Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 3, 776–784

© 2011 American Psychological Association
0278-7393/11/$12.00 DOI: 10.1037/a0022531

RESEARCH REPORT

Seeing a Phrase "Time and Again" Matters: The Role of Phrasal Frequency in the Processing of Multiword Sequences

Anna Siyanova-Chanturia
University of Modena and Reggio Emilia and
University of Nottingham

Kathy Conklin and Walter J. B. van Heuven
University of Nottingham

Are speakers sensitive to the frequency with which phrases occur in language? The authors report an eye-tracking study that investigates this by examining the processing of multiword sequences that differ in phrasal frequency by native and proficient nonnative English speakers. Participants read sentences containing 3-word binomial phrases (*bride and groom*) and their reversed forms (*groom and bride*), which are identical in syntax and meaning but that differ in phrasal frequency. Mixed-effects modeling revealed that native speakers and nonnative speakers, across a range of proficiencies, are sensitive to the frequency with which phrases occur in English. Results also indicate that native speakers and higher proficiency nonnatives are sensitive to whether a phrase occurs in a particular configuration (binomial vs. reversed) in English, highlighting the contribution of entrenchment of a particular phrase in memory.

Keywords: multiword sequences, phrasal frequency, mental lexicon, eye-tracking, bilinguals

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

It is well established that the frequency with which words occur in a language influences how quickly they are recognized (Balota & Chumbley, 1984; Monsell, Doyle, & Haggard, 1989; Rayner & Duffy, 1986). A view that is gaining popularity is that, in addition to this word frequency effect, language users are sensitive to frequency information at the sublexical, phrasal, and clausal levels. This means that the frequency of morphemes, syllables, words, multiword phrases, and clauses may all influence processing.

Although frequency effects have been widely reported in the word-processing literature, only a few studies have investigated frequency effects for units larger than a word, such as two-word combinations (Bell et al., 2003; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Mondini, Jarema, Luzzatti, Burani, & Semenza, 2002; Sosa & MacFarlane, 2002) and larger syntactic structures (Arnon & Snider, 2010; Frazier & Fodor, 1978; Reali & Christiansen, 2007). Although the evidence is somewhat limited, it has been used to support the view that the frequency with which multiword sequences occur affects their representation and processing. For example, Sosa and MacFarlane (2002) had participants monitor for *of* in two-word collocations (*sort of*) varying in frequency. They found that reaction times to *of* in high-frequency phrases were significantly slower than in low-frequency ones, indicating that frequent phrases were treated as unitary entities, which hindered access to their individual components. Although *of* was identical across all conditions, lexical properties of the constituent words were not controlled for. Thus, one must be cautious about drawing strong conclusions from this study.

Mondini et al. (2002) investigated the production of two-word compounds (*natura morta* "still life") and novel combinations (*natura bella* "beautiful nature") by two aphasic patients. Mondini and colleagues found that their participants performed better on compounds than on novel noun–adjective combinations. This was taken to indicate that for novel phrases, participants retrieved the adjective and noun separately and then applied agreement rules. Compounds, however, were retrieved as a unit, and, therefore, no morphosyntactic operations, such as number and gender agreement, were necessary. Because the study only investigated two brain-damaged participants, it is difficult to draw any far-reaching conclusions.

A number of studies have looked at the processing of compounds and their constituents using an eye-tracking paradigm (Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009). Such research has found evidence for parallel access to full forms (*blackboard*) and their constituents (*black* and *board*). Crucially, the effect of compound frequency was observed as early as the first fixation, which suggests that the more entrenched the full form of a compound is, the earlier the processing benefits for it appears (Kuperman et al., 2008).

Although a large number of studies have investigated the processing of one- and two-word compounds, only a few studies have examined the processing of phrases or sequences longer than two

This article was published Online First February 28, 2011.

Anna Siyanova-Chanturia, Department of Biomedical Sciences, University of Modena and Reggio Emilia, Modena, Italy; Kathy Conklin, School of English Studies, University of Nottingham, Nottingham, England; Walter J. B. van Heuven, School of Psychology, University of Nottingham.

This research was supported by Economic and Social Research Council Grant PTA-030-2006-00151 awarded to Anna Siyanova-Chanturia. We thank Harald Baayen and Denis Drieghe for their helpful comments on previous versions of this paper. We are also grateful to Harald Baayen for his invaluable help with the mixed-effects modeling using R.

Correspondence concerning this article should be addressed to Anna Siyanova-Chanturia, Department of Biomedical Sciences, University of Modena and Reggio Emilia, Modena, Italy. E-mail: anna.siyanova@unimore.it

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

words. Arnon and Snider (2010) investigated the role of phrasal frequency in the comprehension of four-word phrases (*don't have to worry*) and found that frequent phrases were processed reliably faster than infrequent ones, even though the frequency of constituent words was matched. Thus, the authors concluded that language users notice, learn, and store frequency information not only about words but also about compositional phrases. Notably, in the study, participants performed a phrasal decision task, in which they decided as quickly as possible whether the target phrases were possible in English. The authors argue that this is comparable to a lexical decision task. However, this is not necessarily the same as making a lexical decision and is not a well-established task in the literature. Furthermore, in the study, low- and high-frequency phrases had different semantics and different expectancies about upcoming information, and some of their phrases could be used in isolation (*We have to talk*), whereas others could not be (*We have to say*). Therefore, the conclusions drawn from their results must also be viewed with caution.

