Are Lexical Bundles Stored and Processed as Single Units?

Antoine Tremblay, Bruce Derwing, and Gary Libben

University of Alberta
antoinet@ualberta.ca
www.ualberta.ca/~antoinet

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 4- and 5-word lexical bundles and their controls, which did not contain lexical bundles, were presented to participants in a word-by-word, chunk-by-chunk, and sentence-by-sentence fashion. The stimuli were controlled for token and bundle frequency, transitional probabilities and morphological and phonological complexity. In the word-by-word experiment, lexical bundle sequences were not read significantly faster than non-lexical bundle strings, thus suggesting that word-by-word presentation disrupts the facilitatory effect of bundles. However, lexical bundles and sentences containing lexical bundles were read significantly faster than their controls in the other two self-paced reading experiments as predicted by the theory.

Keywords: Psycholinguistics; Lexical-bundles; Word-by-word, chunk-by-chunk, and sentence-by-sentence self-paced reading.

1   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, identified as such with the help of corpora. 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 out with him*), and idioms -- *to put one's finger in the dike* (Croft 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). He was not the only 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 existence of multi-word units stored in the mind of speakers noting that language would be too difficult to manage if one had to remember every individual item separately. According to Bloomfield (1933: 181), "many forms lie on the border-line between bound forms and words, or between words and phrases". Firth, for his part, considers that the units of speech are phrases (1937/1964), and that for one to characterize a certain community's speech, one has to list the usual collocations used by its speakers, that is, the set of words that frequently recur with a particular word (1957/1968). Miller (1956) argues that our short-term memory span is limited to seven, plus or minus two units. Nonetheless, we manage to circumvent this severe limitation by organizing information into chunks. Thus, while the number of units we can process at any given time remains constant, we can significantly increase the amount of information contained in each unit and therefore increase the total amount of information we can manage. For Hymes (1962: 41), a large part of communication involves the use of recurrent patterns, that is, of "linguistic routines". Bolinger (1976: 1) maintains 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) writes that knowing how to use formulaic utterances makes up a large part of a speaker's ability to successfully handle language. During the Chomskyan era, which started in the 1950s, formulaic language other than non-compositional idioms were 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 occur 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 texts. 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. Unfortunately, very few psycholinguistic studies have considered the question of how they are stored and processed in the mind

(Schmitt 2004: viii), which, moreover, 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 according to which 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 significantly more accurately than non-formulaic sequences. In the face of such few and mixed results, the question of whether formulaic sequences are stored and processed holistically in the mind 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 processed holistically. We approached the question by conducting three self-paced reading experiments. The reasoning behind them is as follows. Consider for instance a sentence composed of 9 units (words). If a 4-word LB is stored and processed as a whole, a participant should merely have to compute 6 units when reading the sentence. However, non-lexical bundles (NLBs) are not stored and processed holistically, and a participant will thus have to compute 9 units. We thus compared sentences that contained LBs to equivalent sentences that did not. Our prediction was that sentences containing LBs would be read more quickly than those that did not contain LBs. In order to determine whether context is necessary for holistic processing and if so, how much of it is needed, the stimuli were presented word-by-word (experiment 1), chunk-by-chunk (experiment 2), and sentence-by-sentence (experiment 3). The three experiments are described in sections 2, 3, and 4 respectively.

2   Experiment 1

Schmitt and Underwood (2004) 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 are 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 several function words)

(iii)   the sequences did not finish 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 the 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 for each story were devised so as 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 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 in the following lines.

2.1   Participants

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

2.2   Materials

Lexical bundles (LB) were taken from Biber et al. (1999). Their respective frequency was checked against 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).[1] 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). LBs for which an appropriate NLB control string could be found were kept (20 in total). By appropriate NLB, it is meant NLB strings that met the criteria described in (1)-(4) below.

