"That Sounds About Right"—Lexical Bundle Naturalness Intuitions in Japanese Learners of English

Julian Northbrook and Kathy Conklin
School of English, University of Nottingham

Abstract

The current study examines the perceived naturalness of lexical bundles learned from early-stage teaching materials in Japanese learners of English. Naturalness ratings of 24 native English speakers and 23 non-native speakers with Japanese as their first language were compared in relation to corpus derived frequencies from the British National Corpus (BNC) and a corpus of Japanese secondary-school English textbooks—the "Junior High School English Textbook Corpus" (JHSETC, Northbrook & Conklin, 2018). The rating scores of both groups were significantly predicted by lexical bundle frequency. However, the groups were sensitive to different metrics; the performance of native speakers was best predicted by the BNC, and that of the Japanese speakers by the JHSETC. This is taken as evidence that learner intuitions are affected by the input they receive from teaching materials, and that these intuitions may stay with learners long term.

Keywords: naturalness intuitions, Japanese learners of English, lexical bundles

1 Introduction

Anyone who has learned a second language has likely had the experience of being told something they said or wrote is, *"Okay, but not how a native speaker would say it."* In Hymes' (1972) taxonomy of situational competence, this relates to the fourth parameter, namely, that language should be *performed.* In other words, language needs to be "done" in the real world, not simply be possible, well formed (feasible) or appropriate (the first three parameters). Arguably, this is one of the greatest challenges language learners face: not distinguishing between grammatical and non-grammatical utterances but distinguishing between *possible* utterances and *likely utterances.* Early work in the area of formulaic language highlighted this fact and proposed a new way of looking at language—not as simply a set of syntactic rules and words, as generative grammar approaches had proposed, but as a memory-based system using lexical elements (Bolinger, 1975). Sinclair (1991) referred to this as a system that considers both an open choice principle and the *idiom* principle. The "open choice" principle refers to speakers' ability to construct unique and creative utterances using a system of rules and connected items. However, Sinclair argued that this is not the way in which we primarily produce language. Rather, we default to the *idiom principle*—that is, where possible we

use pre-constructed formulaic—or, "performed"—language that is entrenched in long-term memory (see, for example, Conklin & Schmitt, 2012).

Following this, Pawley and Syder (1983) coined the terms "nativelike fluency" and "nativelike selection," arguing that simply exercising the creative potential of syntactic rules is not sufficient to sound "nativelike" when speaking a language. Although native speakers could say *"Pleasant morning."* or, *"Could you aid me in this task?"* they tend not to do so. This is not because these phrases are ungrammatical but rather because there are more common, socially institutionalised phrases available (e.g., *"Good morning."* and *"Could you help me with this?"*). Most people would likely agree that these more common alternatives sound more natural, are understood more readily and are preferred by native speakers. At the same time, there are utterances such as, *"Thanks very much."* that are highly frequent expressions and commonly used by native speakers, but which are, in fact, ungrammatical (the expression being an amalgamation of *"Thank you very much"* and the archaic *"many thanks to you"* and, therefore, does not conform to typical grammatical conventions). We must conclude, then, that communicative competence in a language is not simply the ability to produce grammatically correct utterances. It also encompasses the ability to select frequent and conventional patterns that are *likely* in the language (see Biber et al., 1999; Cowie, 1998; Ellis, 2001; Howarth, 1998; Skehan, 1998, among others, for further discussion), and therefore, whether learners are able to do this is an important issue in English as a Foreign Language (EFL) and Second Language Acquisition. Before we consider studies that have looked at learner intuitions of formulaic sequences, it is first necessary to define the concept of formulaicity in the context of the current study.