Tremblay and Baayen (2010) used behavioral and electrophysiological measures to investigate the processing of four-word sequences (*in the middle of*). They found that the probability of occurrence of the four-word sequences, and constituent and trigram frequency, affected participants' recall (in the behaviural task) and their event-related potentials (in the electroencephalograph experiment). The key finding of the study was that already in a very early time window (110–150 poststimulus), there was a frequency effect for the four-word sequence. These early effects were discussed in terms of the P1, N1, and P2, which are usually associated with attentional processes (although the P2 has also been associated with sentential constraint and expectancy of a given word). Thus, although these results were taken to suggest that multiword forms are stored both as parts and wholes, it is unclear whether they support such a view or are indicative of more general attentional processing.

In a self-paced reading study, Tremblay, Derwing, Libben, and Westbury (in press) compared the processing of sentences containing lexical bundles (*don't worry about it*) and matched control phrases. They found that sentences containing lexical bundles were read faster than control sentences, and were more likely to be remembered and recalled correctly than sentences with novel phrases, suggesting that the more frequent a unit is, the more likely it is to leave memory traces in the brain.

Finally, a number of studies revealed robust differences in the production of frequent multiword sequences versus novel ones. Van Lancker, Canter, and Terbeek (1981) found that instances of novel language had a longer duration because they contained more and longer pauses and that their constituents were spoken more slowly. Similarly, Bybee and Scheibman (1999) found that *don't* was phonetically reduced when it was part of a frequent phrase (*I don't know*). Likewise, Bell et al. (2003) found that words were phonetically reduced when they were predictable in a given context (*middle of the*).

Although the evidence is somewhat incomplete, the above findings suggest that frequent multiword sequences may be processed differently from less frequent ones by native speakers. However, no study has addressed the issue of phrasal frequency with respect to second-language speakers. If frequency influences whether a multiword sequence is represented or not, one might expect that with increased exposure, second-language learners will have not only single words represented in their lexicon but also instances of frequent multiword sequences. If it is the case that frequency of exposure determines what is represented in the mental lexicon, we would expect native speakers, who have accumulated a sufficient amount of experience with frequent expressions, to show a robust processing advantage for them. Crucially, the inclusion of nonnative participants in our study allowed us to investigate the relationship between exposure to a phrase and phrasal frequency. Thus, the main question that the present study aimed to answer was: Are native and proficient nonnative speakers sensitive to phrasal frequency during online language comprehension?

To investigate phrasal representation and processing by native and proficient nonnative speakers, we used three-word phrases, called *binomial expressions*. Binomials are phrases formed by two content words from the same lexical class connected by a conjunction, where one word order is more frequent than the other (*bride and groom/groom and bride*). For the purposes of the present study, we define *binomials* as recurrent (frequent), familiar (conventional) expressions. Thus, we do not consider infrequent combinations with no word-order preference (*green and yellow*) to be binomial expressions. Binomials are ideal for studying phrasal comprehension for a number of reasons. First, they are more frequent than idioms, which are considered to be the prototypical example of multiword phrases. Second, unlike idioms, binomials' components contribute overtly to the overall meaning of the expression.[1] Although readers cannot compute the meaning of the idiom *ring a bell* ("sound familiar"), they can compute the meaning of the binomial *bride and groom*. Finally, in more idiosyncratic expressions, such as idioms, changes are rarely permitted (*kick the bucket* vs. **the bucket was kicked*). Because in the majority of binomial expressions the word order can be reversed without any meaning change,[2] we were able to investigate whether such expressions have a processing advantage over reversed forms, which only differ in phrasal frequency.

In the present study, we used eye tracking to investigate the processing of binomial expressions by native and proficient nonnative English speakers. If frequency of exposure plays an important role in what is represented in the lexicon, we would expect native English speakers, who have accumulated a sufficient amount of experience with frequent expressions, to show a processing advantage for binomials over their reversed forms. Nonnative speakers, whose exposure to English will not have been as rich, may not show such an advantage. Specifically, we may observe higher proficiency nonnatives performing similarly to native speakers, whereas those having a lower proficiency may differ from native speakers in their processing of binomials versus

[1] The majority of binomials are regular expressions that are used literally. However, some binomials can be used both literally and figuratively. In our set of stimuli, there were four items that had an additional figurative interpretation (*bread and butter*, *day and night*, *black and white*, *cat and mouse*). The inclusion or exclusion of these items did not change the pattern of results.

[2] In some binomials, where the order of events plays a role, the meaning may change if the expression is reversed (*trial and error*, *cause and effect*). Analyses were done with and without these two items, and the pattern did not change.

their reversed forms. Nonnative speakers thus allowed us to investigate the role of frequency of exposure more explicitly.

Method

Participants

Twenty-eight native and 28 proficient nonnative English speakers took part in the study. All participants were students at the University of Nottingham, United Kingdom (mean age = 21.1 years). They received course credit or £ 6 (about U. S. $9.27) for their participation. The nonnative speakers came from various first-language backgrounds. On average, they had spent 21 months in the United Kingdom (ranging from 2 months to 7 years), and their first contact with English was at the age of 7.8 years. Their self-rated proficiency for speaking, reading, writing, and listening comprehension on a 5-point Likert scale (ranging from 1 [*very poor*] to 5 [*excellent*]) was 3.9, 4.2, 3.8, and 4.1, respectively.

