**Is what you put in what you get out? – Textbook-derived lexical bundle processing in beginner English learners**

JULIAN NORTHBROOK and KATHY CONKLIN
University of Nottingham

**Abstract**

Usage-based approaches to second language acquisition put a premium on the linguistic input that learners receive and predict that any sequences of words that learners encounter frequently will experience a processing advantage. The current study explores the processing of high-frequency sequences of words known as ‘lexical bundles' in beginner learners of English in Japanese secondary school. To do this we use a phrasal judgment task, testing students on items taken directly from their teaching materials. Students responded to lexical bundles that occurred in their textbooks significantly faster and more accurately compared with non-lexical bundles, and were sensitive to the frequency of occurrence in their textbooks. This study shows, perhaps for the first time, that even very low level, beginner secondary students are sensitive to the frequency of lexical bundles which appear in the input they receive from teaching materials, a finding that has important implications for teaching and material design.

Keywords: lexical bundles, junior high school textbooks, language processing, usage-based theories, Japanese English education

**Introduction**

It seems obvious that if we try to build something with poor materials rather than the best ones that the end product will be of lesser quality. This sentiment is nicely encapsulated in Computer Science by the saying *garbage in, garbage out*, and refers to the fact that the input computers receive wholly determines the quality of their output. However, is this true in second language learning? More precisely what is the relationship between learners’ input and their subsequent processing? The current research will address this question by examining the processing of ‘lexical bundles’ by Japanese secondary school students learning English – an as of yet largely unexplored age and proficiency group in formulaic language research.

Loosely defined, lexical bundles are sequences of words that occur frequently in language. They are generally corpus derived and may constitute a complete utterance (*how’s it going*) but often do not (*a lot of the*). They fall under the broad heading of formulaic language, which is referred to by many different names in the literature (see Wray 2002, p. 8 for a list of more than 50 such terms). In applied linguistics, there is a growing body of research that suggests formulaic language more broadly, and lexical bundles specifically, are processed more quickly than novel or less frequent sequences of words (Conklin & Schmitt 2012). Importantly, formulaic language is thought to contribute to fluent and accurate language use (Pawley & Syder 1983) that facilitates real-time communication for both speaker and listener (Boers, Eyckmans, Kappel, Stengers, & Demecheleer 2006; Kuiper 2004; Pawley & Syder 1983) and reduces demands on working memory (Conklin & Schmitt 2008; Kuiper & Haggo 1984). However, important questions remain about how formulaic language becomes ‘formulaic’, thereby facilitating processing, as well as how early formulaicity starts to emerge in second language speakers. One view is that language learning, representation and processing are all reliant on exposure, which is the position taken by ‘usage-based’ approaches.

Usage-based approaches put a premium on the linguistic input that learners receive. More specific to lexical bundles, a usage-based approach would say that any sequence of words that learners encounter frequently will, all else being equal, experience a processing advantage. Thus, if junior high school students repeatedly encounter a phrase like *how many CDs do you have* in their English textbooks, this sequence will experience a processing advantage, despite the fact that junior high school students today are unlikely to play music using CDs. In a usage-based approach to first and second language acquisition, experience of and exposure to language results in high frequency, repetitive sequences of words being stored in long-term memory (e.g. Barlow & Kemmer 2000; Bybee 1998, 2006; 2013; Ellis 2002, 2003; Tyler 2010). Evidence for this from the field of formulaic language in general will be discussed in detail below.

Northbrook and Conklin (in press) looked at the lexical bundles that Japanese secondary school students encounter in their English textbooks and compared the most frequent 50 lexical bundles at different bundle sizes. We would expect items that are frequent in the textbooks to be frequent in native language, and items that are low frequency in the textbooks to be low frequency in native language also. The authors found that this was the case for three-word lexical bundles, but larger sizes bore little to no resemblance to lexical bundles in the native corpus. Importantly, a usage-based view of language learning would predict that students will extract and store in memory the lexical bundles that they encounter frequently in their textbooks (as well as any frequently occurring formulaic language that they encounter in- and outside of the classroom) and will therefore demonstrate a processing advantage for textbook lexical bundles, regardless of whether they are authentic or not. Thus, if we take frequency of occurrence as a measure of experience, and students are indeed sensitive to the frequency of occurrence of lexical bundles in their textbooks, we would expect them to respond to lexical bundles derived from their textbooks faster and more accurately than novel sequences of words which do not appear in their books, irrespective of whether they represent ‘authentic’ native-speaker lexical bundles. Before turning to the current study, we will take a closer look at how non-idiomatic, corpus derived formulaic sequences such as lexical bundles, collocations and binomials are thought to be acquired and processed.

**Formulaic sequences in non-native speakers**

In native speakers, a processing advantage for formulaic language is well attested (see Conklin & Schmitt 2012; Wray 2012). However, whether this extends to second language speakers (L2) remains a somewhat open question, in particular for non-idiomatic formulaic language such as lexical bundles. As will be taken up below, several studies have tried addressing this question, but they have not unequivocally answered it. Further, to our knowledge, no studies have attempted to ascertain whether or not very low proficiency teenage learners have a processing advantage for formulaic language.

 Three self-paced reading experiments explored whether lexical bundles would be read faster compared to non-lexical bundles by native speakers (Tremblay, Derwing, & Libben 2009; Tremblay, Derwing, Libben, & Westbury 2011) and non-native speakers (Babaei, Taleb Najafabadi, & Fotovatnia 2015). All of them conducted multiple versions of a self-paced reading task, testing word-by-word reading, chunk-by-chunk reading and full sentence reading. Tremblay et al. (2011) found advantages for sentences containing lexical bundles in all three tasks with native speakers. Tremblay et al. (2009), on the other hand, only found a lexical bundle advantage in the chunk and sentence reading tasks with native speakers, while Babaei et al. (2015) found no advantage for lexical bundles compared with non-lexical bundles in any tasks for moderately proficient non-native speakers in an EFL setting.