1.1 Formulaic Sequences in Second Language Learning

There are many different types of formulaic language: idioms: *kick the bucket*, collocations: *plastic surgery*, binomials: *bride and groom*, phrasal verbs: *rely on*, lexical bundles: *I don't want to*, to name but a few. The current study focuses on lexical bundles, which are non-idiomatic sequences of words that are defined purely based on their frequency of occurrence (Biber, 2009; Biber et al., 1999; Stubbs & Barth, 2003) and can be thought of as extended collocations (Biber et al., 1999). At the early stages of second language acquisition, learners are thought to use memorised formulaic language, which allows them to express themselves above and beyond their current abilities (Myles et al., 1998, 1999; Nattinger & DeCarrucio, 1992; Wray, 2002). It is not necessary, for example, to understand that the phrase "*That sounds great*" is made up of a pronoun acting as a noun phrase, which is modified by a verb phrase consisting of a third-person singular verb in the present tense combined with an adjective—or indeed, to worry that the aforementioned, *"Thanks very much"* is, technically speaking, ungrammatical. Learners can simply use them as is. As proficiency advances, however, learners tend to construct much of their speech using rules rather than lexicalised routines (Foster, 2001; Nesselhauf, 2003; Skehan, 1998), perhaps believing that high-proficiency language requires originality. This is mistaken, however, because studies show that the use of formulaic language, far from indicating that a learner is a "beginner," is, in

fact, a mark of nativelikeness and, therefore, proficiency (e.g., Boers et al., 2006; Dickinson, 2012).

1.2 Exploring Naturalness in Second Language Learners

Dickinson (2012) trained students on specific formulaic sequences and found that those who practised and actively used them gave presentations that were more engaging and easier to understand than those who did not. In a similar study, Taguchi (2007) found that instructing and drilling formulaic sequences with learners of Japanese as a second language led to a higher sensitivity to discourse features and was useful for developing learners' oral ability. Wray and Fitzpatrick (2008) similarly concluded that by acquiring formulaic sequences that are common in conversations, learners not only increase their fluency and naturalness in the language but also increase their confidence in using the language. In another study, Boers et al. (2006) trained two groups of students for 22 hours each—one using a traditional words-and-rules approach; the other was trained to use formulaic sequences. Subsequently, two judges rated the group that had formulaic training as having higher proficiency.

Despite these findings, very little has been said about how well non-native speakers are able to judge the naturalness of sequences they encounter. Does instruction help learners to differentiate between utterances such as *can you aid me with this task?* and *can you help me with this?* Which, in turn, would allow them to better assess the formulaicity of their own speech and "select" un-nativelike expressions? In one of the few studies looking specifically at learners' ability to judge formulaic sequences, Edmonds (2013) administered an online naturalness judgement task to French natives and two groups of learners of French living in France (half long stay and half short stay). All groups judged conventional expressions as being more natural and did so faster than control phrases. Importantly, this included the short-stay French learners, which, as Edmonds concludes, indicates that even learners who have little relative exposure to the target language are "well on their way to nativelike selection" (p. 95). In two studies that looked at adjective-noun collocation intuitions, Siyanova and Schmitt (2008) and Siyanova-Chanturia and Spina (2015) asked native and non-native speakers to judge the frequency of collocations, which we take here to be related to their naturalness. Similar to Edmonds (2013), Siyanova-Chanturia and Spina (2015) found that both native and non-native speaker intuitions correlated strongly with corpus frequency. This was, however, only for high-frequency items. Neither group were able to judge medium- or low-frequency pairs, although the native speakers were able to accurately judge very low-frequency items. Siyanova and Schmitt (2008) had somewhat similar findings. They showed that native speaker intuitions correlated strongly with BNC frequency scores. However, the non-native group displayed poor collocational intuitions in comparison with the native speakers.

Overall, it seems that while learners may be able to accurately judge formulaic sequences, their intuitions may not be entirely accurate or in tune with authentic language use. An explanation for this may be found in usage-based approaches to second language acquisition, which put a premium on the linguistic input that learners receive. In such an approach, experience of and exposure

to language result in high-frequency, repetitive sequences of words being stored in long-term memory (Barlow & Kemmer, 2000; Bybee, 1998, 2006, 2013; Ellis, 2002a, 2002b, 2003), which would influence learners' intuitions regarding formulaic sequences. Thus, the question arises: is it that the second language learners cited above are not sensitive to collocation, or is it that their intuitions are simply *different*? That is, do they treat formulaic language as compositional and not as chunks, or is it simply that they have, through exposure to "un-nativelike" English (i.e., "textbook English") developed un-nativelike intuitions?