The actual stimuli consisted of 6 practice trials, 20 target sentences containing LBs paired with 20 control sentences where one word in the target LB string was changed so that it did not constitute an LB (see Appendix). That is, the control sentences differed from the target sentences in only one word. By way of example, consider the following target sentence: *But honestly, I don't think he ran away*. 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: *But honestly, I do think he ran away*. The two differ only in one

---

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

word (which is called the Pivot Word, PW, in bold here), that is, "don't" in the target and "do" in the control. 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 the sentences was 8.6 words (SD 0.8). In an attempt to avoid confounds, various other constraints were also built into each target LB and control NLB pair:

(1)   the token frequency of the PW in the NLB was greater than that of the PW in the LB;

(2)   both the morphological and phonological complexity of the PW in the NLB string was equal to or simpler than that of the PW in the LB;

(3)   the preceding-word-to-PW probability was on average 2.4 times (SD 3.6) greater in the NLB string; and

(4)   in the perfect world, the PW-to-following-word probability would have been greater in the NLB than in the LB across the board. However, it was only possible to construct 4 NLB strings that met this criterion (on average 1.3 times greater, SD 1.3). In the other 16 pairs, the PW-to-following-word probability was on average 19.1 times greater in the LB than in the control NLB (SD 23.5).

In addition, none of the control strings contained any LBs. An effort was also made so that the rest of the sentences did not contain LBs.

2.3   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 (there were 20 "yes" answers and 20 "no" answers).

2.4   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 sentence (Position: centred, Font: Arial, Size: 48); (iii) once the participants had finished reading the sentence, they pressed a key; (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); and (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 (5).

(5)   Word-by-word 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*

In (5), the underlined portion of each example corresponds to the LB and NLB strings and the word in bold is the PW.

2.5   Results

Only the LB and NLB strings were taken into account. Trials involving wrong answers to the questions were eliminated as well as those with reading times three standard deviations above or under the mean. Overall, 5.2% (86/1,665 trials) were thrown out. Regarding the by-subject analysis, a 2 x 2 repeated measures analysis of variance (ANOVA) was performed on the trimmed data for the 1st and 2nd sets (group x lexical-bundlehood). The results of the analyses are given in Table 1.

Table 1
Experiment 1, 1st and 2nd sets: By-subject repeated measures ANOVA.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	576	(383)			
1	NLB	610	(390)	34	8.1	< 0.05
2	LB	376	(191)			
2	NLB	380	(197)	4	0.7	> 0.05

In the 1st set, the 34 ms difference in summed reading times between LB segments and NLB ones was significant [$F1(1, 18) = 8.1, p < 0.05$]. The 4 ms difference in summed reading times in the 2nd set, however, did not reach significance [$F1(1, 18) = 0.7, p > 0.05$]. Moreover, the interaction between group and lexical-bundlehood in the 1st set was not significant [$F1(1, 18) = 0.6, p > 0.05$] but the interaction was significant in the 2nd set [$F1(1, 18) = 4.4, p < 0.05$] (not shown in Table 1).

Regarding the by-item analysis, a paired *t*-test was performed on the trimmed data (LBs vs. NLBs) for each set. The results of the analyses are shown in Table 2.

Table 2

Experiment 1, 1st and 2nd sets: By-item paired t-test analysis.

Set	LBhood	Mean RT	(SD)	Difference	$t(19)$	p
1	LB	583	(38)			
1	NLB	613	(83)	38	-1.5	> 0.05
2	LB	381	(43)			
2	NLB	379	(50)	2	0.1	> 0.05

There were no significant differences in reading times in either set [$t2(19) = -1.5$, $p > 0.05$; $t2(19) = 0.1$, $p > 0.05$ respectively].

Given that the 2nd set by-subject analysis and both the 1st and 2nd set by-item analyses reveal insignificant differences in reading times, it is considered that a type II error was committed in the significant 1st set by-subject analysis (i.e., we failed to reject the null hypothesis when it was false). In brief, the statistical tests presented here show that LB sequences are not read significantly faster than NLB strings.

2.6   Discussion

Though there were no differences in reading times between LB and NLB sequences, it is possible that an LB facilitatory effect occurs in the word following the LB/NLB strings. Let [W] be the word that a participant is currently reading, [W+1] the word following it, and [W–1], [W–2], …, [W–n] words occurring before word [W]. Possibly, participants press a key to see word [W+1] even though they have not finished integrating word [W] to the previous word sequence [W–n, W–2, W–1]. That is, if there is processing spill-over and if LBs are processed faster than NLBs, then there should be less spill-over at word [W+1] and these words should be read faster after LBs than NLBs. It is important to mention that word [W+1] is the same after both LBs and NLBs (cf. (6)), where the LB/NLB appears in bold and the word following them is underlined).

(6)   a.   Target: If workers **don't worry about it** <u>nothing</u> will happen.
       b.   Control: *If workers **don't know about it** <u>nothing</u> will happen.*

A by-subject repeated measures ANOVA and a by-item paired t-test — summarized in Tables 3 and 4 — were performed on RTs associated with the word following LB/NLB strings for both sets.

