungrammatical conditions.

Accuracy data for the different conditions were also analyzed. (The accuracy data were analyzed only for the 36 subjects whose reaction time data were used in calculating the means above.) For the frequent word combination condition, participants produced an average of 2.5 erroneous responses out of 30 trials. For the infrequent phrase condition, on average 4.0 trials were erroneous (again, out of 30). A similar figure was obtained for the
ungrammatical condition, where participants produced an average of 4.2 erroneous trials out of 30. Statistically, these differences were not significant ($F(2, 70) = 1.01, p = .371$).

Importantly, these accuracy data provide evidence that there was no speed/accuracy trade-off in the experiment; the participants were most accurate on the condition for which they were also the fastest to respond, namely the frequent word combination condition.

The results of this experiment are consistent with the hypothesis that, given sufficient frequency of use, a multi-word utterance can come to have lexical representation. If this is the case, in processing such phrases, an individual needs only retrieve one morphosyntactically pre-assembled item from lexical memory, thereby reducing processing time in comparison with a grammatically analogous but less frequent word combination. This phenomenon is hypothesized to be reflected in the difference in reaction times on the grammaticality judgment task for the frequent and infrequent phrases. Stated differently, when an individual encounters a (relatively) novel chunk of language, s/he employs productive mechanisms (rules or rule-like processes) to analyze the linguistic material, identifying morphemes, syntactic relations, etc., and retrieve the individual words from lexical memory. The online processing time required for this analysis would be greater for the infrequent phrases in this experiment, resulting in longer reaction times versus frequently occurring word combinations that are stored holistically.

This is a similar result (and interpretation) to the one obtained by Swinney and Cutler (1979) in their study that compared idioms to other multi-word expressions. Recall that in a grammaticality judgment task, idioms were found to be processed more quickly than syntactically comparable novel phrases. Swinney and Cutler's interpretation (shared by many researchers working in this domain subsequently) was that the observed difference was due to the representational status of the idioms, namely the fact that they are stored as wholes in memory. In the context of work on idioms, holistic lexical representation of these items is typically motivated by the fact that they are not semantically transparent, i.e. their meaning is not predictable based on their component words and/or grammatical structure (note the parallel to irregular verb forms, typically assumed to be stored as separate lexical items because their morphological form is not predictable). Importantly, the present work focuses on phrases that have a significant degree of semantic transparency. Thus, in this case, storage in memory as a single unit may not be logically necessary, as is the case for idioms. Rather, it appears to be a result of high frequency of use.
At least one alternative explanation for these data is available, however. Making explicit grammaticality judgments is not a ‘natural’ task, in that speakers are not required to perform it in the course of normal communication. Such psycholinguistic tasks are constructed in the hope that in performing them, participants will call on more natural, automatic processes, whose effects can be inferred from the experimental results. Ultimately, however, the actual cognitive strategies used to accomplish the task (and thus the extent to which performance reflects ‘true’ psycholinguistic processing) remain opaque. In the idiom processing literature, for example, one area of debate is whether comprehenders have a special idiom-processing mode to deal with non-literal language (Titone and Connine, 1999). If this is the case, in the context of a task where participants are asked to process stimuli that include idioms, they may eventually get into ‘idiom set’. It is at least theoretically possible that for the present study, participants adopted some type of post-lexical cognitive strategy or mental set specific to frequent phrases and that this strategy underlies, at least in part, the observed effect.

To assess the presence of such an effect, the reaction times in the grammaticality judgment task were re-examined, this time separating the trials by order of presentation. As explained above, stimuli were blocked in groups of 30 (10 items from each condition per block). Table 7 shows the means for the different blocks (Block 1 representing the first 30 items presented to the participants, Block 2 the next 30, etc.). As can clearly be seen, the pattern whereby participants took less time to analyze the frequent word combinations is present even for the first block. This was confirmed by statistical analysis, which showed no effect of block when this was added as a factor in the analysis ($F (2, 105) = .037, p = .964$). The block-by-block analysis thus suggests that the observed effect of phrase frequency on reaction time is more likely related to the impact of their frequency on lexical representation, and not an artifact of a post-lexical, task-specific strategy.

<table>
<thead>
<tr>
<th>Block</th>
<th>Frequent</th>
<th>Infrequent</th>
<th>Ungrammatical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>965</td>
<td>1391</td>
<td>1323</td>
</tr>
<tr>
<td>2</td>
<td>951</td>
<td>1381</td>
<td>1288</td>
</tr>
<tr>
<td>3</td>
<td>982</td>
<td>1471</td>
<td>1249</td>
</tr>
<tr>
<td>Overall mean</td>
<td>972</td>
<td>1415</td>
<td>1296</td>
</tr>
</tbody>
</table>

Table 7 - Group means for reaction time by block
The data from this first experiment – which required subjects to read and analyze visually-presented phrases - are consistent with the hypothesis that frequent, semantically transparent word combinations can be stored as single lexical units. A second experiment, using a different task and communicative modality – speech production – was also carried out. Clearly, converging results from different tasks and modalities would compound the strength of empirical evidence. The next section describes the speech production experiment.

7. Experiment Two

7.1 - Overview

The second experiment examines the spoken production of word combinations of varying frequency. The studies of speech production presented above suggest that frequent phrases have distinguishing characteristics (e.g. different types and levels of phonological reduction), but as they are taken from general corpora of spontaneous speech, the researchers were not able to control directly subjects' production of the target lexical items. As such, the range of data, in terms of both types and tokens of the relevant structures, is limited.

The methodology for this experiment is based on a series of studies by Wheeldon and Lahiri (1997, 2002), who used a “prepared speech production task” in which subjects had to produce a variety of utterances as quickly as possible after hearing an auditory cue. Wheeldon and Lahiri (1997) found that the latency to produce a sentence was a function of the number of phonological, not morphosyntactic, words in the sentence. Wheeldon and Lahiri (2002) used this finding to investigate the representational status of compound words in lexical memory. Specifically, the authors were interested in whether compound words - which are themselves single morphosyntactic entities, despite being composed of two or more simple words - are treated as single phonological words during speech production, as evidenced by a shorter speech preparation latency vs. non-compounds (novel word combinations) of similar length and phonological form. In fact, their results revealed that compounds had shorter speech preparation latency than novel word combinations, and it is this finding that is relevant to the subject at hand. Stated generally, Wheeldon and Lahiri
were able to show that chunks of language composed of multiple words (i.e. compound nouns, in this case) can come to be stored and processed as single lexical items, at least at the level of speech encoding. The latency required at this encoding step is less for lexicalized multi-word utterances vs. novel word combinations, despite being phonologically comparable, e.g. in terms of number of syllables. If it can be shown that the frequently collocated words under consideration in the present research also exhibit shorter speech preparation latencies than comparable, less-frequent word combinations, this would be a key piece of evidence in favour of holistic storage for these constructions.

The speech preparation latency task used here is a modification of the one used by Wheeldon and Lahiri. Participants saw a phrase (e.g. next week) on the computer screen, and the task was to produce this phrase in the context of a carrier phrase (“I said X”). The dependent variable of interest is the time between the visual presentation of the phrase and the onset of speech (participants were instructed to produce the answer as soon as possible upon seeing the phrase). The relevant independent variable is the frequency of the visually presented phrase.

For this experiment, two types of control condition were necessary. First, there is the comparison of frequently vs. infrequently collocated terms; this is, of course, the crucial experimental contrast. For example, in the sample stimuli below, latency to produce a sentence with next week, a frequent word pair, would be compared with the production of a sentence containing a less frequently collocated word pair, such as our week. It is also necessary, however, to control for the inherent differences in phonological form between these two conditions. Simply put, it is possible that, given two different word pairs, production latency may not be the same due to differing levels of articulatory complexity. To isolate any such effect, two additional control conditions were created. The stimuli for these two conditions consist of word pairs that have one word in common (a different word than the one the frequent and infrequent conditions share, i.e. week in the example above). The other words in these phrases – i.e. the word in the phrase that they do not have in common – are the same words that differ across the frequent/infrequent stimuli (again, to continue with the same example used above, next/our). In contrast to the frequent/infrequent word pair conditions, however, both word pairs are infrequently collocated and thus any significant difference between them should be due solely to phonological/articulatory
differences. Examples of the four phrase conditions used in this experiment are provided in the “Materials” section below.

7.2 - Method

7.2.1 - Design
This experiment used a within-subject, repeated measures design, with collocation frequency of the two-word stimuli as the relevant independent variable. The dependent variable was the “speech preparation latency” (i.e. how long it took the participants to initiate speech production of the target stimuli), defined as the time between visual presentation of the target phrase and the onset of speech.

7.2.2 - Participants
The same individuals who participated in the first experiment (described above) also completed the present experiment.

7.2.3 - Materials
The frequent two-word combinations and infrequent grammatical controls from the first experiment were also used in the present experiment. However, as this experiment examined speech production, some additional phonological control conditions were necessary. As explained above, based on previous experimental work examining the production of compound words, the hypothesis underlying this task is that production of frequent word combinations that are stored holistically in lexical memory will require less preparation time. However, the relative articulatory complexity of an utterance may also affect its production latency. Thus, any observed differences in this variable between the frequent and infrequent conditions could be due to representational status in lexical memory or different levels of articulatory complexity between the stimuli (or both). To isolate any effect of this latter phenomenon, phonological control stimuli were constructed as follows.