In order to answer this question, it is necessary to consider the input English learners receive. In the case of EFL students, this is, as previously stated, primarily classroom instruction and materials (Meunier, 2012; Nunan, 1991). Studies focusing on secondary school textbooks have found the dialogues in them to be inauthentic (Nguyen & Ishitobi, 2012; Northbrook & Conklin, 2018; Römer, 2004, 2005, see also Gilmore, 2004), in that they do not accurately represent the situations students need to be prepared for. Looking at formulaic language more specifically, Northbrook and Conklin (2018) examined lexical bundles in English textbooks for Japanese secondary school students and found them to occur frequently, although they were qualitatively different to those found in naturally produced English. Northbrook and Conklin (2019) found that junior high school students had a processing advantage on a phrasal judgement task for lexical bundles from their textbooks but not for matched controls. It appears that the learners were exposed to un-nativelike formulaic language and had an advantage for it, irrespective of native-speaker frequency distributions. Crucially, the lexical bundles tested in Northbrook and Conklin (2019) were selected *because* they occurred in the textbook series that the participants had used during their 3 years of junior high school, and the students were tested while in their final year. Subsequent exposure to English might help students to overcome their initial exposure and become more nativelike.

The current research is a small-scale study that investigates whether adult Japanese learners of English intuitions are based on the frequency distributions of the classroom English from their junior high school textbooks or whether they have become more nativelike. To address this, we asked adult native speakers of English and adult non-native, Japanese learners of English to rate the naturalness of sentences containing lexical bundles in an offline judgement task and examine whether their ratings were related to a native speaker corpus (the British National Corpus, BNC) or the Junior High School English Textbook Corpus (JHSETC).

2 Methodology

2.1 Participants

Participants for this study included 24 native speakers from the University of Nottingham and 23 native Japanese speakers. The native speakers participated for course credit. Initially, we had planned to recruit the Japanese English learners on-site and run the experiment in a lab setting, replicating Edmonds's (2013) study. However, due to the difficulty of recruiting appropriate participants in the area, we decided to modify the research design and run the experiment online

(see Procedures below). The Japanese speakers were all studying English as a second language and were recruited from an online email newsletter associated with a self-study e-learning platform targeted at intermediate English learners. We sent an email to a segment of the newsletter that represented the best fit for the study (Japanese English learners living and studying in Japan) asking for volunteers to participate. The group were, therefore, a sample of convenience. The Japanese participants were asked to provide their background information, as summarised in Table 1. In addition to what can be seen in the table, we asked the participants to indicate their profession: five were university students, and the remaining were in full-time employment (10 working in an office, three as engineers, one as a lawyer, one as a medical officer and one as a waitress). All of the Japanese participants reported that they lived in Japan, had never lived in an English-speaking country and had attended a typical Japanese-medium school.

2.2 Materials

We created 17 passages consisting of a line of text to provide context, followed by a short dialogue. One portion of each dialogue contained either a high-frequency lexical bundle or a low-frequency control phrase matched for meaning (see below for lexical bundle and control selection criteria). There were two versions of each passage (a total of 34 possible items). Participants saw either version A or B, but not both, so each participant judged 17 items. Lexical bundles are ideal for looking at questions of naturalness, because unlike idioms, they are not easily identifiable when embedded into a sentence. For example, items such as, *"James' aunt kicked the bucket / died last week,"* might make it clear that the focus is on idioms and potentially influence participants' responses. In contrast items like, *"Really, I don't know why/ I'm not sure why I agreed to this,"* should make the focus on formulaic language less apparent. Each item included a context line and dialogue that served to contextualise and mask the lexical bundle. In the following example, everything other than the underlined portion is supporting the context:

Context: Andy is going on a double date with his friends.

Lexical bundle: "Really, I don't know why I agreed to this," he grumbled to his friends Simon and Alice. "You know this isn't my sort of thing."