 Using eye-tracking, Valsecchi, Künstler, Saage, White, Mukherjee and Gegenfurtner (2013) investigated the reading of lexical bundles by native and lower-proficiency non-native speakers. They found that while native speakers had a processing advantage for lexical bundles compared with non-lexical bundles, non-native speakers did not. The authors concluded that this null result was due to lack of proficiency and that a processing advantage would emerge later with higher proficiency. Importantly, in the only work on low-proficiency EFL students, Isobe (2011) conducted a phrasal judgment task where participants were instructed to judge whether the word order of a lexical bundle was correct or not. The author found lexical bundles were responded to significantly faster and with greater accuracy. Further, in a second experiment, involving a timed familiarity judgement task, the author found familiarity and phrasal frequency to be important in predicting response times for lexical bundles. Similarly, Jiang and Nekrasova (2007) conducted a phrasal judgement task and found that both native and non-native speakers judged corpus-derived units faster and more accurately than novel control sequences. Ellis, Simpson-Vlach, and Maynard (2008) similarly found that non-native speakers were sensitive to the frequency of corpus-derived sequences. Both of these studies were conducted with high-proficiency students in an ESL setting at American universities.

Although the literature shows a somewhat mixed picture, the majority of evidence suggests that lexical bundles do proffer processing benefits to non-native speakers, particularly at the higher levels of proficiency. Importantly, if exposure to lexical bundles underpins the processing advantage for them, we need to have a good measure of the input learners receive. Generally, studies examining the processing of lexical bundles identify them in native speaker corpora (e.g. Isobe 2011; Babaei et al. 2015, and Tremblay et al. 2009, 2011 all used items from Biber, Johansson, Leech, Conrad, & Finegan 1999’s study of the BNC corpus, and Jiang & Nekrasova 2007 used items from the Bank of English). While this may be a valid method of identifying lexical bundles for high proficiency speakers, especially those living in L1 environments, it is likely that lower proficiency speakers in EFL contexts where input is mainly limited to classroom study will have a different experience of the L2. Put another way, we may find a processing advantage for lexical bundles in lower proficiency speakers if we test them on lexical bundles they encounter in their primary source of input – their textbooks. One goal of this study is to test this assumption by matching experimental stimuli to the population under study using items derived directly from their materials (discussed briefly below, but see Northbrook & Conklin, in press, for detailed discussion of these materials, and how they differ from native speaker language).

**How does a processing advantage for lexical bundles emerge?**

Optimisation of language processing can be thought of in terms of a computer's internet browsing cache – by storing information pertaining to frequently accessed webpages, we browse the internet faster and more efficiently without having to load large data files such as images every time they need to be accessed. In the same way, formulaic language is said to reduce processing load by ‘caching’ language into chunks that can be accessed quickly and easily without needing to retrieve and compute their individual components (Goldberg 2006; Wray 2002, 2008; Wray & Perkins 2000). In a usage-based approach, when language users encounter utterances, they store them and look for patterns amongst them (Tomasello 2003). More specifically, if we encounter the examples of *in front of the* in 1., 2. and 3. below, taken from the SUBTLEXus (Brysbaert & New 2009), each instance will be stored and the linguistic pattern will be detected on subsequent encounters. Importantly, the first time a sequence of words is encountered, it will be processed in a word-by-word manner, however having done so it leaves an imprint in memory that is strengthened by subsequent exposure (Logan 1988). Thus, when reading 1. there would be no difference between *in front of the* and any other sequence of words in the same sentence when reading it for the first time. It is a novel, original string of words. However, when reading 2. and subsequently 3. there should be a ‘boost’ in processing speed because we already have an imprint of this ‘cached’ in memory.

1. Tomorrow you're going in front of the Commissioner.
2. I didn't want to bring this up in front of the kids...
3. Stand up in front of the windows.

This is important because our online processing resources are quite limited, while our memory capacity is vast (Crick 1979). A seminal finding by Miller (1956) demonstrated that our working memory can hold seven plus or minus two items, which is often referred to as the ‘magical number seven’ or *Miller’s Law*. Subsequent research has disputed how much information can be held in working memory (e.g. Cowan 2001; Feldman 2012) and while there is disagreement on the precise amount of information that can be held in working memory, there is complete agreement that there are severe constraints on it. This has serious implications when we consider a task like reading sentences. To comprehend a sentence, readers need to hold the individual words in working memory and integrate the meaning of the sentence with their developing understanding of the text. If the median sentence length in a text is more than 15 words, for instance (cf. Kornai 2008 discussing journalistic prose), readers will very quickly exceed the capacity of their working memory. In contrast to working memory, long-term memory is a relatively abundant resource. Thus, to compensate for a lack of working memory, the brain may make use of long-term memory by storing frequently used strings of words.

As we have seen, usage-based models of language learning predict that repeated occurrences of linguistic patterns, like lexical bundles, lead to their entrenchment in memory. A detailed description of the differences between the various usage-based approaches is beyond the scope of this paper, however as Higham (1997) points out, a common factor among all of them is that chunking performance can be predicted by a metric derived from the frequency at which features occur within a given stimulus. Some models (e.g. Goldberg 2006; Wray 2002) predict that only items that are frequent enough are stored in memory, while others (e.g. Bybee 1998, 2006; Bybee and Hopper 2001; Mclaughlin, Rossman and McLeod 1983) predict that every exposure to a phrase influences subsequent entrenchment and processing (see Arnon & Snider 2010). The frequency with which units of language occur, then, is a driving force behind chunking processes and, all else being equal,1 each exposure to a sequence of words or sounds will affect its subsequent processing (Bod, Hay and Jannedy 2003; Bybee 1998, 2006; Bybee & Hopper 2001; Mclaughlin et al. 1983). This has important implications for language learning material design—something which, in the case of EFL contexts, often constitutes students’ primary experience of the language—and has an important consequence: *if we expose students to language that is not sufficiently representative of the target language, it is this non-nativelike language that will become entrenched in memory.* Thus, we need to ask whether the adage of ‘garbage in, garbage out’ is as relevant to our classrooms as it is to computer programming.