Recall that the pairs of frequent and infrequent word combinations that constitute the relevant experimental contrast differ by one word (the infrequent condition stimuli were created from the frequent word combinations by changing one word for another of the same
part of speech and the same or greater frequency). Table 5 from the “Experiment 1 - Materials” section is reproduced below:

<table>
<thead>
<tr>
<th>Frequent bigram</th>
<th>Frequency</th>
<th>Infrequent control bigram</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>next week</td>
<td>80644</td>
<td>our week</td>
<td>43</td>
</tr>
<tr>
<td>very good</td>
<td>217544</td>
<td>very even</td>
<td>84</td>
</tr>
</tbody>
</table>

In these examples, the frequent word combinations, next week and very good, share one word with their corresponding infrequent control phrase (i.e. week and very). For the present experiment, for each frequent/infrequent stimulus pairing, two phonological control phrases were created. These phrases match the frequent/infrequent phrases in terms of phonological/articulatory distance, in that they share one word and differ by the same two words as in the original stimuli. Table 8 illustrates this:

<table>
<thead>
<tr>
<th>Frequent bigram</th>
<th>Infrequent control bigram</th>
<th>Phonological Control 1</th>
<th>Phonological Control 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>next week</td>
<td>our week</td>
<td>next dog</td>
<td>our dog</td>
</tr>
<tr>
<td>very good</td>
<td>very even</td>
<td>quite good</td>
<td>quite even</td>
</tr>
</tbody>
</table>

To illustrate, consider the first example in the table, the frequent word combination next week and its various control phrases. our week is the infrequent control (also used in Experiment One); the difference between the frequent and infrequent phrases is the next – our alteration (both phrases share the word week). The two phonological controls mirror these similarities and differences: one contains the word next and the other contains the word our and they both share one word, dog (which parallels week in the original two items). Thus, if the articulatory-phonological difference between next and our results in a difference in speech production latency, this difference should be present for both pairs of phrases (next week/our week and next dog/our dog).
As was the case for Experiment 1, the stimuli were counterbalanced across subjects. The 120 trial items (30 frequent phrases, 30 grammatical but infrequent phrases, 30 phrases each in the two phonological control conditions) were separated into 4 blocks and the order of presentation of the blocks was varied across participants.

7.3 - Procedure

Participants were tested during one session that lasted approximately 45 minutes. (As noted earlier, during this session, they participated in both of the experiments described in this document.) The experiments were conducted on a Dell PC running the Windows XP operating system, using the Brown Lab Interactive Speech System (BLISS) software (Mertus, 2005). A head-mounted condenser microphone manufactured by AKG Acoustics was connected directly to the computer. The experimental software recorded the voice input directly from the operating system.

For this task, the participants were told they would see a phrase appear on the computer screen. Following Wheeldon & Lahiri's (1997, 2002) methodology, they were instructed to say this phrase in the context of the carrier sentence “I said...”. In other words, if the stimulus phrase was next week, the subject would say: “I said next week”. They were asked to initiate production of this sentence as quickly as possible. (The full text of the instructions given to subjects is included in Appendix B.)

Before commencing the actual experiment, participants completed a series of 4 randomly presented practice items (one from each condition) and were given an opportunity to ask questions regarding experimental procedure (practice items were different from experimental items). Reaction times were equal to the latency between the appearance on the screen of the phrase and the initiation of a vocal response from the participant (i.e. the beginning of the “I said X” utterance). After 2900 ms without a response, the trial was considered a “no response” and discarded during result analysis. The inter-trial interval was 3000 ms. As explained above, stimuli were blocked and counterbalanced across subjects. For each participant, the software assigned a different random order of presentation within each block at run time. The stimuli were presented in 24 point font on a single line in the centre of the computer screen and remained on the screen until the participant initiated a vocal response.
7.4 Results and Discussion

In this experiment, the variable of interest was “speech preparation latency”, defined as the amount of time between the visual presentation of a phrase and the participant’s initiation of production of this phrase (in the context of a carrier sentence). Previous research has shown that speech preparation latency is influenced by the representational status of linguistic material. The speech preparation latencies for the four conditions were averaged and are presented below. In calculating the mean speech preparation latencies for the different conditions, any trial where a participant’s response time was greater than two standard deviations from his/her mean for the relevant condition was removed. One participant’s data had to be excluded entirely due to a hardware error.
The relevant comparisons in the figure above are between the frequent and infrequent phrase conditions and between the two phonological control conditions. Recall that it is hypothesized that if the frequent phrases are stored as single units in lexical memory, speech production latency should be shorter for these items versus the novel (infrequent) phrases. As can be seen from Figure 2, the mean speech production latency for the frequent phrases was indeed shorter than that of the infrequent phrases (1245 ms vs. 1260 ms). Turning to the phonological control conditions, recall that any differences in the speech production latency between the frequent and infrequent conditions that is due simply to differential articulatory complexity should be mirrored across these two control conditions. However, as can be seen, the means for the two phonological control conditions are quite close, differing by just 3 milliseconds.

Repeated measures ANOVAs were performed on these data, with phrase frequency as the independent variable. This analysis confirmed a significant main effect of phrase frequency on speech production latency both by subject ($F(1, 114) = 9.77, p < .001$) and
by item (F2 (3, 116) = 9.87, p < .001). Post-hoc analysis using the Newman-Keuls procedure revealed significant (p < .01) differences between the frequent and infrequent conditions. No difference was found between the phonological control conditions.13

As was the case for the first experiment, the significantly shorter reaction times for the frequent phrases are consistent with the claim that such phrases are stored in memory. If lexicalized frequent word combinations are retrieved as single units during speech production, online cognitive effort related to morphosyntactic productivity (e.g. inflecting words, applying syntactic rules or rule-like mechanisms to produce well-formed phrases, etc.) is reduced and the processing time required to prepare such an utterance should be shorter. In addition, to the extent that motor programs for lexical items (and thus for lexicalized word combinations) are stored in memory (this may underlie some of the effects observed in the speech production studies reviewed above), retrieving an already prepared routine should be faster than building one on the fly.

As this particular experiment examined speech production, differing levels of articulatory complexity across stimulus items represented a potential confound. However, there was no difference in speech preparation latency between the two phonological control conditions, which were constructed to mirror the phonological differences between the frequent and infrequent phrase conditions. Thus, the observed differences between these two latter conditions appear to be related to the effects of the differing frequencies of the stimuli in the experimental conditions.

As was mentioned in the discussion of the results from Experiment 1, examining the first block of stimuli presented to the participants in relation to later stimuli may provide insight into whether the experimental results reflect, at least in part, the development of task-specific cognitive strategies that are independent of lexical access. A block-by-block examination of means, similar to the one done for Experiment 1, was also carried out for this experiment. The results are summarized in Table 9 below.

---

13 As can be seen in Figure 2, the frequent and infrequent stimuli both had shorter speech preparation latencies than the phonological control conditions. No particular theoretical claim is advanced to explain this. One possibility is the fact that the participants had already been exposed to the frequent/infrequent stimuli in the first experiment (all participants completed the experiments in this order), thereby lowering activation thresholds for the lexical items comprising these expressions.
Table 9 – Group means for speech preparation by block

<table>
<thead>
<tr>
<th>Block</th>
<th>Frequent</th>
<th>Infrequent</th>
<th>Phonological match - Frequent</th>
<th>Phonological match - Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1238</td>
<td>1242</td>
<td>1292</td>
<td>1281</td>
</tr>
<tr>
<td>2</td>
<td>1232</td>
<td>1256</td>
<td>1240</td>
<td>1261</td>
</tr>
<tr>
<td>3</td>
<td>1241</td>
<td>1274</td>
<td>1296</td>
<td>1294</td>
</tr>
<tr>
<td>4</td>
<td>1274</td>
<td>1268</td>
<td>1277</td>
<td>1286</td>
</tr>
<tr>
<td>Overall means</td>
<td>1245</td>
<td>1260</td>
<td>1277</td>
<td>1280</td>
</tr>
</tbody>
</table>

These figures are somewhat more variable than those observed in Experiment 1. Importantly, however, the speech preparation latency for the frequent phrases remained short (actually below the overall mean) for the earliest-presented stimuli. This fact suggests that the participants did not develop task-specific “frequent phrase” modes or strategies, since the relevant result is present even for the first stimuli. Statistical analysis again bears this out; an ANOVA showed no significant effect for block number when this was added as a variable in the analysis ($F(4,210) = .59, p = .67$).

A potential limitation of this experiment was the fact that the same participants performed both the experiments described here (all participants were asked to complete the grammaticality judgment and the speech production experiments in that order). This means that when performing the second experiment, they had recently been exposed to some stimulus items (the frequent phrases and their infrequent controls were used in both experiments). This likely explains why the speech production latencies are faster for the frequent and infrequent conditions, relative to the two phonological control conditions, the latter consisting of entirely novel (i.e. not previously presented) phrases. This repetition could have, in theory, introduced a biasing effect, since the participants responded more quickly to the recently-presented stimuli. However, because both the frequent and the infrequent phrases were used in both experiments, any such effect (i.e. of recent exposure) should apply equally to both types of stimuli. Thus, as in Experiment 1, the differences between these two conditions are likely related to their disparate frequencies. No direct comparisons were performed between the frequent and infrequent phrases used in both experiments and the phonological controls used exclusively in Experiment 2. Overall, then, as in Experiment 1, the data remain compatible with the hypothesis that the frequent phrases are stored and retrieved as single units.
8. General discussion

The results from both of the experiments presented above appear to support the hypothesis that frequent word combinations can come to be stored in lexical memory. Two different methodologies - one involving grammaticality judgments, the other speech production - were used to show that frequent phrases behave differently than novel word combinations, even when the two are linguistically comparable (i.e. except for frequency).

