More than words: Multiword frequency effects in non-native speakers

Mireia Hernández[a1], Albert Costa[a,b], & Inbal Arnon[c]

[a] Center for Brain and Cognition. Universitat Pompeu Fabra, Barcelona, Spain.

[b] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

[c] The Language Learning & Processing Lab, Department of Psychology, The Hebrew University of Jerusalem, Israel

[1] Corresponding author. E-mail: mireiahp@gmail.com

Abstract

Native speakers are sensitive to the frequencies of multiword phrases: they are faster to process higher frequency phrases, after controlling for all part frequencies (e.g., high: don't have to worry vs. low: don't have to wait). Here, we ask whether intermediate-advanced late (English) learners are also sensitive to the distributional properties of large language units: four-word combinations. Using a phrasal-decision task, we show that learners process multiword phrase frequency like natives do. This is not restricted to higher frequency phrases, but occurs across the frequency continuum: as natives, learners show multiword frequency effects even when comparing a low frequency phrase to a higher (but still low) frequency phrase. In addition, we show that the effect is not modulated by the type of English exposure (immersion vs. classroom). These results indicate that late language users develop sensitivity to distributional properties of large language units at native-like level.

Keywords: multiword; frequency; lexical processing; second language; learning setting

Introduction

Are late non-native speakers able to develop native-like lexical representations and processes? One way to address this question is by asking if non-native learners show frequency effects comparable to those of native speakers. The bulk of this literature, to date, has focused on word frequency effects and has shown that non-native learners exhibit lager word frequency effects than native speakers – a pattern often attributed to low frequency words being more weakly represented in the second language (L2) lexicon compared with the first language (L1) one: the "lexical entrenchment account"; e.g., see Diependaele, Lemhöfer, and Brysbaert (2012) for an overview of word frequency effects in L2. Frequency effects, however, are not restricted to single words. Speakers are sensitive to frequency information at multiple grain-sizes, including that of sound combinations, morphemes, syntactic constructions, and two-word combinations (see Diessel (2007) and Ellis (2002) for reviews). Recent work has shown that native language users (children and adults) are also sensitive to the frequency of compositional multiword sequences larger than two words (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Tremblay & Baayen, 2010). For instance, more frequent phrases (e.g., Don't have to worry) are processed faster than less frequent phrases (e.g., Don't have to wait) after controlling for all substring frequency (*worry* is as frequent as *wait*, *to worry* is as frequent as *to wait*, and *have to worry* is as frequent as *have to wait*, Arnon & Snider, 2010). Such findings highlight the parallels in the processing words and larger sequences and suggest that both are represented by similar cognitive mechanisms in native speakers (Elman, 2009; McClelland et al., 2010; Snider & Arnon, 2012).

Here, we go beyond existing findings to ask whether late non-native speakers are able to develop sensitivity to the distributional properties of compositional phrases (i.e., at least four-word phrases) to the same extent as native speakers and whether they do so in both classroom and immersion settings. This question has implications for our understanding of the L2 lexicon and the degree to which it is similar to that of native speakers. While much work has examined single word frequency in L2 speakers (e.g., Diependaele et al., 2012; Gollan, Montoya, Cera, & Sandoval, 2008), less work has looked at the processing of larger patterns, and most of it has focused on bigram information, which does not require sensitivity to multiword frequency, and on the processing of formulaic language – which is often seen as a 'special' kind of linguistic stimuli (see more details below). In fact, it has been proposed that L2 speakers will not be sensitive to multiword frequency because they process words separately without detecting that some words are more likely to appear together than others (Wray, 2002). We currently know little about L2 speaker's sensitivity to the frequency of compositional phrases, even though such findings are important for (a) assessing the similarity between native and non-native processing and (b) evaluating the claim that L2 learners differ from native speakers particularly in their sensitivity to multiword frequency. Examining L2 speakers' sensitivity to multiword frequency also bears on the more general question of the representation of words and larger patterns: finding that L2 speakers show frequency effects for both would enhance a single-system view of language where words and larger patterns are processed and represented by the same cognitive mechanism (Bybee, 1998; McClelland et al., 2010).

In the current study, we ask if intermediate-advanced L2 speakers of English are sensitive to the frequency of both single words and four-word phrases (after controlling for all part frequencies). We examine whether this occurs only for very high frequency phrases – which non-native speakers may attend to as a way to speed up lexical processing – or whether, as is the case for native speakers, non-native speakers are sensitive to multiword frequency across the continuum, including to that of lower frequency phrases. In addition, we explore whether the type of L2 exposure (immersion vs. classroom) will influence non-native speakers' sensitivity to multiword frequency. Since the sequences we look at are compositional (and not idiomatic), they are more likely to be acquired implicitly than through explicit instruction. It is possible that immersion settings – which typically promote more implicit learning – will lead to greater sensitivity to multiword frequency in non-natives compared to classroom settings. Finally, looking at the effect of both word and multiword frequency on L2 speakers allows us to see if lexical entrenchment affects larger patterns as well. If it does, we would expect L2 speakers to show stronger frequency effects than natives for both words and phrases. Taken together, the findings will further our understanding of the structure and nature of L2 lexical representations and their similarity to native processing. Before turning to the experiment, we review the relevant literature in more detail below.