In the case of Japanese junior high school students, the teaching materials used in class, in other words what students are exposed to, will (primarily) dictate what they gain a processing advantage for. A good example that illustrates the problem with the language of the textbooks is *do you hear* and *do you play*. In the SUBTLEXus corpus *do you hear* occurs 1176 times (23 times per million) and in the Japanese junior high school English textbooks it does not appear at all. In contrast, *do you play* occurs 136 times (2.7 times per million) in the SUBTLEXus and 56 times (395 times per million) in the junior high school English textbooks (based on data from the JHSETC – see below). Thus, in native speaker English *do you hear* is much more frequent and should elicit a processing advantage relative to *do you play*, at least for speakers who encounter nativelike input. For Japanese students learning English, the other pattern is more frequent and should therefore experience the processing advantage. According to a usage-based model, this should lead to a non-nativelike processing advantage in Japanese junior high school students.

Because the input has serious implications for processing, exploring the input that students receive is an important endeavour. Previous research has analysed the language contained in English textbooks for Japanese lower secondary school students (Northbrook & Conklin, in press). These textbooks were intended to be ‘communicative’, and as such should contain examples of authentic language – a hallmark of the communicative approach. The authors created a corpus of the 16 textbooks for junior high school English approved by the Japanese Ministry of Education called the ‘Junior High School English Textbook Corpus, or the ‘JHSETC’ (Northbrook & Conklin, in press). The corpus consisted of 152,966 words, and contained a total of 141,608 lexical bundle tokens compared with 80,707 tokens in a matched reference corpus, indicating that the language contained in the books is indeed highly repetitive, and should provide optimal conditions for the acquisition of ‘chunked’ language. However, the lexical bundles in the textbooks, although abundant, were found to be qualitatively different (Northbrook & Conklin, in press) to and follow very different distributional patterns to the reference corpus – in other words, to authentic native-speaker language. As with similar studies (e.g. Nguyen & Ishitobi 2012; Ogura 2008), this research concluded that the language contained in textbooks does not serve the purpose of exposing students to authentic language. In the current study, we will explore whether the frequent lexical bundles in the textbook corpus (e.g. *do you play*) elicit a processing advantage relative to possible sequences of words that do not appear in the textbooks (e.g. *do you hear*).

## Methodology

*Experimental stimuli*

Twenty three-word lexical bundles covering a range of frequencies were selected from the “New Horizon” (New Horizon English Course 2013) sub-corpus of the JHSETC, consisting of three textbooks, which we refer to as the New Horizon Corpus (NHC). The whole corpus draws on all of the Ministry approved textbooks for secondary schools in Japan. However, a school will only use one textbook series. Therefore it would be inaccurate to base stimuli on the whole corpus. As the students in the current study used the “New Horizon” series, the lexical bundles and associated frequency information comes from these books. The language presented in the textbooks is, for the most part, presented as oral language in the form of dialogues and monologues. Thus, the SUBTLEXus was used as a reference corpus, which is a corpus of subtitles taken from US films and television series, and as such usually involves people in social interactions (Brysbaert & New 2009). Further, while the language of the SUBTLEXus is scripted, it is meant to represent natural language in the same way that the language in textbooks is intended to. The US version of the SUBTLEX was chosen because the language used in the Japanese textbooks is based on standard American English.

In order to retrieve the three-word bundles, Antconc’s (Anthony 2014) "Clusters" function was used with a cut off frequency of four times in the NHC. Items were then selected from this list based on several criteria. First, items were chosen from a range of frequencies of occurrence. This was done to ensure the results were representative of the fact that lexical bundles were more or less frequent in the textbooks. Second, because we were interested in the role *teaching materials* play, and not the effects of teaching, we selected lexical bundles that were not explicitly taught in the textbooks. This was ascertained by consulting the word lists in the books, and rejecting any lexical bundles which appeared. The students’ (participants’) teachers were also asked about how they approached teaching from the textbook, and none of them reported teaching any phrases that were not listed as items for explicit instruction. A third, and important criterion, was that it had to be possible to create a control sequence by changing only one word in each lexical bundle, using only words that appeared in the textbooks to create a three-word lexical bundle of zero frequency in the NHC. That is, each individual word in both the lexical bundle and control item appeared in the NHC, but only the original lexical bundle appeared as a string of words. This was done to ensure any differences in reading times found between lexical bundles and control phrases are due to whole string processing and not simply because a student has not been exposed to a given word. For example, *do you play* was selected as a lexical bundle, and had16 occurrences in the NHC and 136 in the SUBTLEXus*.* We changed the word *play* to *hear* to create the control item *do you hear*, which had zerooccurrences in the NHC but 1183 in the SUBTLEXus. Crucially, the word *hear* occurred in the NHC. All controls were matched as closely as possible for individual word frequency with their respective bundles. This was calculated as the mean frequency of all three words in the lexical bundle and was confirmed with a *t*-test. There was no significant difference between the word frequency of main items and control items (*t*(41.98) = 0.53, *p* = 0.59), the main items and ungrammatical items (*t*(41.98) = -0.52, *p* = 0.60) or the controls and ungrammatical items (*t*(42) = -1.04, *p* = 0.31). Any differences in response time during the experiment, then, can be assumed to be a result of whole phrase frequency, not of the individual words contained in each lexical bundle.

We selected 20 items that fit these three criteria, and created 20 matched control phrases for them. Because the main task involved a “yes”/“no” decision about whether a phrase was possible, we also created another matched set of items by scrambling the words to produce a set of ungrammatical phrases (e.g. *you with do)* which had zero frequency in both the textbook corpus and the reference corpus. This resulted in a total of 60 experimental items. Finally, 190 filler phrases were added to mask any similarities between the NHC derived lexical bundles and control phrases. We selected three-word lexical bundles from the NHC, again from a range of frequency bands, and scrambled the word order of half of them to make 95 grammatical fillers (e.g. *play the piano* – four occurrences in the NHC), and 95 ungrammatical fillers (e.g. *he yes does –* zero occurrences in the NHC). The ungrammatical items were checked in both the NHC and the SUBTLEXus to ensure that they had a frequency score of zero. In total, the experiment consisted of 250 items.

