Advantage in Reading Lexical Bundles is Reduced in Non-Native Speakers

Matteo Valsecchi
Department of General Psychology,
Justus-Liebig-University of Giessen, Germany

Viktoria Künstler
The Centre for Media and Interactivity,
Justus-Liebig University of Giessen,
Germany

Sven Saage
The Centre for Media and Interactivity,
Justus-Liebig-University of Giessen,
Germany

Brian J. White
Centre for Neuroscience Studies,
Queen's University Kingston, Ontario,
Canada

Joybrato Mukherjee
Department of English, Justus-Liebig-
University of Giessen, Germany

Karl R. Gegenfurtner
Department of General Psychology,
Justus-Liebig-University of Giessen, Germany

Formulaic sequences such as idioms, collocations, and lexical bundles, which may be processed as holistic units, make up a large proportion of natural language. For language learners, however, formulaic patterns are a major barrier to achieving native like competence. The present study investigated the processing of lexical bundles by native speakers and less advanced non-native English speakers using corpus analysis for the identification of lexical bundles and eye-tracking to measure the reading times. The participants read sentences containing 4-grams and control phrases which were matched for sub-string frequency. The results for native speakers demonstrate a processing advantage for formulaic sequences over the matched control units. We do not find any processing advantage for non-native speakers which suggests that native like processing of lexical bundles comes only late in the acquisition process.

Keywords: Formulaic sequences, reading, non-native speakers, language learning.

Introduction

It is widely accepted that natural language is to a great extent made up of multi-word expressions or formulae. Invariable idiomatic expressions such as *a piece of cake* or less fixed but frequent lexical bundles such as *come to terms with* perform a range of pragmatic and discourse organizing functions contributing to fluent and native-like language use (e.g. Nattinger & DeCarrico, 1992; Schmitt, 2004; Sinclair, 1991; Wray, 2002). Depending on the definition of what constitutes a formula,[1] studies, for example, by Erman and Warren (2000) and Foster (2001) suggest that about 30 to 60 per cent of language is made up of formulaic chunks. Corpus-based studies were especially able to uncover and to describe many forms of formulaic expressions like idioms, collocations and lexical bundles in spoken and written language (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Moon, 1998), and also in different subtypes of discourse, for instance, in academic language (Biber, Conrad, & Cortes, 2004; Hyland, 2008; Simpson & Mendis, 2003).

Given that the use of formulaic language is key to native-like performance, formulaic language has also received a great deal of attention in Second Language Acquisition (SLA) research. A lot of research has addressed, for example, the presence of formulaic language in the output of learners showing that formulae constitute a major challenge to non-native speakers of English. While learners often rely heavily on prefabricated language at the early stages of language learning, they fail to use a native-like amount of formulae at the later stages. Based on corpus data, it has been shown that advanced learners of English overuse certain formulae and underuse others in comparison to native speaker (e.g. De Cock, 2004; Durrant & Schmitt, 2009b; Gilquin, 2007; Granger, 1998; Nesselhauf, 2003). A major reason why formulaic language is a difficult area for non-native speakers of English probably lies in the nature of formulaic language itself. Pawley and Syder (1983), Kjellmer (1990) and Wray (2002) claim that unlike native speakers, who store, process and retrieve formulaic language as holistic units, learners tend to compose such units word by word and have to make a conscious effort to acquire formulaic sequences. In the light of the results obtained from corpus-based studies, the holistic hypothesis for the processing of formulaic language seems very appealing. Yet, corpus data alone does not allow us to automatically draw any conclusions about how formulaic sequences are processed and needs to be complemented by experimental data. Multi-word units units have received little attention so far and only a few studies have addressed the online processing of formulae by learners of English.

One example is a study by Jiang and Nekrasova (2007) who conducted two online judgment experiments to investigate whether formulaic language is processed holistically by native and non-native speakers of English. Participants were required to judge as quickly and as accurately as possible the well-formedness of the presented phrases. The test materials were made up of formulaic sequences taken from pre-selected lists published in previous corpus linguistic studies (e.g. Biber et al., 1999), matched control sequences and ungrammatical sequences. Based on the assumption that holistic storage and retrieval of formulaic language should require less or no syntactic parsing, they analyzed the participants' reaction times and error rates. They found that both native and non-native speakers responded to formulaic phrases faster and with fewer errors than to the control sequences.

An eye-tracking study by Underwood and colleagues (2004) and a follow-up experiment using a self-paced reading task (Schmitt & Underwood, 2004) also looked at the processing of formulaic language by native and non-natives. Using the same set of items in both experiments, they studied the terminal words in formulaic sequences and the same lexical items in a non-formulaic context in the eye-tracking study and were then able to measure the reading times for individual words in the follow-up self-paced reading task. While significant advantages for the processing of formulaic language in the native group were revealed, the results in the non-native group failed to support the processing advantage.