In the remainder of this document, some potential implications of this finding will be considered from several different angles. First, a hypothesis about the psychological mechanism(s) underlying this phenomenon will be advanced. Next, some complementary data from other domains (language acquisition, aphasia, and studies of text corpora) are presented and argued to lend further support to the central hypothesis of holistic storage for frequent phrases. Finally, some general considerations for models of lexical memory – in light of the present findings – are discussed; it is proposed that models of lexical knowledge need to reflect the prevalence of lexical co-occurrence in language, as well as the demonstrated impact of usage factors such as frequency in shaping representation.

8.1 - Psychological mechanisms

As discussed above, there is a certain practical necessity for semantically non-transparent idioms to be stored as wholes. For the phrases used in the current research, this is not the case. What, then, would drive them to be stored in lexical memory?

The answered proposed to this question highlights my preference for the term “lexical memory”. Specifically, I propose that the storage of these high frequency phrases is a consequence of how memory consolidation takes place generally. In the course of language use (speech production and perception, reading, writing), linguistic information passes through working memory. With repetition, memory traces for specific bits of information are strengthened, facilitating consolidation in long-term memory. Importantly, my claim is that this consolidation can take place for chunks of linguistic information of varying lengths and descriptions, from the single-word level to multi-word expressions of various kinds (including the type of phrases examined here, but also, for example, the modal verb + negative contractions discussed by Bybee (1985) above, non-propositional fixed expressions such as pleased to meet you, etc.). In essence, I am extending the claim about word-level
representations made by Schreuder and Baayen (1995) (presented in section 3.4) to linguistic material generally. Recall that Schreuder and Baayen argue that "any word leaves a memory trace, irrespective of its morphological properties. Thus, we do not exclude that, due to repeated exposure, fully regular complex words... may develop their own lexical representations" (p. 133). It is hypothesized that in fact the preceding statement could be reformulated, replacing "word" with "utterance" (i.e. "any utterance leaves a memory trace, irrespective of its linguistic properties..."). A model of lexical representation that limits lexical status largely to single words would seem to imply the existence of a neuropsychological gating mechanism that screens memories, in essence saying to the brain, "You can remember this bit of information, because it corresponds to one morphosyntactic word, but you're not allowed to store this phrase, even though you say it and hear it all the time." Such a mechanism seems implausible.

Amongst the models of lexical representation presented above, the so-called "usage-based" models most readily accommodate a view of the lexicon as a structurally heterogeneous repository of memorized linguistic routines. Indeed, this description corresponds quite closely to Langacker's (1987) proposal. While less obviously concerned with phrase-length linguistic material, Bybee's (1985) model also seems to be able to accommodate this view, given its focus on the frequency-based notion of lexical strength (and not morphological structure) as a prime determinant in lexical representation. (Bybee has subsequently explicitly endorsed the notion of multi-word expressions in the lexicon, for example in Bybee and Scheibman (1999), reviewed above.)

A model of lexical memory that provides for the storage of frequent multi-word expressions can account for the data from the present research, as well as the various studies reviewed in section 3 above. Models that postulate storage only for material that cannot be generated by rules end up making some very strong (and arguably psychologically implausible) claims, namely that there is either a selective mechanism blocking storage of linguistic units that do not fall into one of the atomic categories (e.g. noun, verb, etc.) conceived of as feeding morphosyntactic rules or that general principles of memory do not apply to linguistic information. Abstracting beyond the experimental data presented in this document, however, is the central hypothesis of holistic storage for frequent phrases consistent with what we know about language based on other domains of inquiry? Below, I will (briefly) consider data from three distinct sources (studies of language development in
children, language impairment in adults, and corpus studies of collocation in written texts) that all seem to offer support to the hypothesis.

8.2 - Speech and language development

It has often been observed that children in the early stages of language development sometimes memorize multi-word phrases as single chunks (e.g. Brown, 1973; MacWhinney, 1978; Peters, 1983; Plunkett, 1993). Plunkett (1993) claims that “children may ascribe lexical status to whole sequences of words and use these lexical chunks with the same distributional properties as adult lexical items” (44). These lexicalized chunks have been referred to as “amalgams” by MacWhinney (1978) and “formulaic expressions” by Peters (1983). Various distributional, phonological and frequency-based criteria have been suggested for the identification of these lexicalized chunks. For example, one piece of evidence comes from the observation that the morphosyntactic knowledge necessary to produce the phrases is not readily apparent elsewhere in the child's output (Hickey, 1993; Plunkett, 1993; Locke, 1997). For example, an English-learning child may say “what's that?” before s/he has a general rule for producing wh- questions, or before s/he has acquired the copula is/’s. This phenomenon has been observed cross-linguistically (see Brown, 1973 for English; Plunkett, 1986, 1993 for Danish; Hickey, 1993 for Irish) and it is has been argued by Locke (1997) that this type of “utterance accumulation” underlies grammatical development, in that the child first stores (i.e. commits to memory) a critical mass of data from which productive patterns can subsequently be extracted. These language acquisition data suggest that children may store complex chunks of linguistic material in long-term memory, in this case, perhaps as part of a learning strategy. However, Peters (1983) claims that this phenomenon is not restricted to children developing language, but that (adult) speakers continue to store multi-word units and access them during speech production. The data from the experiments presented above lend support to Peters' claim. More generally, the fact that two very different sources of data (child language acquisition and psycholinguistic studies of adult language processing) point to a similar psycholinguistic mechanism strengthen the notion that entrenchment in lexical

---

14 Ellis (2002) makes a similar claim about L2 acquisition, namely that to become a competent speaker of a second language, one must master multi-word “constructions”. He writes that “the knowledge underlying fluent use of a language is not grammar in the sense of abstract rules or structure but a huge collection of memories of previously experienced utterances...” (p. 166) and that “the acquisition of grammar is the piecemeal learning of many thousands of constructions and the frequency-based abstraction of regularities within them” (p. 168). See also section 8.6.2 below for more discussion of the idea of exemplar-based lexical memory.
memory of frequent chunks of speech is a robust, central phenomenon in the psychology of human language.

8.3 - Language impairment in adults

The way that cognitive processes break down subsequent to neurological damage can inform us about the way they typically function in neurologically intact individuals (Coltheart, 2001). Specifically, in the study of language, researchers often look at the way that communication is impaired in aphasic individuals to inform theories of language representation and processing (Obler and Gjerlow, 1999, pp. 1-2). In the context of the present discussion about the representational status of frequently occurring phrases, several aspects of language impairment in aphasia are potentially informative. First, various studies (see Duffy and Coelho, 2001, for a review) demonstrate that the frequency of lexical items can be reflected in the patterns of impairment exhibited by aphasic individuals. For example, Schuell et al. (1961) tested the auditory comprehension of 48 aphasic subjects and found that low-frequency words were associated with decreased auditory comprehension. This effect has also been found at the sentence level. Shewan and Canter (1971) showed that decreasing frequency of words in sentences was correlated with decreased accuracy and speed in processing those sentences. Frequency effects are also found in studies of speech production by aphasic individuals. For example, various researchers have shown that performance on naming tasks is correlated with lexical frequency, with high-frequency items being easier to produce (Shuell et al., 1964; Gardner, 1973; Williams and Canter, 1982). Thus, all in all, these studies indicate that, as has been shown repeatedly for non-brain damaged individuals, word frequency has a pronounced psycholinguistic effect in aphasic patients.

Crucially for the subject at hand, frequency effects on processing in persons with acquired language impairment are not limited to the individual word. For aphasic individuals, especially in the case of nonfluent aphasia, linguistic material alternately described as “automatic”, “non-propositional” and/or “overlearned” (to list just a few) has often been observed to be relatively less impaired versus other types of language (Van Lankcer Sidtis, 2004), so much so that evaluation of it is specifically targeted in subtests of standard aphasia assessment tools (e.g. The Boston Diagnostic Aphasia Examination (Goodglass, Kaplan and Barresi, 2001) and The Western Aphasia Battery (Kertesz, 1982)). For example, some aphasic individuals can produce greetings, expletives, conversational fillers, or ‘automatic sequences’
(numbers, days of the week, poems or song lyrics, etc.) with performance that is better than would be expected based on their production of novel utterances. In the production of this type of linguistic material, verbal fluency or articulatory accuracy may be greater, lexical access problems may be fewer (or absent), and/or otherwise problematic grammatical morphemes may be used.

There are undoubtedly a number of different neuropsychological and neurolinguistic factors that underlie the sparing of automatic sequences (and the like) in aphasia and clearly, a memorized poem (for example) is likely a qualitatively different type of lexical knowledge from, say, a verb. For example, many memorized sequences (e.g. numbers, lists, the alphabet, the days of the week) are non-propositional and have no obvious syntactic structure, and thus impairment to syntax may have a relatively minimal impact on this type of linguistic material. In contrast, more ‘traditional’ linguistic units that are stored in lexical memory (e.g. words, phrases) may be the objects of productive linguistic operations (inflection, insertion into larger syntactic structures, etc.) in a way that a poem or list may not. However, in the context of the present discussion, this phenomenon is presented to highlight, once again, the fact that speakers, based at least sometimes on frequency of occurrence, commit to memory various multi-word stretches of speech. The ability to recall, for example, an entire poem, or to state the months of the year, when accessing single words in spontaneous speech is generally impaired due to brain damage, illustrates the heterogeneous nature of the linguistic material stored in long-term memory and the impact that frequency can have on psycholinguistic processes.