Or:

Control: "Really, I'm not sure why I agreed to this," he grumbled to his friends Simon and Alice. "You know this isn't my sort of thing."

Table 1. Summary of Non-native Speakers' Age and Self-rating of English Proficiency for Participants

	Age	Age of first contact	Speaking	Listening	Reading	Overall proficiency
Mean	31.15	12.55	3.23	3.66	3.29	3.36
SD	6.38	0.91	1.06	1.16	0.99	0.95

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

In lexical bundle studies, controls are normally made by switching a single word for another matched for frequency. For example, "*in the middle of the*" (1,513 occurrences in the BNC) might become "*in the front of the*" (47 occurrences in the BNC). This was not possible here because the lexical bundles and controls needed to have the same meaning so that they could be used in the same contextual sentences. Therefore, we first selected 17 lexical bundles with a minimum frequency of over 3 per million occurrences in the BNC. We then selected synonymous expressions, ensuring that their frequency was as low as possible in comparison with the lexical bundles. For example, the only difference between the two versions in the example above is the lexical bundle, "*I don't know why*" (frequency = 651 in the BNC) and "*I'm not sure why*" (15 occurrences in the BNC). Thus, any difference between the two can be attributed to the presence or absence of a lexical bundle.

2.3 Procedure

The passages were shown to participants in an offline naturalness judgement task. The participants saw each passage individually in the form of a survey and were asked to rate the underlined portion for how natural they thought it sounded on a seven-point scale (7 = very natural and 1 = very unnatural). The items were counterbalanced so that participants saw an equal number of lexical bundles and control items, but never both versions of the same item. It should be noted that the decision to use a seven-point scale was arbitrary and other scales would be possible, which will be taken up in the discussion of limitations of the current study.

2.4 Analysis and Results

Initially, we planned to model the data using the seven-point naturalness rating scale in its raw form; however, in order to use a simpler form of analysis, more suited to the (small) sample size (discussed below), we made the decision to collapse the scale into a binary, "High" or "Low" metric where a score of 5, 6 or 7 constituted a high level of naturalness and a score of 1, 2 or 3 constituted low naturalness. We removed scores of 4 from the analysis as they represent a middle ground, and therefore, likely represent a "do not know" decision (this resulted in a loss of 11% of the data from the Japanese speakers and 9% from the native speaker group. Table 2 provides a summary of responces.

A chi-square analysis reveals that the native English speaker group was significantly more likely to give items a "high" rating, irrespective of lexical bundlehood

Table 2. Summary of Responses

		High	Low
Japanese	Lexical bundle	114 (29%)	62 (16%)
	Control	81 (21%)	91 (23%)
Native	Lexical bundle	154 (38%)	34 (8%)
	Control	115 (28%)	70 (17%)

Note: *Percentages are calculated from the total number of responses across both conditions by participant group.*

(lexical bundles: $X^2 (1) = 76.59$, $p \leq 0.001$, controls: $X^2 (1) = 10.95$, $p \leq 0.001$), and that the Japanese participants were more likely to give lexical bundles a high rating ($X^2 (1) = 15.36$, $p \leq 0.001$); however, there was no significant difference in the number of items they rated as high or low ($X^2 (1) = 0.58$, $p = 0.45$). This indicates that the Japanese English learners are sensitive to higher frequency lexical bundles; however, in this case we would also expect them to rate control items significantly lower, which we would also expect from the native speaker group, as discussed below.

