ON-LINE PROCESSING ADVANTAGES OF LEXICAL BUNDLES*

Antoine Tremblay[1], Bruce Derwing[1], Gary Libben[1], and Chris Westbury[2]
University of Alberta
[1]department of linguistics, [2]department of psychology

Abstract

This paper explores the tenability of the hypothesis that lexical bundles (i.e., frequently recurring strings of words that often span traditional syntactic boundaries) are stored and processed holistically. Three self-paced reading experiments were conducted to test the hypothesis, where sentences containing lexical bundles and their controls were presented to participants in a word-by-word, portion-by-portion, and sentence-by-sentence fashion. Lexical bundles and sentences containing lexical bundles were read faster than their controls in all three experiments. The self-paced reading experiments were followed up by two word and sentence recall experiments in the visual and auditory modalities. In experiment 4, lexical bundles were remembered more accurately and more words were recalled after the sentences in which they occurred. In experiment 5, however, only accuracy of recall was significant.

Keywords: lexical bundles; formulaic sequences; self-paced reading; word and sentence recall; auditory; visual; English.

Introduction

The term ‘lexical bundle’ comes from the field of corpus linguistics. It first appeared in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999), a monumental work entirely based on the British National Corpus of 100 million words. Lexical bundles are very common continuous multi-word strings, which may span phrasal boundaries. Some instances are *I don’t know whether*, *don’t worry about it*, and *in the middle of the*. The concept of lexical bundles, however, goes back at least to Salem (1987) and the research he carried out on a corpus of French government texts. Butler (1997) and Altenberg (1998) subsequently employed the notion in their investigations based on Spanish and English corpora. Lexical bundles are part of a larger family of multi-word strings (continuous or discontinuous) known as formulaic sequences, which are commonly thought to be stored and processed in the mind as holistic units. Examples include greeting formulae (*how do you do?*), back-channelling formulae (*yes, I see*), phrasal verbs (*to show up*), and other constructions/patterns of different sorts, ranging from the very schematic *Subject-Verb-Object* construction (*He kicked the ball*) to the less schematic *Verb Noun into V–ing* pattern (*He talked her into going*), and idioms (*to put one’s finger in the dike)* (Croft

* This paper is supported by a Major Collaborative Research Initiative Grant (#412-2001-1009) and a Doctoral Fellowship (#752-2006-1315) from the Social Sciences and Humanities Research Council of Canada (SSHRC). We wish to thank the participants of the *Symposium on Formulaic Language*, held at the University of Wisconsin Milwaukee, 18-21 April 2007, for their insightful comments. We also wish to thank R. Harald Baayen and John Newman. Correspondence concerning this article should be addressed to Antoine Tremblay, Department of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Edmonton Alberta, T6G 2E7, Canada. Internet: antoinet@ualberta.ca; www.ualberta.ca/~antoinet.

2001; Erman and Warren 2000; Hunston and Francis 2000; Pawley and Syder 1983; Titone and Conine 1999; Wray 2002 and references cited therein; Schmitt 2004 and references cited therein).

Wray (2002) gives us a nice overview of the history of formulaic sequences in linguistics. Their existence was noticed at least as early as the mid-nineteenth century by John Hughlings Jackson, who observed that aphasics could fluently recall rhymes, prayers, greeting formulae and so forth, whereas they could not produce novel sentences (cited in Wray 2002: 7). Jackson was not the only early scholar to detect such linguistic peculiarities. Ferdinand de Saussure (1916/1966) talked of agglutinations, that is, the unintentional fusion of two or more linguistic signs that frequently recur together into a single unanalyzed unit so as to form a short cut for the mind. Jespersen (1924) acknowledged the likely existence of multi-word mental units, arguing that language would be too difficult to manage if one had to remember every individual word separately. Bloomfield (1933: 181), too, wrote that "many forms lie on the border-line between bound forms and words, or between words and phrases." Firth, for his part, considered that the basic syntactic units of speech were phrases (1937/1964), and that in order to characterize a certain community's speech, one had to list the usual collocations used by its speakers, that is, the set of words that frequently co-occured with a particular word (1957/1968). In a similar vein, Hymes (1962: 41) observed that a large part of communication involved the use of recurrent patterns, that is, of "linguistic routines", and Bolinger (1976: 1) maintained that "our language does not expect us to build everything starting with lumber, nails, and blueprints, but provides us with an incredibly large number of prefabs." Finally, Fillmore (1979) wrote that knowing how to use formulaic utterances makes up a large part of a speaker's ability to successfully handle language. However, during the Chomskyan era, which started in the late 1950s, formulaic language was largely marginalized, and only recently has "the idea of holistically managed chunks of language" resurfaced (Wray 2002: 8).

As just mentioned, a number of researchers recognized that certain words systematically occurred with one another. However, their observations were based on perceptual salience and a number of highly frequent lexical sequences went unnoticed. Nowadays, linguists have powerful tools that enable them to reliably identify lexical sequences that recur across increasingly large amounts of spoken and written text. More importantly, "corpus-based techniques enable investigation of new research questions that were previously disregarded because they were considered intractable" (Biber and Conrad 1999: 181). Owing to corpus-based approaches, we are not only realizing "how extensive and systematic the pattern of language use" is, but also apprehending how such "association patterns are well beyond the access of intuitions" and how they are "much too systematic to be disregarded as accidental" (Biber et al. 1999: 290). Given this systematicity, one may wonder whether formulaic sequences are stored and processed holistically.

This brings us to the psycholinguistic domain. During the last three decades, a great deal of psycholinguistic research has focused on the mental lexicon. As Libben (1998: 30) points out, there are at least two reasons for this. The human ability to store and access a large number of words is central to language and thus needs to be better understood. Another fundamental question to be answered is how we process language, or, more specifically, what is the trade-off between storage and computation in the representation and processing of linguistic signs. One of the key theoretical arguments in favour of "chunking" as a strategy for linguistic processing occurred in Miller (1956), which is the classic paper on the "seven, plus or minus two" limitation of short-term memory capacity.[1] In that paper, Miller also developed the thesis that one way for humans to circumvent this limitation on information processing is to organize information into larger and larger units (or chunks), thereby increasing the total amount of information conveyed without having to increase the number of units involved. In language processing, then, the various kinds of formulaic sequences that have been hypothesized might be viewed as examples of the use of this strategy (see section 2.3 below for further discussion). Unfortunately, very few psycholinguistic studies have considered the question of how formulaic sequences are stored and processed in the mind (Schmitt 2004: viii), and even these have produced mixed results. Let us briefly review these studies.