One final study that is especially relevant was carried out by Van Lancker and Kempler (1987). They compared aphasic individuals’ comprehension of idiomatic expressions with novel utterances which were matched to the idioms on length, syntactic structure, and frequency of the individual words (in some respects this mirrors the stimuli used in the two experiments described above). They found that comprehension was superior for the idiomatic expression versus the novel phrases. The authors attributed this result to the fact that the idiomatic expressions were stored as lexical units; holistic representation facilitates lexical access in that the individual needs only retrieve one item from memory.
8.4 - Empirical studies of large corpora

In recent years, the increasing speed and processing capacity of computers combined with the proliferation of textual information in electronic form has allowed researchers to examine linguistic patterns across large amounts of real-world data. One of the more robust findings in the field of corpus linguistics has been the importance of collocation, convention (speech communities tend to communicate particular thoughts and concepts in conventionalized ways) and lexically-base patterns.

An example of this type of research is presented by Hunston and Francis (2000), who describe a lexically-based “pattern grammar”. Their basic observation is that most (if not all) words tend to occur in lexicalized local patterns involving other words and/or grammatical frames. Hunston and Francis illustrate this point with the noun matter, in the sense of the word occurring in the following phrases (the lemma matter is, of course, ambiguous):

- a matter of developing skills
- a matter of learning a body of information
- a matter of becoming able to reason coherently...

They point out that this sense of the word essentially only occurs in the frame a matter of + -ing, and argue that “there is little point in treating matter, in the sense in which it is used here, as a single lexical item that can be slotted into a general grammar of English. Rather, the word comes, as it were, with its attendant phraseology, which in this case consists of ‘a __ of –ing’” (2). They go on to claim, more generally, that “a large amount of language encountered is not constructed from ‘basic’ structures and a lexicon, but occurs in sequences of morphemes that are more or less fixed in form” (7).

Stubbs (2001), also using corpus data, comes to similar conclusions about the prevalence of lexicalized word combination patterns. He writes that “corpus research shows that words are typically used in routine phrases, and that even the most frequent words have typical collocates and typical uses” (58). To cite but one of Stubbs’ many examples, he provides a corpus-based analysis of the collocation patterns of the word commit. He found that “of over 3000 occurrences of the word-form commit in the 200-million-word corpus, 15 per cent co-occur with the single word suicide, and nearly 30 per cent co-occur with one of

---

15 In Hunston and Francis (2000), they are analyzing a particular chunk of text from which these examples are drawn.
only three lemmas [suicide, crime(s), murder]" (64). While his book is mainly a corpus-based investigation of lexical semantics, Stubbs does make some claims about what this type of information implies for psycholinguistics, drawing a direct link between this type of empirically observable collocation pattern and the important fact of idiomaticity in human language (he quotes Searle’s (1975) conversational maxim “Speak idiomatically unless there is some special reason not to” (59)). He states:

“It is implausible that [idiomatic or fixed phrases] are created individually on each occasion of use. There are conventional ways of saying conventional things...Many other perfectly grammatical ways of saying the same things are conceivable, but people just don’t say them...Every native speaker has thousands upon thousands of multi-word units stored in memory” (58-59).

The notion of linguistic conventionalization is considered in more detail in the next section.

8.5 - The importance of constrained lexical co-occurrence in language competence

This document presents ample evidence that multi-word units are of interest in the study of language. However, dominant theories of linguistic competence have generally overlooked this aspect of linguistic knowledge. For example, in North America, one of the most influential approaches to linguistics is Generative Grammar, a school of thought which has also heavily influenced much contemporary work in psycholinguistics and neurolinguistics (even if researchers working in these fields do not explicitly endorse generative theory). Generative Grammar conceives of linguistic competence largely as a set of abstract, symbolic rules designed to produce novel utterances. Research focuses on defining and refining rules to account for the range of grammatical/morphological/phonological structures observed\textsuperscript{16} in a language, or in language generally. In fact, however – and the current work should be one more piece of evidence in

\textsuperscript{16} Actually, because research in this paradigm is not typically carried out via systematic examination of data (cf. the corpus-based studies reviewed above), but rather via speaker intuition about targeted structures, “observed” is perhaps not the most accurate word here. It might be more accurate to say: “Research focuses on defining and refining rules to account for the range of grammatical/morphological/phonological structures hypothesized to exist in a language, or language generally”.

57
this direction – a huge component of linguistic competence is knowledge of how specific lexical items combine in idiomatic ways, above and beyond abstract structure.

Consider the following examples. *hearty*, in the context of *hearty drinker/smoker* is an intensifier meaning, roughly, “someone who does a lot of [action denoted by noun, i.e. drinking, smoking]”. However, *hearty* cannot be always be used in this context (i.e. to modify deverbal nouns with the -er ending). While at the syntactic level, ‘adjective + noun’ is a licit structure, one does not say *hearty flyer* to mean someone who frequently travels by plane (in fact, there is another lexicalized ‘intensifier + noun’ phrase to denote this concept: *frequent flyer*), *hearty gardener* to refer to someone who spends a lot of time cultivating vegetables, etc.

Another example is the “light” or “delexicalized” senses of the verbs *take*, *make* and *have*, in expressions such as *take a trip, make/take (British) a decision, take/ have a shower, make a choice, have a heart attack*, etc. While these verbs have the same syntactic distribution (they take a noun phrase as a direct object) and have the same lexico-semantic role (they verbalize a noun that denotes an act or process), the specific combinations are lexicalized. In English, it is unidiomatic to say *have a trip, make a heart attack, take a choice*, etc.17

Various frequent phrases have, in theory, perfectly grammatical, roughly synonymous expressions that are rarely or never used. For example, consider the following pairs:

<table>
<thead>
<tr>
<th>weapons of mass destruction</th>
<th>arms of large-scale damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>separation of church and state</td>
<td>division of religion and government</td>
</tr>
<tr>
<td>it’s a pleasure to meet you</td>
<td>it’s pleasant to meet you</td>
</tr>
</tbody>
</table>

In the left hand column are frequent expressions of various types. In the right hand column are phrases that are, in theory, synonymous with the frequent phrases on the left. However, the phrases on the right are rarely or never used18. To be a competent speaker of English, one needs to know that particular combinations of lexical items are conventionally used to express certain concepts and/or in certain situations. And, clearly, those conventionalized expressions must be stored in memory (the capacity to list synonymous

17 Further evidence that these combinations are lexicalized in English and form part of a speaker’s competence is cross-linguistic variation; In French, for example, one “makes a heart attack” (*faire une crise cardiaque*) – it is not idiomatic to say “have a heart attack” (*avoir une crise cardiaque*). Erroneous use of light verbs is a hallmark second language intrusion in non-native speakers’ language production.

18 A quick and dirty way to get a handle on this (and to avoid reliance on intuition) is to enter these phrases into Google or any other web search engine that provides statistics on number of pages returned (the phrases should be entered surrounded by quotation marks, to ensure that only documents that contain the exact phrase will be returned). For example, on June 7, 2006, “weapons of mass destruction” returned 27,000,000 hits, whereas “arms of large-scale damage” returned zero.
expressions – i.e. as is done above in the right-hand column – illustrates the logical necessity of storing the idiosyncratic conventional expressions).

The phenomena described above – namely, lexicalized intensifiers and light verbs, and conventional expressions – are examples of what Mel'čuk (1984, 1988, 1995) has referred to as “constrained lexical co-occurrence”\(^\text{19}\). Mel'čuk has proposed a linguistic theory, Meaning-Text Theory, which has as a central component the formal lexicographic description of constrained lexical co-occurrence for each individual lexical item in a given language. This description is accomplished via the construction of “explanatory combinatorial dictionaries” that list “lexical functions” (and their values) for the entries in the dictionary. For example, the intensifier relationship expressed by heavy in the example above (heavy smoker) is encoded by the lexical function Magn. In the explanatory combinatorial dictionary, the entry for smoker would contain the value for this lexical function, i.e. Magn (smoker) = heavy. Numerous types of lexical functions exist, expressing various productive semantic relations for the different parts of speech. For example, consider the following examples (adapted from Mel'čuk, 1984).

\[\text{Table 10 – Examples of constrained lexical co-occurrence (after Mel'čuk, 1984)}\]

<table>
<thead>
<tr>
<th>Lexical Function</th>
<th>Gloss</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magn</td>
<td>very, intensely, to a high degree</td>
<td>Magn (smoker) = heavy Magn (desire) = burning Magn (poverty) = abject</td>
</tr>
<tr>
<td>Excess</td>
<td>Function excessively, to the point of being abnormal</td>
<td>Excess (motor) = race Excess (heart) = race, pound</td>
</tr>
<tr>
<td>Pos</td>
<td>Conventionalized expression of positive evaluation</td>
<td>Pos (review) = positive Pos (reward) = handsome Pos (trial) = fair</td>
</tr>
<tr>
<td>Son</td>
<td>Produce a typical sound</td>
<td>Son (dog) = bark Son (floor) = creak Son (stomach) = growl</td>
</tr>
</tbody>
</table>

\[^{19}\text{My translation of the original French term “la cooccurrence lexicale restreinte”}\].
In theory, the full set of lexical functions\footnote{Mel'čuk (1984) lists 36 basic lexical functions, but many of them can combine with one another in productive ways, to produce new functions.} can exhaustively describe all the lexical co-occurrence patterns in the language.

Meaning-Text is a linguistic theory rooted in lexicography. It is not a theory of language processing or psycholinguistic representation. However, in the context of the present discussion, it serves to highlight the breadth of the knowledge speakers have about specific combinations of lexical items, and the centrality of this knowledge in the language system. Simply put, constrained lexical co-occurrence is part of a native speaker’s competence and, as such, necessitates the storage of lexical co-occurrence patterns in memory. The experiments described in this document lend empirical support to the hypothesis that speakers do in fact store frequent word combinations.