The data were further analysed using generalised linear mixed effects models (GLMER) with the package lme4 (Bates et al., 2015). The data were modelled with the collapsed naturalness metric as the response variable with condition, phrasal frequency and sequence length (calculated as the number of characters in a sequence) as fixed effects. For the Japanese learners, we also included age and age of first acquisition. Because we are primarily interested in differences in how the native speakers rate items when compared with Japanese English learners, we needed two separate phrasal frequency metrics tailored to each of the two groups. As a "native" metric we used values from the BNC. For the Japanese speakers, selecting a representative frequency metric is difficult as we know little about the relative exposure to English of each participant. What we do know, however, is that each participant learned English at school from the age of 13 years using Ministry of Education approved textbooks. Frequency scores from a corpus of these textbooks—the JHSETC (Northbrook & Conklin, 2018)—proved a robust indicator of online processing in a task with junior high school students (Northbrook & Conklin, 2019). Therefore, we used the JHSETC as an indication of participants' exposure to lexical bundles (and the controls). In order to account for idiosyncratic differences between participants (see Winter, 2014), participants were included as a random-effect factor in the model. 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, BNC phrasal frequency was correlated with JHSETC frequency, so BNC phrasal frequency was fitted as the response variable in a model with JHSETC frequency as a predictor. Other correlated factors were fitted in a similar manner. The resulting residuals all correlated significantly with their related variables (<0.001): BNC phrasal frequency ($r = 0.72$) and length ($r = 0.77$). All other continuous predictors were centred to avoid having a change in slope that might correlate with a change in intercept (see Baayen, 2008).

For both native and non-native speakers, a simple model was fit with residual JHSETC frequency, and with subject as a random effect. Complexity was then built into this model by adding new predictors in a step-by-step manner. These were also tested as adjustments for the participant random effect (as recommended by Barr et al. 2013). Each iteration of the model was tested against the previous best fit to see whether the change was significant (likelihood ratio test, $p \leq 0.05$). This process was continued until all predictors and interactions had been tested as both fixed and random effects. Table 3 summarises the best fit for both models.

Whether an item was classified as a lexical bundle or a control phrase made no difference for either native ($X^2 (1) = 2.09$, $p = 0.15$) or non-native speakers ($X^2 (1) = 0.29$, $p = 0.59$) indicating that lexical bundlehood cannot be reduced to

Table 3. Summary of Fixed and Random Effects for the Best-fit Model

Predictors	Native speakers					Japanese speakers				
	b	*CI*	*SE*	*X²*	*p*	*b*	*CI*	*SE*	*X²*	*p*
(Intercept)	5.37	2.22–18.54	3.47	-	0.001 **	1.19	0.00–1.69	0.04	-	0.083
BNC	1.7	1.08–1.56	0.12	8.25	0.004 **	1.13	0.90–1.25	0.09	0.49	0.48
JHSETC	0.81	0.57–1.23	0.16	0.79	0.38	1.91	1.23–2.33	0.28	10.69	0.001 **
Length	1.55	1.07–1.30	0.06	10.71	0.001 **	1.2	0.98–1.17	0.05	2.41	0.12
LB type	0.44	0.15–1.33	0.25	2.09	0.15	1.3	0.53–3.23	0.6	0.29	0.59
Age of FA	-	-	-	-	-	1.22	0.89–1.73	0.21	1.24	0.26
Age	-	-	-	-	-	1.25	0.99–1.08	0.02	1.61	0.21

Subject	Variance¹		Variance	
	0.24		0.06	

Native Speaker Model formula: Naturalness Binary ~ Residual BNC Frequency + Residual JHSETC Frequency + Residual Length + Lexical Bundle Type + (1 | Subject). Participants = 24, Observations = 373.
Japanese Speaker Model formula: Naturalness Binary ~ Residual BNC Frequency + Residual JHSETC Frequency + Residual Length + Lexical Bundle Type + Age of First Acquisition + Age + (1 | subject). Participants = 23, Observations = 348.

a simple binary metric. However, there were significant effects based on lexical bundle frequency. The data was modelled with both native-speaker frequency values and frequency values from the JHSETC. Native speaker ratings were predicted best by the BNC phrasal frequency, and there was no significant effect of the JHSETC values ($X^2 (1) = 0.79$, $p = 0.38$). Japanese speaker ratings, however, were predicted best by the JHSETC phrasal frequency but not the BNC values ($X^2 (1) = 0.49$, $p = 0.48$). This is an important finding and will be taken up in the Discussion section. Phrasal length was a significant predictor for the native speakers, but not for the Japanese learners ($X^2 (1) = 2.41$, $p = 0.12$). Neither age ($X^2 (1) = 1.61$, $p = 0.21$) nor age of first acquisition ($X^2 (1) = 1.24$, $p = 0.26$) were significant factors for the Japanese speakers. None of the random effect adjustments improved either of the models.