L1 Multiword frequency effects: Evidence for parallels between words and phrases

In the domain of L1 processing, the question of how language users process multiword phrases has been the object of much debate. The two main views in this debate differ in whether they assume that single words and larger phrases are processed via the same or different cognitive mechanisms. The "words-and-rules" approach (Pinker 1991, 1999; Pinker & Ullman, 2002; Prince & Pinker, 1988) promotes a dual-system view of language that distinguishes between units that are stored in the lexicon and ones that are computed by grammar. In this perspective, words and compositional multiword phases are processed by different cognitive mechanisms: words are "stored" while multiword phrases are computed by combining words using rules or constrains. The "words-and-rules" approach argues that frequency effects only arise for "stored" but not for "computed" forms (Ullman & Walenski, 2005), giving rise to the prediction that words will show frequency effects but multiword phrases will not. In contrast, the "emergentist" perspective (Bybee, 1998; McClelland et al., 2010) proposes a single-system view of language whereby all linguistic experience – be it words or larger sequences – is processed by the same cognitive mechanism and subject to similar processing constraints. Consequently, language users are expected to develop sensitivity to the frequency of both words and multiword phrases.

Finding multiword frequency effects is more compatible with single-system views of language since they suggest that language users do learn frequency information about larger patterns, as they do for words.[1] Bannard and Matthews (2008) were the first to report the existence of multiword phrase frequency effects in a phrase-production task with two- and three-year old children. Arnon and Snider (2010) extended these findings to adults, and showed that such effects are found even for low frequency phrases, that is, whenever a higher (but still low) frequency phrase was compared to a lower one. This indicated that the effect of multiword frequency is continuous and is not limited to very high frequency phrases, which may be stored as a way to facilitate processing, as suggested by the "frequency-threshold" account (Biber et al., 1999; Goldberg, 2006; Wray, 2002). Thereafter, many other studies have shown that children and adults are sensitive to the frequency of multiword phrases, and that this sensitivity affects their language production, comprehension, and learning (Arnon & Clark, 2011; Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Ellis, Simpson-Vlach, & Maynard, 2008; Frank & Bod, 2011; Reali & Christiansen, 2007; Siyanova-Chanturia, Conklin, & van Heuven., 2011; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011).

Are non-native speakers sensitive to multiword frequency?

It has been proposed that L2 learners have difficulty in learning larger distributional patterns (e.g., Ellis, 2006; Wray, 2002). Wray (2002), in particular, suggests that, in contrast with child L1 learners, late L2 learners do not detect that some words tend to be encountered together, a pattern affected by both social and cognitive factors. From a social perspective, the claim is that non-native learners (especially classroom learners) are not faced with a pressing need to communicate, which may hinder the development of native-like processing of larger units. Cognitively, the fact that L2 learners are most often literate may direct their attention to words as the basic unit of processing, at the expense of learning multiword combinations. Such accounts predict that L2 learners will not be as sensitive as native speakers to multiword frequency. The question of L2 speakers' sensitivity to multiword frequency is also relevant for the debate on whether the mechanisms used for learning L1 and L2 are fundamentally different and require separate models and theories (e.g., Bley-Vroman, 2009; Clashen & Muysken, 1986; Paradis, 2004; Ullman, 2004) or whether similar processes underlie both learnings (MacWhinney, 2008). Finding that L2 speakers are similar to L1 speakers in their sensitivity to multiword frequency would suggest that, at least on the level of lexical processing, there are parallels in processing between the two.

To date, most experimental research on the question of multiword processing has been conducted with multiword sequences that are rather fixed in form and meaning. The most fixed type of multiword sequences are idioms (e.g., *kick the bucket*), which native speakers typically process faster than non-idiomatic novel phrases (e.g., Gibbs, 1980; Gibbs & Gonzales, 1985; Swinney & Cutler, 1979; Tabossi, Fanari, & Wolf, 2009; Van Lancker, Canter, & Terbeek, 1981). Such a facilitatory effect does not seem to be present in non-natives (e.g., Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Underwood, Schmitt, & Galpin, 2004). Idioms, however, are not the most appropriate type of multiword sequences to test Wray's (2002) hypothesis because it is not clear they are processed on a word-to-word basis: their non-compositional nature has led researchers to claim that they may be treated as unanalyzed wholes (Pinker, 1999; for literature not supporting this non-compositionality see, for example, Gibbs, 1992, 1993; Gibbs, Nayak, & Cutting, 1989; Tabossi, Fanari, & Wolf, 2008; Titone & Connine, 1994).

Experimental evidence with more compositional combinations does not support Wray's (2002) prediction: it seems that late non-native learners are able to detect and learn the co-occurrences of word sequences and know that some word combinations are more likely than others. One source of evidence comes from lab-based training studies on learning word-to-word co-occurrences where L2 learners are exposed to word combinations during training (i.e., collocations; e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013; Szudarski & Conklin, 2014). For example, Durrant and Schmitt (2010) exposed non-native learners of English to adjective-noun combinations (e.g., medical boat) and then asked them to perform a cued recall test to see if the target nouns facilitated the retrieval of their paired target adjectives. The results suggested that non-native learners were able to learn adjective-noun co-occurrences, even if they only saw them twice during the training session. These effects, however, are not long lasting: the benefits of such training last two weeks after training (Sonbul & Schmitt, 2013), but decline after six weeks (Szudarski & Conklin, 2014).

Other studies examine multiword sensitivity in a more naturalistic setting by comparing the magnitude of two- and three-word frequency effects for L2 learners to that of native speakers (Siyanova-Chanturia et al., 2011b; Sonbul, 2015; Wolter & Gyllstad, 2013). For example, in an eye-tracking experiment, Siyanova-Chanturia et al.