Possible explanations for the mixed results in these studies might be the different research paradigms used, but also the different definitions of formulaic language. While the former study by Jiang and Nekrasova (2007) looked at multi-word expressions "that occur as phrases and as coherent semantic units at a relatively high frequency", as, for example, *in other words* or *take a look at*, the latter studies used a wide definition of formulaic language and included a range of multi-word expressions, such as lexical phrases, transparent metaphors, and idioms. Although the authors did not concentrate on the identification of formulae, this approach runs the risk of masking possible processing differences between these types of formulaic language. Indeed, Schmitt (2005) and Columbus (2010) draw attention to different types of formulae or multi-word units (MWUs) . The eye-tracking study by Columbus (2010) finds that various types of formulaic language are psycho-linguistically valid, and although all types of formula (in the case of the Columbus study, restricted collocations, figurative idioms and lexical bundles) are read faster than compositional phrases, differences in processing times between the subtypes should not be underestimated and merit further research. A couple of studies, however, have focused on particular types of formulaic language.

Two studies which focus on a specific type of multi-word units are two eye-tracking studies by Siyanova-Chanturia and colleagues (Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011). In the first study, they address the processing of idioms and especially their figurative and literal meaning and find that both figurative as well as literal ones are processed faster than novel phrases by native speakers. The non-native speaker group, on the con-

trary, read idioms and compositional phrases at the same speed and processed figurative uses of idioms significantly slower than literal ones.

In the second study, Siyanova-Chanturia, Conklin and van Heuven (2011) focused on binomial expressions. Binomial phrases are made up of two content words from the same lexical class and a conjunction such as *black and white*, *day and night*, and are reversible in the word order in the majority of cases. Presenting the participants with frequent binomials[2] and their less frequent reversed form, they show that both native and non-native speakers process frequent binomials faster. Advanced non-native speakers and native speakers are also sensitive to the binomial order, processing binomials faster than the reversed forms.

Finally, a study by Ellis and colleagues (2008) which combined corpus linguistic methods and metrics for identification and selection of items and reading and production tasks closely examined the processing of academic lexical bundles with respect to length, frequency and association measures (Mutual Information-MI) by native speakers and second language learners. Their results suggest that a fast processing of lexical bundles in native and non-native speakers is triggered by different factors. Whereas the MI-score seems to be more relevant for native speakers, non-native speakers are influenced by the frequency of the lexical bundle. One has to keep in mind, however, that the study was limited to lexical bundles occurring in academic language.

The general picture that emerges from the review of these studies and other similar ones (e.g. Conklin & Schmitt, 2008; Tremblay, Derwing, Libben, & Westbury, 2011) is that formulaic language clearly shows a processing advantage in native speakers. Yet, processing advantages appear to apply to a variable extent to various types of formulaic language and it remains unclear which determinants affect the processing of formulaic language. Similarly, the issue of formula processing in non-native speakers is far from resolved. The above-mentioned studies provided us with first insights into processing by non-native speakers.

Inspired by these previous findings, the goal of our study is to add further insight into the processing of formulaic language by combining corpus-derived information with an eye-movement experiment, as a window into the online processing of formulae. We focus on lexical bundles and posit the following hypotheses:

- Assuming holistic storage and processing of lexical bundles, we expect to find shorter overall fixation times in lexical bundles compared to non-formulaic control sequences.

- We also predict that contrary to native speakers this processing advantage is absent or at least less pronounced in non-native speakers.

Methods

Item Selection

In order to identify our test sequences we used the operational definition for lexical bundles (LB) by Biber and colleagues (1999), who define them as a combination of n-word bundles that recur and co-occur very frequently in a given register, and that can be identified empirically on the basis of frequency in a corpus. Additional requirements were imposed by our eye-tracking paradigm. First of all, we limited ourselves to four-word bundles. The goal was to obtain a number of within-region fixations, because shorter sequences such as 2 and 3-grams usually contain function words which are often skipped. Secondly, the high frequency of 2- and 3-grams made it undesirable to include them in the present study, because we wanted to reduce the data set to a manageable size. Thirdly, we did not adhere to the categorical frequency threshold criterion of at least 10 per million words (pmw) formulated by Biber and colleagues (1999), which allowed us to include four-word sequences which fall beyond this cut-off value. Recent research, for example, by Tremblay and Baayen (2010) or by Arnon and Snider (2010) suggests that phrasal frequency effects hold for less frequent multi-word units as well, and occur rather in a continuous than categorical manner. Thus, the cut-off values or frequency ranges remain questionable.