8.6 - Wider implications for models of language and psycholinguistics

8.6.1 – Increasing the role of lexical memory in linguistic competence – Impact on linguistic theory

Wray (2002) points out that the findings reviewed above (i.e. that a large amount of speech is “formulaic”, that speakers have a great deal of knowledge about lexical co-occurrence patterns and conventional expressions, that many lexical items occur in restricted, particular patterns and constructions, etc.) represent a challenge for Generative Grammar and related theories of language, which have long viewed linguistic “creativity” and the mechanisms for creating novel utterances as essentially the only interesting objects of study. As well, as should be clear from the discussion of such models above, theories of lexical memory inspired by Generative-style linguistics focus on the storage of single words and morphemes and the rules used to generate complex word forms; knowledge of constrained lexical co-occurrence and lexically-based grammatical patterns is simply not in the picture.

Interestingly, it could also be argued that Hunston and Francis (2000) and Stubbs’ (2001) observations about lexical co-occurrence patterns in corpora are part of a larger trend in linguistic theory towards a “strong lexicalist” position, where increasing amounts of linguistic competence are located in the lexicon, and the grammatical component is a more stripped-down, general-purpose computational system. While Chomskyan generative
grammar maintains a strict modular approach, with grammar and lexicon as separate components, other linguistic theories are increasingly collapsing lexis and grammar (Bates and Goodman, 1997). Hunston and Francis' pattern grammar is one such theory. Construction Grammar (Fillmore, Kay & O'Connor (1988), Goldberg (1995)) is another, where "all elements of linguistic form are represented within a heterogeneous lexicon that contains bound morphemes, free-standing content and function words, and complex phrase structures without terminal elements" (Bates and Goodman, 1997, p. 2). Bates and Goodman present data from language acquisition, aphasia and language processing experiments with normal individuals, with the goal of arguing for the "inseparability" of grammar and the lexicon in theories of psycholinguistics, neurolinguistic and language development.

All in all, there appears to be movement from various points of departure (psycholinguistics, computational linguistics, linguistic theory), based on data gathered with diverse methodologies, towards the consensus that humans store large amounts of their linguistic knowledge in lexical memory. The results of the present study support and strengthen this idea. Importantly, they also introduce the notion that usage factors, such as frequency, may play a role in this phenomenon.

8.6.2 - Lexical memory – Exemplar-based?

The nature of exemplar-based models of language

In preceding sections, I have reviewed previous experimental results and presented new data suggesting that frequent word combinations can come to be stored holistically in lexical memory. A parallel was drawn between this finding and evidence demonstrating that complex word forms with regular morphology are stored in memory, provided they are sufficiently frequent. It was hypothesized that consolidation of linguistic material (of various types) in long-term memory happens automatically with sufficient exposure (i.e. frequency of use) and that there is no neuropsychological or neurolinguistic mechanism in place to block storage of utterances having specific linguistic characteristics (e.g. those with multiple morphemes or multiple words). Subsequently, a broader discussion of frequent phrases and other lexicalized multi-word utterances was undertaken, and it was shown that a) memorization of such sequences is a salient feature in language development in children, b) this type of material can be relatively less affected in acquired language impairment in adults.
(perhaps due in part to its high frequency) and c) corpus studies of large amounts of real-world linguistic data highlight the prevalence of lexical co-occurrence patterns. Finally, a brief review was given of a linguistic theory (Meaning-Text Theory) that has as a central objective the exhaustive description of lexical co-occurrence in languages. Research carried out in the Meaning-Text framework demonstrates the extent of this co-occurrence information in linguistic competence.

From the claims and data just summarized, an important role for memory emerges. Stated generally, competent speakers/hearers store large amounts of complex linguistic material in long-term memory. This notion is at odds with the Generative-style theories of the lexicon that I have examined critically throughout the present work. In light of the data presented here, these theories would appear to overestimate the role of novel computation in language production and perception, and underestimate the role of memory. Lamb (1998) (cited in Wray, 2002) writes: “It is...easy to overlook the role of human memory in information processing. Linguists seem to underestimate the great capacity of the human mind to remember things while overestimating the extent to which humans process information by complex processes of calculation rather than by simply using prefabricated units from memory” (p. 169). In the context of such claims, it is appropriate to speculate about the nature of memory as it relates to models of lexical representation. Specifically, I suggest that the data presented in this document lend support to an exemplar-based model of lexical memory.

In exemplar-based, or “episodic”, models of the lexicon, each individual usage episode (be it during perception or production) is stored in memory, at least initially (like any memory, the trace may decay with time if it is not reactivated and reconsolidated). These stored traces are the basis of a speaker’s linguistic knowledge; similar exemplars are stored in clusters and with the accumulation of a sufficient number of similar traces, regularities or patterns emerge, allowing, for example, for the production of novel utterances. Importantly, usage information - be it details of phonetic realization, or lexical co-occurrence information - can be associated with particular lexical items (Bod, 2006; Pierrehumbert, 2001). Another interesting aspect of exemplar-based systems relates to their ability to account for the demonstrated effects of frequency on representation. As Pierrehumbert (2001) points out, because this type of model links specific memory traces to specific linguistic material,
frequency is inherently encoded, in that frequency directly impacts memory (i.e. repeated exposure lowers activation thresholds, strengthens traces, etc.).

The notion of linguistic knowledge consisting of a large amount of stored examples has, of course, an important impact on the type of assumptions one makes about the psychological processes underlying language use. For example, Goldinger (1998) applied this model to the question of speech perception. He hypothesized that when an individual hears a word, “all stored traces are activated, each according to its similarity to the stimulus. The most activated traces connect the new word to stored knowledge, the essence of recognition” (p. 251). This contrasts with the more common view in psycholinguistics, namely that details of individual instances of perception are not retained, and that phonological representation in the lexicon consists instead of abstract, canonical representations of words extracted from a sea of irrelevant acoustic variability (see Gaskell and Marslen-Wilson, 1996 for one such model).

Interestingly, as Goldinger (1998) points out, exemplar-based and canonical/prototype-based models “arose as philosophically opposite accounts of common data” (p. 263). In the context of speech perception, both approaches attempt to account for what is sometimes referred to as “speaker normalization”, namely the fact that hearers are typically able to understand novel instances of speech, produced by any speaker, despite variance in the actual acoustic realization of words, differences in voice quality, the presence of extraneous noise, etc. (Hawkins, 1999). Mainstream canonical representation models handle this signal variability by postulating a prototypical lexical representation that retains an abstract phonological core (e.g. as a sequence of phonemes or phonological features), independent of incidental phonetic detail. A problem for this type of model, however, is the finding that people in fact do remember details about specific instances of speech and language use, including word presentation modalities and details (such as font in visually-presented words), exact wording of sentences, voice, intonation and pitch (Goldinger (1998) reviews relevant studies for these various parameters).

Evidence from speech production is arguably even more compelling in demonstrating that speakers recall specific phonetic details about particular lexical items. In arguing for an exemplar-based model of speech production and perception, Pierrehumbert (2001) writes that “speakers have detailed phonetic knowledge of a type which is not readily modeled using categories and categorical rules of phonological theory” (p. 137). Specifically,
she points to two revealing linguistic phenomena as evidence for her claim: the existence of "lexical classes" and frequency effects on phonetic realization.

The notion of lexical class is an important one in sociolinguistics, dialectology and historical linguistics. It refers to a phenomenon whereby a phonological process or phonetic variation applies to some but not all lexical items meeting certain phonological criteria. For example, in the dialect of English spoken in Boston, words that have the orthographic sequence [or+Consonant] fall into one of two lexical classes21, depending on their phonetic realization. In one class are words that consistently have the [or] sequence pronounced as /o/, i.e. with the characteristic r-lessness of the dialect; for example, border is pronounced /braʊ̯ ðə/ (monosyllabic words also have a schwa off-glide, so that e.g. fort is realized as /foʊt/). Words in the other lexical class, on the other hand, have a 'dialectal pronunciation' where the [o] vowel is sometimes realized as /aw/, i.e. a low, back, slightly-rounded vowel, (again, monosyllabic words have a schwa off-glide). This class contains words such as short, produced as /ʃɔːt/ (a low, back, slightly-rounded vowel), which actually makes short and shot homophonous (Laferriere, 1979).

Table 11 (adapted from Laferriere, 1979) shows some examples of words of varying phonological shapes in these two lexical classes.

<table>
<thead>
<tr>
<th>Type of consonant in orC sequence</th>
<th>Class I</th>
<th>Class II</th>
</tr>
</thead>
<tbody>
<tr>
<td>nasals</td>
<td>torn</td>
<td>horn</td>
</tr>
<tr>
<td></td>
<td>/toʊn/</td>
<td>/hawɔn/</td>
</tr>
<tr>
<td>voiceless stops</td>
<td>pork</td>
<td>cork</td>
</tr>
<tr>
<td></td>
<td>/poʊk/</td>
<td>/kawɔk/</td>
</tr>
<tr>
<td>voiced stops</td>
<td>border</td>
<td>order</td>
</tr>
<tr>
<td></td>
<td>/bɔʊk/</td>
<td>/ɔwɔk/</td>
</tr>
<tr>
<td>fricatives</td>
<td>force</td>
<td>horse</td>
</tr>
<tr>
<td></td>
<td>/foʊs/</td>
<td>/hawɔs/</td>
</tr>
</tbody>
</table>

As can clearly be seen, the /aw/ pronunciation is not conditioned by phonological environment (the [orC] sequence across the word pairs in Table 12, such as torn and horn, and pork and cork, is identical). Yet, speakers of this dialect apply the /o/ -> /aw/ process only to some lexical items (i.e. those in Class II). Laferriere writes:

21 Laferriere (1979) uses the term "phonetic class".
"The membership of an or-word in either the non-alternating or alternating class [i.e. Class I or II in the table above] appears to be phonetically arbitrary, as may be seen from the shared phonetic environments of the two classes. Length and learnedness are also not factors: orchestra, orthopedic, corporation, organization are all in the alternating class. It thus appears that speakers of the Boston dialect learn or words as members of two separate phonetic classes" (604).