Table 3.

Experiment 1, 1st and 2nd sets: By-subject analysis of the word following LB and NLB strings.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	612	(349)			
1	NLB	670	(423)	58	8.3	< 0.05
2	LB	404	(200)			
2	NLB	423	(243)	19	1.7	> 0.05

Table 4.
Experiment 1, 1st and 2nd sets: By-item analysis of the word following LB and NLB strings.

Set	LBhood	Mean RT	(SD)	Difference	$t(119)$	p
1	LB	618	(92)			
1	NLB	684	(141)	66	-2.0	> 0.05
2	LB	404	(67)			
2	NLB	427	(83)	23	-0.8	> 0.05

There was a significant main effect of lexical-bundlehood in the by-subject analysis for the 1st set [$F1(1, 18) = 8.3$, $p < 0.05$]. However, the by-subject analysis for the 2nd set did not yield any significant differences [$F1(1, 18) = 1.7$, $p > 0.05$] neither were there any significant differences in the by-item analyses for the 1st set [$t2(19) = -2.0$, $p > 0.05$] nor the 2nd set [$t2(19) = -0.8$, $p > 0.05$]. Similarly to the analysis performed on the summed RTs for the LB and NLB sequences, it will be considered that a type II error was committed in the 1st set by-subject analysis. In sum, the word following an LB sequence was not read significantly faster than words following NLBs.

What if the integration process was resolved at the last word of the sentences? If this is the case, the LB facilitatory effect might only be apparent in RTs associated with this word. That is, the last word of a sentence would be read faster when the sentence would contain an LB than when it would not. A by-subject repeated measures ANOVA and a by-item paired *t*-test was performed on RTs associated with the last word of the sentences. The results of these analyses are summarized in Tables 5 and 6.

Table 5.
Experiment 1, 1st and 2nd sets: By-subject analysis of the last word of sentences.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	851	(637)			
1	NLB	919	(710)	68	2.6	> 0.05
2	LB	521	(383)			
2	NLB	506	(332)	15	0.6	> 0.05

Table 6.
Experiment 1, 1st and 2nd sets: By-item analysis of the last word of sentences.

Set	LBhood	Mean RT	(SD)	Difference	$t(119)$	p
1	LB	840	(184)			
1	NLB	960	(251)	120	-1.5	> 0.05
2	LB	530	(156)			
2	NLB	481	(104)	49	0.9	> 0.05

Neither the by-subject [1st set: $F1(1, 18) = 2.6$, $p > 0.05$; 2nd set: $F1(1, 18) = 0.6$, $p > 0.05$] or the by-item analyses [1st set: $t2(19) = -1.5$, $p > 0.05$; 2nd set: $t2(19) = 0.9$, $p > 0.05$] reach significant differences in RT between lasts words in sentences that contain LBs and those that do not. The data suggests that LBs do not procure any processing advantage over NLBs.

However, it is possible that the task is not picking up the advantage in processing time of LBs by virtue of the sentences being presented in a word-by-word fashion. If this is right, pivot words (PWs) in control strings should not be read faster than pivot words (PW) in target sequences. Indeed, it is well known that more frequent words are processed faster than

less frequent words (Taft 1979; Reichle et al. 2003). Note that all of the control PWs were more frequent than target PWs [on average 2.5 times more frequent; mean frequency PWs in LBs = 3,144 per million, SD = 4,108; mean frequency PWs in NLBs = 7,927 per million, SD = 6,239]. A by-subject repeated measures ANOVA and a paired t-test for each set were performed on the RTs associated with the PW. The summary of the analyses is given in Tables 7 and 8.

Table 7.
Experiment 1, 1st and 2nd sets: By-subject analysis of the pivot word.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	573	(356)			
1	NLB	619	(417)	46	6.3	< 0.05
2	LB	384	(201)			
2	NLB	372	(208)	15	1.4	> 0.05

Table 8.
Experiment 1, 1st and 2nd sets: By-item analysis of the pivot word.