Additionally, we did not want to rely on frequency alone, because the identification of formulae based on statistical frequency only usually yields many grammatically incomplete units without an identifiable function such as the fact that the or the end of the. Thus, following Ellis and colleagues (2008), who argue for a combination of frequency and association measures for the identification of recurrent formulae, we calculated the mutual information score (MI-score) and the t-score for our possi-

ble test items. Both scores measure the collocational strength of a given word bundle, but provide us with different insights. The MI-score is calculated as the logarithm of the ratio between the relative frequency of a MWU within a corpus and the probability of observing the collocation if the words in the unit were distributed independently (which is given by the product of their individual frequencies). The MI-score assesses the strength of association between words in the formula, and thus favors less frequent but strong collocations, which often perform a distinct function (cf. Ellis et al., 2008). The t-score is instead calculated as the ratio between the difference between the observed and predicted frequency of a MWU and the square root of the observed frequency (Church, Gale, Hanks, & Hindle, 1991). Although both the MI-score and the t-score are meant to index the collocational strength of a MWU, in practice the t-score emphasizes frequent collocations, being to a large extent correlated with the frequency of the MWU (r=.97 for the formulaic sequences in our sample of formulaic sequences). The large correlation between t-score and the frequency of occurrence of the sequences also limits the possibility to consider them as separate determinants of the reading performance. The items we extracted for our study have an MI-score higher than 8 (mean 14.5) and a t-score higher than 5 (mean 13.2).[3] Finally, we agreed on further criteria which were required by the eye-tracking methodology. Whenever possible, lexical bundles had to start and to end with a word containing at least three letters. Mean log frequencies of the constituents of the formulaic sequences were matched to ensure that a possible processing advantage did not occur due to the frequency effects on the word level.

Having determined the requirements for the identification of four-word bundles, we automatically extracted a list with lexical bundles from a 29 million word subcorpus of the British National Corpus (BNC). Because there are considerable genre differences in the types of formulae, we restricted it to the newspaper section of the BNC, which represents a written and rather formal genre, but is not as specialized as an academic or business type of text. The program Collocate (Barlow, 2004) calculated the phrasal frequency and the MI-score. In a next step, we manually analyzed the extracted list of formula and chose 25 four word lexical bundles (with frequencies ranging from 1 pmw to 74 pmw; MI-score from 8 to 21), which were in line with our additional criteria. A list of descriptors of the items is reported in Table 1 (also see the

item list in Appendix 1). Nonetheless, all formulaic sequences constitute fully grammatical phrases, which in the majority of cases, function as referential expressions (from time to time), discourse organizers (for the time being) or stance expressions (there is no need). We then embedded each formulaic test sequence into the middle of a sentence (see Columbus, 2010; Tremblay et al., 2011 for lexical bundles processing effects in sentential contexts), and tried to control for the position of the lexical bundles in the sentence and the sentence length to ensure that faster processing is not triggered by these variables. We also controlled for the mean log frequencies of the exchanged words in the test and control sequences and balanced the length of each sequence. The exchange resulted in a compositional unit which constitutes a rare but not an artificial structure.[4] An example of a target sentence containing a lexical bundle is

(1) *Winning an Oscar has 'nothing to do with' having much talent,*

compared to a control sentence

(2) *Ben realized that Tom had seen 'nothing of the film' last night.*

Participants

Two groups of twenty participants took part in the experiment. One group consisted of native speakers of English, and the other group of non-native speakers of English.

Native speakers. The native speaker participants were tested at Queen's University in Kingston, Ontario (Canada). Eighteen out of the 20 participants were university students, 13 were female, the mean age was 21.7 years. Thirteen participants reported being able to speak a language other than English (including Cantonese, Chinese, French, Korean, Macedonian, Romanian, Gujarati, Spanish).[5]

Non-native speakers. The non-native speaker participants were tested at Justus Liebig University Giessen, (Germany). All of the participants were students at the university level, 15 were female, the mean age was 23.9 years. All the participants were native speakers of German, six participants reported being able to speak a language different from English and German (including Arabic, Chinese, French, Russian and Turkish).

Descriptor	Formulaic Sequences	Control Items
Frequency of sequence (PMW)	8.89 ± 16.46	.03 ± .06
Average frequency of words in unit (PMW)	18926 ± 7158	19912 ± 6366
MI-score	14.47 ± 3.35	
t-score	13.22 ± 9.25	
Unit length (chars)	66.84 ± 9.25	65.36 ± 7.65
Naturalness score (overall)	-.043 ± .315	.043 ± .33
Naturalness score (native speakers)	-.03 ± .447	.03 ± .44
Naturalness score (non-native speakers)	-.056 ± .519	.056 ± .589

Table 1

Descriptive statistics of the experimental items (mean ± standard deviation). 18 of the control sequences never occur in the corpus, so the MI- and t- scores are undefined. Notice that, apart from those derived from collocation frequency, the values descriptors that could be associated with a reading advantage (higher average frequency of words in unit, shorter unit length, higher naturalness) are observed in control items.