(2011b) explored this issue using binomials – pairs of words joined by a conjunction (e.g., safe and sound, loud and clear). They found that only natives and higher proficiency non-natives (but not lower proficiency non-natives) were sensitive to the most typical binomials' configuration: processing was facilitated if binomials were presented in their frequent configuration (e.g., safe and sound) as opposed to a less frequent but still correct reversed form (sound and safe). However, both natives and non-natives were faster to process higher frequency binomials compared to lower frequency ones. In another study, Wolter and Gyllstad (2013) found that non-natives showed sensitivity to the frequency of collocations (e.g., human rights, bottom line), regardless of the L1-L2 congruency of the collocational pairs (i.e., whether the L2-L1 literal translation was felicitous or infelicitous). Effects of proficiency were observed in those collocational pairs with an infelicitous L2-L1 translation, and were restricted to accuracy rates.

Finally, in a study combining behavioral and eye tracking measures, Sonbul (2015) used adjective-noun collocations (like *fatal mistake*) that were more flexible relative to those used in the two previously mentioned studies. That is, although the word combinations used in previous studies are more flexible than idioms (because, as opposed to idioms, meaning is indeed predicted from that of the constituent words), their word order is usually fixed: for instance, the reversed combination of *safe and sound* sounds unnatural, even if one can still understand its meaning. In contrast, an adjective-noun pair such as *fatal mistake* is not as fixed because it does not exclude other (less frequent) combinations that still sound natural (e.g., awful mistake). Sonbul's (2015) eye tracking measures showed early frequency effects in both natives and non-natives. Behavioral data also showed frequency effects in both groups, which correlated with proficiency in the case of non-native language users. Therefore, Sonbul's results indicated that non-natives are not merely sensitive to unmodifiable and, thus, highly predictable two-word configurations, but that they also develop sensitivity to the relative frequency of word combinations.

Taken together, the current findings on L2 learners suggest that they are sensitive to multiword frequency like native speakers. However, all studies to date have only explored this sensitivity with rather fixed word sequences. The only study to date that has explored this issue with more compositional word combinations has used two-word combinations (Sonbul, 2015). However, finding that L2 learners show bigram frequency effects does not entail that they are sensitive to multiword frequency: two-word frequency effects can arise by simply representing the relations between two individual words. Word and bigram frequency effects can be easily accommodated via links between words (or a non-symbolic representation of them). But frequency effects beyond the bigram (e.g., phrase-frequency effects) call for the representation of larger chains of relations (sequential information), not only between single words but also between word strings of varying sizes. Put differently, L2 learners may be sensitive to bigram frequency (as in adjective-noun combinations, or the order of the two content words in a binomial) without developing sensitivity to the distributional patterns of larger sequences, as native speakers to. To examine this possibility, we need to examine frequency effects for fully compositional multiword combinations beyond the bigram in L2 learners. An additional open question is whether the effect of multiword frequency on L2 processing is continuous (not limited to high frequency phrases), as it is for native speakers, or whether it is limited to very high frequency phrases, which may be stored as a way to facilitate processing, as suggested by the “frequency-threshold” account (Biber et al., 1999; Goldberg, 2006; Wray, 2002). To address this question, one needs to look at sequences across the entire frequency continuum (not only comparing high and low frequency stimuli).

One additional aspect when addressing this question is to consider how the type of exposure (immersion vs. classroom) may affect L2 learners' sensitivity to multiword frequency. Immersion settings are generally thought to prompt more implicit learning, while classroom settings lead to more explicit learning (e.g., Batterink & Neville, 2013; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Explicit learning provides L2 learners with information about L2 grammar rules or gives them directions to find rules, whereas such instruction is not given in implicit learning settings (Norris & Ortega, 2000). Implicit learning is thus more similar to the way a child learns a first language and might therefore lead to more native-like proficiency. However, the relative effectiveness of immersion/implicit versus classroom/explicit training for different linguistic abilities remains unclear. For example, in one electrophysiological study where participants were taught an artificial language using either explicit or implicit training, Morgan-Short et al., (2012) found that only the implicit learning conditions resulted in native-like activation patterns during syntactic processing. In contrast, Batterink & Neville (2013) taught native English speakers a reduced set of syntactic rules in an unknown language (French) using either implicit or implicit training. Both training groups (implicit and explicit) showed similar electrophysiological responses to those newly acquired syntactic rules. In another electrophysiological study, Nickels, Opitz, and Steinhauer (2013) showed that L2 learners in a classroom setting (with minimal immersion L2 exposure) are able to achieve native-like brain processing of speech prosody. In other studies, the relation between proficiency and type of exposure makes it difficult to differentiate the effect of proficiency from that of learning setting on the development of native-like processing. For example, Bowden, Steinhauer, Sanz, and Ullman (2013) found that L2 learners with both classroom and immersion exposure showed native-like electrophysiological responses to syntactic word-order violations, while the group with only classroom L2 exposure did not show native-like activation. However, L2 proficiency was lower in the classroom group, making it difficult (as the authors acknowledged) to disentangle the potential effects of the type of exposure with those of proficiency in achieving native-like processing. In sum, while it is not clear which setting is more facilitative overall, native-like processing of at least certain aspects of the L2 can be achieved in both learning settings.