Bod (2001), using a lexical-decision task, has shown that high-frequency three-word sentences such as *I like it* were reacted to faster than low-frequency sentences such as *I keep it*. Underwood, Schmitt and Galpin (2004) used an eye-tracking paradigm to examine the processing of formulaic sequences such as *a stitch in time saves nine* and *as a matter of fact*. They found that the terminal words in formulaic sequences were processed more quickly than the same words appearing in non-formulaic contexts. These results provide evidence supporting the view that formulaic sequences (including high-frequency three-word sentences) are stored and processed holistically. Nevertheless, other studies failed to find processing discrepancies between formulaic and non-formulaic sequences. Schmitt and Underwood (2004) conducted a self-paced reading experiment using the same stimuli used in the Underwood, Schmitt and Galpin study, where words were flashed on the screen one-by-one. Contrary to the eye-tracking experiment, the terminal words in formulaic sequences were not processed more quickly than the same words appearing in non-formulaic contexts. Finally, in their oral recall experiment, Schmitt, Grandage, and Adolphs (2004) did not find that formulaic sequences were recalled more accurately than non-formulaic sequences. In the face of such limited and mixed results, the question of whether formulaic sequences are stored and processed in the mind as wholes remains unresolved. If we are to elucidate this question, more research needs to be done.

In this paper, we wish to advance our understanding of the mental lexicon by addressing the question of whether lexical bundles (LBs) are stored and/or processed holistically. We approached the question by conducting three self-paced reading experiments and two word and sentence recall experiments. We reasoned that if LBs are equivalent to one unit they would be read more quickly than a comparable non-lexical bundle (equivalent to more than one unit). Similarly, LBs would be remembered more accurately and more words would be recalled after sentences that contain them. By way of example, consider the target sentence *If workers don't worry about it nothing will happen*, where the LB is *don't worry about it*, and its control *If workers don't know about it nothing will happen*, where the non-lexical bundle (NLB) is *don't know about it*. The target sentence would be processed more efficiently than the control one, given that the former sentence only comprises 6 units (i.e., *If* + *workers* + *don't worry about it* + *nothing* + *will* + *happen*), whereas the latter one contains 9 (i.e., *If* + *workers* + *don't* + *worry* + *about* + *it* + *nothing* + *will* + *happen*). We begin our discussion with a description of the word-by-word, portion-by-portion, and sentence-by-sentence self-paced reading experiments and then move on to the auditory and visual word and sentence recall experiments, respectively.

[1] Based on converging data, Cowan (2000) argues that short-term memory capacity is only four plus or minus one.

Experiment 1

Experiment 1 was modeled after Schmitt and Underwood (2004), who investigated the processing of formulaic sequences such as *by the skin of his teeth* by running a word-by-word self-paced reading experiment. They reasoned that if formulaic sequences were stored and processed holistically, the terminal word of a formulaic sequence would be read faster than the same word in a non-formulaic sequence text. They chose 20 formulaic sequences that met the following criteria:

(i) the sequences had a relatively high frequency in the *British National Corpus* and the *Cambridge and Nottingham Corpus of Discourse in English;*
(ii) the sequences had a relatively obvious beginning (i.e., they did not begin with multiple function words);
(iii) the sequences did not end with a function word;
(iv) the sequences were 4-8 words long;
(v) the sequences were relatively predictable from their initial components.

The sequences were embedded in extended contexts. Each story had one sequence and one terminal word from a formulaic sequence from another passage. Each passage was subjected to a frequency analysis in *The Compleat Lexical Tutor v.2* to ensure that low frequency vocabulary was kept to a minimum. Finally, simple comprehension questions were devised for each story to ensure that participants actually read the passage. They compared reading times between terminal words appearing in formulaic sequences and terminal words occurring in non-formulaic sequence text. However, they did not find any significant difference in reading latencies. According to these authors (2004: 187), their failure to find positive results might be due to the "word-by-word nature of the task [which] disrupts the holistic processing of formulaic sequences". Alternatively, it is possible they did not find any differences because they did not directly compare reading times between formulaic sequences and equivalent non-formulaic sequences. It is also possible that factors such as transitional probabilities — that is the probability of word W2 occurring after word W1 — washed out the gain in reading speed that formulaic sequences would have provided. We thus ran a word-by-word self-paced reading experiment where reading times for lexical bundles were directly compared to nearly equivalent non-lexical bundles. This experiment is described below.

Participants

Twenty undergraduate students at the University of Alberta were paid to participate in the self-paced reading task. They were all native speakers of North American English.

Materials

Lexical bundles (LB) were taken from Biber et al. (1999). Their frequencies were taken from the spoken subcorpus of the *British National Corpu*s using the *Variations in English Words and Phrases* search engine. Any 4-word string with a frequency of occurrence of at least 10 times per million words and any 5-word string with a frequency of occurrence of at least 5 times per million were retained as LBs (e.g., *the end of the*; frequency 112 per million).[2] Any string of words with a frequency below this threshold was considered to be a non-lexical bundle (NLB; e.g., *I see what you*; frequency 7 per million). The control sentences differed from the target LB sentences by only one word, which will be referred to as the pivot word (PW). We endeavoured to ensure that, if anything, control sentences had a processing advantage over target sentences. The substituted control PWs were chosen to be more frequent and on average shorter in length than target PWs. We initially selected 20 LBs for which appropriate NLB control strings could be found. By way of example, consider the following target sentence: *I sat in the **middle** of the bullet train*. The underlined portion of the sentence corresponds to the LB. Compare this sentence to the following control sentence, where the underlined portion does not constitute an LB: *I sat in the **front** of the bullet train*. The two differ only in the PW (in bold): "middle" in the target and "front" in the control (see Appendix A for a complete stimulus list). Finally, we excluded 5 pairs (i.e., items 3, 4, 6, 12, and 17) based on the unnaturalness of the sentence. A summary of characteristics of the 15 target/control stimuli is given in Table 1. Statistical tests were conducted using R 2.5.1.

Table 1
Summary of stimuli characteristics

Measure	LB	NLB	Difference	Test	Statistic
Length	14	13	1	paired *t*(12)	3.5***
PW	3514	6942	-3428	*W*	46*
AB	1108	1312	-204	*W*	62.5
BC	375	133	242	*W*	132*
CD	344	302	42	*W*	90
DE	1042	1042	0	*W*	4.5
ABC	163	6	157	*W*	166***
BCD	61	5	56	*W*	164***
CDE	17	6	11	paired *t*(2)	1.9
ABCD	26	1	25	*W*	169***
BCDE	16	3	13	*W*	8
ABCDE	23	1	22	paired *t*(2)	2.5

Note. We used a paired *t*-test for variables which did not significantly depart from the normal distribution, as indicated by a Shapiro-Wilk normality test (not shown here). For those that did depart from the normal distribution, we used a Wilcoxon rank sum test (*W*) instead. PW = Pivot Word. The capital letters A, B, C, D, and E refer to individual words within LB/NLB sequences. For instance, A is the first word of the multi-word string, B is the second, and so forth. The strings AB, BC, ABC, CDE, etc. refer to portions of LB/NLB sequences. For example, AB designates the portion composed of the first and second word, while CDE designates the third, fourth, and fifth word of a sequence. Values are frequencies in number of occurrences per million words except for length, which is in number of characters.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

In most cases, the LB and NLB strings were embedded after the second word of the sentence and were followed by two words – the mean length of sentences was 8.5 words (SD = 0.7). Finally,

[2] This arbitrary threshold originates from Biber et al. (1999: 992-3).

an effort was made so that the portions occurring before and after the target and control strings did not contain LBs.