In order to exclude advanced speakers of English from our non-native speaker pool, we did not allow students of English and students who had spent more than one semester in an English-speaking country to participate.

Non-native speakers filled a questionnaire where they reported their final mark in English in high-school and gave an estimate of the weekly time they were exposed to English (e.g. listening, reading and talking).

Ethics Statement. Participants in the study provided written informed consent in agreement with the Declaration of Helsinki. Methods and procedures were approved by the local ethics committee LEK FB06 at Giessen University (proposal number 2009-0008).

Procedure

In both eye-tracking laboratories, 23-inch CRT monitors were used to show the stimuli. The items were presented in white letters on a black background. The sentences were presented in Arial font subtending 0.66° degrees of visual angle.

Each item stayed on the screen for a maximum of ten seconds but participants were encouraged to press a key to continue with the next item after they had finished reading. After each of the 8 practice items and after 12 out of the 50 experimental items, a question was presented, which required a yes-no response.

Eye-movement recording and data analysis

In both laboratories eye-movements were recorded monocularly at 500Hz using Eyelink II eye trackers (SR

Research, Missisauga, ON, Canada). Fixations were identified using the online Eyelink parser.

Given that our items were defined at the four-gram level, we limited our analyses to Region of Interest (ROI) eye-movement parameters. We computed two measures describing participants´reading behavior: first-pass reading time and total reading time (Siyanova-Chanturia, Conklin, & van Heuven, 2011).[6]

Data were analyzed by means of linear mixed models, implemented using the R package lme4 (Bates & Maechler, 2011). Subjects and items were introduced as random effects. Various predictors were tested, including Condition (CND: formulaic sequence vs. control) , Group (GROUP: native speaker vs. non-native speaker), average log-transformed lexical frequency of words in ROI (AVLF) and ROI length (ROIL). Furthermore, for each item we calculated the residual MI- and t-scores (RMIS and RTS) after subtracting the effect of Condition (computing the residuals was not necessary for average log-transformed lexical frequency and ROI length because those predictors were balanced while constructing the items). Finally, limited to the non-native speakers we considered the average weekly exposure to English (EXP) and the School Grade (SGD)[7] as predictors. All three dependant variables were log-transformed and all continuous predictors were centered before the analysis.

The models were selected in a forward fashion starting with a minimal model including only the random ef-fects and subsequently adding each simple effect and interaction and keeping it in the model if a likelihood test indicated an improved (p<.05) fit. Where appropriate, the final model was further refined by removing the predictors whose coefficient did not significantly differ from 0 (p>.05 computed from 10,000 samples), which in no case introduced a significant decrease in the model fit.

Separate models were fitted on two dependent variables (first-pass reading time, total reading time). The two pairs of predictors Residual MI-score and Residual t-score and Exposure and School Grade were not independent in principle, so separate models were fitted to the three dependent variables including only one of the predictors at a time. In no case Residual MI-score and Residual t-score were significant at the same time so only the model with the significant effect is reported but we report separate analyses for Exposure and School Grade (only applied to the data from the non-native speaker group as for the native speakers the predictors are undefined). We also applied a model to the data of the native speakers excluding Exposure and School Grade.

For each model we will present the estimate of each predictor´s slope, together with the mean and upper and lower bounds of the corresponding 95% Highest Posterior Density interval derived from 10,000 Markov Chain Monte Carlo samples (MCMC mean, HPD95lower and HPD95upper) and the associated p value (pMCMC).

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.4775	6.477	6.3767	6.5667	0.0001
SGD : RMIS	0.0121	0.012	0.0019	0.0212	0.0148

Table 2
Model of First-pass reading time in non-native speakers with School Grade (SGD) and Residual MI-score (RMIS). Residual t-score did not contribute to the model fit when tested instead of Residual MI-score.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.4775	6.4772	6.3891	6.5582	0.0001
EXP	-0.0324	-0.0326	-0.0557	-0.0086	0.0078
AVLF	-1.1170	-1.1080	-2.0358	-0.1496	0.0240

Table 3

Model of First-pass reading time in non-native speakers with Exposure (EXP). Reading time decreased both as a function of Exposure and Average Lexical Frequency (AVLF). None of the main effects or interactions of residual MI-score and residual t-score were significant.

Results

Response accuracy

On average the accuracy of the responses to the questions was 82.5% for native speakers (NS) and 75.8% for non-native speakers (NNS). This difference was not significant ($t(38)=1.476$, $p=.148$).