3 Discussion

Although this study did not find any difference between lexical bundles and non-lexical bundles (*I don't know why* vs. *I'm not sure why*), within the category of "lexical bundle" native and non-native speakers display some interesting differences. The lexical bundles that native and non-native speakers rate as more or less natural are not the same. The native speaker group rate items that are of high frequency in the BNC as more natural (although the effect size for this is relatively small; discussed below), and there is no JHSETC frequency effect. Native English speakers have not been exposed to Japanese English textbooks, and therefore, we would not expect frequency in the JHSETC to be a significant predictor of their behaviour, which was, indeed, the case. In contrast, the Japanese learners of English are sensitive to the frequency of occurrence in their (early) English textbooks (JHSETC); that is, items with a higher frequency in the JHSETC are judged

as more natural and those with a lower frequency, less so. BNC frequency metrics did not predict the Japanese speakers' naturalness ratings.

As mentioned briefly in the Introduction section, the detailed analysis of the lexical bundles in the JHSETC from Northbrook and Conklin (2018) demonstrated that they are quantitatively and qualitatively different from those in a native-speaker corpus. The current results provide some additional evidence that the lexical bundles in the JHSETC do not represent native-speaker English; if they did, the native speakers would rate ones occurring more frequently in the JHSETC as more natural. Notably, the Japanese speakers were exposed to lan- guage from the JHSETC, on average, 15 years before the current study. The fact that we see an effect of the JHSETC, a corpus that reflects the learners' earliest experience with English, demonstrates the importance of teaching materials tar- geted at beginners. In addition, other studies focusing on secondary school text- books have found the dialogues in them to be inauthentic (Nguyen & Ishitobi, 2012; Römer, 2004, 2005), in that they do not accurately represent the situations students need to be prepared for. Based on the findings of the current study, not only is this problematic because we are not preparing our students for English use in the real world, but arguably presenting inauthentic, "unnatural" English to stu- dents is potentially also damaging in the long term as evidenced by their intuitions about the naturalness of English lexical bundles.

Native speakers rated items significantly higher than non-native speakers, irrespective of whether the item was a lexical bundle or matched control phrase. Given this, it is not surprising that lexical bundlehood was not a significant factor in the model, at least for the native speakers. However, rather than showing evidence of a binary distinction (lexical bundle vs. control), native speakers' intuitions about the naturalness of the phrases was related to their frequency in a native-speaker corpus (BNC), which is consistent with the previous study findings (e.g., Siyanova & Schmitt, 2008). It is important to note, however, that for the native speakers, the effect size for fixed effects was very small in relation to the intercept, indicating that such a conclusion should be treated with caution. This was not the case for the Japanese learners, where the intercept was not significant in the model and JHSETC frequency was the best predictor of their rating behaviour.

It is important to note that the current study was relatively small in scale and so included fewer observations than can be considered best practice for mixed effects modelling. Brysbaert and Stevens (2018) stated that a properly powered experiment with repeated measures should have at least 1,600 observations per condition. Clearly, this is considerably more than the 373 observations that this study has (for the native speaker group; the Japanese group included 348 observa- tions). We do not believe this to be a problem given the relative simplicity of the models, however. Furthermore, we do not believe that this detracts from the core finding of the study, particularly regarding the Japanese speaker group. Results from the current study indicate that the participants are influenced by the mate- rials they have studied, which we would expect from a usage-based approach and are in line with previous studies (discussed below). However, in light of the small effect size, notably regarding the native speaker group, subsequent research cer- tainly needs to include a greater number of participants and experimental stimuli. Doing so would allow us to not only replicate the current findings but also build

more complexity into the statistical models, and thus, account for a wider range of potential factors that may influence intuitions of naturalness.