Experimental Design

The stimuli were split into two counter-balanced lists: list A and list B. Participants from group A first saw list A, had a 30-40 minute break (they did other experiments) and then saw list B, whereas participants in group B saw list B before and list A after. Note that the first list participants saw is referred to as the 1st set in the remainder of the text, and the second list as the 2nd set. The sentences were presented in a pseudo-randomized fashion. That is, the order of presentation was randomly determined but then kept constant across groups and lists. Each trial was paired up with a simple yes-no question specific to the sentence to ensure that the participants actually read and processed the sentences. The right answers to the questions were balanced (20 "yes" answers and 20 "no" answers).

Procedure

The practice and experimental trials were presented to the participants visually using Psyscope version 1.2.5. Each practice and experimental trial consisted of the following: (i) The participants heard a beep and saw an asterisk in the centre of the screen (Font: Arial bold, Size: 100); (ii) when ready, the participants pressed a key to see a the first word of a sentence (Position: centred, Font: Arial, Size: 48); (iii) once the participants had finished reading it, they pressed a key to see the next word until they read the whole sentence; (iv) then the participants heard a beep and saw three asterisks in the centre of the screen for 1000 ms (Position: centred, Font: Arial, Size: 48); (v) the word "Question:" appeared in the centre of the screen (Font: Arial bold; Size: 48) for 1000 ms, and then the question as such appeared (Position: centre; Font: Arial bold; Size: 36); fianlly (vi) the participant answered the question by using the "y" key for *yes* and "n" key for *no*. Once they pressed either the "y" or "n" key, the next trial started. An example of a trial block is shown in (1).

(1) Word-by-word presentation
  a. Target: *He's - glad - you – **don't**- want - to – dig - tunnels.*
     Question: *He doesn't want to dig tunnels, right*
     Answer: *Yes*
  b. Control: *He's - glad - you – **do**- want - to – dig - tunnels.*
     Question: *Is he sad?*
     Answer: *No*

In (1), the underlined portion of each example corresponds to the LB and NLB strings and the word in bold is the PW.

Results

Reaction times for words constituting LB and NLB strings were summed for each participant and each item; only these portions only were considered for analysis. Trials for which questions were incorrectly answered were eliminated as well as those with reading times 2.5 standard deviations above or under the mean. Overall, 6% (36/600 trials) were eliminated. We performed a linear mixed-effects regression analysis using R 2.5.1. Subjects and items were treated as random effects, while lexical bundlehood entered the model as a fixed effect (the factors set and group did not reach significance). Results are summarized in Table 2.

Table 2
Linear Mixed-Effects Regression for Word-by-Word Self-Paced Reading Experiment

Random Effects			
	Variance	*SD*	
Subject	1,060,103	1,030	
Item	22,683	151	
Residual	498,421	706	
Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	1,948	237	8.2***
NLB	160	60	2.7**

Note. Estimates and standard errors are in milliseconds (msec). N = 564. Model log-likelihood = -4,536.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

To paraphrase Table 2, after accounting for once subject and item variability, LBs (M = 1,948 msec) are read 160 msec faster ($t(563) = 2.7$, $p < .01$, $d = 0.14$, $r = 0.07$) than NLB (M = 2,108 msec). Note that effect sizes are Cohen's *d* estimates.[3] Contrary to Schmitt and Underwood's (2004) self-paced reading experiment, our experiment shows on-line processing advantages for LBs over NLBs. Word-by-word reading, however, is somewhat unnatural. We thus designed a self-paced reading experiment where LBs and NLBs were presented as a whole instead of one word at a time.

Experiment 2

In everyday life, one rarely, if ever, reads sentences one word at a time. The portion-by-portion self-paced reading experiment reported in this section aims at determining whether the LB on-line processing facilitatory effect holds in a more naturalistic reading setting.

Participants

Same as in experiment 1. None of them had done experiment 1.

Materials

Same as in experiments 1.

Experimental Design

[3] Cohen's *d* effect sizes are hesitantly defined as "small, $d = 0.2$", "medium, $d = 0.5$", and "large, $d = 0.8$".

Same as in experiments 1.

Procedure

Same as experiments 1, except that the stimuli were presented in a portion-by-portion fashion. An example of a trial block is shown in (2).

(2) Portion-by-portion presentation

a. Target: *He's glad* - *you* ***don't*** *want to* - *dig tunnels.*
   Question: *He doesn't want to dig tunnels, right?*
   Answer: *Yes*
b. Control: *He's glad* - *you* ***do*** *want to* - *dig tunnels.*
   Question: *Is he sad?*
   Answer: *No*

As before, the underlined portion of each example corresponds to the LB and NLB strings and the word in bold to the PW.

Results

Trials for which questions were incorrectly answered were eliminated as well as those with reading times 2.5 standard deviations above or under the mean. Overall, 12.4% (224/1800 trials) were eliminated. We carried out a linear mixed-effects regression as in experiment 1. Results are summarized in Table 3.

Table 3
Linear Mixed-Effects Regression for Portion-by-Portion Self-Paced Reading Experiment

Random Effects			
	Variance	*SD*	
Subject	118,044	344	
Item	16,696	129	
Residual	213,250	462	
Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	1,060	88	12.0***
NLB	218	41	5.3***

Note. Estimates and standard errors are in milliseconds (msec). N = 520. Model log-likelihood = -3,954.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

Table 2 shows that, similarly to experiment 1, after subject and item variability have been accounted for, LBs (*M* = 1,060 msec) are read 218 msec faster ($t(519) = 5.3$, $p < .001$, $d = 0.35$, $r = 0.17$) than NLBs (*M* = 1,278 msec). It is therefore predicted that sentences containing LBs in a sentence-by-sentence self-paced reading experiment would be read more quickly than sentences that do not contain them. We investigate this in the following experiment.

Experiment 3

In experiments 1 and 2, we found that the "LBhood" of a sequence of words, that is, whether it is an LB or not, predicts reading speed. The sentence-by-sentence self-paced reading experiment reported in this section will enable us to determine whether the effect is also present when LBs and NLBs are presented in full sentences.

Participants

Same as in experiments 1 and 2. None of them had done experiments 1 or 2.

Materials

Same as in experiments 1 and 2.

Experimental Design

Same as in experiments 1 and 2.

Procedure

Same as experiments 1 and 2, except that the stimuli were presented in a sentence-by-sentence fashion. An example of a trial block is shown in (3).