Set	LBhood	Mean RT	(SD)	Difference	$t(119)$	p
1	LB	577	(47)			
1	NLB	614	(99)	37	-1.6	> 0.05
2	LB	389	(51)			
2	NLB	374	(56)	15	0.7	> 0.05

As in previous analyses, the 1st by-subject ANOVA is significant [$F1(1, 18) = 6.3$, $p < 0.05$] whereas the 2nd set by-subject [$F1(1, 18) = 0.7$, $p > 0.05$] as well as the 1st and 2nd sets by-item [1st set: $t2(19) = -1.6$, $p > 0.05$; 2nd set: $t2(19) = 0.7$, $p > 0.05$] analyses are not significant. Again, given that the significance of the by-subject analysis in the 1st set is not backed-up by the other analyses, it will be considered that a type II error was committed. In brief, though PWs in control sentences should have been read faster than PWs in target sentences by virtue of their being more frequent, this well know effect was not observed here.

In conclusion, the word-by-word self-paced experiment failed to produce reliable significant differences in reading times between LB and NLB sequences as well as between words immediately following them and sentence-final words.[2] According to the results presented here, LBs do not procure any processing advantage over NLBs. Nonetheless, PWs in controls were not read significantly faster than PWs in targets, as would have been expected, thus indicating that word-by-word self-paced reading experiments are not reliable techniques when it comes to measuring frequency effects. It is true that reading sentences word-by-word is very unusual and this precisely might be the reason why no effect was found (if there is one to be found). What if the stimuli were presented in a more natural manner, such as sentence-by-sentence or even in a slightly less natural fashion, chunk-by-chunk? Experiments 2 and 3 investigate this question.

---

[2] By reliable it is meant that the majority of by-subject and by-item analyses are significant, where a completely reliable significant difference would mean that the by-subject and by-item analyses in both sets are significant. The difference in reading time between LBs and NLBs in experiment 1 were not reliable because only the by-subject analysis of the 1st set was significant, whereas the by-subject analysis of the 2nd set and the two by-item analyses in the two sets were not significant.

3   Experiment 2

No differences in reading time between LBs and NLBs were found in the word-by-word self-paced reading experiment. Apparently, the nature of the task was at fault. The chunk-by-chunk self-paced experiment reported in this section aims at determining simultaneously (i) whether there is an LB facilitatory effect in on-line sentence processing, and (ii) whether the lack of positive findings in experiment 1 is really due to the manner of presentation of the stimuli.

3.1   Participants

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

3.2   Materials

Same as in experiments 1.

3.3   Experimental design

Same as in experiments 1.

3.4   Procedure

Same as experiments 1, except that the stimuli were presented in a chunk-by-chunk fashion. An example of a trial block is shown in (7).

(7)   Chunk-by-chunk 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.

3.5   Results

Only the LB and NLB chunks were taken into account. Trials involving wrong answers to the questions were eliminated as well as those with reading times three standard deviations above or under the mean. Overall, 13.4% (107/1,665 trials) were thrown out. Regarding the by-subject analysis, a repeated measures 2 x 2 ANOVA was performed on the trimmed data for the 1st and 2nd sets (group x lexical-bundlehood). The results of the analysis are given in Table 9.

Table 9

Experiment 2, 1st and 2nd sets: By-subject repeated measures ANOVA.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	1,269	(632)			
1	NLB	1,508	(542)	239	6.8	< 0.05
2	LB	914	(317)			
2	NLB	1,153	(525)	239	11.9	< 0.05

LB chunks were read 239 ms faster than NLB chunks both in the 1st [$F1(1, 18) = 6.8$, $p <$ 0.05] and 2nd sets [$F1(1, 18) = 8.1$, $p < 0.05$]. Moreover, the interaction between group and lexical-bundlehood in the 1st set did not reach significance [$F1(1, 18) = 0.6$, $p > 0.05$] but the interaction was significant in the 2nd set [$F1(1, 18) = 5.3$, $p < 0.05$] (not shown in Table IX).

Regarding the by-item analysis, a paired *t*-test analysis was performed on the trimmed data (LBs vs. NLBs) for each set. The results of the analysis are shown in Table X.

Table 10.

Experiment 2, 1st and 2nd sets: By-item paired t-test analysis.

Set	LBhood	Mean RT	(SD)	Difference	$t(119)$	p
1	LB	1,265	(400)			
1	NLB	1,505	(415)	240	-1.7	< 0.05
2	LB	923	(140)			
2	NLB	1,157	(304)	234	-4.0	< 0.05

The 240 ms difference between LB and NLB chunks in the 1st set did not reach significance, though there is a trend for LB chunks to be read faster than NLB chunks [$t2(19) = -1.7$, $p > 0.05$]. However, the 234 ms difference in the 2nd set did reach significance [$t2(19) = -4.0$, $p < 0.05$].