A further limitation of the current study is the use of the seven-point scale to rate naturalness. As mentioned previously, the original intention was to analyse the data using the raw ratings from this scale. However, doing so required a more complex form of analysis that was not appropriate given the (small) number of observations. This meant that we had to make a decision on how to use the data, which almost certainly affected the results. It is not clear whether a score of "3," for example, is really different to a score of "4" (which we took as a middle ground, and therefore, removed from the analysis). Ideally, we would have only kept the bottom and top end of the scale, removing everything between a score of 3 and 5; however, doing so would have resulted in considerable data loss. In terms of further research, this scale certainly needs further validation. In a future iteration of this study, we will either follow Edmonds (2013) and use a simple, "High," "Low" or "Can't Decide" scale, or else use a longer nine-point scale, which would have a clearer low, middle and high end.

Finally, the current study is limited in that it only focuses on a single group of non-native speakers: Japanese English learners. We chose this population because it continues on from and adds evidence to findings from previous studies in the author's research (Northbrook & Conklin, 2018, 2019). Furthermore, corpora such as the JHSETC are essential for generating phrasal frequency metrics tailored to a specific group's experience of English. In order to test the role of early input in other populations, similar corpora would be needed (which will be the topic of future research, specifically with Taiwanese secondary-school learners of English, who use very similar government-mandated textbooks to Japanese secondary-school students). However, the current research provides an indication of the lasting impact of the language materials that young learners are exposed to. In the intervening time between exposure to the textbook language and the current study, the learners will have been exposed to English in other ways (including their current English study in the e-learning platform from which they were recruited). Crucially, this additional exposure was not reflected by naturalness ratings that were in line with the native speaker ones and related to a native speaker corpus. It may be that frequency metrics from another source, say, for example, the Corpus of Contemporary American English (COCA) (Davies, 2008) or the SUBTLEXus (Brysbaert & New, 2009), may better predict Japanese speakers' rating scores because the Japanese learners may have had more exposure to American English, or indeed scores taken directly from materials used in the e-learning environment. This warrants further research, particularly to address whether very recent exposure to materials optimised with high-frequency lexical bundles (i.e., ones that could be embedded in their e-learning platform) would "overwrite" the (un-nativelike) lexical bundles that learners' have entrenched in long-term memory. This is the topic of a subsequent study by the authors (Northbrook et al., 2021).

Overall, as with Northbrook and Conklin's (2019) previous research, the results from this study align well with usage-based accounts of language acquisition, which maintain that language develops in reaction to use and exposure (Barlow & Kemmer, 2000; Bybee, 2006; Ellis et al., 2013; Tyler, 2010). Northbrook and Conklin (2019)

demonstrated a clear processing advantage on a phrasal judgement task by Japanese junior high school students for the lexical bundles that they had encountered in their textbooks. The students were sensitive to the phrasal frequency of items embedded in their early textbooks, and responded to them faster and more accurately. Such findings indicate that exposure to language underpins entrenchment in memory—even from the very earliest stages of learning. That frequency information from these textbooks predicts the naturalness ratings of *adult* Japanese learners of English suggests that the processing advantages they developed early on may be retained over time, and even in some cases—for example where there is a large discrepancy between the frequency distribution of an item in the textbooks and in naturally-occurring language—influence their judgement and prevent them from making (fully) "nativelike" evaluations of lexical bundles. These findings once again highlight the need for carefully considering the input students are given, even at the earliest stages of learning, as they may have a far-ranging influence.

4 Conclusion

This study, while small in scale, demonstrates that language learners may develop intuitions of English based on the actual English they have learned and have been exposed to, which supports a usage-based view of language acquisition. If the input learners receive is not nativelike, their *intuitions* will not be nativelike. Crucially, once formed, these intuitions may become entrenched in memory and stay with learners long term; the current findings of this study indicate an influence of English materials that the learners encountered, on average, 15 years previously. Notably, these findings highlight the need for carefully considering the input students are given and for textbook designers to create contents that reference native speaker English, rather than simply creating examples to demonstrate grammar points without consideration of the usage of the examples in which they are embedded.