#### Participants

Participants for this study consisted of 35 grade 9 students from a private secondary school in Japan. A summary of the participants is shown in Table 1. All students spoke Japanese as their first language. No students reported reading, hearing or speaking impairments. Students’ English proficiency was assessed by asking them to rate themselves on a seven-point scale across four categories – speaking ability, reading ability, listening ability and ability to use grammar accurately. We took the average of these four ratings to create a ‘proficiency’ metric. There were some differences in the age at which students were first exposed to English. Eleven students (31%) reported that they currently studied English outside of school at ‘juku’, or ‘cram school’. We conducted a *t*-test to assess if there was a difference in proficiency between the students who did and did not study English outside of school – those who did scored significantly higher for proficiency *t*(797.78) = 7.45, *p* < .001 and therefore this variable should largely be captured by proficiency. We also asked students to report on use of English outside the classroom. Twenty-four students (69%) reported nothing, seven students reported watching films in English and two students had been abroad. This was also significantly related to proficiency *t*(3070.9) = 22.81, *p* < .001 and therefore should fundamentally be captured by this variable in the analysis.

####

**[Table 1 NEAR HERE]**

#### Procedure

Students were tested at their school. They completed a phrasal judgement task first, followed by a language background survey which included questions related to the students’ use English outside of school, as well as a familiarity rating task for all of the lexical bundles. Ungrammatical items were assumed to have a familiarity of zero. All instructions were given in Japanese, and the experiment was run using Inquisit (2014). Two hundred and fifty three-word sequences were displayed for a simple “yes” or “no” judgement task. Students were instructed to press a “yes” key if the phrase was possible in English, and a “no” key if it was not. For each trial a “\*” appeared in the centre of the screen for 700 milliseconds, followed by the target lexical bundle. Students were instructed to respond as quickly as possible. They were given 20 practise items with feedback telling them whether they were correct or not.

#### Analysis and Results

The data was analyzed using mixed effects modelling, with the package lme4 (Bates, Maechler, Bolker, & Walker 2015). Response time was modelled as the response variable, with textbook frequency, mean word frequency, familiarity, proficiency, phrase length (the number of characters in a given lexical bundle), trial number and age of first contact as fixed effects. Participants and items (the lexical bundles) were included as random-effect factors in the model. Reaction times under 400 milliseconds or over 1999 milliseconds were removed, which accounted for less than 0.7% of the data. Inaccurate responses were removed for the response time analysis (described below). Because there were a large number of errors (23% of the data), we also modelled the data using accuracy as the response variable and used the same fixed and random effects. Table 2 summarizes mean response times and accuracy scores.

**[Table 2 NEAR HERE]**

Some predictors were correlated and so were residualised by fitting linear models and creating a new metric consisting of the residual (i.e. information not accounted for by correlating factors, Baayen 2008). For example, textbook frequency was correlated with phrase length and familiarly, and so textbook frequency was used as the response variable in a model with phrase length and familiarity as predictors. Similar models were made for any correlated factors, including previously residualised response variables. The resulting residuals all correlated significantly with their related variables (*p* < .0001): textbook frequency (*r* = .89) mean word frequency (*r* = .80), proficiency (*r* = .92) and familiarity (*r* = .99). All other continuous predictors were centred to avoid having a change in slope that might correlate with a change in intercept (see Baayen 2008). The accuracy model also included response time as a predictor, which we will label ‘response speed’ to avoid confusion with the same metric as used in the response time model. It was not correlated with any other predictors.

Response time was modelled using a Liner Mixed Effects model. Because accuracy is a binary variable, accuracy was modelled using a Generalized Linear Mixed Effects model. For both response time and accuracy, a simple model was fit by residual phrase frequency, and with subject and lexical bundle as random effects. Complexity was then built into this model by adding new predictors in a step-by-step manner. Each new predictor was tested as a discrete factor, and for interactions with all other predictors. These were also tested as adjustments for both the subject and lexical bundle random effects (as recommended by Barr, Levy, Scheepers, & Tily 2013). Each iteration of the model was tested against the previous best fit to see if the change was significant (likelihood ratio test, *p* ≤ .05). This process was continued until all predictors and interactions had been tested both as fixed and random effects. Table 3 summarises the best fit for both of the models.

**[Table 3 NEAR HERE]**

Response time. The best fit for the model of response times included residual textbook frequency, residual proficiency, phrase type, phrase length and trial number. Predictors and interactions were tested in the random effects structure. The model fit was significantly improved by adjusting subject by residual textbook frequency (*X*2 (2) = 10.0, *p* = .007). As predicted, the most prominent effect is phrase type. A post-hoc analysis of the differences of LSMEANS shows that the difference between all phrase types is also significant (lexical bundle vs. control phrase, *t*(61.4) = 2.98, *p* = 0.004; lexical bundle vs. ungrammatical phrase, *t*(55.4) = -7.49, *p* = <.0001; control phrases vs. ungrammatical, *t*(61.9) = -4.03, *p* = <.0001).

Residual proficiency and phrase length were significant factors in the model, such that more proficient participants responded more quickly than less proficient ones and shorter items were responded to more quickly than longer ones. Trial number was also a significant factor meaning that participants got faster over the course of the experiment which is a normal phenomenon (Baayen, Davidson, & Bates 2008). Importantly, there was also a significant effect of residual textbook frequency (*X*2 (1) = 5.65, *p* = .018), which indicates that more frequent sequences were responded to more quickly. There was no significant effect of age of first contact (*X*2 (1) = 0.06, *p* = .80) or residual familiarity (*X*2 (1) = 1.92, *p* = .17). There was also no significant effect of mean word frequency (*X*2 (1) = 0.13, *p* = .72), indicating that students are making use of phrasal frequency but not the frequency of individual words. This is not surprising, as items were controlled for this.