First-pass reading time

Items which had a high MI-score relative to their condition (formulaic or control sequences) imposed a particularly strong cost in terms of first-pass reading time to non-native speakers with a lower proficiency as indicated by the high value of the grade. More expert non-native speakers were unaffected by the Residual MI-score (Table 2).

First-pass reading time decreased as a function of both Exposure and Frequency, indicating that the non-native speakers who were more exposed to English read the items faster and that items containing high-frequency words were read faster (Table 3).

None of the fixed factors had an effect on first-pass reading time in native speakers (no model presented).

The first-pass reading time in both groups as a func-

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.1642	6.165	6.0699	6.2560	0.0001
GROUP	0.3133	0.313	0.1885	0.4331	0.0001

Table 4

Model of First-pass reading time in both groups Native Speakers read the items faster. Neither the main effect nor the interactions of Condition were significant.
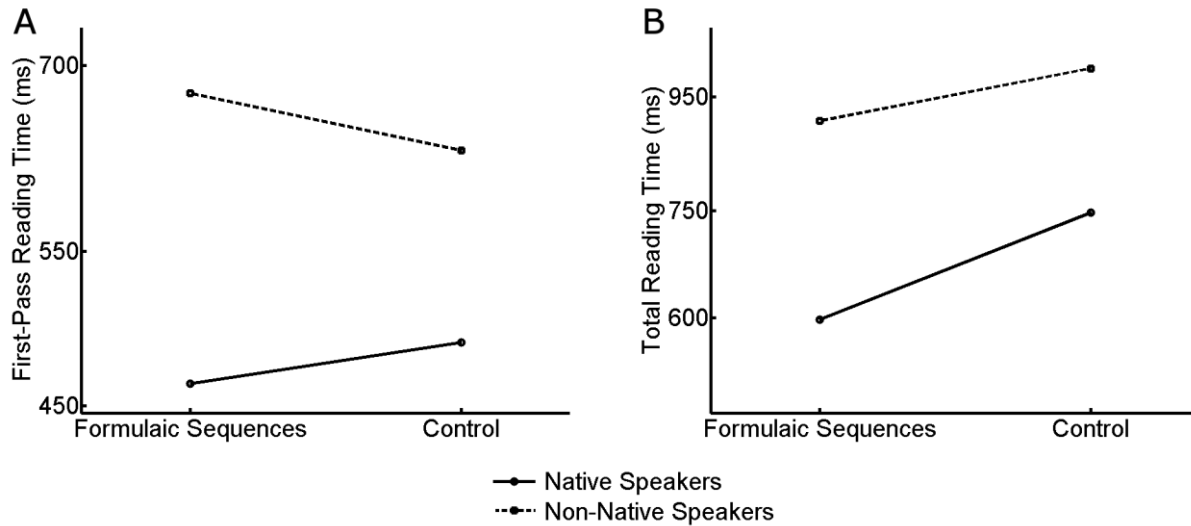
Figure 1. First-Pass reading time (Panel A) and Total Reading Time (Panel B) in both Groups as a function of Condition. Both First-Pass and Total reading times was larger in non-native speakers. Only native speakers showed an advantage in terms of Total Reading Time when reading Formulaic Sequences.

tion of Group and Condition is depicted in Figure 1a. Native speakers read the sequences much faster and they tended to do so particularly in the case of formulaic sequences (the difference between the conditions is larger by 88 ms in native speakers as compared to non-native speakers), although this interaction did not turn out sig-

nificant (see Table 4), indicating that our analysis of first-pass reading times was quite underpowered.

Total reading time

Total Reading time in non-native speakers decreased as a function of average lexical frequency, increased as a

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.8606	6.8608	6.7646	6.9609	0.0001
ROIL	0.0029	0.0029	0.0011	0.0048	0.0026
EXP	-0.0552	-0.0550	-0.0819	-0.0276	0.0006
AVLF	-2.3082	-2.3159	-3.3527	-1.2823	0.0001

Table 5

Model of Total Reading Time in Non Experts with Exposure (EXP). Reading time increased as a function of ROI length (ROIL) and was smaller in items with a higher Average Lexical Frequency (AVLF). None of the main effects or interactions of Residual MI-score and Residual t-score were significant.

function of the ROI length and decreased as a function of the observers´exposure to English (Table 5).

The analysis of Total Reading time was evidently more powerful than the analysis of first-pass reading time. This was revealed in the fact that the effect of Condition as well as the effect of average lexical frequency were significant in native speakers (Table 6). Furthermore, when we applied our analysis to the data of both groups, we were able to show significant main effects of Group, Condition and, crucially, a significant interaction of Group and condition (Table 7). Native speakers took less time in total to read the items as compared to non-native speakers, and formulaic sequences were read faster than control sequences. Furthermore, the advantage for reading formulaic sequences increased from 297 ms in non-native speakers to 357 ms in native speakers (Figure 1b).