Materials

The British National Corpus (BNC) was used to find a set of 42 binomial expressions and their reversed forms (see the Appendix). By definition, binomials and their reversed forms are matched in frequency of the individual words, length, and part of speech. Crucially, they differ in phrasal frequency: 247.3 occurrences in the BNC (per 100 million words) for binomials, and 27.4 occurrences for the reversed forms. Two types of fillers were also selected. The first set contained 42 meaningful and grammatically correct phrases that were matched with the binomials and their reversed forms in word length and part of speech (*fluid and fumes*). The second group of fillers was composed of 63 low-frequency meaningful and grammatically correct phrases (*tennis and badminton*). These were not matched with the binomials and reversed forms on any of the above properties. The syntactic structure of both filler types was identical to that of binomials and reversed forms (*X and X*). The two groups of fillers were used in order to prevent participants from noticing the binomials and, in particular, their reversed forms, which might be marked due to their low frequency.

To ensure that any processing advantage for binomials over their reversed forms was not due to the first word (*bride*) serving as a better prime for the third word (*groom*) than the other way around (*groom* for *bride*), the Edinburgh Associative Thesaurus database (http://www.eat.rl.ac.uk/) was used to check that binomials and their reversed forms were matched in semantic association strength as closely as possible. The mean strength of the forward association was 0.29, whereas the backward association was 0.25, which was not significantly different, $t(37) = 0.73$, $p = .47$. Finally, two groups of 10 native English speakers, who had not participated in the eye-tracking experiment, were asked to provide a completion for "Word 1 + and" (*bride and*) or "Word 2 + and" (*groom and*). If seven out of 10 participants were able to provide the "correct" completion of a phrase, then it was given a score of 7 on the completion test. The mean score for binomials was 6.9, whereas it was 4.7 for their reversed forms.

Procedure

Binomials and their reversed forms were presented across two presentation lists. Thus, no participant saw both versions of the same phrase. In each list, experimental items were intermixed with 21 fillers of the first type and all 63 fillers of the second type. Binomials and their reversed forms were embedded in identical sentence contexts (*John showed me pictures of the bride and groom*/*groom and bride both dressed in blue*).

Eye movements were recorded using a SMI EyeLink I (SR Research Ltd., Mississauga, Ontario, Canada). Participants were given a verbal explanation of the procedure. A 9-point grid calibration procedure was done before the experiment. Participants first completed a practice session. Each trial started with a fixation point that appeared in the middle of the screen. After participants fixated it and a calibration check was conducted, a sentence appeared across one line in the middle of the screen, which participants were instructed to read as quickly as possible for comprehension. One quarter of the sentences were followed by a comprehension question. The rest were followed by "Ready?" After the experiment, nonnative participants completed a language background questionnaire, assessing their self-reported English speaking, reading, writing, and comprehension.

Results

We performed the analyses on 30 binomials and their reversed forms (60 items in total).[3] Single fixation durations shorter than 100 ms and longer than 800 ms were excluded, because short fixations reflect oculomotor programming, and fixations longer than 800 ms are due to momentary track loss or blinks (Morrison, 1984). The missing data accounted for 2.6% of the total data for nonnative speakers and 1.2% for native speakers. Because the items are multiword sequences that are longer than single words (and hence may be prone to having a bimodal distribution), we excluded cumulative fixation durations shorter than 200 ms and longer than 1,500 ms (per phrase). This resulted in the loss of 12.4% of the data for total reading time and 16.2% of the data for

[3] For the present study, we selected 42 binomials and their 42 reversed forms (84 items in total). However, four of the binomials had an additional idiomatic interpretation, two binomials had a different meaning in the reversed condition, two binomials and their reversed forms did not occur in the Edinburgh Associative Thesaurus, due to experimental error two binomials and their reversed forms were not included in the completion test, and *silver and gold* (reversed form) is famous as a Christmas song from the animated movie "Rudolf the Red Nosed Reindeer." Additionally, because semantics and the idea of salience has been used to account for word order in binomials (Benor & Levy, 2006), *knife and fork* was excluded. In the case of the binomial *knife and fork*, one might assume that knife precedes fork because we hold it in the right hand, which is the dominant hand for most humans. However, one might also argue that it is possible to eat without a knife but not without a fork (and, in fact, many people do exactly that), which should make a fork a more central or salient entity than a knife. For the above reasons, 12 binominals and their reversed forms (24 items) were excluded from the analyses reported in the present article, leaving us with 60 items in total (30 binomials and 30 reversed). However, when these 24 items are included, the pattern of results remains exactly the same for the total reading time. For the first-pass reading times and the fixation count, the interaction between phrase type and proficiency became a trend. All other fixed effects in the final models remained significant.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

first-pass reading time.[4] For the fixation count measurement, we also excluded fixation counts of 10 or more (one datapoint). Means for each of the eye-tracking measurements for binomials and reversed forms for nonnative and native speakers are presented in Table 1. The participants had no difficulty answering the comprehension questions (natives 94.5% correct, nonnatives 89.9% correct).