In other words, speakers of this dialect must store information relating to membership in a lexical class in memory.

Pierrehumbert (2001) also discusses research (Hooper, 1976; Bybee, 2000) indicating that lenition processes (specifically, schwa reduction and deletion and t/d deletion) are more prevalent in high-frequency words (section 3.2 above reviews other studies showing an effect of frequency on phonetic output). In the case of both lexical classes and frequency-correlated phonetic variance, Pierrehumbert’s point is the same: Modular generative models, in which speech production is conceived of as consisting of the retrieval of an abstract phonological form from the lexicon and the feeding of this form to a phonetic/articulatory implementation component which applies rules and constraints, cannot account for these data. This is because “[t]he phonetic implementation component applies in exactly the same way to all surface phonological representations, and the outcome depends solely on the categories and prosodic structures displayed in those representations...there is no way in which the phonetic implementation can apply differently to some words than to others” (138). An exemplar-based lexicon does not run into this problem, as acoustic templates and/or articulatory programs for particular words are stored in memory.

Johnson (2004) also uses speech production data to argue for an exemplar-based lexicon. He presents data from an 88,000 word corpus of conversational American English showing that “massive reduction” of words is quite prevalent (Johnson defines “massive reduction” as “phonetic realization of a word [that] involves a large deviation from the citation form such that whole syllables are lost and/or a large proportion of the phones in the form are changed” (3)). Table 12 gives some examples of attested massively reduced forms in the corpus.
Table 12 - Massive reduction in conversational speech (after Johnson, 2004)

<table>
<thead>
<tr>
<th>Word</th>
<th>Citation form</th>
<th>Massively reduced form</th>
</tr>
</thead>
<tbody>
<tr>
<td>because</td>
<td>/bikəz/</td>
<td>/kəz/</td>
</tr>
<tr>
<td>apparently</td>
<td>/əpˈhɛrəntli/</td>
<td>/pʰɛrəl/</td>
</tr>
<tr>
<td>hilarious</td>
<td>/ˈhɪlərɪəs/</td>
<td>/hɪləres/</td>
</tr>
<tr>
<td>particular</td>
<td>/pʰəˈtɪk ˈjələr/</td>
<td>/pʰɪk ˈjər/</td>
</tr>
</tbody>
</table>

Johnson (2004) found that, overall, 60% of the words in the corpus deviated from their citation form by at least one phone, while 28% of all words deviated by two or more phones. 20% of all words had at least one segment deleted and 5% had two or more deleted. 6% of all content words in the corpus had at least one syllable deleted, while 22% of four to six-syllable words had two or more syllables deleted.

Given the prevalence of variance from the hypothesized canonical form, Johnson argues for the rejection of “traditional dictionary assumptions about the mental representation of auditory lexical form”, namely (1) the “segmental assumption” that “lexical forms are composed of phonetic segments that are analogous to the letters of an alphabetic writing system” and (2) the “single-entry” assumption that “lexical forms are stored in a single prototypical or underlying representation” (22). Instead, he argues for an exemplar-based model of lexical memory, which is, by nature, nonsegmental (because it consists of stored memory traces, not abstract representations) and has multiple entries. Importantly, exemplar-based models allow for “word-specific patterns of phonetic variation” (31), such as the massively reduced forms observed in his study (as well as the lexical classes presented above).

An exemplar-based model best accommodates the present data

I propose that the conclusion drawn from the research presented above – that frequent word combinations are stored in lexical memory – is, like these data from speech production and perception, best understood in the context of an exemplar-based model of the lexicon. First, as already discussed, the model of lexical memory argued for throughout the present work is one that could be termed memory intensive; in other words, competent speakers store large amounts of knowledge about specific patterns of lexical co-occurrence. At the same time, however, lexical status is not granted to every combination of words ever encountered. Rather, at least in the case of the data presented here, lexicalization appears to be related to frequency. Thus, an optimal model of lexical memory is one that allows for the
storage recurrent usage patterns associated with particular lexical items. Exemplar-based memory stores particular instances of usage (presumably, as mentioned above, including lexical co-occurrence) and encodes frequency, allowing for multi-word constituents to become entrenched in memory, if they are encountered often enough. Finally, the central finding of the research presented in this document is that frequently occurring phrases can come to have holistic lexical representation, even though they do not fit into the traditional syntactic categories of which the lexicon is typically argued to be composed. It is clear that the type of system that can best accommodate these characteristics is not one which views lexical knowledge as a maximally austere, abstract dictionary-like structure. Rather, it requires a system that stores large amounts of information about actual language use and, not being constrained by a fixed formal architecture, can accommodate the storage of a heterogeneous set of lexical items. An exemplar-base model of memory conforms to these system requirements (and has the added bonus of explaining the range of empirical data reviewed above, i.e. lexical classes, frequency-modulated phonetic differences, and massive variance from canonical phonological form in conversational speech).

8. Conclusions, future directions

8.1 - A quick recap
The goal of the present work was to investigate whether frequently occurring stretches of language might sometimes be stored in the mind as single chunks, even if, in theory, they could be produced by generative mechanisms within the linguistic system. In considering this question, I began by examining some common ways of thinking about the way that humans store lexical information in their minds. Specifically, I highlighted the widespread assumption of a sort of “mental dictionary”, namely a list of individual words from the traditional parts of speech. My claim was that while such as assumption is convenient for theories positing symbolic rules for the generation of complex morphological and syntactic structures, psycholinguistic experimentation has demonstrated that factors that are neither purely linguistic or structural in nature, such as word frequency, have an impact on the psychological representation of lexical information. To back up this claim with empirical evidence, one particularly robust finding was explored, namely the observation that frequently occurring multimorphemic word forms are processed similarly (for example, in
lexical decision tasks) to monomorphemic words; various researchers have interpreted this finding as suggesting that frequent complex words are stored in their non-decomposed form in lexical memory.

The question of longer complex utterances - multi-word phrases, in this case - was then raised. Some "usage-based" models of language (Langacker, 1987; Bybee, 1985) predict that such phrases can come to be stored as single units in memory. Furthermore, various studies of language production and perception have shown that frequently occurring word sequences behave differently from more novel utterances. For example, in speech production, certain phonological processes (e.g. vowel reduction, segment deletion) are more prevalent in frequent phrases. Psycholinguistic studies of language processing using different methodologies have also shown reliable differences between frequent and infrequent phrases; it has been argued that these differences reflect the representational status of the phrases (namely, that the frequent phrases are stored as wholes in lexical memory, while the infrequent ones are not). Given these theoretical claims and empirical findings, two experiments were described which were designed to directly test the hypothesis that frequent word combinations may sometimes be stored as single lexical items, even when they are semantically transparent.

In the first of these experiments, the participants made a grammaticality judgment about visually-presented two word phrases of varying frequency. It was found that they required significantly less time to make this judgment for high-frequency phrases. This finding paralleled work done on idioms with the same type of task (Swinney and Cutler, 1979), in that idioms were found to be processed more quickly than novel word combinations having the same syntactic structure. It was suggested that the present results are explained by the fact that frequent word combinations (like idioms) are stored as single chunks in lexical memory and are thus retrieved as single units during lexical access (thereby decreasing processing time relative to novel utterances where each individual word must be retrieved from the lexicon). An important difference between idioms and the frequent word combinations studied in the present work is the level of semantic transparency. While the holistic storage of idioms is perhaps logically necessary given their lack of semantic transparency (similarly to the lack of formal transparency of irregular word forms, often adduced in support of the notion that such forms are listed in lexical memory), such is not the case with the type of two-word phrases examined here; in theory, they could be
constructed on the fly via productive syntactic rules. Thus, the finding that they, too, may be lexical units is interesting, as it suggests a role for frequency, independent of linguistic content or structure, in shaping lexical memory. It was suggested that this result parallels the experimental finding that regular complex morphological forms may be stored holistically if they are sufficiently frequent.

The second experiment used a different experimental task to examine the representational status of frequent phrases. Participants were shown two-word phrases on the computer screen and asked to read them aloud (in the context of a carrier phrase) as quickly as possible. It was found that production of frequent phrases was initiated significantly more quickly than production of syntactically comparable but less frequent phrases. Following work by Wheeldon and Lahiri (2002), in which the authors showed that speech production of compound words (assumed to be lexically represented as single units) was initiated more rapidly than production of novel compound-like combinations, these results were interpreted as suggesting (like the data from the first experiment) that the frequent phrases are stored in lexical memory. Similarly to the hypothesized explanation of the participants' performance on the grammaticality judgment task, it was argued that the faster speech preparation latencies for the frequent phrases reflect the fact that only one lexical item needed to be accessed in this case.

These experimental results and their suggested interpretations were subsequently considered in more general terms. To begin, some potential implications for models of lexical memory were explored. It was argued that the so-called usage-based models accommodated these findings most easily, as they view the lexicon as a repository of memorized linguistic routines whose storage is independent of internal linguistic structure (in other words, linguistic material of various sorts, from morphemes to multi-word expressions, can be lexicalized). Furthermore, such models postulate that frequency can be a driving force in shaping lexical representation, a claim that is consistent with the findings of the present work, as well as the studies of storage of frequent, regularly-inflected words reviewed earlier.

Given the apparent role of frequency – independent of linguistic structure – in shaping lexical representation, a preliminary hypothesis regarding the psychological processes underlying this phenomenon was advanced. Specifically, it was argued that chunks of linguistic material that frequently pass through working memory are likely to be
automatically consolidated to long-term memory; this being a function of how memory functions generally, linguistic information such as part of speech, phrase structure, etc. does not necessarily shape or constrain this process.