Furthermore, the analysis of Total Reading Time indicated that observers took less time to read shorter items and items with a larger average lexical frequency. In both cases the effects are more pronounced in non-native speakers as compared to native speakers (Table 7).

Discussion

In this study, we investigated the processing of 4-grams by native speakers and non-native speakers by measuring first pass reading time and total reading time in an eye-tracking experiment. The results confirmed our hypotheses and allow for several observations. First of all, native speakers read 4-grams faster than the matched control sequences, a finding which is consistent with previous research into the processing of formulae by native speakers (e.g. Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011). Given that the sub-string frequency was matched and other factors such as sentence length and sentential position were controlled for, the reading advantage we observed can only be due to the sequence-level co-occurrence of the words.

Different explanations can be suggested for the reading advantage of formulaic sequences. On one side it is possible that particularly experienced and native speakers process formulaic sequences in a holistic way (Kjellmer, 1990; Pawley & Syder, 1983; Wray, 2002). On the other side it is quite possible that this effect is mediated by the transitional probability of the words in a co-occurring sequence (e.g. McDonald & Shillcock, 2003), and in general it is clear that the co-occurrence of words in a sentence can help generate predictions about the future word input (Elman, 1990). This means that it is quite likely that the predictability of the words constituting formulaic sequences is higher, expecially for expert speakers. Word predicatability is a key determiner of reading time in many models of oculomotor control in reading (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998)

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.393	6.3928	6.2802	6.4994	0.0001
AVlogFR	-1.107	-1.1127	-2.0760	-0.0934	0.0320
CND	0.222	0.2226	0.1286	0.3164	0.0001

Table 6

Model of Total Reading Time in Native Speakers. Reading time was faster for Formulaic Sequences and decreased as a function of Average Lexical Frequency (AVLF) None of the effects of Exposure or School Grade were significant. None of the main effects or interactions of Residual MI-score and Residual t-score were significant.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.3952	6.3939	6.2850	6.5069	0.0001
AVLF	-1.0196	-1.0153	-1.8745	-0.1036	0.0248
GROUP	0.4134	0.4143	0.2736	0.5537	0.0001
CND	0.2186	0.2194	0.1328	0.3004	0.0001
GROUP:ROIL	0.0031	0.0031	0.0015	0.0048	0.0010
GROUP:AVLF	-1.4109	-1.4096	-2.1667	-0.6882	0.0004
CND:RTS	-0.0787	-0.0785	-0.1487	-0.0099	0.0272
GROUP:CND	-0.1142	-0.1146	-0.1813	-0.0466	0.0008

Table 7

Model of Total Reading Time in both groups. Residual MI-score did not contribute to the model fit when tested instead of Residual t-score.

The role of frequency effects in processing merits further attention. Although we included only 8 items with a frequency of at least 10 pmw, we were still able to find a processing advantage for 4-grams in our analysis. The results of our study are consistent with those by Siyanova-Chanturia and colleagues (2011), who were able to find a processing advantage for binomials of different frequency ranges. It should be mentioned that the majority of extracted formulae are classified as high MI-grams and also have a t-score of 2 and higher. Like the study by Ellis and colleagues (2008), this suggests that the processing advantage for native speakers might be attributed not only to sequence frequency, but to other factors such as collocational strength or fixedness, pragmatic function or even to a combination of frequency and these factors. This assumption, however, would go against usage-based models (e.g Bybee, 2006; Ellis, 2002; Tomasello, 2003) in which frequency is the main predictor for facilitated processing of language.

Another possible explanation for this observation could be the definition of high frequency along with the identification of formulae by corpus analysis. On the one hand, Tremblay and colleagues (2011) point out that it remains unclear whether the BNC or other corpora are able to provide us with real frequency values. Single corpora might, for instance, underestimate or overestimate frequencies (as is the case for *take a look at* in this study). Corpus-derived frequency values simply might not represent the degree of familiarity that speakers have with certain chunks (cf. Gries, 2010). Furthermore, recent research suggests that the frequency cutoff of 10 pmw or any other frequency should be taken with caution because frequency effects might appear rather in a continuous way (cf. Arnon & Snider, 2010).

When looking at the processing of 4-grams and control sequences by non-native speakers, we did not find any processing advantages for 4-grams in the non-native speaker group. This result is in contrast to the previous study by Jiang and Nekrasova (2007), who were able to

show a processing advantage for both native and non-native speakers. We did not find any differences between the total reading time for non-native speakers which were shown by Schmitt and Underwood (2004). In addition, we found that formulae with a high t-score which tend to be overused by learners in text production (cf. Durrant & Schmitt, 2009a) do not show a processing advantage for non-native speakers in our study.