In brief, given that the 1st and 2nd by-subject and the 2nd set by-item analyses were significant, and given the trend seen in the 1st set by-item analysis in favour of LB chunks, it is concluded that LB chunks were read significantly faster than NLB chunks. These findings also suggest that word-by-word presentation of stimuli in self-paced reading experiments does disrupt the advantage in processing time LBs have over NLBs.

Given these findings, it is predicted that sentences containing LBs in the sentence-by-sentence self-paced reading experiment will be read more quickly than sentences that do not contain LBs. But how reliable and how large will the effect be compared to the one found in experiment 2 (if there is one to be found)?

4   Experiment 3

4.1   *Participants*

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

4.2   *Materials*

Same as in experiments 1 and 2.

4.3   Experimental design

Same as in experiments 1 and 2.

4.4   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 **Erreur ! Source du renvoi introuvable.**.

(8)   Sentence-by-sentence 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*

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

4.5   Results

As in the other two experiments, trials involving wrong answers to the questions were eliminated as well as those with reading times three standard deviations above or under the mean. Overall, 8.5% (68/800 trials) were thrown out. Regarding the subject analysis, a 2 x 2 repeated measures ANOVA was performed on each set (group x lexical bundlehood). The results are summarized in Table 11.

Table 11
Experiment 3, 1st and 2nd sets: By-subject repeated measures ANOVA.

Set	LBhood	Mean RT	(SD)	Difference	$F(1, 18)$	p
1	LB	4,335	(1,194)			
1	NLB	5,031	(1,631)	696	17.6	< 0.05
2	LB	3,406	(1,080)			
2	NLB	4,097	(1,487)	691	8.5	< 0.05

The results show that sentence containing LBs were read 696 ms faster than sentences that did not contain LBs in the 1st [$F1(1, 18) = 17.6$, $p < 0.05$] and 691 ms faster in the 2nd sets [$F1(1, 18) = 8.5$, $p < 0.05$]. Though the interaction between lexical-bundlehood and group in the 1st set was significant [$F1(1, 18) = 4.6$, $p < 0.05$], it did not reach significance in the 2nd set [$F1(1, 18) = 0.3$, $p > 0.05$] (not shown in Table 11).

Regarding the by-item analysis, a paired *t*-test analysis was performed on the trimmed data (LBs vs. NLBs) for each set. The results of the analyses are shown in Table 12.

Table 12
Experiment 3, 1st and 2nd sets: By-item paired t-test analysis.

Set	LBhood	Mean RT	(SD)	Difference	$t(119)$	p
1	LB	4,332	(987)			
1	NLB	5,122	(1,087)	790	-2.7	< 0.05
2	LB	3,393	(477)			
2	NLB	4,192	(1,192)	799	-2.6	< 0.05

In the 1st set, the 790 ms difference between sentences that contain LBs and those that do not was significant [$t2(19) = -2.7$, $p < 0.05$], as well as the 799 ms difference in the 2nd set [$t2(19) = -2.6$, $p < 0.05$]. In sum, the by-subject and by-item analyses reveal that LBs provide an advantage in on-line sentence processing.

4.6   Discussion

Counter to our earlier prediction, the magnitude of the LB facilitatory effect in experiment 3 was about 3 times bigger (mean difference = 744 ms) than the one found in experiment 2 (mean difference = 238 ms). This is unexpected given that words before and after LBs and NLBs should have been read equally fast given that they are exactly the same. This suggests the existence of some kind of synergy between LBs and natural, full-sentence context.

5   General discussion

These results obtained in experiments 2 and 3 parallel results obtained for other self-paced reading experiments that tried to determine 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 (to appear) also 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.