Subsequently, the central claims of the research presented here were discussed in relation to findings from other language-related domains, namely language acquisition, acquired language impairment in adults, and studies of large textual corpora. Essentially, the goal was to scrutinize the 'face validity' of these claims, i.e. to question whether they are to be expected given other facts we know about language and psycholinguistic processes. For example, it was shown that evidence from language development in a variety of languages suggests a central role for memorization of multi-word chunks by children. The presence of this phenomenon across the lifespan is arguably one piece of evidence for the psycholinguistic importance of entrenchment of frequent linguistic routines. Some data from studies of individuals with aphasia were also examined, again with an eye to highlighting the apparent special status of frequent, multi-word expressions. While not exactly the same as the two-word phrases studied in the present work, it was pointed out that aphasic individuals often have comparatively less difficulty producing highly frequent bits of 'automatic' speech, such as greetings, sequences (numbers, days of the week), expletives, etc. In terms of comprehension, a study by Van Lancker and Kempler (1987) was reviewed, in which it was found that aphasic speakers understood common idioms better than novel utterances with the same syntactic structure and similar frequencies of individual words. The language production and comprehension data from aphasia point again to the fact that speakers memorize chunks of linguistic material of varying length and complexity (song lyrics and idioms are clearly not novel utterances built creatively on the fly) and, more generally, to the impact of frequency on language processing and storage. Finally, a totally different source of data was considered: corpus studies of collocation and lexical co-occurrence patterns in texts. The large amounts of real-world linguistic data available in modern corpora allow for the discovery of patterns of all sorts, including the way words tend to combine. As it turns out, in recent years, researchers working in this domain have been nearly unanimous in highlighting the salience of identifiable patterns of word co-occurrence as a fundamental characteristic of human language. Simply put, much of language production appears to consist of combining specific lexical items in conventionalized ways. If this is the case, it follows, then, that competent speakers of a language must store this knowledge in lexical
memory. It was thus argued that this data represents one more indication that lexical knowledge is much more than a list of single words waiting to be fed into an independent morphosyntactic module. Rather, knowledge about frequent ways that words combine is an integral component of linguistic competence.

The final portion of the present work was devoted to considering the data presented in terms of its implications for models of language and psycholinguistic representation. The phenomenon of ‘constrained lexical co-occurrence’ - introduced by the discussion of corpus-based studies above - is given a central role in Meaning-Text Theory, a theoretical model advanced by Mel’čuk (1984, 1988, 1995) and which was briefly presented. This theory has developed an extensive formal architecture, as well as actual specialized dictionaries, devoted to describing conventionalized combination of lexical items. Meaning-Text theory and its focus on constrained lexical co-occurrence were described to reinforce the notion that such information is part of a speakers’ linguistic competence. The extrapolation made from this observation is that speakers’ knowledge of words must include the specific way they pattern with other words. Thus, the finding that frequent phrases are stored in lexical memory perhaps reflects, in part, the psycholinguistic instantiation of such knowledge.

A final theoretical issue addressed was the nature of lexical memory. It was claimed that the data suggest an exemplar-based model, similar to proposals (based on different types of evidence) made by Goldinger (1998), Pierrehumbert (2001) and Johnson (2004). It was argued that an exemplar-based memory system best accommodates a frequency-sensitive model of lexical representation, where linguistic patterns associated with specific lexical items can be extracted from memorized usage episodes. This view of memory contrasts with models that postulate symbolic representation that has been abstracted from specific exemplars and in some respects this theoretical debate parallels the one between the minimalist dictionary and the heterogeneous usage routines discussed earlier.

8.2 - Some future lines of inquiry

The data presented here have been used to argue in favour of a somewhat non-traditional view of how humans store lexical information. Needless to say, substantial revisions to dominant models would need to be further motivated and informed by more research. Some brief observations about what such research could consist of are offered here.
First, the two-word noun and adjective phrases studied for the present work are likely just the tip of the iceberg in terms of lexicalized multi-word expressions; various other types of lexical co-occurrence were discussed above, such as idioms (Swinney and Cutler, 1979), auxiliary - verb combinations (Bush, 2001), noun - preposition combinations (Sosa and MacFarlane, 2002), sentence/clause-length phrases (Bybee and Scheibman, 1999), and ‘automatic’ expressions and sequences such as greetings, expletives, poems, numbers, etc. Thus, at a general level, this type of experimentation needs to be done with a wider range of stimulus types. More specifically, a logical next step would be to investigate processing and production of a) longer semantically-transparent phrases and b) phrases that are not easily classifiable as syntactic constituents.

The significance of investigating longer word combinations lies in the opposition of such linguistic material to the common view of an austere, maximally-economical lexicon. In other words, lexicalization of such phrases goes against the grain of this view of word storage and evidence showing that semantically-transparent strings of several words (i.e. > 3) can be stored as lexical units makes an even more dramatic point in calling the dominant view into question.

The usefulness of investigating word combinations that do not map neatly onto syntactic units (XPs, parts of speech) again relates to the ability of this material to provide a stark contrast to common assumptions. Many of the stimuli used in the present work (particularly the NPs) are syntactically comparable to traditional compound words, which are less controversially assumed to be stored as single lexical units. However, the notion that strings like I don’t know, would you, or I’d like you to meet X can also be stored as chunks in lexical memory is clearly less compatible with traditional views and thus such cases could perhaps represent more compelling evidence for the holistic storage hypothesis.

Another potentially fruitful angle of inquiry would be more targeted studies of speech production. Various studies reviewed in section 3.2 indicate that word co-occurrence frequency impacts the production of certain phrases. However, these studies all drew from corpora of spontaneous speech and thus the data consisted of the exemplars the speakers happened to produce. Eliciting production of a targeted set of utterances, where frequency and syntactic structure could be manipulated, control items designed, etc., would allow for the collection of a greater amount of data and thus perhaps the detection of further patterns related to frequent word combinations.
8.3 – The last word

The research presented here had several goals. The main objective was to produce data that might increase scientific understanding of the nature of lexical representation. To this end, an analysis was offered that sought to challenge some frequently held conceptions about this topic. However, an additional, perhaps more implicit goal was to advocate for psychologically realistic models of language and language processing. In the context of this discussion of lexical memory, a frequently made observation was that assumptions and analysis techniques that are useful for the elaboration of formal linguistic theories are not necessarily transferable to psycholinguistic models. More specifically, a relevant example that emerges from this work is, of course, the observation that lexical memory may not consist primarily of single words and morphemes, waiting to be slotted into rules of the type described by linguists, but rather of a heterogeneous set of language routines, varying in length and structure. It was suggested that this may be the case due to properties of the general cognitive mechanisms which participate in language processing, such as memory.

Whether these particular claims turn out to be accurate will, of course, be determined only by further research. However, the general cautionary exhortation to researchers to constrain theories of language processing based on cognitive and/or psychological considerations is still relevant. Fortunately, as our knowledge of these latter domains expands, it will be easier and easier to posit psychologically plausible models of linguistic representation.
References


Van Lancker Sildis, D.R. (2004). When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process model of language.