How could the type of exposure influence learning of compositional multiword sequences? Unlike idiomatic expressions – which are often taught explicitly – Ellis (2002) proposes that learning word-to-word co-occurrence patterns is more likely to be subject to implicit rather than explicit processes of acquisition. The results of the previously mentioned Durrant and Schmitt‘s (2010) study – where L2 learners acquired adjective-noun co-occurrences they saw embedded in sentences – is in line with Ellis' proposal. Participants were not instructed to pay (conscious) attention to those adjective-noun combinations, yet they learned adjective-noun co-occurrences by just reading the adjective-noun pairs in a sentence context. Based on this, one could predict that sensitivity to multiword frequency will be learned better through immersion settings (which give rise to more implicit learning processes) than through classroom settings (which utilize more explicit learning processes).

The current study

The current study has several goals. The first and main goal is to explore whether late non-native speakers of English are able to develop sensitivity to compositional multiword phrase frequency across the continuum. Prior studies indicate that non-natives show frequency effects for relatively flexible two-word combinations. However, this does not necessarily mean that they are also sensitive to the distributional properties of larger chunks of language. To address this goal, we will use the same phrasal-decision task as in Arnon and Snider (2010, Experiment 1): we compare reaction times to pairs of four-word phrases that differ only in one word, are controlled for part frequency and plausibility but differ on phrase frequency (e.g., Don't have to worry vs. Don't have to wait). If L2 learners are sensitive to multiword frequency, they should respond faster to the higher frequency variants. We use items across the frequency continuum to see if non-native speakers are sensitive to multiword frequency even at low frequency ranges, as is the case for native speakers.

As a secondary goal, we examine the possibility that sensitivity to multiword frequency is affected by the type of English exposure (immersion vs. classroom). Prior studies are inconclusive about the extent to which immersion (as opposed to classroom) settings benefit the achievement of native-like processing. However, it is possible that immersion settings – where learning is more implicit – will lead to better learning of the distributional properties of multiword sequences that are compositional – like the ones we examine here. To test this possibility, we compared multiword frequency effects between two types of non-native speakers with intermediate-advanced levels of English proficiency who differed in the setting in which they used English at the time of testing: participants in the "immersion group" were living in an English speaking country, whereas participants in the "classroom group" lived in their country of origin and were exposed to English only through their studies (all were getting a Translation and Interpreting degree with English as their 'A' language).

In addition to the experimental task (i.e., the phrasal-decision task), we also included a word lexical decision task (LDT) for two reasons. First, we want to ensure that our non-natives are sensitive to the occurrence of smaller units of language for which frequency effects have been repeatedly reported in literature (i.e., words): it would be hard to expect phrase frequency effects in non-native language users if they did not show the typical word frequency effect. Second, we want to see if the finding that L2 speakers show a stronger frequency for words than native speakers carries over to multiword phrases.

Method

Participants

Seventy-nine participants took part in the study (27 native, and 52 non-native speakers of English). Native participants had English as a first and dominant language. This native group was composed of undergraduate, Master's, and PhD students, as well as individuals who were not enrolled in any university program but who had jobs requiring a degree. Non-native participants had English as an L2 and they had never used this language in any context except for classroom instruction (i.e., English lessons) before adolescence. All non-natives were classified as having a Common European Framework (CEF) English proficiency level of upper intermediate (CEF = B2) or lower advanced (CEF = C1). This means that they were able to use English with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations in practical, social and professional topics. This proficiency level was estimated through the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012): an English vocabulary size test that is often used to estimate the Common European Framework (CEF) proficiency level.

Non-native speakers were divided into two groups according to their type of English exposure: there were 27 participants in the immersion exposure group (18 females and 9 males) and 25 participants in the classroom exposure group (21 females and 4 males). Participants in the immersion exposure group had a variety of languages as their L1(Spanish (n=9), Catalan (n=3), Mandarin (n=3), Dutch (n=2), Italian (n=2), Creole (n=1), Korean (n=1), Nyanja (n=1), Polish (n=1), Portuguese (n=1), Romanian (n=1), and Turkish (n=1)). All participants in this group were studying (degree, Master, and PhD programs) or working (post-doctoral research or non-academic jobs requiring a degree) in the US at the time of testing. Participants in the classroom exposure group had either Catalan or Spanish as an L1.[2] Most of them were studying towards a BA in Translation and Interpreting at the *Universitat Pompeu Fabra* (UPF, Barcelona, Spain).

Table 1 reports the comparison between the three groups of participants in age and educational attainment. There were no significant differences between the native and the immersion exposure groups either in age ($t_{52} = 1.51$, $p < .14$) or educational attainment ($t_{52} = 1.44$, $p < .16$). Participants in the classroom exposure group were 7.6 years younger ($t_{50} = 7.3$, $p < .0001$) and had lower educational attainment ($t_{50} = 3.14$, $p < .003$) than those in the immersion exposure group. Participants in this classroom exposure group were 5.73 years younger than those in the native group ($t_{50} = 6.76$, $p < .0001$), but these two groups did not differ in educational attainment ($t_{50} = 1.44$, $p < .16$). Age and education attainment differences between the classroom exposure group and the other two reflect the fact that most participants in the classroom exposure group were still undergraduate students of the Translation and Interpreting degree.

(Table 1 about here)

Table 2 reports details on age of English acquisition, exposure, and proficiency of the two non-native groups. Participants in the classroom exposure group started acquiring English and became fluent in that language earlier than participants in the immersion exposure group. What is relevant, however, is that all non-natives started acquiring English during childhood (ranging from 3 to 14 years old), but none of them became fluent in English before adolescence. Since participants in the classroom exposure group were students of Translation and Interpreting, they were regularly exposed to English, but only in classroom settings. These participants had never lived in an English speaking country more than three consecutive weeks, and it had been at least one year since their last stay (if any) in an English speaking country. Participants in the immersion exposure group were exposed to English regularly in a native context because they worked or studied in an English-speaking country (i.e., US). The time spent in the US varied between three months and seventeen years.