One possible explanation for the fact that LBs are read faster than NLBs is that LBs — like idioms and other formulaic sequences such as *everything but the kitchen sink* — are stored and processed holistically. This, however, is only one of other possible explanations. Storage, as Harald Baayen (p.c.) would say, might simply be combinatorial knowledge, that is, knowledge of what goes with what. Therefore, it is possible that transitional probabilities — that is, the higher likelihood of occurrence for a lexical item after one or more previous lexical items, or as Gary Libben (p.c.) would say, knowing where one is going given knowledge of where one was — underlies the facilitatory effect. First-order transitional probabilities (i.e., W1 → W2 → W3) were controlled for and this variable did not affect reading times, otherwise NLBs would have been read faster than LBs. Similarly, in their study of idioms Swinney and Cutler (1979) did not find a transitional probability advantage

for idioms over controls. However, it is possible that second-order transitional probabilities (i.e., W1, W2 → W3 or the probability of occurrence of word W3 after word W1 and word W2 have occurred) or even third-order transitional probabilities (i.e., W1, W2, W3 → W4 or the probability of occurrence of word W4 after word W1, word W2, and word W3 have occurred) underlie the LB effect. Yet another possibility is that LBs are stored both with and without internal structure.

That LBs are holistically stored without internal structure finds support from the first-language acquisition literature (e.g., Ellis 1996, 1998; Wray 2002). 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. 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 without internal structure; otherwise she would probably replace the 1st person subject with a 2nd person subject, change the verb from "am" to "are", and replace the 2nd person object with a 1st person object so as to say something like "you're gonna catch me". The point to be made here is that at least some more or less complex linguistic units we have in our mental lexicon, which were acquired during the (very) early stages of our lives, have to have a holistic entry node (if that's how things are stored in our brain) without internal structure. Otherwise, Erin would not have used the sequence "you're gonna catch me" to initiate the game.

Nonetheless, humans seem to be very powerful pattern finders (Bowers, Davis, and Hanley 2005), and in order to find patterns, we must decompose, analyse things. 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 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 in processing time between an LB and a NLB? Instead of saying that LBs are holistically retrieved as opposed to NLBs — in other words, that NLBs are decomposed but not LBs — Libben (2005b: 276) would suggest that the difference is attributable to LBs being processed less than NLBs. 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. In other words, the less predictable a sequence was, the more activation there was in Wernicke's area and its right homologue, whereas more predictable sequences correlated with less activation of these areas. Assuming that predictability is tantamount to frequency of occurrence, this finding relates to LBs in the following way. LBs are highly recurrent (predictable) strings of words, while NLBs are non-recurrent and unpredictable (at least compared to LBs). The on-line processing advantage LBs have over NLBs might reside

in that LBs incur less activation of Wernicke's area than NLBs. Though Wernicke's area and its right homologue are possibly related to the processing of LBs and NLBs alike, it is still unclear in what respect LBs are processed to a lesser degree than NLBs.

Baayen (2003: 266) mentions that "if the brain does indeed make use of probabilities, then it must somehow keep track of (relative) frequency information". In other words, each time a linguistic unit is encountered, the brain (automatically) reinforces it, thus increasing its initial activation (i.e., its frequency). Processing differences might be attributable (at least in part) to this reinforcement process, which might be the building of neural pathways. Assuming that a particular linguistic unit needs to be accessed in order for it to be reinforced, less frequent linguistic units should be harder to reinforce than more frequent ones given that less frequent words are processed slower than more frequent ones (Wingfield 1968; Taft 1968; Reichle et al. 2003). If this has any claim to reality, then, LBs would demand a lesser quantity of blood to flow to relevant parts of the brain in order to reinforce them than NLBs and this difference in blood quantity would correlate with a difference in reading time.

6   Conclusion

The results of the chunk-by-chunk and sentence-by-sentence self-paced reading experiments reported here show that LBs and sentences containing LBs are read faster than NLBs and sentences that do not contain LBs. This suggests that LBs facilitate on-line sentence processing. Note that these results occurred despite the fact that NLB control strings should have been read faster according to measures known to facilitate processing such as token frequency, morphological and phonological complexity, and transitional probabilities. The decisive factor here was lexical-bundlehood, that is, whether a string of words was an LB or not. In the word-by-word experiment, however, there were no significant differences in reading time between LBs and NLBs, similarly to Schmitt and Underwood's (2004) word-by-word self-paced reading experiment. This indicates that word-by-word presentation of LBs disrupts their facilitatory effect.