(3) Sentence-by-sentence presentation

a. Target: *He's glad <u>you **don't** want to</u> dig tunnels.*
Question: *He doesn't want to dig tunnels, right?*
Answer: *Yes*

b. Control: *He's glad <u>you **do** want to</u> dig tunnels.*
Question: *Is he sad?*
Answer: *No*

The underlined portion in each example corresponds to the LB and NLB strings and the word in bold to the PW.

Results

As in the other two experiments, trials for which questions were incorrectly answered were eliminated as well as those with reading times 2.5 standard deviations above or under the mean (7.5% (56/750 trials) were eliminated). We performed a linear mixed-effects regression analysis as in experiments 1 and 2. Results are summarized in Table 4.

Table 4
Linear Mixed-Effects Regression for Sentence-by-Sentence Self-Paced Reading Experiment

Random Effects			
	Variance	*SD*	
Subject	902,692	950	
Item	189,515	435	
Residual	1,585,464	1,259	
Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	3,633	231	15.7***
NLB	495	96	5.1***

Note. Estimates and standard errors are in milliseconds (msec). N = 689. Model log-likelihood = -5,931.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

In experiment 3, LBs (M = 3,633 msec) are read 495 msec faster ($t(688) = 5.1$, $p < .001$, d = 0.29, r = 0.14) than NLBs (M = 4,128 msec) after accounting for subject and item variability.

Discussion

The three self-paced reading experiments reported here show that LBs have an on-line processing facilitatory effect over equivalent NLBs. In principle, however, NLBs should have been read faster given that they were on average shorter and more frequent, two factors know to facilitate on-line processing (Staub and Rayner 2007). These findings parallel results obtained for other self-paced reading experiments aimed at determining whether formulaic sequences (other than LBs) were stored and processed holistically. Ortony, Schallert, Reynolds, and Antos (1978) found that idioms used figuratively were understood more quickly than idioms used literally, thus suggesting that the meaning of an idiom is stored like the meaning of single words. Gibbs, Bogdanovich, Sykes, and Barr (1997) also found in their self-paced reading paradigm that idioms were read faster than control non-idioms. Finally, Conklin and Schmitt (2007) used a self-paced reading experiment in their study of formulaic sequences such as *everything but the kitchen sink* and *a breath of fresh air*, which were embedded in passages and presented in a line-by-line fashion. They found that formulaic sequences were processed faster than non-formulaic sequences.

Though the results reported here lend support to the holistic storage hypothesis, they do not preclude the possibility that the reading advantage of LBs over NLBs reflects over-practiced processing tied to higher frequency of occurrence. Storage, as Harald Baayen (p.c.) would say, might simply be combinatorial knowledge, that is, knowledge of what goes with what, or as the third author would say, knowing where one is going given knowledge of where one was.

In addition, the LB facilitatory effect might be task specific. Much like the lack of significant results in the Schmitt and Underwood study that likely arose from comparing reading times for the same word occurring in formulaic and non-formulaic contexts, our significant results might be solely due to the self-paced reading paradigm. We addressed these issues by running two word and sentence recall experiments, which are described in the following section.

Word and Sentence Recall Experiments

Miller (1956) distinguishes between *bits* of information and *chunks* of information. The term *bit* of information refers to an amount of information whereas the notion of chunk can be defined as the unit formed by organizing or grouping bits of information. He argues that the span of working memory, or the number of chunk it can accommodate, is equal to seven, plus or minus two. The number of bits, however, that can be contained in these chunks is in principle unlimited. For example, let a phoneme be worth one bit of information. If one chunk is equal to one bit of information, then we can store in our working memory roughly seven phonemes, for instance ə, l, ɛ, d, ʌ, k, and a. In this case, we are storing seven chunks each worth one bit for a total of seven bits of information. However, phonemes can be grouped together into larger chunks called words. For example, the words *man*, *gone*, *did*, *hail*, *hack*, *lip*, and *pan* all contain three phonemes and are thus worth three bits of information. If we store all seven words in our working memory, we will be storing 21 bits of information instead of seven, even though the number of chunks stored remains constant.

Savin and Perchonock (1965) used a sentence and word recall paradigm to determine working memory requirements of different types of sentences such as actives, negatives passives, and questions. They reasoned that participants would reconstruct correct sequences such as sentences from abstract schemes that underlie groupings of words. Indeed, Miller and Selfridge (1950), Tulving and Patkau (1962), and Maher and Skovengaard (1988) have shown that there is a positive correlation between the number of words recalled and the amount of structure that exists between them. Considering that only one scheme underlies active sentences, and that passives and questions, for example, require additional schemes, they predicted that participants would recall more words after simple-scheme sentences than more complex-scheme sentences. Thus, more working memory resources ought to be allocated to remembering a negative passive sentence such as *the cake was not eaten by the dog* than an active sentence such as *the dog ate the cake*. To test this hypothesis, Savin and Perchonock presented participants with sentences followed by a list of words. Participants were then asked to recall the sentence and as many words as they could remember. It turned out that the more complex the structure of a sentence was, the fewer words participants remembered, thus supporting their assumption.

Though this technique was used to bring empirical support to the transformational hypothesis of the 50s and 60s, it can also be used to determine whether LBs are stored as holistic chunks. That is, if LBs are stored holistically, sentences that contain them should take up less space in working memory than equivalent sentences that do not. Therefore, more additional words should be recalled after sentences that contain LBs. We ran two word and sentence recall experiments, one in the auditory modality and one in the visual modality. We first describe the word and sentence recall experiment in the visual modality and then turn to the one in the auditory modality.

Experiment 4

In this section, we report on the word and sentence recall experiment in the visual modality.

Participants

Same as in experiments 1, 2, and 3. None of the participants had done experiments 1, 2, or 3.

Materials

We used the same LB and NLB sentence pairs as in experiments 1, 2, and 3. In addition, there were 240 mono-morphemic 3-segment words taken from the *British National Corpus*, which were presented after the sentences.

Experimental Design

Six words taken from the list of mono-morphemic words were randomly selected for each of the 40 sentences. These six words and their order of presentation remained constant across participants. Two examples are given in (4) (the LB and NLB strings are underlined; the PW appears in boldface).

(4) Target: *His friends <u>got **nothing** to do</u> next Friday. Catch, cheque, hair, knock, lane, road*.
Control: *Ron thinks <u>you **want** it to do</u> another one. Arm, case, crew, heat, team, tool.*

These sentences and the additional words following them formed a trial block. The trial blocks were then separated into 2 balanced lists of 20 trials.

Procedure

Participants were split into groups A and B. Participants in group A saw one list and participants in group B saw the other list. The stimuli were presented on a Mac using the Actuate stimuli presentation software developed by Chris Westbury. The sentence appeared in the centre of the screen for 3,000 ms and then disappeared. Subsequently, each additional word was individually shown for 1,500 ms. Finally, participants were prompted to type in the sentence and as many words as they could remember. Each trial (i.e., a sentence in addition to the six words following it) was presented in random order.