All non-natives answered an English exposure questionnaire with a twofold purpose. First, it served to rule out that the two non-native groups differed in the amount of English exposure through other language-related activities – i.e., watching TV, listening to the radio/music, classroom instruction (i.e., English lessons not including those in the Translation and Interpreting degree in the case of participants in the classroom exposure group), or self-instruction. Second, it served to make sure that participants in the classroom exposure group were not exposed to English in immersion-like settings with family and friends. In this questionnaire, participants rated from 0 (none at all) to 10 (very much) their daily exposure to English in different contexts/activities (family, friends, reading, formal education, self-education, watching TV, and listening to the radio/music). The total amount of English exposure (i.e., 70, collapsing the seven contexts/activities) did not differ significantly between the classroom and the immersion exposure groups. However, and unsurprisingly, participants in the immersion group had more exposure to English in their daily life because they were living in an English-speaking country. Participants in the immersion group were exposed to more English with friends than participants in the classroom exposure group. There was not much exposure to English in a family context in either group. But, probably due to some participants in the immersion contexts having English-speaking partners, the amount of exposure to English in a family context was slightly higher in this immersion exposure group relative to the classroom exposure group. Participants in the immersion exposure group were also slightly more exposed to English from reading, probably due to having access to journals, newspapers, and books mostly in that language. Perhaps the higher exposure to English of the immersion group contributed to participants in this group showing slightly larger vocabulary size as measured through LexTALE than their classroom counterparts. Despite this, the two non-native groups did not differ in self-assessed English proficiency in any language domain (speaking, reading, and listening). This was measured through the *Interagency Language Roundtable (ILR) language skill self-assessment* questionnaire *(*United States Federal Government; http://www.govtilr.org/): non-natives answered “yes” or “no” to 90 statements about English skills (e.g., I rarely, if ever, have to ask speakers to paraphrase or explain what they have said).

(Table 2 about here)

Tasks

Word Lexical Decision Task (LDT)

In this task, participants had to judge whether each of the eighty presented strings of letters was a real word in English (e.g., bread) or not (i.e., pseudoword, e.g., solk). Half of the items were words and the other half were pseudowords. Participants were instructed to respond as quickly and accurately as possible using the keyboard. The task was presented in a single block of about 5 minutes that participants completed in between the two blocks of the experimental task (i.e., the phrasal-decision task). Each trial began with the presentation of a fixation point for 500 milliseconds (ms) followed by the presentation of the string of letters for 500 ms. The next trial began after participants responded or after 1500 ms had passed. The task was run using DMDX (Forster & Forster, 2003).

Materials

Words. Words were divided into twenty high frequency words and twenty low frequency words using CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). High and low frequency words did not differ in syllable, phoneme, or letter length. High frequency words had a larger number of senses relative to low frequency words (see Table 3).

(Table 3 about here)

Pseudowords. All pseudowords were constructed by replacing one or more letters from the original set of words. The number of letters replaced depended on the length of the original word.

Phrasal-Decision Task

Participants had to judge if the four-word English phrases that appeared on the screen were possible sequences in English or not. Impossible sequences were incorrect due to scrambled word order (e.g., I saw man the) or inappropriate prepositions (e.g., Jump during the pool). Phrases appeared at once (i.e., in their entirety) in the middle of the screen and participants were instructed to respond as quickly and accurately as possible using a keyboard. The task was divided in two blocks of about 5 minutes each, which were separated by the LDT task. The order of presentation of the blocks was counterbalanced between participants. Each trial began with the presentation of a fixation point for 500 ms. The phrase was then presented and stayed visible on the screen until participants responded or until 3000 ms had passed. The task was run using DMDX (Forster & Forster, 2003).

Materials

Experimental phrases. We used the same materials as those in Arnon and Snider (2010, Experiment 1), which were constructed using a 20-million word corpus resulting from the combination of two spoken corpora, the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992) and the Fisher corpus (Cieri, Miller, & Walker, 2004). These two corpora were based on American's phone conversations, which guaranteed that the experimental set of phrases was typically used in spontaneous speech. This experimental set was composed of 26 pairs of phrases. Phrases in each pair (a) differed only in the final word (e.g., Don't have to worry vs. Don't have to wait), (b) differed in phrase-frequency (high vs. low), but (c) they did not differ in the frequency of the final word, bigram, or trigram, or in plausibility (see Appendix A for complete item list). In order to reduce possible priming effects from seeing two very similar phrases, only one variant appeared in each block and block order was counter-balanced across participants. To avoid incomplete intonational phrases, the last word in each phrase was never a determiner. Similarly, no phrase ended with a demonstrative (e.g., that), which could be interpreted as a modifier (e.g., part of *that* boy). In addition, to increase the reliability of the frequency estimates for low frequency phrases, all first 3-grams (e.g., Don't have to) had a frequency over 30 per million, and the last word in the 4-gram (the one differing between the two phrases in each pair) always had a frequency of at least 50 per million.