To answer the question posed in the title of this paper, the LB effect observed in experiments 2 and 3 can be accounted for by holistic storage. Nonetheless, it is still unclear how exactly is the term 'stored' defined. On the one hand, 'stored' could mean that the words making up the LB are individual items that are linked together through combinatorial knowledge (i.e., knowing that they go together), whereas NLBs would not benefit from such knowledge. From this perspective, a stored LB, for instance *in the middle of the*, would look something like the following: [in→the→middle→of→the]. On the other hand, 'stored' could also mean that an LB has no internal structure and would look something like [inthemiddleofthe]. Further research is needed to determine what exactly is storage. Perhaps, the best way to discriminate between holistic and non-holistic storage is through speech production. I have heard a few times people repetitively say "The problem is is that"; it seems that the string "the problem is" is retrieved as a non-analysed, holistic chunk. Otherwise, they would have felt the ungrammaticalness of the double "is" and omitted one of them in subsequent productions. It seems, however, that they did not feel this sequence was problematic given that they kept using it.

One last point needs mentioning. 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 the discourse functions of LBs found in the conversation and academic prose subcorpora of the *British National Corpus*. However, the LBs 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. This suggest that even though LBs might bear more often than not a set of specific discourse functions, there is no inherent association between the two.

Acknowledgments

Appendix . Stimuli

Item	Code	4 or 5 word Seq.	LB = 1; NLB = 2	Practice Trials	Question	Correct Answer
pt1	pt1		2	I had the flu but I still went to class.	Was I sick?	y
pt2	pt2		2	Information can travel across the globe in seconds.	Can information can't travel across the globe?	y
pt3	pt3		2	A year ago, Mr. Jones started to sell electronic games.	Was it Mr. Smith who started selling games?	n
pt4	pt4		2	The man was found in a hotel off route 99.	Did they find the man?	y
pt5	pt5		2	I bought this book in the new bookstore.	Did he buy a knife?	n
pt6	pt6		2	The other day I slept like a log.	Does the sentence mention anything about a rabbit?	n
				Group A 1st Set		
1	1nlb1	4	2	His friend's got one to do next Friday.	Is there a weekday mentioned in the sentence?	y
2	1nlb2	4	2	If workers don't know about it nothing will happen.	Does the sentence mention workers?	y
3	1lb3	5	1	Ron thinks you want me to do another one.	Is there an animal mentioned in the sentence?	n
4	1lb4	4	1	Yeah although you might as well buy one.	Should he borrow one?	n
5	2lb1	4	1	Tell me when you want me to return it.	Does the sentence say anything about returning something?	y
6	2nlb2	4	2	Would he like to stop and have to look inside it?	Does the sentence mention anything about stopping?	y
7	2lb3	4	1	Yes, everything I said to her was sacred.	Was everything he said ordinary?	n
8	2lb4	5	1	I sat in the middle of the bullet train.	Did I sit in an underground train?	n
9	2nlb5	5	2	But unfortunately all the top of it kept burning.	Did it keep burning?	y
10	3nlb1	4	2	He's glad you do want to dig tunnels.	Is he sad?	n
11	3lb10	5	1	Now, must I tell you what I discovered yesterday?	Is the sentence about walking?	n

12	3nlb11	4	2	I might, I do if you seriously care.	Is there a possibility that he will do it?	y
13	3lb2	4	1	But honestly, I don't think he ran away.	Does the sentence mention anything about eating?	n
14	3nlb3	4	2	Yeah, maybe I'll get you what these guys want.	Is the sentence about getting something?	y
15	3nlb4	4	2	He believes you do know what David did.	Does he believe you know?	y
16	3nlb5	4	2	Sam assumes you know where you begin singing.	Is the sentence about donuts?	n
17	3lb6	4	1	Indeed, whatever you think about it feels weird.	Is the thing pleasant?	n
18	3lb7	4	1	I confess I don't know what Smith wants.	Did the person confess?	y
19	3nlb8	4	2	I admit I do know whether Jack cheated.	Does the person know whether Jack cheated?	y
20	3lb9	5	1	I realize I don't know how research is done.	Is the sentence about skiing?	n