One first difference between the present study and the previous ones may lay in the sequences which were used. Most of the sequences used by Schmitt and Underwood (2004) were in fact idiomatic expressions, but if anything one would expect non-native speakers to have larger difficulties with idiomatic formulas, which cannot be interpreted without a specific knowledge. Jiang and Nekrasova (2007) on the other hand used non-idiomatic sequences very similar to the ones we used (although they mostly tested tri-grams). We would rather suggest that the discrepancy with the previous reports is due to the characteristics of our non-native speakers. The previous studies had tested rather advanced non-native speakers of English in an English-speaking environment. In our study, we limited the non-native speaking sample to less proficient students of English in order to clearly contrast prototypical native speakers with prototypical foreign-language learners of English. The different results may thus be explained by the participant selection. Indeed, Siyanova-Chanturia and colleagues (2011) tested non-native participants with a range of proficiency levels and found that low-proficiency individuals tended not to show any advantage when reading formulaic sequences. Contrary to what Siyanova-Chanturia and colleagues found (2011), our analyses do not provide evidence that a processing advantage for formulaic sequences would appear in the more proficient non-native speakers of English, or in those who were more exposed to the language. We can speculate that a level of exposure produced by leaving in a completely English-speaking environment has to be achieved before the processing of formulaic sequences is facilitated. These results could indicate that the mastery of formulaic sequences comes only later in the acquisition process of a foreign language. Simultaneously, it is also possible to argue that learners' performance is influenced by the frequency of our test items, because learners might be more sensitive to frequency than to other factors such as t-score and MI-score (cf. Ellis et al., 2008). Finally, our results reveal high variability in the non-native speaker group in the case of for-

mulaic language, which suggests that individual differences and other factors, e. g. exposure to and teaching of formulaic language in the classroom, have to be taken into consideration as well.

References

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*, 67-82.

Barlow, M. (2004). *Collocate 1.0: Locating collocations and terminology*. . Houston: Athelstan.

Bates, D., & Maechler, M. (2011). lme4: Linear mixed modeling using S4 classes (Computer program and manual):http://cran.r-project.org/web/packages/lme4/index.html.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at... : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371-405.

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

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711-733.

Church, K. W., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon* (pp. 115-164). Hillsdale: Lawrence Erlbaum.

Columbus, G. (2010). Processing MWUs: Are MWU subtypes psycholinguistically real? In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 194-212). London/New York: Continuum.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers. *Applied Linguistics, 29*, 72-89.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures*225-246.

Durrant, P., & Schmitt, N. (2009a). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics 47*, 157-177.

Durrant, P., & Schmitt, N. (2009b). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics, 47*, 157-177.

Ellis, N. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 297–339.

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language ppeakers: Psycholinguistics, corpus Linguistics, and TESOL. *TESOL Quarterly, 42*, 375-396.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*, 179-211.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777-813.

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

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.), *Research pedagogic tasks: Second language learning, teaching and testing* (pp. 75-93). Harlow: Longman.

Gilquin, G. (2007). To err is not all: What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik, 55*, 273-291.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and application.* (pp. 145-160). Oxford: Oxford University Press.

Gries, S. (2010). Disperions and adjusted freqeuncies in corpora: further explorations. . In S. Gries, S. Wulff & M. Davies (Eds.), *Corpus linguistic applications: current studies, new directions.* (pp. 197-212). Amsterdam: Rodopi.

Hunston, S. (2002). *Corpora in Applied Linguitics*. Cambridge: Cambridge University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal, 91*, 433-445.

Kjellmer, G. (1990). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik.* (pp. 111–127). London: Longman.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735-1751.

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications.* (pp. 79-100). Oxford: Clarendon.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching. .* Oxford: Oxford University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*, 223–242.

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

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

Schmitt, N. (2005). Formulaic language: Fixed and varied. . *Estudios de Linguistica Inglesa Aplicada, 6*, 13-39.

Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Philadelphia: John Benjamins.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a selfpaced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 173–189). Philadelphia: John Benjamins.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *Tesol Quarterly, 37*, 419-441.

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

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27,* 251-272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (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,* 776-784.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge: Harvard University Press.

Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. . In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). London/New York: Continuum.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning, 61,* 569-613.

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

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Footnotes

[1] For a detailed discussion see Wray (2002).

[2] Note that the authors do not define frequency and provide only a total frequency number for binomials (274.3 pmw) and their reversed forms (27.4 pmw). The binomials and their reversed forms range in frequency and are neither grouped into frequency ranges nor are any binomials excluded on the basis of a frequency cut-off (e.g. 10 pmw by Biber and colleagues (1999). The study, however, shows clear phrasal frequency effects for frequent binomials over less frequent ones and over their reversed forms which might render categorical thresholds untenable (see Section 2 and 4).