We used mixed-effects modeling with crossed random-effect factors for subjects and items (Baayen, Davidson, & Bates, 2008) to analyse the three eye-tracking measures (first-pass reading times, total reading times, and fixation count). We conducted the analyses with R version 2.11.1 (R Development Core Team, 2010) and the R package lme4 (Bates & Maechler, 2010). We log transformed the dependent variables (total reading times, first-pass reading times, and fixation count) to reduce the skewness in the distributions. The following predictors were considered: the phrasal frequency, the frequency of Content Word 1 as an isolated word, the frequency of Content Word 2 as an isolated word. The frequencies were obtained from the BNC (counts based on occurrences per 100 million words) and were log transformed. The next predictor was phrase type (binomial vs. reversed). We also considered phrase length (number of letters in the phrase), the association strength (forward and backward association between Content Words 1 and 2 based on the Edinburgh Associative Thesaurus database), and the score on the completion test. Proficiency was also considered as a predictor. Because a dichotomous proficiency predictor (native vs. nonnative speakers) leads to reduction of power (Baayen, 2010; MacCallum, Zhang, Preacher, & Rucker, 2002), we used a continuous proficiency variable that was based, for the nonnatives, on their subjective proficiency ratings (the average rating on a 5-point scale for reading, speaking, writing, and listening). For native speakers, the proficiency ratings were assumed to be at ceiling (and were given the maximum score of 5). The trial number of the presentation of the phrase in the experiment was considered as a predictor to account for the longitudinal effect of the experimental task on the behavior of the participants.

Participants and items were random-effect factors in the models. In order to address the issue of the collinearity between the predictors, we orthogonalized phrasal frequency by fitting a linear model in which phrasal frequency was predicted by phrase type. The residuals of this model (ResidPhrasalFrequency) were then used as our predictor of phrasal frequency (effects of phrase type are thus partialed out). These residuals correlated with phrasal frequency ($r = .67$ $p < .0001$). The same was done with the completion test predictor that correlated significantly with phrase type. The residuals of this linear model (ResidCompletionScore) correlated with the original completion test values ($r = .91$, $p < .0001$). The association strength correlated significantly with ResidCompletionScore, and therefore another linear model was created in which ResidCompletionScore predicted the association strength. The residuals of this model (ResidAssociationStrength) correlated highly with the original association strength values ($r = .96$, $p < .0001$). Finally, we calculated the residuals of the frequency of the first content word and the second content word because the first content word correlated significantly with phrase length, completion test score, and the frequency of the second content word, whereas the second content word correlated significantly with phrase length and frequency. Again, the residuals of these models correlated highly with the original variable (Content Word 1 and ResidWord1: $r = .86$, $p < .0001$; Content Word 2 and ResidWord2: $r = .87$, $p < .0001$). A summary of the continuous variables is presented in Table 2. We also investigated the need for by-subject random slopes for predictors tied to items and by-item random slopes for predictors tied to subjects. To avoid having a change in slope that might correlate with a change in intercept (see Baayen, 2008), all continuous predictors were centered.

For each dependent variable, we started with a simple mixed-effects model with subjects and items as random-effect factors that included trial number, phrasal length, and proficiency as predictors. In a step-by-step forward model selection procedure, we first looked at whether interactions between proficiency and trial number and proficiency and phrase length improved the model, and then we conducted an investigation including other predictors and their interactions with proficiency in the model. We also looked at the interactions between phrase type and phrasal frequency, and between phrase type and the frequency of Content Word 1 and the frequency of Content Word 2. Predictors and interactions between predictors were only included in the model if the model fit was significantly better (likelihood ratio test, $p < .05$) compared with the previous, more simple model. The coefficients of the fixed effects, their 95% highest posterior density intervals, p values based on 10,000 Markov Chain Monte Carlo samples of the posterior samples of the parameters of the final models, and the p values obtained with the t test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom for the three eye-tracking measures are presented in Table 3. We also used a backward model selection procedure starting with a model with all predictors and interactions tested in the forward selection procedure. This resulted, for all eye-tracking measurements, in final models that were exactly the same as the final models of the forward model selection procedure. Including in the final models by-subject and by-item random slopes for predictors tied to items and subjects did not significantly improve any of the final models except for the total reading time model in which trial number as a by-subject random slope improved the model significantly. However, this did not change the significance of the fixed effects in the model.

The mixed-effects modeling revealed that eye-tracking measures were significantly affected by trial number, phrase length, proficiency, phrase type, and phrasal frequency. Furthermore, a significant interaction was found between proficiency and phrase type in all measures (Figure 1 illustrates this interaction for the total reading time). This interaction indicates that proficiency plays a crucial role in phrasal processing. Namely, although the processing of binomials versus reversed forms differs significantly in native and higher proficiency nonnative speakers, their processing is similar in lower proficiency nonnatives. Importantly, independent of this interaction, the data revealed that phrasal frequency significantly influenced the eye-tracking measures, which is not

[4] We also conducted the mixed-effects modeling without removing any outliers. The results showed an identical pattern with and without outliers for the models of first-pass reading times and fixation count. The model of total reading times was slightly different because the interaction between phrase length and proficiency was no longer significant. However, everything else remained significant.