Results

Two dependent variables are of concern: Accuracy of recall of the LB/NLB sequence and number of words recalled. Accuracy of recall was calculated using the Damerau-Levenshtein algorithm edit distance between the recalled sequences and the target ones; the resulting distance was then divided by the length of the LB/NLB strings. The data was analysed using linear mixed-effects regression with a quasi binomial distribution, where subjects and items entered the model as random effects and lexical bundlehood as a fixed effect. The results are summarized in Table 5.

Table 5
Linear Mixed-Effects Regression for Accuracy of Recall in Visual Word and Sentence Recall Experiment

Random Effects			
	Variance	*SD*	
Subject	0.0006	0.02	
Item	0.0009	0.03	
Residual	0.0089	0.09	

Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	0.97	0.01	21.1***
NLB	-0.03	0.008	-3.8***

Note. Estimates and standard errors are correspond to the ratio of accuracy of recall . N = 600. Model log-likelihood = -49.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

To paraphrase Table 5, LBs ($M = 0.97$) are recalled more accurately ($t(599) = -3.8$, $p < .001$, $d = 0.32$, $r = 0.16$) than NLB ($M = 0.94$) after subject and item variability have been accounted for.

We now turn to the number of additional words recalled after sentences. Words that were not part of the target list were rejected (minor misspellings such as inversions and accidental omissions were tolerated). We performed the same linear mixed-effects regression analysis as for accuracy of recall. Results are given in Table 6.

Table 6
Linear Mixed-Effects Regression for Number of Words of Recalled in Visual Word and Sentence Recall Experiment

Random Effects			
	Variance	*SD*	
Subject	0.51	0.72	
Item	0.08	0.29	
Residual	1.45	1.20	

Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	2.8	0.1	19.5***
NLB	-0.3	0.1	-3.4***

Note. Estimates and standard errors are correspond to number of words recalled . N = 600. Model log-likelihood = -1014.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

In brief, Table 6 shows that, once subject and item variability removed, participants recall more words after LBs than after NLBs ($M_{LB} = 2.8$, $M_{NLB} = 2.5$, $t(599) = -3.4$, $p < .001$, $d = 0.2$, $r =$

0.1). These results parallel findings made by Watkins and Watkins (1977) who report greater probability of recall for high-frequency four-syllable words in serial recall tasks. However, given the processing divergence between the visual and auditory modalities (Watkins and Watkins 1977; Hue, Fang, and Sue, 1990; Beaman 2002; and Cowan, Saults, and Brown 2004), it is possible that the reduction in working memory load observed with LBs does not carry over to the auditory modality. We addressed this issue by running the word and sentence recall experiment in the auditory modality.

Experiment 5

Participants

Same as in experiments 1, 2, 3, and 4. None of the participants had done experiments 1, 2, 3, or 4.

Materials

The same as in experiment 4. For each sentence and additional word, an audio file was created with Natural Reader version 6.3 (a text-to-speech synthesizer) using the English Neo voice "Kate". This was done so as to factor out any prosodic cues that might be associated with LBs (but not with NLBs).

Experimental Design

Same as in experiment 4.

Procedure

Same as in experiment 4. In addition, the stimuli were read at a speed of approximately 3.7 words per second (speed –2 in Natural Reader). There was a 1,500 ms break between the end of the sentence and the first additional word, and a 1,500 ms break between each additional word to be recalled.

Results

As in experiment 4, accuracy of recall of LB/NLB sequence and number of words recalled were of interest. Accuracy of recall was calculated as in experiment 4. Results of the linear mixed-effects regression are provided in Table 7.

Table 7
Linear Mixed-Effects Regression for Accuracy of Recall in Auditory Word and Sentence Recall Experiment

Random Effects			
	Variance	*SD*	
Subject	0.0002	0.01	
Item	0.0015	0.04	
Residual	0.0074	0.09	

Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	0.96	0.01	22.8***
NLB	-0.01	0.007	-1.8*

Note. Estimates and standard errors are correspond to the ratio of accuracy of recall . N = 649. Model log-likelihood = -49.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

Table 7 indicates that after accounting for subject and item variability, LBs (M = 0.96) are recalled more accurately ($t(648) = -1.8$, $p < .05$, $d = .14$, $r = .07$) than NLB (M = 0.95). We proceeded as in experiment 4 regarding the data analysis for number of additional words recalled after sentences (results are shown in Table 8).

Table 8
Linear Mixed-Effects Regression for Number of Words of Recalled in Auditory Word and Sentence Recall Experiment

Random Effects			
	Variance	*SD*	
Subject	0.50	0.71	
Item	0.16	0.40	
Residual	1.01	1.00	

Fixed Effects			
	Estimate	*SE*	*t* value
Intercept (LB)	3.25	0.15	21.3***
NLB	-0.04	0.08	.5

Note. Estimates and standard errors are correspond to number of words recalled . N = 649. Model log-likelihood = -990.
* = $p < .05$. ** = $p < .01$. *** = $p < .001$.

Contrary to visual experiment, LBs in the auditory experiment did not contribute to decreasing working memory load ($M_{LB} = 3.3$, $M_{NLB} = 3.3$, $t(648) = .5$, $p > .05$).

Discussion

The recall process precluding over-practised processing, the LB facilitatory effect observed in the visual modality as well as in accuracy of recall in the auditory modality can be taken as evidence of holistic storage. To our surprise, however, there were no differences between LBs and NLBs in the auditory modality in terms of number of words recalled. Previous research has shown that formulaic sequences are produced with an integral intonation contour (Aijmer 1996; Raupach 1984; Wiese 1984). In order to factor out any facilitatory effect LBs that might be related to this, the audio files were created with a text-to-speech synthesizer. It would seem, however, that intonation contour is necessary for listeners to recognize LBs as such. In addition, the lack of differentiation between LBs and NLBs with respect to intonation might be reflected in the relatively small effect size found for accuracy of recall in this modality. In order to ascertain whether lack of intonation contour was at fault, it will be necessary to replicate this study using natural human voices.

As a final note, a substantial body of research has demonstrated a superiority effect in serial recall for the auditory modality (Watkins and Watkins 1977; Hue, Fang, and Sue, 1990; Beaman 2002; and Cowan, Saults, and Brown 2004). This effect was found for number of words recalled ($W = 145{,}202$, $p < .001$, $d = .46$, $r = .22$) but not for accuracy of recall ($W = 192{,}981$, $p > .05$). Again, this is possibly due to the fact that the audio files were created with a text-to-speech synthesizer.

Conclusion

A significant body of research suggests that (at least some) children first learn chunks and then decompose them at a later stage into smaller units (e.g., Ellis 1996, 1998; Wray 2002). By way of example, a friend's one-and-a-half year old named Erin loves a game where her mother Tracy tries to catch her. Every time Tracy would initiate the game she would say *I'm gonna catch you* and then run after Erin. After some time, Erin developed the habit of initiating the game herself. She would go up to her mother and say *I'm gonna catch you* and then run away in the hopes that Tracy would play the game. Clearly, Erin has associated the whole string of sounds *I'm gonna catch you* with the meaning "let's play the game where you try to catch me". She has evidently learned it and uses it as a non-decomposed, holistic unit. The point to be made here is that at least some more or less complex linguistic units contained in our mental lexicon, the ones that were acquired during the (very) early stages of our lives, are necessarily stored as a whole.