Accuracy. The best fit for the Accuracy response variable included residual textbook frequency, residual proficiency, residual familiarity and response speed. There was also a significant interaction between phrase type in the ungrammatical control condition and response speed. For lexical bundles and control phrases, accuracy decreases as speed of response becomes slower. This is interesting in itself, as we would expect a speed-accuracy trade-off whereby responding faster leads to making more mistakes. For ungrammatical items, however, we see a slight increase in accuracy as response time increases. Although there is no interaction between lexical bundle frequency and response time (*X*2 (1) = 0.28, *p* = .59) or familiarity (*X*2 (1) = 3.00, *p* = .08), given the fact that frequency and familiarity also predict accuracy in the model it seems likely that the lexical bundles responded to faster are the higher frequency and more familiar items. This makes sense as we would not expect any facilitation for frequency or familiarity with ungrammatical items. It is possible that for ungrammatical items, ones that are responded to quickly are more prone to mistakes. This would also account for the slower response time for these items. All other predictors were insignificant – phrase length (*X*2 (1) = 0.11, *p* = .74), residual mean word frequency (*X*2 (1) = 0.03, *p* = .86), trial number (*X*2 (1) = 1.31, *p* = .25), age of first contact (*X*2 (1) = 1.66, *p* = .19) or phrase type (*X*2 (2) = 4.86, *p* = .08), although the later does approach significance. The model was significantly improved by adjusting subject by phrase type (*X*2 (5) = 70.93, *p* = < .0001), suggesting that students did not respond to the different phrase types in the same way.

### Discussion

This study shows, for the first time, that even very low proficiency, beginner secondary students are sensitive to recurring lexical bundles from their input – the textbooks. Importantly, they demonstrate an effect of frequency such that lexical bundles that appear more frequently in the books are responded to more quickly and accurately than less frequent ones. This indicates that beginners are sensitive to the frequency of lexical patterns in their input, which highlights the importance of presenting learners with authentic formulaic language in their textbooks.

As predicted there was a significant difference in response time between lexical bundles and matched control phrases. The controls were possible using only words from the textbooks but did not appear as a whole string in them. Because the students had been exposed to all the words in the lexical bundles and control items to the same degree, if they were processing lexical bundles in a word-by-word fashion we would not expect any difference between the two conditions. Further, including individual word frequency did not improve the fit of the model for either response time or accuracy. This finding is in line with similar studies (Ellis et al. 2008; Isobe 2011; Jiang & Nekrasova 2007) showing that L2 learners responded to lexical bundles faster in a phrasal judgement task, and supports the hypothesis that *even* low proficiency non-native learners are sensitive to the exact sequences of words that they are exposed to – as predicted by a usage-based approach.

Interestingly, phrase type does not significantly predict accuracy. However, we do see a significant interaction between item type and response speed whereby the relationship between speed of response is different for ungrammatical control phrases and that of lexical bundles and grammatical control phrases. In terms of the response time analysis, the students are faster at responding to sequences from their textbooks than what to them are novel sequences, and finally ungrammatical items, in that order. For accuracy, however, correct responses are increased for lexical bundles and control phrases that are responded to quickly, and decreases as speed of response decreases. For ungrammatical items, we see the opposite: accuracy increases as response speed decreases. One explanation may be found in the concepts of ‘nativelike fluency’ and ‘nativelike selection’ as put forward by Pawley and Syder (1983), where the difference between a likely utterance and a possible utterance is one of formulaicity. In this context, fluency can be thought of as an unconscious process, whereas nativelike selection, or more simply whether a phrase sounds natural or not to the test taker, can be thought of as a conscious, explicit decision-making process that involves collocational knowledge. It may be that students are sensitive to these aspects unequally, and if a given item is not immediately recognised (nativelike fluency) a participants’ first reaction may be to reject it – however where conscious processing takes over, participants then judge the item based on form and collocational knowledge, or to put it another way, whether the sequence sounds likely to them (i.e. nativelike selection). Where conscious processing does not take over in time, we are likely to see many errors, which would account for the relationship between speed of response and accuracy. This may also go some way towards explaining the fact that familiarity is only significant in the model for accuracy – it may be that familiarity, at least in this context, is related to conscious decision making processes, such as judging the likelihood of a phrase being grammatical or not based on collocational knowledge, or put simply, whether it sounds natural or not.

Students were sensitive to the frequency of lexical bundles, which was reflected in both response time and accuracy. Put simply, more frequent items were responded to faster and more accurately. This is important because it indicates that not only are students sensitive to whether or not items appeared in their books, but to the degree that they are exposed to a given item. In the current study, experimental items were selected across the frequency continuum and results suggest a gradual decrease in processing speed in accordance with the increased frequency of a given item. This finding is noteworthy because it provides strong evidence that the input given to students matters. Studies have shown explicit instruction is effective for formulaic sequence learning (e.g., Laufer & Girsai 2008; Webb & Kagimoto 2009), however the items studied here were neither explicitly taught, nor necessarily ‘highly’ frequent. Webb & Kagimoto (2009) as well as others, for example, Szudarki & Carter (2014) and Pellicer-Sánchez (2017), found evidence for collocational knowledge in language learners only at the level of recognition and recall. Further, in the case of the studies by Webb & Kagimoto (2009) and Szudarki & Carter (2014), learning gains resulted from focused, explicit instruction. Pellicer-Sánchez (2017) found evidence of incidental learning similar to our findings, however, no significant differences were found between four and eight exposures, whereas we found clear differences across the frequency continuum. One explanation for this may be that exposure in our study is over a three year period, whereas in Pellicer-Sánchez's study all repetitions occurred in a short space of time (during the experiment); we know from memory research that memory effects become more robust over time (Ebbinghaus 1885). Here, in the current study, we see evidence of nativelike processing for lexical bundles learned incidentally on a continuous scale – lower frequency items are responded to slower and with more errors, while higher frequency items are responded to increasingly faster and with fewer mistakes. This means that the actual input—even if it is only intended to provide supporting context in demonstrating a particular word, item or grammatical construction selected for instruction—is attributing to a student's fluency in the language.