Table 1
Nonnative and Native Speakers' First-Pass Reading Times (Means), Total Reading Times (Means), and Fixation Count With Standard Error (SE) for Binomials and Reversed Forms

	Nonnative speakers			Native speakers		
	Phrase type			Phrase type		
Measure	Binomial	Reversed	Diff.	Binomial	Reversed	Diff.
First-pass reading time	564 (9.4)	577 (9.7)	13	317 (5.8)	355 (7.7)	38
Total reading time	602 (10.0)	616 (10.0)	14	342 (6.1)	399 (8.8)	57
Fixation count	2.50 (0.04)	2.50 (0.04)	0	1.80 (0.03)	2.0 (0.04)	.20

Note. Diff. = Difference.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

due to phrase type because the effect of phrase type was partialed out.

The completion test score was not included as a predictor in the final models of the three eye-tracking measurements, which indicates that the completion test score does not add anything to the model after phrase type is partialed out. This suggests that predictability of the second content word (*groom*) from the first content word plus *and* (*bride and*), as measured by the completion test scores, does not explain the effect of phrase type. However, we could not rule out the possibility that phrase type and the completion test score are measures of the same thing. Therefore, we reanalysed the data by first partialing out the completion test score from phrase type using a linear model. Phrasal frequency was included in this model as well because it correlated with phrase type. The residuals (PhraseTypeResid) of this linear model correlated with phrase type ($r = .65$, $p < .001$). Importantly, these residuals now reflect something above and beyond what is measured by the completion test. In this analysis, the same procedure for reducing collinearity, as described earlier (using residuals from linear models), was conducted before we analysed the data using mixed-effects modeling with items and subject as random-effect factors with a forward model selection procedure. The results again revealed significant effects of phrasal frequency for the first-pass reading time (pMCMC = 0.0004, $pr(>|t|) = 0.0001$), the total reading time (pMCMC = 0.0001, $pr(>|t|) = 0.0000$), and fixation count (pMCMC = 0.0001, $pr(>|t|) = 0.0000$), and a significant interaction between phrase type and proficiency for the first-pass reading time (pMCMC = 0.0144, $pr(|t|) = 0.0149$) and the total reading time (pMCMC = 0.0106, $pr(|t|) = 0.0100$). Note again that the completion test score was not included in any of the final models. This strongly suggests that the processing advantage observed for binomials over their reversed forms is not a mere reflection of the differences in their predictability, as measured by the completion test. Furthermore, it implies that the processes engaged extend above and beyond predictability alone and that it is the phrasal configuration and the phrasal frequency that play a crucial role in phrasal processing.

Discussion

These results reveal two major findings. First, native speakers and nonnative speakers, across a range of proficiencies, are sensitive to the frequency with which phrases occur in English. Second, native speakers and higher proficiency nonnatives are sensitive to whether a phrase occurs in a particular configuration (binomial vs. reversed), highlighting the contribution of entrenchment of a particular phrase in memory. Crucially, the frequencies of the first and second content word of the binomials and the reversed forms were not significant predictors of reading speed. This shows that it is the frequency of the entire phrase, and not the frequency of the individual words, that influences reading speed.

The significant interaction between the phrase type and proficiency suggests that binomial versus reversed form processing depends on proficiency, with more proficient nonnative speakers and native speakers reading binomials significantly faster than the reversed forms, and less proficient nonnative speakers exhibiting comparable reading speeds for both phrase types. The significant main effect of

Table 2
Summary of Continuous Variables

Variable	Range (adjusted range)	*SD*	*Mdn*
TrialNum	4–129 (−62.59–62.41)	36.4	−0.59
Proficiency	3–5 (−1.54–0.46)	0.66	0.46
PhraseLength	10–22 characters (−4.3–7.7 characters)	3.33	−0.80
ResidPhraseFrequency	0–1,956 (−1.96–3.57 log units)	1.42	−0.35
ResidWord1	119–142,377 (−3.73–2.43 log units)	1.02	0.10
ResidWord2	119–142377 (−3.19–2.71 log units)	1.03	−0.03
ResidAssociationStrength	0–0.85 (−0.35–0.57)	0.23	−0.04
ResidCompletionScore	0–10 (−6.93–5.53)	2.71	0.07

Note. The second column shows the range of the variables. The adjusted range after transformation, partialing out correlated predictors and/or centering, is presented in parentheses. Standard deviations and medians refer to the predictor values in the models. All variables are centered, and therefore their means are zero.