Nonetheless, humans are very powerful pattern finders (Bowers, Davis, and Hanley 2005), and in order to find patterns, we must decompose, analyse. For example, Libben (1994, 1998, 2005b) and Libben and de Almeida (2002) have found that compound-word decomposition is automatic and obligatory. Coming back to our earlier example, Erin has certainly decomposed the string *I'm gonna catch you* but at the time she had insufficient data and experience with the language so as to build up in her mental lexicon the smaller interrelated entries *I'm*, *gonna*, *catch*, *you*. Nonetheless, with time and experience these entries would eventually be created and she would know that these words go together. If every linguistic unit is automatically decomposed, whether it be holistically stored or not, then how can we account for differences between an LB and a NLB? Instead of saying that LBs are holistically retrieved as opposed to NLBs, Libben (2005b: 276) would suggest that the difference is attributable to LBs being processed less than NLBs. There is physiological evidence lending support to this view.

Using functional magnetic resonance imaging (fMRI), Bischoff-Grether et al. (2000) found a negative correlation between activation of Wernicke's area and its right homologue and predictability of nonverbal sequences. That is, less predictable sequences were associated with increased activation of these areas than more predictable ones. Similar findings are reported for perceptual priming tasks, where previously presented stimuli correlate with decreased activation in visual areas compared to new items (Koutstaal et al. 2001); for dot pattern classification tasks (Reber et al. 1998), where a decrease in activation of BA 18 was observed for test items generated from a prototype compared to randomly generated items; and for overt picture naming (Graves et al. 2007) and word reading (Fiez et al. 1999) tasks, where increased activation of the posterior superior temporal gyrus (Wernickes' area) correlated with low-frequency words. This, however, is only part of the story, as activation decrement of one part of the brain seems to entail increased activation of other neural networks.

Recent findings indicate a competitive relationship during task performance between areas related to processing and areas associated with holistic/fact retrieval. In an artificial grammar learning tasks, where grammaticality and frequency of occurrence of two- and three-letter chunks (i.e., chunk strength) were controlled for, Lieberman et al. (2004) report a strong negative correlation between activation of the caudate nucleus for grammatical versus non-grammatical items, and activation of the hippocampus and medial temporal lobe for high rather than low chunk strength items. Similar patterns of activation are reported for weather prediction tasks (Poldrack et al. 2001).

There is also evidence from event-related brain potential recordings during arithmetic problem-solving tasks (Jost et al. 2004) pointing towards competitive dual-route processing. Given differences in amplitude and topography of the left-anterior negativity deflection elicited by zero, small, and large multiplication problems, Jost et al. propose that default fact retrieval strategies are replaced by distinct rule-based processing in the event of retrieval failure.[4] In consideration of the brain imaging studies discussed here, it can be conjectured that the hippocampus and medial temporal lobe are associated with fact retrieval, whereas the caudate nucleus and posterior superior temporal gyrus are linked to rule-based processing. Thus, upon stimulus presentation areas related to fact retrieval would first be activated then followed by activation of networks associated with rule-based processing in the event of retrieval failure accompanied by deactivation of cell assemblies linked to fact retrieval. Coming back to lexical bundles, it is plausible that the processing advantage LBs have over NLBs stems from this very same process. That is, NLBs would take longer to read because their processing entails more second stage activation (i.e., rule-based processing) than first stage activation (i.e., fact retrieval), whereas LB processing involves more fact retrieval than rule-based processing, if any at all.

One last point needs consideration. It is widely assumed in the formulaic sequences literature that these entities are linked to specific discourse functions and usually appear in certain positions in a sentence. For instance, the sequence *you know what I mean* would appear at the end of a sentence in order to request feedback. Regarding lexical bundles more specifically, Biber et al. (2003) have developed a taxonomy to classify discourse functions of LBs found in

---

[4] The left-anterior negativity is a brain potential normally associated with working memory and processing load (Müller, King, and Kutas 1997, 1998; Vos et al. 2001; Kutas and Schmitt 2003; Hagiwara et al. 2007).

the conversation and academic prose. The present study brings mitigating evidence suggesting that LBs are not intrinsically related to specific discourse functions. Indeed, the stimuli used here were not embedded in their usual place within a sentence and as such did not carry the discourse functions they have been said to portray, if any at all. It would thus seem that even though LBs might bear more often than not a set of specific discourse functions, there is no inherent association between the two in the mental lexicon.

References

AIJMER, K. 1996. *Conversational Routines in English: Convention and Creativity*. London: Longman.

ALTENBERG, B. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (ed.). *Phraseology: Theory, Analysis, and Applications* (pp. 101-22). Oxford: Clarendon Press.

BEAMAN, PHILIP B. 2002. Inverting the modality effect in serial recall. *The Quartely Journal of Experimental Psychology*, 55A, 371-389.

BIBER, DOUGLAS. 1999. Investigating language use through corpus-based analyses of association patterns. In M. Barlow and S. Kemmer (eds.). *Usage Based Models of Language* (pp. 287-313). Stanford, CA: CSLI Publications.

BIBER, DOUGLAS AND CONRAD, SUSAN. 1999. Lexical bundles in conversation and academic prose. In H. Hasselgård and S. Oksefjell (eds.). *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181-190). Amsterdam and Atlanta: Rodopi.

BIBER, DOUGLAS, CONRAD, SUSAN, AND CORTES, V. 2003. Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, and Tony McEnery (eds.). *Corpus Linguistics by the Lune* (pp. 71-93). Frankfurt/Main: Peter Lang.

BIBER, DOUGLAS, JOHANSSON, STIG, LEECH, GEOFFREY, CONRAD, SUSAN AND FINEGAN, EDWARD. 1999. *Grammar of Spoken and Written English*. Harlow: Longman.

BISCHOFF-GRETHE, AMANDA, PROPER, SHAWNETTE M., MAO, HUI, DANIELS, KAREN A., AND GREGORY S. BURNS. 2000. Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *The Journal of Neuroscience*, 20, 1975-81.

BLOOMFIELD, LEONARD. 1933. *Language*. New York: Henry Holt and Company.

BOD, RENS. 2001. *Sentence Memory: Storage vs. Computation of Frequent Sentences*. Retrieved on October 29, 2006 at http://staff.science.uva.nl/~rens/cuny2001.pdf.

BOLINGER, D. 1976. Meaning and memory. *Forum Linguisticum*, 1, 1-14.

BOWERS, JEFFREY S., DAVIS, COLIN J., AND HANLEY, DEREK A. 2005.Automatic semantic activation of embedded words: Is there a "hat" in "that"?. *Journal of Memory and Language*, 53, 131-43.