In the original study (Arnon & Snider, 2010), the 26 phrases were divided into two frequency bins, a high bin (15 phrases) and a low bin (11 phrases) that differed in their cutoff: while the cutoff point between the high and low frequency variants was 10 per million in the high frequency bin, it was one per million in the low frequency bin. High and low frequency phrases were equally plausible in the high frequency bin (high: 6.7, low: 6.7, $W = 113.5$, $p > .5$), and in the low frequency bin (high: 6.6, low: 6.4, $W = 43.5$, $p > .1$) – see Arnon and Snider (2010) for details about the ratings on plausibility. The division into bins was done to test the prediction that speakers show frequency effects across the continuum and not only for high frequency phrase. Since that has already been demonstrated, and since treating frequency as a binary variable may lead to a loss of information, in the current study, we collapse the two bins and treat all item pairs as belonging to one set that ranges from very low frequency (0.2 per million) to high frequency (27 per million). The important feature of the items is that they span the frequency continuum and are paired to minimize the effect of part frequencies.

Fillers. In addition to the experimental phrases, there were 84 fillers. Sixty-eight of these fillers were grammatically incorrect phrases (e.g., I saw man the, Jump during the pool). The remaining sixteen fillers consisted of grammatically correct phrases that were added to have the same number of grammatically correct and grammatically incorrect phrases.

Results

Word Lexical Decision Task (LDT)

Responses under 200 ms and over two SD from the mean per condition (high frequency vs. low frequency) were excluded. This resulted in loss of small percentages of data per frequency condition in each group of participants (≤ 5.42%). Participants in all three groups were highly accurate in their responses across conditions (see Table 4).

(Table 4 about here)

We analyzed the results using mixed-effect models. Following Barr, Levy, Scheepers, and Tily (2013), we use the maximal random effect structure justified by the data. We used log(response times) as the predicted variable to reduce the skewness in the distribution of response times. All reported models had low collinearity [vif's < 1.8]. To further reduce the effect of any remaining collinearity, we tested the effect of our variables of interest through model comparisons and included only ones that significantly improved the model. To test the effect of lexical frequency on native and non-native speakers, we ran a model with frequency (high vs. low) and group (native, L2-classroom, L2-immersion) as fixed effects, we well as the interaction between frequency and group – to see if the effect of frequency differs between native and non-natives. We had number of senses and trial-number as additional fixed effects (this was done to control for the difference in mean number of senses between the high and low frequency items). The model had participant and item as random effects, as well as a by participant random slope for frequency (high vs. low).

Significance was obtained using the lmerTest package in R [Formula: f1 <- lmer (log (RT_clean) ~ freq*group + dmdx_order + num.senses + (1+freq|subject) + (1|item), data=subset (use))]. As predicted, the effect of lexical frequency was significant: across groups, participants were faster to respond to the high frequency words (518 ms, SD = 79) compared to low frequency ones (607 ms, SD = 130, $\beta = -.08$ (SE=.01), $p < .001$). However, the effect of lexical frequency was stronger in the non-native groups (natives: 45 ms, SD = 27; immersion: 105 ms, SD = 74; classroom: 117 ms, SD = 63), as reflected in the significant interactions of frequency and the two L2 groups (frequencyXimmersion: $\beta = .07$ (SE=.01), $p < .001$; frequencyXclassroom: $\beta = .09$ (SE=.01), $p < .001$). Both non-native groups were overall slower than the native speakers (immersion: β = .08 (SE=.01), $p < .01$; classroom: β = .01 (SE=.03), $p < .001$). The effects of trail number and number of senses were not significant (trail number: β = -.001 (SE=.001), $p > .09$; number of senses: β = -.001 (SE=.001), $p > .1$).

In summary, these results showed that non-natives (as well as natives) are sensitive to the frequency of occurrence of single words (Figure 1), and that they show a larger lexical frequency effect than native speakers, as has been found repeatedly in the past (e.g., Diependale et al., 2012). This is of importance for our study because it would be hard to expect phrase-frequency effects in the absence of the extremely robust effect word frequency effect found for L2 learners.

(Figure 1 about here)

Phrasal-Decision Task

Responses under 200 ms and over two SD from the mean were excluded. This resulted in loss of small percentages of data in each group of participants (≤ 6.3). Participants in all three groups were highly accurate in their responses to phrases and fillers (see Table 5).

(Table 5 about here)

We analyzed the results using mixed-effect models. All models had the maximal random effects structure justified by the design (Barr et al., 2013). We used log(response times) as the predicted variable to reduce the skewness in the distribution of response times. We added the log frequency of the final word, final bigram and final trigram as controls (we used log frequencies to correct for non-normal distributions). These were the only substrings that differed between the high and low variant of each pair, and while their frequency was matched, we wanted to make sure that any effect of frequency was caused by phrase-frequency and not substring-frequency. These frequencies were calculated from the same corpus used to select the target items. In all analyses, we checked for collinearity between the fixed effects (e.g., between phrase-frequency and the frequency of the final bigram) and reduced it by regressing one of the collinear factors (the factor of interest, if one was involved) against the collinear covariates, and using the residuals of these regressions instead of the original variables in the final models we report. All reported models had low collinearity [vif's < 2.5]. To further reduce the effect of any remaining collinearity, we tested the effect of our variables of interest through model comparisons and included only ones that significantly improved the model.