Importantly, proficiency is also a significant predictor for both response time and accuracy, yet appears to be unrelated to frequency. Siyanova-Chanturia, Conklin, and van Heuven (2011) found that sensitivity to frequency of formulaic language, in this case for binomials, only emerged at higher proficiency levels. Such a finding could indicate that a processing advantage only emerges after a learner encounters formulaic language a certain number of times – an argument also put forward by Babaei et al. (2015) and Valsecchi et al. (2013) to explain the null-result in their studies. However, it may be that these studies would have found a processing advantage with a different set of experimental stimuli – i.e. items matched to the input those participants had encountered. The current study addresses this concern by using items taken directly from the students’ textbooks, and therefore we know they have encountered them. By taking frequency information from the textbook, we should have a fairly accurate estimate as to how often the students have encountered the lexical bundles in the study. A problem with Siyanova-Chanturia et al. (2011), Babaei et al. (2015), Valsecchi et al. (2013), and indeed all studies to date with non-native speakers, is that we cannot be sure of what degree participants had encountered all of the items tested, if at all. For high-level non-natives who have spent considerable time in an ESL environment, native speaker norms may be justified – but can we say the same for lower-level non-native speakers? Evidence that frequency information is related to an individual learner’s exposure to the language may come from the fact that adjusting the subject random effect by lexical bundle frequency significantly improves the fit of the response time model. This provides an indication that each individual’s relative exposure underlies the frequency advantage that he/she demonstrates.

This study also highlights a further important consideration, this time not in terms of methodology, but in terms how English education is being approached in Japan. There is a commonly held belief that students do not spend enough time learning English, which underpins their eventual attainment that is often considered deficient because it is not sufficiently ‘nativelike’. This has led to a steady increase in time allocated to English first by including English in the primary school curriculum and then by increasing the number of hours in English at secondary school from 105 hours to 140 hours (announced by Ministry of Education 2008a; 2008b and implemented from 2011). However, increasing exposure is only one part of the equation. The input that students receive must also be considered. Importantly, the current study demonstrates that beginning learners in an ELF context show a very nativelike formulaic benefit to lexical bundles. Thus, they are showing the onset of a phenomenon that is evident in native speakers and advanced non-natives. Then the key question is what are they showing a formulaic advantage to – is this actually authentic language? If they are simply showing an advantage to genre-specific textbook language that is not reflective of actual native speaker English, this is hugely problematic and points to an important role for textbook developers. This is especially true since, as was demonstrated in Northbrook and Conklin (in press), we do not necessarily need to make large changes to materials to make them more authentic.

**Conclusion**

This study is important because it demonstrates a clear processing advantage for the lexical bundles that Japanese junior high school students encounter in their textbooks. The findings show that students are sensitive to the phrasal frequency of items embedded in their textbooks, and respond to them faster and more accurately. Our results are indicative of a usage-based approach to language learning, in which exposure to language underpins entrenchment in memory and processing speed, *and* is evident in early EFL learners, a population that is underrepresented in the literature. Unfortunately, it also demonstrates the computer adage – *what we put in is what we get out*. The junior high school students receive input—the lexical bundles in their textbooks—and they show a clear processing advantage for them; it is simply that these lexical bundles are not representative of native speaker formulaic language. Crucially, these findings highlight the need to carefully consider the input students are given.

**Acknowledgements**

We would like to express our gratitude to Masao and the other teachers at the school where this research was conducted for their gracious assistance. I would also like to thank David Allen, as well as the anonymous reviewers for their valuable suggestions and insights which contributed a great deal towards improving an earlier draft of the paper.

**Notes**

1 It should be noted that while frequency of exposure is a key metric in usage-based approaches, it is not the only property of the input that contributes to learning. While each exposure to a stimuli can be expected to affect subsequent processing, this may occur to a greater or lesser degree depending on other factors such as recency, Zipfian distribution and saliency (see for example Ellis (2006, 2009, 2012) for detailed discussions on this, as well as Ebbinghaus (1885) and more recently Pavlik & Anderson (2008) for a more general discussions of recency effects in memory and learning). Although we do not wish to overly simplify what we know about usage-based language development, when textbooks are the ‘source' of language exposure, it would be very difficult, if not impossible, to disentangle these factors. Therefore, in the current study we take properties such as recency, distribution and, although we do also attempt to control for it by only including non-taught lexical bundles, saliency, to be captured by frequency of occurrence and familiarity.

**References**

**Arnon, I.** and **N. Snider.** 2010. ‘More than words: frequency effects for multi-word phrases,’ *Journal of Memory and Language* 62/1: 67-82.

**Anthony, L.** 2014. ‘AntConc (Version 3.4.3) [Computer Software]. Retrieved from <http://www.laurenceanthony.net/>.

**Baayen, R. H.** 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

**Baayen, R. H.,** **D. J.** **Davidson** and **D. M. Bates.** 2008. ‘Mixed-effects modeling with crossed random effects for subjects and items,’ *Journal of Memory and Language* 59/4: 390-412.

**Babaei, S.,** **F. Taleb N.** and **Z. Fotovatnia.** 2015. ‘Processing of lexical bundles by Persian speaking learners of English,' *The Journal of Teaching Language Skills* *(JTLS)* 6/4: 1-18.

**Barlow, M.** and **S. Kemmer.** 2000. *Usage-Based Models of Language*. United States: CSLI.

**Barr, D. J.,** **R. Levy.,** **C. Scheepers** and **H. J. Tily.** 2013. ‘Random effects structure for confirmatory hypothesis testing: keep it maximal,’ *Journal of Memory and Language* 68/3.

**Bates, D.,** **M. Maechler.,** **B. Bolker** and **S. Walker.** 2015. ‘Fitting linear mixed-effects models using lme4,’ *Journal of Statistical Software* 67/1: 1-48.

**Bybee, J.** 2013. ‘Usage-based theory and exemplar representation’ in Hoffman, T. and G. Trousdale (eds.): *The Oxford Handbook of Construction Grammar* 49-69. Oxford University Press.