BUTLER, C.S. 1997. Repeated word combinations in spoken and written text: Some implications for Functional Grammar. In C.S. Butler, J.H. Connolly, R.A. Gatward, and R.M. Vismans (eds.). *A Fund of Ideas: Recent Developments in Functional Grammar* (pp. 60-77). Amsterdam: IFOTT, University of Amsterdam.

CONKLIN, KATHY AND SCHMITT, NORBERT. 2007. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 1-18.

COWAN, NELSON. 2000. The magical number 4 in short-term memory: A reconsideretion of mental storage capacity. Behavioral and brain sciences, 24, 87-185.

COWAN, NELSON, SAULTS, SCOTT J., AND BROWN, GORDON D.A. 2004. On the auditory modality superiority effect in serial recall: Separating input and output factors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 639-644.

CROFT, WILLIAM. 2001. Radical Construction Grammar. Syntactic Theory in Typological Perspective. Oxford: Oxford University Press.

CURTIS, CLAYTON E., AND D'ESPOSITO, MARK. 2006. Functional neuroimaging of working memory. In Roberto Cabeza and Alan Kingstone (eds.). *Handbook of Functional Neuroimaging of Cognition*, pp. 269-306. Cambridge, MA: MIT Press.

ELLIS, NICK C. 1998. Emergentism, connectionism and language learning. *Language Learning*, 48, 631-64.

ELLIS, NICK C. 1996. Sequencing in SLA. Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91-126.

ERMAN, BRITT AND WARREN, BEATRICE. 2000. The idiom principle and the open choice principle. *Text*, 20, 29-62.

FIEZ, JULIA A., BALOTA, DAVID A., RAICHLE, MARCUS E., AND PETERSEN, STEVEN E. 1999. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, 24, 205-218.

FILLMORE, CHARLES. 1979 On fluency. In Charles Fillmore, D. Kempler, and S.-Y.W. Wand (eds.). *Individual Differences in Language Ability and Language Behaviour* (pp. 85-101). New York: Academic Press.

FIRTH, J.R. 1937/1964. *The Tongues of Men and Speech*. London: Oxford University Press.

FIRTH, J.R. 1957/1968. A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis: Special Volume of the Philological Society* (pp. 1-32). Oxford: Blackwell. Reprinted in F.R. Palmer (ed.). *Selected Papers of J.R. Firth 1952-59*, (pp. 168-205). Harlow: Longman.

GIBBS, RAYMOND W., BOGDANOVICH, JOSEPHINE M., SYKES, JEFFREY R., AND BARR, DALE J. 1997. Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, 141-154.

GRAVES, WILLIAM W., GRABOWSKI, THOMAS J., MEHTA, SONYA, AND GORDON, JEAN K. 2007. A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*, 19, 617-631.

HAGIWARA, HIROKO, SOSHI, TAKAHIRO, ISHIHARA, MASAMI, AND IMANAKA, KUNIYASU. 2007. A topographical study of the event-related potential correlates of scrambled word order in Japanese complex sentences. *Journal of Cognitive Neuroscience*, 19, 175-193.

HUE, CHIH-WEI, FANG, DAVID, AND HSU, KUNG-YU. 1990. Immediate serial recall of Chinese characters: A study of input modality effect. *Acta Psychologica*, 75, 213-223.

HUGHLINGS JACKSON, JOHN. 1866. Notes on the physiology and pathology of language. In J. Taylor (ed.). *Selected Writings of John Hughlings Jackson, vol. 2* (pp. 121-128). London: Staples Press.

HUNSTON, SUSAN AND FRANCIS, GILL. 2000. *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam and Philadelphia: John Benjamins.

HYMES, DELL H. 1962. The ethnography of speaking. In *Anthropology and Human Behaviour*, (pp. 13-53). Washington, D.C.: The Anthropological Society of Washington.

JERPERSEN, OTTO. 1924. *The Philosophy of Grammar*. London: George Allen and Unwin.

JOST, KERSTIN, BEINHOFF, ULRIKE, HENNIGHAUSEN, ERWIN, AND RÖSLER, FRANK. 2004. Facts, rules, and strategies in single-digit multiplication: Evidence from event-related brain potentials. *Cognitive Brain Research*, 20, 183-193.

KOUTSTAAL, W., WAGNER, A.D., ROTTE, M., MARIL, A, BUCKNER, R.L., AND SCHACTER, D.L. 2001. Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, 39, 184-199.

KUTAS, MARTA, AND SCHMITT, BERNADETTE M. 2003. Language in microvolts. In Marie T. Banich and Molly Mack (eds.). *Mind, Brain, and Language. Multidisciplinary Perspectives*, pp. 171-209. Mahwah, NJ: Erlbaum.

LIBBEN, GARY. 2005a. Why study compound processing? An overview of the issues. In Gary Libben and Gonia Jarema (eds.). *The Representation and Processing of Compound Words*, pp. 11-37. Oxford: Oxford University Press.

LIBBEN, GARY. 2005b. Everything is psycholinguistic: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics*, 50, 267-84.

LIEBERMAN, MATTHEW D., CHANG, GRACE Y., CHIAO, JOAN, BOOKHEIMER, SUSAN Y., AND KNOWLTON, BARBARA J. 2004. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, 16, 427-438.

MAHER, BRENDAN A. AND SKOVENGAARD, JAN. 1988. Contextual constraint: a Danish language replication of Miller-Selfridge with methodological improvements. *Scandinavian Journal of Psychology*, 29, 194-9.

MILLER, GEORGE A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97. Reprint found at http://www.well.com/~smalin/miller.html.

MILLER, GEORGE A., AND SELFRIDGE, JENNIFER A. 1950. Verbal context and the recall of meaningful material. *The American Journal of Psychology*, 63, 176-85.

MULLER H.M., KING, J.W., AND KUTAS, M. 1997. Event-related potentials to relative clause processing in spoken sentences. *Cognitive Brain Research*, 5, 193-203.

MULLER H.M., KING, J.W., AND KUTAS, M. 1998. Electrophysiological analysis of spoken sentence processing as a function of working memory requirements. *Klinische Neurophysiologie*, 29, 321-30.

ORTONY, ANDREW, SCHALLERT, DIANE L., REYNOLDS, RALPH E., AND ANTOS, STEPHEN J. 1978. Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal Language and Verbal Behaviour*, 17, 465-477.

PAWLEY, ANDREW AND SYDER, FRANCES HODGETTS. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Jack C. Richards and Richard W. Schmidt (eds.). *Language and communication* (pp. 191-226). New York: Longman.

POLDRACK, R. A., CLARK, J. PARÉ-BLAGOEV, E.J., SHOHAMY, D., CRESO MOYANO, J., MYERS, C., AND GLUCK, M.A. 2001. Interactive memory systems in the human brain. *Nature*, 414, 546-550.