Sensitivity to multiword frequency in native and non-native speakers

To test the effect of multiword frequency on native and non-native speakers, we ran a model with logged (phrase-frequency), frequency-type (high vs. low) and group (native, L2-classroom, L2-immersion) as fixed effects, we well as the two-way interaction between logged (phrase-frequency) and group – to see if non-native speakers are as sensitive to multiword frequency as native speakers across the continuum. We also had trial-number, block (first vs. second), number of letters, and the relevant substring frequency measures (unigram4, bigram3, trigram2) as additional fixed effects. The model included random intercepts for participant and item, as well as a by-subject random slope for logged (phrase-frequency).

As predicted, the effect of multiword frequency was significant: across groups, participants were faster to respond to the high-frequency variants (see Table 6): higher frequency led to shorter reaction times (high-frequency = 1023 ms, SD = 378; low-frequency = 1073 ms, SD = 400; $\beta = -.03$ (SE=.01), $p < .01$), leading to a phrase frequency effect of 50 ms, which is similar to that reported in Arnon and Snider (2010).

Unlike the word frequency data, the magnitude of the effect did not differ between native and non-native speakers, or between the two groups of non-native speakers (non-significant interactions between logged(phrase-frequency) and L2-classroom group: $\beta = -.002$ (SE=.005), $p > .7$, L2-immersion group: $\beta = .001$ (SE=.05), $p > .7$). Interestingly, there was no difference in overall reaction times between the immersion learners (1027 ms, SD = 161) and native speakers (958 ms, SD = 125; $\beta = .06$ (SE=.01), $p > .28$), but the classroom group (1210 ms, SD = 187) had slower reaction times compared to native-speakers ($\beta = .02$ (SE=.05), $p < .01$). Unsurprisingly, having more letters led to slower reaction times ($\beta = .05$ (SE=.005), $p < .001$). Participants were faster overall in the second block ($\beta = -.06$ (SE=.007), $p < .001$), but reaction times slowed down with increasing trial number ($\beta = .003$ (SE=.001), $p < .05$). The effect of bigram and trigram frequency was not significant (bigram3: $\beta = .007$ (SE=.01), $p > .5$, trigram2: $\beta = .02$ (SE=.01), $p > .15$), but the effect of the final word frequency was marginal in the unexpected direction: participants were somewhat slower when the final word was more frequent ($\beta = .01$ (SE=.009), $p = .055$). Note however, that the items were tightly controlled for part frequencies and were not designed to span the lexical frequency continuum.

(Table 6 about here)

These results replicate the ones found in Arnon and Snider (2010) for native speakers – adults are sensitive to multiword frequency after controlling for all substring frequencies – and extend them to two groups of non-native speakers. Non-native speakers showed multiword frequency effects similar to those shown by native speakers in both learning settings. To further explore the effect of multiword frequency on non-native speakers, and to see if they are affected by the length of time spent in an English-speaking country, we conducted an additional analysis looking only at the immersion group. Since there was variability in the time spent in the US by participants in this group (ranging from three months to seventeen years) we added length of stay, and its' interaction with phrase-frequency as additional factors (see Table 7). As in the previous analysis, participants were faster to respond to higher frequency phrases (β =.-03 (SE=.01), $p < .01$). While learners who had spent longer in the US were faster overall (β = -.02 (SE=.01), $p < .05$), this did not interact with phrase-frequency (β = .004 (SE=.001), $p > .9$).

(Table 7 about here)

We also inspected whether non-natives' phrase frequency effects were associated with variables particularly relevant for foreign language users such as proficiency level or amount of current L2 exposure. The magnitude of the frequency effects for phrases was not correlated with any of the proficiency measures (not correlated with LexTale or with any of the ILR self-assessment scales; all *r*s < 0.22, all *p*s > .11).

In sum, our analyses show that (a) native and non-native speakers are sensitive to multiword frequency across the frequency continuum, and (b) that this effect is similar in both learning settings. These results replicate Arnon and Snider's (2010) study showing that language users are sensitive to the frequency of word combinations at the level of multiword phrases, and that this occurs even at low frequency ranges – i.e., whenever a higher (but still low) frequency phrase was compared to a lower one. Crucially for our purposes, non-native language users, in both the immersion exposure group and the classroom exposure group, showed the same pattern of results as natives. That is, the frequency effects for phrases were of similar magnitude in the three groups of speakers, and that occurred regardless of one group (i.e., the classroom exposure group) showing overall slower response latencies relative to the other two (the native and the immersion exposure groups).

Discussion

The objective of the present study was to determine whether late L2 learners learn distributional properties of relatively large language sequences as efficiently as L1 speakers. After confirming that both native and non-native speakers of English were sensitive to the frequency of occurrence of single words, we compared their sensitivity to multiword frequency using a phrasal-decision task (Arnon & Snider, 2010). In addition, we also examined whether any potential phrase frequency effect in non-native speakers would be restricted to the highest frequency pairs – i.e., whenever a very high frequency phrase is compared to a lower (but still very high) one. Finally, we compared the magnitude of the effect between two groups of non-native speakers (immersion vs. classroom) to examine the effect of learning setting on multiword processing. The results indicated that non-native speakers were as sensitive as natives to the frequency of multiword phrases along the continuum (i.e., also for lower frequency phrases), regardless of the type of exposure (immersion vs. classroom): there was no difference in the magnitude of the frequency effect between the three groups of speakers (native, immersion-L2, classroom-L2).