Note

1. Variance (ICC) indicates between-person differences with a score ranging from 0 to 1 where a low score indicates a higher level of between-participant differences. There is a higher level of variance in rating behaviour in the Japanese participants than the native speaker group. This is expected given that the Japanese speakers were a sample of convenience and the native speakers from a single student population, as well as because we expect second language learners to have a more varied experience of English than monolingual speakers for whom English is a first language.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to Statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Barlow, M., & Kemmer, S. (2000). *Usage based models of language.* CSLI Publications.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275–311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245–261. https://doi.org/10.1191/1362168806lr195oa

Bolinger, D. (1975). Meaning and memory. *Forum Linguisticum, 1*, 2–14.

Brysbaert, M., & New, B. (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. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

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. https://doi.org/10.1353/lan.2006.0186

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49–69). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.001.0001

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61. https://doi.org/10.1017/S0267190512000074

Cowie, A. P. (1998). Phraselogical dictionaries: Some east-west comparisons. In A. P. Cowie (Ed.), *Phrasology: Theory, analysis and applications* (pp. 209–228). Oxford University Press. https://doi.org/10.1353/lan.2000.0046

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. https://corpus.byu.edu/coca/

Dickinson, P. (2012). Improving second language academic presentations with formulaic sequences. *Bulletin of Niigata University of International and Information Studies Department of Information Culture, 15*, 25–36.

Edmonds, A. (2013). Conventional expressions. *Studies in Second Language Acquisition, 36*(1), 69–99. https://doi.org/10.1017/S0272263113000557

Ellis, N. C. (2001). Memory for language. In P. J. Robinson (Ed.), *Cognition and second language acquisition* (pp. 33–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 297–339. https://doi.org/10.1017/S0272263102002140

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 63–103). Blackwell. https://doi.org/10.1002/9780470756492

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning, 63*(1), 25–51. https://doi.org/10.1111/j.1467-9922.2012.00736.x

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 75–93). Longman.

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *English Language Teaching Journal, 59*(4), 363–374. https://doi.org/10.1093/elt/58.4.363

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phrasology: Theory, analysis and applications* (pp. 161–186). Oxford University Press.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics, 32*(2), 111–129. https://doi.org/10.1017/S0267190512000128

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48*(3), 323–363. https://doi.org/10.1111/0023-8333.00045

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition, 21*(1), 40–80. https://doi.org/10.1017/S0272263199001023

Nattinger, J. R., & DeCarrucio, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press. https://doi.org/10.1017/S0272263100013000

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Nguyen, H. T., & Ishitobi, N. (2012). Ordering fast food: Service encounters in real-life interaction and in textbook dialogs. *Japan Association for Language Teaching Journal, 34*(2), 151–186. https://doi.org/10.37546/JALTJJ34.2-2

Northbrook, J., Allen, D., & Conklin, K. (2021). "Did you see that?"—The role of repetition and enhancement on lexical bundle processing in English learning materials. *Applied Linguistics.* https://doi.org/10.1093/applin/amab063

Northbrook, J., & Conklin, K. (2018). "What are you talking about?" An analysis of lexical bundles in Japanese junior high school textbooks. *International Journal of Corpus Linguistics, 23*(3), 311–334. https://doi.org/10.1075/ijcl.16024.nor

Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out?—Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics, 40*(5), 816–833. https://doi.org/10.1093/applin/amy027

Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. Prentice Hall.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards (Ed.), *Language and communication* (pp. 191–225). Longman.

Römer, U. (2004). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 151–168). John Benjamins. https://doi.org/10.1075/scl.17

Römer, U. (2005). *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.18

Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review, 64*(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning, 65*(3), 533–562. https://doi.org/10.1111/lang.12125

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language, 10*(1), 61–104. https://doi.org/10.1075/fol.10.1.04stu

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*(4), 433–457. https://doi.org/10.1177/1362168807080962

Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics, 30*, 270–291. https://doi.org/10.1017/S0267190510000140

Winter, B. (2014). *A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)*. http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123–147). John Benjamins. https://doi.org/10.1075/z.138