[3] Different critical values have been suggested for the MI-score and the t-score. Hunston (2002), for instance, deems an MI-score of 3 and higher and a t-score of 2 and higher as significant. Others, as Durrant and Schmitt (2009b), and Ellis and colleagues (2008) prefer to classify the items in their studies according to a range of bands. The latter study, for instance, classifies different n-grams into Low MI (3.), Medium MI (6.7) and High MI (11).

[4] The naturalness of the items was assessed by four native speakers of English and four non-native speakers. The normalized scores did not differ between the items containing formulaic sequences and control sequences. this was the case also when both groups of speakers were considered separately.

[5] To ensure that our results would not be biased by the fact that we presented items from a British corpus to speakers of American English, we conducted a frequency analysis for the same test items in the Corpus of Contemporary American English (COCA, containing 425 million words). The item frequencies did not deviate significantly from the frequency values in the sub corpus of the BNC apart from one item only (take a look at). Additionally, Tremblay and colleagues (in press) show that such biases appear to be very unlikely.

[6] For the sake of brevity, we do not report about the fixation count. The pattern of results were in any case completely overlapping with the one we show for total reading time.

[7] Notice that the grades in the German school system vary between 4, i.e. sufficient, and 1, i.e. very good.

Appendix 1: Items

Items marked by a) contain formulaic sequences, whereas those marked by b) contain control sequences. The four-grams are indicated by italics, formulaic sequences are indicated in red.

1a) The movie prize was awarded *for the first time* in Berlin.

1b) Tom was never paid *for the last work* he did for Ben.

2a) Johnny Depp is *one of the most* charismatic actors in the world.

2b) Barbecuing in the summer is *one of the good* things in life.

3a) The film script was written *with the help of* a theatre expert from London.

3b) Ben left the pizza in the fridge along *with the half of* the bagel he couldn't finish.

4a) Winning an Oscar has *nothing to do with* having much talent.

4b) Ben realized that Tom had seen *nothing of the film* last night.

5a) Alice checked her watch *from time to time* waiting for the babysitter.

5b) People were gathering *from up and down* the country.

6a) Federer announced he would be retiring at the *end of the season* in June.

6b) In 1989 Gorbachev acknowledged the *end of the Soviet* empire.

7a) In France marches were planned *all over the country* to demonstrate against war.

7b) Henry hasn't complained at *all about the service* of the new Hilton hotel.

8a) John could not *come to terms with* his wife's death.

8b) The circus has *come to town with* a new artist show.

9a) They expect that by the *turn of the century* the debts will have grown massively.

9b) This morning it was the *turn of the opposition* to make their announcement.

10a) Henry traveled to different *parts of the world* during his studies.

10b) In the end Tom deleted two *parts of the work* that he had already edited.

11a) Lately, there has been an *increase in the number* of people seeking employment.

11b) Recent statistics show an *increase in the market* demand for firearms.

12a) Jane lives in constant pain, but *most of the time* she tries to ignore it.

12b) Job security is what matters *most to the people* in our society.

13a) Jane has been in hospital *since the beginning of* last December.

13b) Tom didn't understand a word *since the language* of instruction was Hebrew.

14a) Mr. Lumbergh says that *there is no need* to work next Saturday.

14b) Ben soon realized that *there was no work* left for his secretary.

15a) In 1985 Poland was still *under the control of* the Soviet Republic.

15b) In 1888 the Eiffel Tower was still *under the course of* construction.

16a) New Orleans won the Super Bowl *with the support of* their loyal fans.

16b) The gangsters escaped *with the money to* the getaway car.

17a) Mary decided to *take a look at* the new science museum.

17b) Jane said it would *take a month at* most to move to Australia.

18a) The eggs were splattered *all over the place* in the kitchen.

18b) The fuss was *all about the part* Mary would play in the new movie.

19a) Tomorrow night Jane will *have the opportunity to* win a cruise.

19b) Tomorrow Henry will *have the evening to* relax and watch TV.

20a) Ben had to *bear in mind that* Jane left him no other choice.

20b) The US soldiers had to *bear the burden of* fighting a pointless war.

21a) People can access the Internet *anywhere in the world* nowadays.

21b) Sarcasm does not appear *anywhere in the work* of Virginia Woolf.

22a) The software had been developed *within the framework of* the course.

22b) Dave was not willing to work *within the restrictions of* the law.

23a) Yesterday Dave took a picture *from the top of* the Empire State Building.

23b) Ben could always distinguish the man of talent *from the man of* genius.

24a) Mary's mid-teen years were made bearable *with the presence of* Dave.

24b) Dave waited for Jane at the cafe *with the present in* his pocket.

25a) Jane was faced *with the problem of* raising a child on her own.

25b) Dave argued *with the manager of* the new nightclub yesterday.