## Appendix A – Frequency data for frequent word combinations and infrequent controls

<table>
<thead>
<tr>
<th>Frequent phrases</th>
<th>Freq. 1st word</th>
<th>Freq. 2nd word</th>
<th>Freq. of bigram</th>
<th>Infrequent controls</th>
<th>Freq. 1st word</th>
<th>Freq. 2nd word</th>
<th>Freq. of bigram</th>
</tr>
</thead>
<tbody>
<tr>
<td>human rights</td>
<td>387033</td>
<td>457186</td>
<td>181769</td>
<td>business rights</td>
<td>852159</td>
<td>457186</td>
<td>54</td>
</tr>
<tr>
<td>next week</td>
<td>1044786</td>
<td>1302475</td>
<td>113170</td>
<td>our week</td>
<td>1100597</td>
<td>1302475</td>
<td>43</td>
</tr>
<tr>
<td>very good</td>
<td>891496</td>
<td>904788</td>
<td>40819</td>
<td>very even</td>
<td>891496</td>
<td>1235141</td>
<td>84</td>
</tr>
<tr>
<td>nuclear weapons</td>
<td>270505</td>
<td>278178</td>
<td>37857</td>
<td>nuclear problems</td>
<td>270505</td>
<td>384302</td>
<td>55</td>
</tr>
<tr>
<td>young man</td>
<td>414104</td>
<td>580062</td>
<td>20821</td>
<td>prime man</td>
<td>604317</td>
<td>580062</td>
<td>1</td>
</tr>
<tr>
<td>heart disease</td>
<td>214954</td>
<td>167492</td>
<td>12119</td>
<td>heart pressure</td>
<td>214954</td>
<td>255479</td>
<td>1</td>
</tr>
<tr>
<td>hard work</td>
<td>340386</td>
<td>905433</td>
<td>11631</td>
<td>past work</td>
<td>615679</td>
<td>905433</td>
<td>182</td>
</tr>
<tr>
<td>vast majority</td>
<td>61761</td>
<td>218248</td>
<td>11310</td>
<td>tight majority</td>
<td>67539</td>
<td>218248</td>
<td>4</td>
</tr>
<tr>
<td>security guard</td>
<td>763126</td>
<td>150756</td>
<td>7285</td>
<td>security goal</td>
<td>763126</td>
<td>282731</td>
<td>10</td>
</tr>
<tr>
<td>younger brother</td>
<td>72897</td>
<td>121032</td>
<td>7134</td>
<td>smaller brother</td>
<td>104486</td>
<td>121032</td>
<td>9</td>
</tr>
<tr>
<td>first base</td>
<td>2697862</td>
<td>261702</td>
<td>7024</td>
<td>first price</td>
<td>2697862</td>
<td>282731</td>
<td>100</td>
</tr>
<tr>
<td>cold weather</td>
<td>122292</td>
<td>160161</td>
<td>4585</td>
<td>cold summer</td>
<td>122292</td>
<td>251085</td>
<td>50</td>
</tr>
<tr>
<td>early retirement</td>
<td>673755</td>
<td>77195</td>
<td>3713</td>
<td>early commitment</td>
<td>673755</td>
<td>86050</td>
<td>68</td>
</tr>
<tr>
<td>far right</td>
<td>542706</td>
<td>721465</td>
<td>3470</td>
<td>best right</td>
<td>589888</td>
<td>721465</td>
<td>119</td>
</tr>
<tr>
<td>special interest</td>
<td>379581</td>
<td>438787</td>
<td>3170</td>
<td>better interest</td>
<td>508268</td>
<td>438787</td>
<td>49</td>
</tr>
<tr>
<td>lethal injection</td>
<td>16628</td>
<td>11068</td>
<td>3024</td>
<td>fatal injection</td>
<td>31632</td>
<td>11068</td>
<td>32</td>
</tr>
<tr>
<td>book review</td>
<td>301123</td>
<td>180738</td>
<td>2794</td>
<td>book earnings</td>
<td>301123</td>
<td>259628</td>
<td>6</td>
</tr>
<tr>
<td>sexually transmitted</td>
<td>20538</td>
<td>41632</td>
<td>2619</td>
<td>physically transmitted</td>
<td>23128</td>
<td>41632</td>
<td>0</td>
</tr>
<tr>
<td>fried chicken</td>
<td>9802</td>
<td>47326</td>
<td>2615</td>
<td>sliced chicken</td>
<td>11390</td>
<td>47326</td>
<td>23</td>
</tr>
<tr>
<td>directly involved</td>
<td>93862</td>
<td>236716</td>
<td>2511</td>
<td>directly given</td>
<td>93862</td>
<td>369661</td>
<td>9</td>
</tr>
<tr>
<td>densely populated</td>
<td>4181</td>
<td>14011</td>
<td>2433</td>
<td>lightly populated</td>
<td>17819</td>
<td>14011</td>
<td>81</td>
</tr>
<tr>
<td>tour guide</td>
<td>241497</td>
<td>55132</td>
<td>2156</td>
<td>tour film</td>
<td>241497</td>
<td>291263</td>
<td>16</td>
</tr>
<tr>
<td>natural disaster</td>
<td>144743</td>
<td>68802</td>
<td>1906</td>
<td>personal disaster</td>
<td>232130</td>
<td>68802</td>
<td>31</td>
</tr>
<tr>
<td>illegal immigrant</td>
<td>147511</td>
<td>23593</td>
<td>1633</td>
<td>illegal employee</td>
<td>147511</td>
<td>56525</td>
<td>1</td>
</tr>
<tr>
<td>end result</td>
<td>1011175</td>
<td>247803</td>
<td>1631</td>
<td>end process</td>
<td>1011175</td>
<td>366947</td>
<td>4</td>
</tr>
<tr>
<td>closely related</td>
<td>72097</td>
<td>98271</td>
<td>1541</td>
<td>nearly related</td>
<td>377784</td>
<td>98271</td>
<td>2</td>
</tr>
<tr>
<td>wide margin</td>
<td>141159</td>
<td>55430</td>
<td>1506</td>
<td>wide highway</td>
<td>141159</td>
<td>66211</td>
<td>8</td>
</tr>
<tr>
<td>diplomatic immunity</td>
<td>112081</td>
<td>21150</td>
<td>1493</td>
<td>official immunity</td>
<td>678648</td>
<td>21150</td>
<td>26</td>
</tr>
<tr>
<td>peer pressure</td>
<td>561900</td>
<td>14663</td>
<td>1421</td>
<td>male pressure</td>
<td>57278</td>
<td>14663</td>
<td>0</td>
</tr>
<tr>
<td>family reunion</td>
<td>7989</td>
<td>255479</td>
<td>1187</td>
<td>family profession</td>
<td>7989</td>
<td>14814</td>
<td>7</td>
</tr>
</tbody>
</table>
Appendix B – Text of instructions given to experimental participants

Experiment 1
In this experiment, you will see a number of phrases on the computer screen. Some of the phrases will be grammatically correct phrases that a native speaker of English might say, for example: *blue sky, bicycle thief, totally useless*, etc. Others will be incorrect or unnatural-sounding phrases that a native speaker of English wouldn’t normally say, such as *basket solid, painful bright, recently knife*. Your task is to read each phrase silently and indicate if you think it is an acceptable English phrase, i.e. a phrase that you yourself might say or write. You should press the “yes” button if you think the phrase is acceptable in English and the “no” button if you think it is not acceptable. You should make this judgment as quickly as possible, while still maintaining accuracy.

If you have questions, please ask now.

Experiment 2
In this experiment, you will again see phrases on the computer screen. This time, however, the task is to read the phrase silently and then say out loud “I said [phrase]”. So, for example, if the phrase *bicycle thief* appeared on the screen, you would say “I said bicycle thief”. You should say this sentence as soon as possible after the phrase appears on the computer screen, while still maintaining accuracy. Speak in your normal voice, at a volume level you would use in normal conversation.

If you have questions, please ask now.
Appendix C – Phonological control items (from Experiment 2)

<table>
<thead>
<tr>
<th>Frequent phrases</th>
<th>Infrequent controls</th>
<th>Phonological controls - Frequent</th>
<th>Phonological controls - Infrequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>human rights</td>
<td>business rights</td>
<td>human control</td>
<td>business control</td>
</tr>
<tr>
<td>next week</td>
<td>our week</td>
<td>next dog</td>
<td>our dog</td>
</tr>
<tr>
<td>very good</td>
<td>very even</td>
<td>quite good</td>
<td>quite even</td>
</tr>
<tr>
<td>nuclear weapons</td>
<td>nuclear problems</td>
<td>psychological weapons</td>
<td>psychological problems</td>
</tr>
<tr>
<td>young man</td>
<td>prime man</td>
<td>young crops</td>
<td>prime crops</td>
</tr>
<tr>
<td>heart disease</td>
<td>heart pressure</td>
<td>serious disease</td>
<td>serious pressure</td>
</tr>
<tr>
<td>hard work</td>
<td>past work</td>
<td>hard experience</td>
<td>past experience</td>
</tr>
<tr>
<td>vast majority</td>
<td>right majority</td>
<td>vast garment</td>
<td>tight garment</td>
</tr>
<tr>
<td>security guard</td>
<td>security goal</td>
<td>important guard</td>
<td>important goal</td>
</tr>
<tr>
<td>younger brother</td>
<td>smaller brother</td>
<td>younger child</td>
<td>smaller child</td>
</tr>
<tr>
<td>first base</td>
<td>first price</td>
<td>permanent base</td>
<td>permanent price</td>
</tr>
<tr>
<td>cold weather</td>
<td>cold summer</td>
<td>damp weather</td>
<td>damp summer</td>
</tr>
<tr>
<td>early retirement</td>
<td>early commitment</td>
<td>lengthy retirement</td>
<td>lengthy commitment</td>
</tr>
<tr>
<td>far right</td>
<td>best right</td>
<td>far side</td>
<td>best side</td>
</tr>
<tr>
<td>special interest</td>
<td>better interest</td>
<td>special children</td>
<td>better children</td>
</tr>
<tr>
<td>lethal injection</td>
<td>fatal injection</td>
<td>lethal reaction</td>
<td>fatal reaction</td>
</tr>
<tr>
<td>book review</td>
<td>book earnings</td>
<td>record review</td>
<td>record earnings</td>
</tr>
<tr>
<td>sexually transmitted</td>
<td>physically transmitted</td>
<td>sexually impaired</td>
<td>physically impaired</td>
</tr>
<tr>
<td>fried chicken</td>
<td>sliced chicken</td>
<td>fried meat</td>
<td>sliced meat</td>
</tr>
<tr>
<td>directly involved</td>
<td>directly given</td>
<td>previously involved</td>
<td>previously given</td>
</tr>
<tr>
<td>densely populated</td>
<td>lightly populated</td>
<td>densely packed</td>
<td>lightly packed</td>
</tr>
<tr>
<td>tour guide</td>
<td>tour film</td>
<td>cheap guide</td>
<td>cheap film</td>
</tr>
<tr>
<td>natural disaster</td>
<td>personal disaster</td>
<td>natural feature</td>
<td>personal feature</td>
</tr>
<tr>
<td>illegal immigrant</td>
<td>illegal employee</td>
<td>handsome immigrant</td>
<td>handsome employee</td>
</tr>
<tr>
<td>end result</td>
<td>end process</td>
<td>latest result</td>
<td>latest process</td>
</tr>
<tr>
<td>closely related</td>
<td>nearly related</td>
<td>closely connected</td>
<td>nearly connected</td>
</tr>
<tr>
<td>wide margin</td>
<td>wide highway</td>
<td>narrow margin</td>
<td>narrow highway</td>
</tr>
<tr>
<td>diplomatic immunity</td>
<td>official immunity</td>
<td>diplomatic procedure</td>
<td>official procedure</td>
</tr>
<tr>
<td>peer pressure</td>
<td>male pressure</td>
<td>peer advice</td>
<td>male advice</td>
</tr>
<tr>
<td>family reunion</td>
<td>family profession</td>
<td>popular reunion</td>
<td>popular profession</td>
</tr>
</tbody>
</table>

82