Table 3
Models for the First-Pass Reading Times, Total Reading Times, and Fixation Count

	Estimate	MCMCmean	HPD95lower	HDP95upper	*p*MCMC	*pr*(>|*t*|)
First-pass RT						
Intercept	6.0675	6.0692	6.0143	6.1201	0.0001	0.0000
TrialNum	−0.0006	−0.0006	−0.0011	−0.0001	0.0092	0.0095
PhraseLength	0.0210	0.0209	0.0143	0.0276	0.0001	0.0000
PhraseType	0.0625	0.0622	0.0198	0.1064	0.0062	0.0064
Proficiency	−0.2803	−0.2789	−0.3503	−0.2055	0.0001	0.0000
ResidPhrasalFrequency	−0.0240	−0.0240	−0.0392	−0.0072	0.0042	0.0037
PhraseType:Proficiency	0.0520	0.0517	0.0051	0.1000	0.0372	0.0300
Total RT						
Intercept	6.1124	6.1138	6.0592	6.1648	0.0001	0.0000
TrialNum	−0.0007	−0.0007	−0.0012	−0.0003	0.0012	0.0013
PhraseLength	0.0178	0.0177	0.0103	0.0248	0.0001	0.0000
PhraseType	0.0735	0.0731	0.0257	0.1190	0.0024	0.0028
Proficiency	−0.2910	−0.2896	−0.3581	−0.2178	0.0001	0.0000
ResidPhrasalFrequency	−0.0261	−0.0261	−0.0437	−0.0097	0.0028	0.0033
PhraseLength:Proficiency	−0.0084	−0.0085	−0.0154	−0.0016	0.0160	0.0164
PhraseType:Proficiency	0.0660	0.0655	0.0211	0.1134	0.0050	0.0048
Fixation count						
Intercept	1.1070	1.1071	1.0686	1.1429	0.0001	0.0000
TrialNum	−0.0006	−0.0006	−0.0009	−0.0003	0.0001	0.0001
PhraseLength	0.0198	0.0198	0.0153	0.0245	0.0001	0.0000
PhraseType	0.0343	0.0341	0.0029	0.0652	0.0348	0.0331
Proficiency	−0.1394	−0.1391	−0.1887	−0.0898	0.0001	0.0000
ResidPhrasalFrequency	−0.0193	−0.0193	−0.0302	−0.0080	0.0020	0.0008
ResidAssociationStrength	−0.0863	−0.0869	−0.1549	−0.0202	0.0124	0.0142
PhraseType:Proficiency	0.0334	0.0332	−0.0001	0.0661	0.0476	0.0474

Note. RT = reading time; MCMC = Monte Carlo Markov chain; HPD95lower = lower boundary of the 95% highest posterior density interval; HPD95upper = upper boundary of the 95% highest posterior density interval; *p*MCMC = *p* values estimated by the MCMC chain method using 10,000 simulations; *pr*(>|*t*|) = *p* values obtained with the *t* test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

phrasal frequency coupled with the absence of an interaction between phrasal frequency and proficiency implies that overall higher frequency phrases are read faster than lower frequency ones by native speakers and nonnative speakers across all proficiency levels.

It is noteworthy that although phrase type and phrasal frequency are overlapping concepts, they are not by any means the same. We can take the example, *east and west*, which occurs 380 times in the BNC, whereas its reversed forms occurs 63 times, and compare this with the binomial *sweet and sour*, which occurs 36 times, whereas its reversed form is unattested. (The binomial constituents by themselves are of different frequencies: *east*, 17,449, *west*, 21,345, *sweet*, 34,80, and *sour*, 623). In this case, the less preferred *west and east* is actually more frequent than the binomial *sweet and sour*. The present results indicate that, first, both native and nonnative speakers read frequent phrases more quickly than less frequent ones. If we continue with the above example, this means that the binomial *east and west* should be read more quickly than its nonpreferred reversed form *west and east*, followed by the binomial *sweet and sour*, which in turn should be read faster than *sour and sweet*. Second, in addition to phrasal frequency, there is an effect of phrase type that interacts with proficiency. This means that it is not just the overall frequency of a phrase that matters for native speakers and highly proficient nonnative speakers, but whether the phrase is in the preferred (binomial) or nonpreferred (reversed) configuration matters as well. Thus, although *sweet and sour* is less frequent than *west and east*, its processing should be speeded. This means that for frequently occurring expressions, something above and beyond simple frequency of occurrence is represented. Whether this can be attributed to predictability is addressed in the paragraphs that follow.

That native adult speakers process frequent multiword sequences faster than low-frequency ones is consistent with existing research (Arnon & Snider, 2010; Sosa & MacFarlane, 2002). Comparable evidence also exists in the child processing literature. Bannard and Matthews (2008) found that young children processed frequent phrases (*a drink of milk*) faster than infrequent ones (*a drink of tea*). This lead them to conclude that children have experience-derived knowledge of four-word utterances, the most frequent of which are stored in their lexicon. They took this as an indication of "complementary representations at different levels of granularity" (Bannard & Matthews, 2008, p. 246). We believe our results point to the same conclusion. Native speakers and higher proficiency nonnatives appear to have representations not only for the words that make up binomials (*bride*, *and*, *groom*) but also for the binomial phrases themselves (*bride and groom*). Frequency may thus lead to a particular form being represented in the mental lexicon. However, if a form has not been encountered frequently enough, as in the case of lower proficiency nonnative speakers, it appears that it may not be well entrenched in memory, leading to similar reading times for phrases like *bride and groom* and *groom and bride*.