**Biber, D.,** **S. Johansson.,** **G. Leech.,** **S. Conrad** and **E. Finegan.** 1999. *Longman Grammar* *of Spoken and Written English*. Essex: Pearson Education Limited.

**Bod, R.,** **J. Hay** and **S. Jannedy.** 2003. *Probabilistic Linguistics*. Cambridge Massachusetts: The MIT Press.

**Boers, F.,** **J. Eyckmans., J. Kappel., H. Stengers** and **M. Demecheleer.** 2006. ‘Formulaic sequences and perceived oral proficiency: putting a lexical approach to the test,’ *Language* *Teaching Research* 10/3: 245–261.

**Brysbaert, M.** and **B. New.** 2009. ‘Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English,’ *Behavior Research Methods* 41/4: 977-990.

**Bybee, J.** 1998. ‘The emergent lexicon,’ *Proceedings of the Chicago Linguistic Society* 34: 421-455.

**Bybee, J.** 2006. ‘From usage to grammar: the mind's response to repetition,’ *Language* 82/4: 711-733.

**Bybee, J.** and **P. Hopper.** 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

**Conklin, K.** and **N. Schmitt.** 2008. ‘Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers?,’ *Applied Linguistics* 29/1: 72-89.

**Conklin, K.** and **N. Schmitt**. 2012. ‘The processing of formulaic language,’ *Annual Review of Applied Linguistics* 32: 45–61.

**Cowan, N.** 2001. ‘The magical number 4 in short-term memory: a reconsideration of mental storage capacity,’ *Behavioral and Brain Sciences* 24/1: 87-114.

**Crick, F. H. C.** 1979. ‘Thinking About the Brain,’ *Scientific American* 241/3: 219-233.

**Ebbinghaus, H.** 1885. *Memory: A Contribution to Experimental Psychology*. *Translated by Henry A. Ruger and Clara E. Bussenius. (1913).* New York: Teachers College Columbia.

**Ellis, N. C.** 2002. ‘Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition,’ *Studies in* *Second Language Acquisition* 24: 143-188.

**Ellis, N. C.** 2003. ‘Constructions, chunking, and connectionism: the emergence of second language structure’ in Doughty and Long (eds.): *Handbook of Second Language* *Acquisition* 63-103. Oxford: Blackwell.

**Ellis, N. C.** 2006. ‘Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning,’ *Applied Linguistics* 27/2: 164-194.

**Ellis, N. C.** 2009. ‘Optimizing the input: frequency and sampling in usage-based and form focussed learning’ in Long M. and C. Dought (eds.): *The Handbook of Second and Foreign* *Language Teaching* 139-158. Oxford: Blackwell Publishing.

**Ellis, N. C.** 2012. ‘Formulaic language and second language acquisition: Zipf and the phrasal teddy bear,’ *Annual Review of Applied Linguistics* 32: 17-44.

**Ellis, N. C.,** **R. Simpson-Vlach** and **C. Maynard.** 2008. ‘Formulaic language in native and second language speakers: psycholinguistics, corpus linguistics, and TESOL,’ *TESOL Quarterly* 42/3: 375-396.

**Goldberg, A. E.** 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

**Higham, P. A.** 1997. ‘Chunks are not enough: the insufficiency of feature frequency based explanations of artificial grammar learning,’ *Canadian Journal of Experimental* *Psychology/Revue Canadienne De Psychologie Expérimentale* 51/2: 126-138.

**Inquisit 4.0.7.0** [Computer software]. 2014. Seattle, WA: Millisecond Software.

**Isobe, Y.** 2011. ‘Representation and processing of formulaic sequences in L2 mental lexicon: how do Japanese EFL learners process multi-word expressions,’ *JACET Kansai* *Journal* 13: 38-49.

**Jiang, N.** and **T. M. Nekrasova.** 2007. ‘The processing of formulaic sequences by second language speakers,’ *The Modern Language Journal* 91/3: 433-445.

**Kornai, A.** 2008. *Mathematical Linguistics*. New York: Springer Verlag.

**Kuiper, K.** 2004. ‘Formulaic performance in conventionalised varieties of speech’ in Schmitt, N. (ed.): *Formulaic Sequences: Acquisition, Processing and Use* 37-54. Amsterdam: John Benjamins.

**Kuiper, K.** and **D. Haggo.** 1984. ‘Livestock auctions, oral poetry, and ordinary language,’ *Language in Society* 13/2: 205-234.

**Laufer, B.** and **N. Girsai.** 2008. ‘Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation,’ *Applied* *Linguistics* 29/4: 694-716.

**Logan, G. D.** 1988. ‘Toward an instance theory of automatization,’ *Psychological Review* 95/4: 492-527.

**Mathy, F.** and **J. Feldman.** 2012. ‘What's magic about magic numbers? chunking and data compression in short-term memory,’ *Cognition* 122/3: 346-362.

**Mclaughlin, B.,** **T. Rossman** and **B. McLeod.** 1983. ‘Second language learning: an information-processing perspective,’ *Language Learning* 33/2.

**Ministry of Education, C. Science, Sports and Technology (MEXT).** 2008a. ‘Shogakkou gakushu shidou yoryo eigoban gaikokugo katsudou,’ *Study of Course Guideline for Foreign Language Activities in Elementary Schools; Provisional Version*. Retrieved from <http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2011/01/05/1234912_010_1.pdf>.

**Ministry of Education, C. Science, Sports and Technology (MEXT).** 2008b. ‘Shogakkou gakushu shidou yoryo kaisetsu gaikokugo katsudouhen,’ *Description for* *Study of Course* *Guideline for Foreign Language Activities in Elementary Schools*. Retrieved from <http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfle/2009/06/16/1234931_012.pdf>.

**Northbrook, J**. and **Conklin, K.** in press ‘“What are you talking about?”An analysis of lexical bundles in Japanese junior high school textbooks,’ *International Journal of Corpus Linguistics*.

**Nguyen, H. T.** and **N. Ishitobi.** 2012. ‘Ordering fast food: service encounters in real-life interaction and in textbook dialogs,’ *JALT Journal* 34/2: 151-186.