R DEVELOPMENT CORE TEAM (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

RAUPACH, M. 1984. Formulae in second language speech production. In H.W. Dechert, D. Mole, and M. Raupach (eds.), *Second Language Productions* (pp. 114-137). Tübingen, Germany : Gunter Narr Verlag.

REBER, P.J., STARK, C.E.L., AND SQUIRE, L.R. 1998. Contrasting cortical activity associated with category memory and recognition memory. *Learning and Memory*, 5, 420-428.

REICHLE, ERIK D., RAYNER, KEITH, AND POLLATSEK, ALEXANDER. 2003. The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences* 26, 445–526.

SALEM, A. 1987. *Pratique des Segments Répétés*. Paris : Institut National de la Langue Française.

SAUSSURE, FERDINAND DE. 1916/1966. *Course in General Linguistics*. New York: McGraw-Hill.

SAVIN, HARRIS B. AND PERCHONOCK, ELLEN. 1965. Grammatical structures and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behaviour*, 4, 348-53.

SCHMITT, NORBERT. 2004. *Formulaic Sequences*. Amsterdam and Philadelphia: John Benjamins.

SCHMITT, NORBERT, GRANDAGE, SARAH, AND ADOLPHS, SVENJA. 2004. Are corpus-derived recurrent clusters psychologically valid? In Norbert Schmitt (ed.). *Formulaic Sequences* (pp. 127-151). Amsterdam and Philadelphia: John Benjamins.

SCHMITT, NORBERT AND UNDERWOOD, GEOFFREY. 2004. Exploring the processing of formulaic sequences through a self-paced reading task. In Norbert Schmitt (ed.). *Formulaic Sequences*, pp. 171-89. Amsterdam and Philadelphia: John Benjamins.

SCHWEIGERT, WENDY A. AND MOATES, DANNY R. 1988. Familiar idiom comprehension. *Journal of Psycholinguistic Research*, 17, 281-296.

STAUB, ADRIAN, AND RAYNER, KEITH. (2007). Eye movements and on-line comprehension processes. In M. Gareth Gaskell (ed.). *The Oxford Handbook of Psycholinguistics* (pp. 325-342). Oxford: Oxford University Press.

TITONE, DEBRA A. AND CONNINE, CYNTHIA C. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, 31, 1655, 1674.

TULVING, ENDEL AND PATKAU, JEANETTE E. 1962. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, 16, 83-95.

UNDERWOOD, GEOFFREY, SCHMITT, NORBERT AND GALPIN, ADAM. 2004. The eyes have it: An eye-movement study into the processing of formulaic sequences. In Norbert Schmitt (ed.). *Formulaic Sequences*, pp. 153-172. Amsterdam and Philadelphia: John Benjamins.

VOS, SANDRA H., GUNTER, THOMAS C., KOLK, HERMAN H.J., AND MULDER, GIJSBERTUS. 2001. Working memory constraints on syntactic processing: An electrophysiological exploration. *Psychophysiology*, 38, 41-63.

WATKINS, OLGA C., AND WATKINS, MICHAEL J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 712-718.

WIESE, R. 1984. Language production in foreign and native languages: Same or different? In H.W. Dechert, D. Mole, and M. Raupach (eds.), *Second Language Productions* (pp. 11-25). Tübingen, Germany : Gunter Narr Verlag.

WRAY, ALISON. 2002. Formulaic Sequences and the Lexicon. Cambridge: Cambridge University Press.

Appendix A1

Stimuli Group A 1st Set

Item	LB = 1; NLB = 2	Trials	Question	Correct Answer
1	2	His friend's got one to do next Friday.	Is there a weekday mentioned in the sentence?	y
2	2	If workers don't know about it nothing will happen.	Does the sentence mention workers?	y
3	1	Ron thinks you want me to do another one.	Is there an animal mentioned in the sentence?	n
4	1	Yeah although you might as well buy one.	Should he borrow one?	n
5	1	Tell me when you want me to return it.	Does the sentence say anything about returning something?	y
6	2	Would he like to stop and have to look inside it?	Does the sentence mention anything about stopping?	y
7	1	Yes, everything I said to her was sacred.	Was everything he said ordinary?	n
8	1	I sat in the middle of the bullet train.	Did I sit in an underground train?	n
9	2	But unfortunately all the top of it kept burning.	Did it keep burning?	y
10	2	He's glad you do want to dig tunnels.	Is he sad?	n
11	1	Now, must I tell you what I discovered yesterday?	Is the sentence about walking?	n
12	2	I might, I do if you seriously care.	Is there a possibility that he will do it?	y
13	1	But honestly, I don't think he ran away.	Does the sentence mention anything about eating?	n
14	2	Yeah, maybe I'll get you what these guys want.	Is the sentence about getting something?	y
15	2	He believes you do know what David did.	Does he believe you know?	y
16	2	Sam assumes you know where you begin singing.	Is the sentence about donuts?	n
17	1	Indeed, whatever you think about it feels weird.	Is the thing pleasant?	n
18	1	I confess I don't know what Smith wants.	Did the person confess?	y
19	2	I admit I do know whether Jack cheated.	Does the person know whether Jack cheated?	y
20	1	I realize I don't know how research is done.	Is the sentence about skiing?	n

Appendix A2

Stimuli Group A 2nd Set

Item	LB = 1; NLB = 2	Trials	Question	Correct Answer
1	1	His friend's got nothing to do next Friday.	Does his friend have something to do?	n
2	1	If workers don't worry about it nothing will happen.	If workers don't worry, will something happen?	n
3	2	Ron thinks you want it to do another one.	Is there a man named Ron in the sentence?	y
4	2	Yeah, although you would as well buy one.	Should he buy one?	y
5	2	Tell me when you see me to return it.	Will he return it next week?	n
6	1	Would he like to stop and have a look inside it?	If he stops, will he buy something?	n
7	2	Yes, everything I was to her was sacred.	Was he sacred to her?	y
8	2	I sat in the front of the bullet train.	Was sitting in the train?	y
9	1	But unfortunately all the rest of it kept burning.	Did it stop burning?	n
10	1	He's glad you don't want to dig tunnels.	He doesn't want to dig tunnels, right?	y
11	2	Now, must I get you what I discovered yesterday?	Is the sentence about a discovery?	y
12	1	I might, I mean if you seriously care.	Does the sentence mention a country?	n
13	2	But honestly, I do think he ran away.	Does he think the man ran away?	y
14	1	Yeah, maybe I'll tell you what these guys want.	Does the sentence mention ducks?	n
15	1	He believes you don't know what David did.	Is the man called Patrick?	n
16	1	Sam assumes you know when you begin singing.	Does the sentence mention something about Sam?	y
17	2	Indeed, whatever you do about it feels weird.	Is the sentence about feelings?	y
18	2	I confess I do know what Smith wants.	Does the sentence mention a brand name?	n
19	1	I admit I don't know whether Jack cheated.	Does he know if he cheated?	n
20	2	I realize I do know how research is done.	Does he know how it is done?	y