Our finding that phrasal frequency affects the ease of processing is of importance for models of language use and processing. In the words-and-rules approach, a distinction is made between the lexicon, a collection of memorized and stored forms, and grammar, a collection of rules that are applied to these forms (Pinker, 1999;

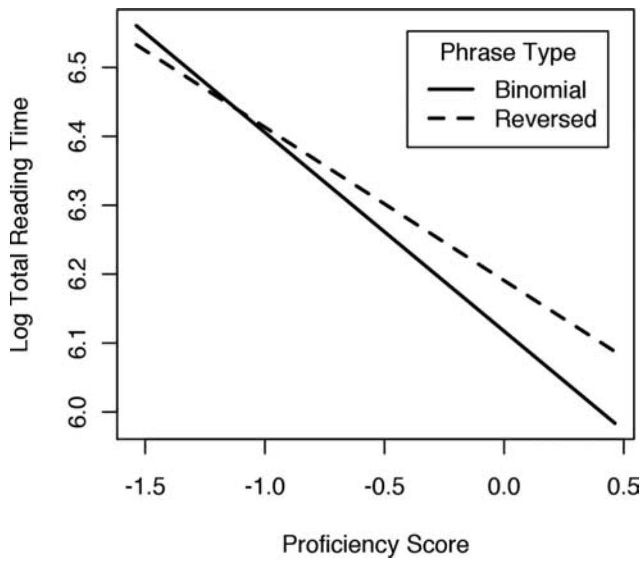

Figure 1. Interaction between phrase type and proficiency in the model for the total reading time.

Pinker & Ullman, 2002). In line with this approach, frequency effects should only be observable in the processing of memorized forms (words). Researchers argue that frequency effects should not manifest themselves in the processing of compositional multiword sequences. Thus, such a model is incompatible with our results.

However, usage-based (Bybee, 1998; Goldberg, 2006; Tomasello, 2003) and exemplar-based models (Abbot-Smith & Tomasello, 2006; Bod, 2006; Pierrehumbert, 2001) propose that the basic unit of language acquisition is a construction and that the task of a language learner is to acquire a set of constructions that vary in size, complexity, and level of abstractness (Goldberg, 2006; Tomasello, 2003). These theories propose that all linguistic information is represented and processed in the same way, and thus it should be similarly affected by frequency. New experiences with a linguistic unit, a word or a phrase, are not decoded and then discarded; rather, they determine memory representations (Bybee, 2006). As Bod (2006) noted, what is represented is based solely on statistics. Thus, language should be viewed not as a set of grammar rules, but as a statistical accumulation of experiences that changes every time a particular utterance is encountered. This view predicts faster processing for all frequent units, words and phrases, over less frequent ones. Our results are in line with such a view. Furthermore, our data are compatible with connectionist approaches to language acquisition and processing, which emphasize statistical properties of the input in language learning (Christiansen & Chater, 1999; Elman, 1990; Rumelhart & McClelland, 1986). In a connectionist approach, units do not exist in isolation; rather, they form and exist in relationships (networks) with each other. The frequency with which various linguistic exemplars occur together is a determining factor in what and how speakers learn and eventually represent in their lexicon. Thus, we take our results to support usage-based, exemplar-based, and connectionist models of language processing.

One could argue that due to their relative fixedness and frequency, multiword sequences have a special status in the lexicon and, as a result, are processed faster than novel language. However, one might also argue that the processing advantage observed for *bride and groom* is the result of a very quick, almost simultaneous activation of *groom* upon encountering *bride*. In line with probabilistic models of language processing, probabilistic information about word co-occurrences forms an integral part of speakers' knowledge of language (Gregory et al., 1999; Jurafsky, 1996; McDonald & Shillcock, 2003). Reichle, Pollatsek, Fisher, and Rayner (1998) and Engbert, Nuthmann, Richter, and Kliegl (2005) hold that eye-movement patterns reflect a reader's experience with language and are thus influenced by frequency and predictability. In the present study, the probability of $Word_{n+2}$ occurring after $Word_{n+and}$ is higher than the probability of $Word_n$ appearing after $Word_{n+2+and}$. Because *bride and groom* is a frequent expression, whereas *groom and bride* is not, one might, therefore, expect to see *groom* after reading *bride and*, which should facilitate reading; no such expectations may exist for *bride* upon reading *groom and*. Thus, the processing difference between binomials and their reversed forms may be due to the difference in their predictability, rather than one being represented in the lexicon and the other one not.

In order to assess the potential effect of predictability on reading times, we looked at whether scores on the completion test predicted reading times. The completion test did not significantly add anything to the models. Importantly, the analyses revealed that predictability and phrase type were not entirely the same. When completion test scores were regressed out from phrase type, phrase type still had a significant effect. Thus, we can conclude that the processing advantage for familiar phrases extends beyond the first word plus *and* (*bride and*) predicting the last one (*groom*). Rather, these findings signal the important contribution of phrasal frequency and entrenchment of a particular phrase in memory. Finally, it is worth pointing out that the "predictability story" per se does not go against a representational account: Each and every instance of a multiword sequence (idiom, binomial, etc.) is a highly predictable word combination in which subsequent words can be predicted from an initial one(s). Thus, being predictable is an *intrinsic* characteristic of a multiword sequence.

In summary, the results of the present study show that language users are sensitive not only to lexical frequencies, as has been widely shown in psycholinguistic research, but also to the frequency of multiword sequences. We take our results to support the view that each and every occurrence of a linguistic form, a word or a phrase, contributes to its degree of entrenchment in a speaker's memory.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23,* 275–290. doi:10.1515/TLR.2006.011

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67–82. doi:10.1016/j.jml.2009.09.005

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, England: Cambridge University Press.

Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon, 5,* 149–157. doi:10.1075/ml.5.1.06baa

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357. doi:10.1037/0096-1523.10.3.340

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19,* 241–248. doi:10.1111/j.1467-9280.2008.02075.x

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–35. Retreived from http://CRAN.R-project.org/package=lme4

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America, 113,* 1001–1024. doi:10.1121/1.1534836

Benor, S., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language, 82,* 233–278. doi:10.1353/lan.2006.0077

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review, 23,* 291–320. doi:10.1515/TLR.2006.012

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society, 34,* 421–435.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82,* 711–733. doi:10.1353/lan.2006.0186

Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics, 37,* 575–596. doi:10.1515/ling.37.4.575

Christiansen, M., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205. doi:10.1207/s15516709cog2302_2

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211. doi:10.1207/s15516709cog1402_1

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. doi:10.1037/0033-295X.112.4.777

Frazier, L., & Fodor, J. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325. doi:10.1016/0010-0277(78)90002-1

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford, England: Oxford University Press.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society, 35,* 151–166.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194. doi:10.1207/s15516709cog2002_1

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23,* 1089–1132. doi:10.1080/01690960802193688

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology, 35,* 876–895.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40. doi:10.1037/1082-989X.7.1.19

McDonald, S. A., & Shillcock, R. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research, 43,* 1735–1751. doi:10.1016/S0042-6989(03)00237-2

Mondini, S., Jarema, G., Luzzatti, C., Burani, C., & Semenza, C. (2002). Why is "red cross" different from "yellow cross"?: A neurophysiological study of non-adjective agreement within Italian compounds. *Brain and Language, 81,* 621–634. doi:10.1006/brln.2001.2552

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43–71. doi:10.1037/0096-3445.118.1.43

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 667–682. doi:10.1037/0096-1523.10.5.667

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam, the Netherlands: Benjamins.

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York, NY: HarperCollins.

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6,* 456–463. doi:10.1016/S1364-6613(02)01990-3

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14,* 191–201.

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Reali, F., & Christiansen, M. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57,* 1–23. doi:10.1016/j.jml.2006.08.014

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. doi:10.1037/0033-295X.105.1.125

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Mechanisms of language acquisition* (Vol. 2, pp. 216–271). Hillsdale, NJ: Erlbaum.

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language, 83,* 227–236. doi:10.1016/S0093-934X(02)00032-9

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

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral 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, England: The Continuum International Publishing Group.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (in press). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning.*

Van Lancker, D., Canter, G., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research, 24,* 330–335.

(*Appendix follows*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix

Set of All 42 Binomials Presented in the Study

alive and well, bride and groom, intents and purposes, king and queen, crime and punishment, mix and match, sweet and sour, bread and butter, stocks and shares, arts and sciences, cause and effect, heart and soul, mother and child, pain and suffering, safe and sound, buy and sell, church and state, war and peace, newspapers and magazines, cat and mouse, profit and loss, right and wrong, food and drink, husband and wife, name and address, research and development, knife and fork, black and white, brothers and sisters, backwards and forwards [sic], mind and body, day and night, trial and error, supply and demand, past and present, east and west, family and friends, men and women, radio and television, flora and fauna, read and write, gold and silver.

Received February 18, 2010
Revision received November 16, 2010
Accepted November 18, 2010 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Call for Papers: Special Section on *Theory and Data in Categorization: Integrating Computational, Behavioral, and Cognitive Neuroscience Approaches*

The Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP:LMC) invites manuscripts for a special section on approaches to categorization, to be compiled by guest editors Stephan Lewandowsky and Thomas Palmeri working together with journal Associate Editor Michael Waldmann.

The goal of the special section is to showcase high-quality research that brings together behavioral, computational, mathematical, neuropsychological, and neuroimaging approaches to understanding the processes underlying category learning. There has been some divergence between approaches recently, with computational-mathematical models emphasizing the unity of category-learning processes while neuropsychological models emphasize the distinction between multiple underlying memory systems. We are seeking articles that integrate cognitive neuroscience findings in designing models or interpreting results, and behavioral studies and modeling results that constrain neuroscientific theories of categorization. In addition to empirical papers, focused review articles that highlight the significance of cognitive neuroscience approaches to cognitive theory—and/or the importance of behavioral data and computational models on constraining neuroscience approaches—are also appropriate.

The submission deadline is **June 1st, 2011**. The main text of each manuscript, exclusive of figures, tables, references, or appendixes, should not exceed 35 double-spaced pages (approximately 7,500 words). Initial inquiries regarding the special section may be sent to Stephan Lewandowsky (stephan.lewandowsky@uwa.edu.au), Tom Palmeri (thomas.j.palmeri@Vanderbilt.Edu), or Michael Waldmann (michael.waldmann@bio.uni-goettingen.de).

Papers should be submitted through the regular submission portal for *JEP: LMC* (http://www.apa.org/pubs/journals/xlm/submission.html) with a cover letter indicating that the paper is to be considered for the special section. For instructions to authors and other detailed submission information, see the journal Web site at http://www.apa.org/pubs/journals/xlm.