**Ogura, F.** 2008. ‘Communicative competence and senior high school oral communication textbooks in Japan,' *The Language Teacher* 32/12: 3-8.

**Pavlik, P. I.** and **J. R. Anderson.** 2008. ‘Using a model to compute the optimal schedule of practice,’ *Journal of Experimental Psychology: Applied* 14/2: 101-117.

**Pawley, A.** and **F. H. Syder.** 1983. ‘Two puzzles for linguistic theory: nativelike selection and nativelike fluency’ in Richards, J. C. (ed.): *Language and Communication*: Longman.

**Pellicer-Sánchez, A.** 2017. ‘Learning L2 collocations incidentally from reading,’ *Language Teaching Research* 21/3: 381-402.

**Siyanova-Chanturia, A., K. Conklin** and **W. J. Van Heuven. 2011.** ‘Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences,’ Journal of Experimental Psychology: Learning, Memory, and Cognition 37/3: 776-784.

**Szudarski, P.** and **R. Carter.** 2014. ‘The role of input flood and input enhancement in EFL learners’ acquisition of collocations,’ *International Journal of Applied Linguistics* 26/2: 245-265.

**Tremblay, A.,** **B. Derwing** and **G. Libben.** 2009. ‘Are lexical bundles stored and processed as single units?,’ *Working Papers of the Linguistics Circle* 19/1: 258-279.

**Tremblay, A.,** **B. Derwing.,** **G. Libben** and **C. Westbury.** 2011. ‘Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks,’ *Language* *Learning* 61/2: 569-613.

**Tyler, A.** 2010. ‘Usage-based approaches to language and their applications to second language learning,’ *Annual Review of Applied Linguistics* 30: 270-291.

**Valsecchi, M.,** **V. Künstler., S. Saage., B. J. White., J. Mukherjee.** and **K. R. Gegenfurtner.** 2013. ‘Advantage in reading lexical bundles is reduced in non-native speakers,’ *Journal of Eye* *Movement Research* 6/5: 1-15.

**Wray, A.** 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

**Wray, A.** 2008. *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

**Wray, A.** 2012. ‘What do we (think we) know about formulaic language? an evaluation of the current state of play’, *Annual Review of Applied Linguistics* 32: 231-254.

**Wray, A.** and **M. R. Perkins.** 2000. ‘The functions of formulaic language: an integrated model,' *Language & Communication* 20: 1-28.

**Webb, S.** and **E. Kagimoto.** 2009. ‘The effects of vocabulary learning on collocation and meaning,’ *TESOL Quarterly* 43/1: 55–77.

**Tables**

**Table 1. Summary of non-native speakers’ age, years of studying English, and self-rating of English proficiency.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Age | Age of first contact  | Years studying English | Speaking | Listening | Reading | Grammar | Overall proficiency |
| Mean | 14.82 | 6.91 | 7.91 | 3.23 | 3.66 | 3.29 | 3.29 | 3.36 |
| SD | 0.38 | 3.49 | 3.50 | 1.06 | 1.16 | 0.99 | 0.96 | 0.95 |

*Note: Speaking, Listening, Reading and Grammar are self-rated on a seven-point scale (1 = very weak, 7 = very strong). Overall proficiency is calculated as the mean of these four categories.*

**Table 2. Mean response times in milliseconds and standard deviations in brackets in each of the three experimental conditions.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lexical Bundles | Non-LBs | Ungrammatical |
| Response time | 1072.61 (325.59) | 1186.65 (362.47) | 1285.59 (368.71) |
| Accuracy | 89% (382.54) | 67% (153.44) | 73% (208.59) |

**Table 3 – Summary of fixed and random effects for the best-fit model**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Response time** |  |
|  | **B** | **CI** | **p** |
| **Fixed Parts** |  |  |  |
| (intercept) | 1193.07 | 1131.41 – 1254.73 | **<.001** |
| Residual Textbook Frequency | -6.34 | -11.49 – -1.19 | **.018** |
| Proficiency | -72.25 | -119.60 – -24.89 | **.005** |
| Phrase Type (Main Item) | -97.59 | -161.73 – -33.45 | **.004** |
| Phrase Type (Ungrammatical Control) | 121.37 | 62.27 – 180.46 | **<.001** |
| Phrase Length | 24.45 | 11.77 – 37.13 | **<.001** |
| Trial Number | -0.36 | -0.54 – -0.17 | **<.001** |
| **Random Parts** |  |  |  |
| σ2 | 68478.713 |  |  |
| τ00, phrase | 4973.613 |  |  |
| τ00, subject | 17412.132 |  |  |
| ρ01 |  |  |  |
| Nphrase | 60 |  |  |
| Nsubject | 35 |  |  |
| ICCphrase | 0.055 |  |  |
| ICCsubject | 0.192 |  |  |
| Observations | 1514 |  |  |
| R2 / Ω02 | .380 / .377 |  |  |
|  |  | **Accuracy** |  |
|  | **OR** | **CI** | **p** |
| **Fixed Parts** |  |  |  |
| (Intercept) | 10.72 | 4.10 – 28.05 | **<.001** |
| Residual Textbook Frequency | 1.09 | 1.01 – 1.18 | **.025** |
| Proficiency | 1.34 | 1.11 – 1.63 | **.003** |
| Residual Familiarity | 1.25 | 1.10 – 1.41 | **<.001** |
| Latency | 1.00 | 1.00 – 1.00 | **.001** |
| Phrase Type (Ungrammatical Control):Response Speed | 1.00 | 1.00 – 1.00 | **.005** |
| **Random Parts** |  |  |  |
| τ00, phrase | 0.705 |  |  |
| τ00, subject | 0.575 |  |  |
| Nphrase | 60 |  |  |
| Nsubject | 35 |  |  |
| ICCphrase | 0.154 |  |  |
| ICCsubject | 0.126 |  |  |
| Observations | 1975 |  |  |
| Deviance | 1549.987 |  |  |