Group A 2nd Set

1	1lb1	4	1	His friend's got nothing to do next Friday.	Does his friend have something to do?	n
2	1lb2	4	1	If workers don't worry about it nothing will happen.	If workers don't worry, will something happen?	n
3	1nlb3	5	2	Ron thinks you want it to do another one.	Is there a man named Ron in the sentence?	y
4	1nlb4	4	2	Yeah, although you would as well buy one.	Should he buy one?	y
5	2nlb1	4	2	Tell me when you see me to return it.	Will he return it next week?	n
6	2lb2	4	1	Would he like to stop and have a look inside it?	If he stops, will he buy something?	n
7	2nlb3	4	2	Yes, everything I was to her was sacred.	Was he sacred to her?	y
8	2nlb4	5	2	I sat in the front of the bullet train.	Was sitting in the train?	y
9	2lb5	5	1	But unfortunately all the rest of it kept burning.	Did it stop burning?	n
10	3lb1	4	1	He's glad you don't want to dig tunnels.	He doesn't want to dig tunnels, right?	y
11	3nlb10	5	2	Now, must I get you what I discoverd yesterday?	Is the sentence about a discovery?	y
12	3lb11	4	1	I might, I mean if you seriously care.	Does the sentence mention a country?	n
13	3nlb2	4	2	But honestly, I do think he ran away.	Does he think the man ran away?	y
14	3lb3	4	1	Yeah, maybe I'll tell you what these guys want.	Does the sentence mention ducks?	n

15	3lb4	4	1	He believes you don't know what David did.	Is the man called Patrick?	n
16	3lb5	4	1	Sam assumes you know when you begin singing.	Does the sentence mention something about Sam?	y
17	3nlb6	4	2	Indeed, whatever you do about it feels weird.	Is the sentence about feelings?	y
18	3nlb7	4	2	I confess I do know what Smith wants.	Does the sentence mention a brand name?	n
19	3lb8	4	1	I admit I don't know whether Jack cheated.	Does he know if he cheated?	n
20	3nlb9	5	2	I realize I do know how research is done.	Does he know how it is done?	y

References

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.

Baayen, R. H. (2003). Probabilistic approaches to morphology. In R. Bod, J. Hay, and S. Jannedy (eds.). *Probabilistic Linguistics*, pp. 229-88. Cambridge, MA: The MIT Press.

Biber, D. (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, D. and Conrad, S. (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, D., Conrad, S., and Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, and T. McEnery (eds.). *Corpus Linguistics by the Lune* (pp. 71-93). Frankfurt/Main: Peter Lang.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Grammar of Spoken and Written English*. Harlow: Longman.

Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., and Gregory S. B. (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *The Journal of Neuroscience*, 20, 1975-81.

Bloomfield, L. (1933). *Language*. New York: Henry Holt and Company.

Bod, R. (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, J. S., Davis, C. J., and Hanley, D. 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, K. and Schmitt, N. (to appear). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*.

Croft, W. (2001). *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48, 631-64.

Ellis, N. C. (1996). Sequencing in SLA. Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91-126.

Erman, B. and Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29-62.

Fillmore, C. (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, R. W., Bogdanovich, J. M., Sykes, J. R., and Barr, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, 141-154.

Hughlings Jackson, J. (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, S. and Francis, G. (2000). *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam and Philadelphia: John Benjamins.

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

Jerpersen, O. (1924). *The Philosophy of Grammar*. London: George Allen and Unwin.

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

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

Ortony, A., Schallert, D. L., Reynolds, R. E., and Antos, S. J. (1978). Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal Language and Verbal Behaviour*, 17, 465-477.

Pawley, A. and Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (eds.). *Language and communication* (pp. 191-226). New York: Longman.

Reichle, E. D., Rayner, K., and Pollatsek, A. (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, F. (1916/1966). *Course in General Linguistics*. New York: McGraw-Hill.

Schmitt, N. (2004). *Formulaic Sequences*. Amsterdam and Philadelphia: John Benjamins.

Schmitt, N., Grandage, S., and Adolphs, S. (2004). Are corpus-derived recurrent clusters psychologically valid? In N. Schmitt (ed.). *Formulaic Sequences* (pp. 127-151). Amsterdam and Philadelphia: John Benjamins.

Schmitt, N. and Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (ed.). *Formulaic Sequences*, pp. 171-89. Amsterdam and Philadelphia: John Benjamins.

Schweigert, W. A. and Moates, D. R. (1988). Familiar idiom comprehension. *Journal of Psycholinguistic Research*, 17, 281-296.

Swinney, D. A. and Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, 18, 522-534.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 263-272.

Titone, D. A. and Connine, C. C. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, 31, 1655, 1674.

Underwood, G., Schmitt, N. and Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.). *Formulaic Sequences*, pp. 153-172. Amsterdam and Philadelphia: John Benjamins.

Wingfield, A. (1968). Effects of frequency on identification and naming objects. *American Journal of Psychology*, 81, 226-234.

Wray, A. (2002). *Formulaic Sequences and the Lexicon*. Cambridge: Cambridge University Press.