The findings replicate the effect of multiword frequency on native speakers (Arnon & Snider, 2010): using the same items, we found that speakers respond faster to higher frequency phrases. These findings further highlight the parallels in processing words and larger phrases – both words and compositional multiword phrases display frequency effects. Such findings pose a challenge for dual-systems views of language, which differentiate between words and compositional phrases (e.g., Pinker, 1999), and are instead supportive of single-systems of language (e.g., Bybee, 1998; Christiansen & Chater, 1999; McClelland et al., 2010) where all linguistic material is processed by the same cognitive mechanisms. It is worth acknowledging, however, that there are other ways to capture frequency effects in addition to the two views presented above. For instance, in the "Adaptive control of though-rational" theory (ACT-R; Anderson et al., 2004), phrase frequency effects are thought to reflect a set of computations that constrain the retrieval of chunks (i.e., elements stored in declarative memory that in our case would be phrases). These computations include the frequency and recency of retrieval of each particular chunk, and also the so-called utility value. This value is determined by how often the production of that chunk was satisfactory. That is, given multiple competitor chunks that can be activated in certain contexts, how often that particular chunk led to satisfactory production in a given context (see Patil, Hanne, Burchert, De Bleser, and Vasishth (2016) for an implementation of this framework for sentence processing). Regardless of the precise model used to capture frequency effects, the results highlight the need of any model to be able to account for both word and multiword frequency effects.

More importantly, the results provided an affirmative answer to our main question: late non-native speakers are able to learn information about the frequency of multiword phrases to the same extent as native speakers. This finding extends prior research that documented frequency effects for rather fixed word combinations in non-native speakers (e.g., Siyanova-Chanturia et al., 2011b; Wolter & Gyllstad, 2013) in several ways. First, unlike Wolter and Gyllstad's work, the phrases used in this study were compositional and non-formulaic, and unlike Siyanova-Chanturia et al.'s, work there were broader in scope than binomials. This means that non-natives are not only sensitive to fixed and highly predictable multiword expressions, but to more fine-grained information about how multiple words combine in a language (e.g., that *I don't like the idea* is more frequent than *I don't fancy the idea*, even though both are grammatical and plausible). Second, this goes in line with results showing that non-native speakers are sensitive to the frequency of relatively flexible adjective-noun pairs (e.g., fatal mistake, Sonbul, 2015), and shows that it occurs beyond the bigram and in a wider range of phrases (not only adjective-noun combinations). Third, the fact that non-natives showed similar frequency effects as natives even at low frequency ranges excludes the possibility that they are only learning those word combinations that occur with extremely high frequency, and can therefore be memorized as unmodifiable wholes. Instead, the findings suggest that non-natives are able to detect and learn that some word combinations are more likely than others. In addition, the fact that our non-natives were “only” at upper intermediate and lower advanced levels of English proficiency means that this ability does not require functionally native levels of English proficiency. Finding that L2 learners at this level of proficiency show multiword frequency effects presents additional evidence against Wray’s (2002) proposal that L2 learners process words separately without detecting that some words are more likely to appear together than others. What’s more, this finding indicates that L1 and L2 learning show parallels in at least certain aspects, such as sensitivity to the distributional properties of multiword sequences – under a single-system view of language, this means that the parallels between the processing of words and phrases for native speakers also holds for non-native speakers. This goes in line with the view that the similarities between L1 and late L2 learning are more striking than the differences and, therefore, that unified theoretical models rather than separate ones are needed to account for the mechanisms used for L1 and L2 learning (MacWhinney, 2008).

One important question in assessing non-natives’ ability to develop native-like lexical processing is how this is affected by proficiency levels. In our study, phrase frequency effects did not seem to be modulated by proficiency levels: the magnitude of the phrase frequency effect did not correlate with any of the L2 proficiency measures (self-reports or the English vocabulary size test LexTALE). Prior studies about the ability to develop sensitivity to multiword co-occurrences (restricted to two-word combinations) also did not show clear proficiency effects. For example, Siyanova-Chanturia et al. (2011b) did not observe proficiency differences in non-natives' sensitivity to the frequency of binomials (e.g., safe and sound), but only in their sensitivity to the binomials' most typical configuration (e.g., *safe and sound* as opposed to *sound and safe*). In Wolter and Gyllstad's (2013) study higher proficiency L2 learners were more accurate only when collocational pairs had an infelicitous L2-L1 translation. Finally, Sonbul (2015) found proficiency effects in non-natives' sensitivity to adjective-noun pairs (e.g., fatal mistake) in the behavioral measures but not in the eye-tracking measures. However, there is one important limitation in the conclusions that can be drawn from both our study and prior studies: all of them use a relatively narrow range of proficiency levels, because leaners have to be able to understand and complete the tasks. Studies testing non-natives' sensitivity to multiword co-occurrences need to test participants that have achieved a relatively advanced level of proficiency. This, in turn, reduces non-natives participants' variability in terms of proficiency levels, which may hinder the detection of any potential proficiency effect. The lack of variability may also impact our ability to detect a relation between single word representation and multiword representation. It is possible that information on how often different words appear together cannot be efficiently learnt if the lexical representation of those words is weak. The lack of correlation between the single word frequency effect (in the LDT task) and the phrase frequency effect (in the phrasal-decision task) suggests a negative answer to this question. However, our non-native speakers may not vary enough in the strength of single word lexical representations in the same way they do not vary enough in proficiency levels.

The comparison between the LDT results and the phrasal-decision task revealed an interesting contrast: while non-natives showed stronger word frequency effects compared to native speakers, they did not show larger multiword frequency effects. The counterintuitive phenomenon of larger lexical frequency effects in L2 speakers has been repeatedly reported in prior studies with single words (e.g., Diependaele et al., 2012; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Gollan et al., 2008; Gollan et al., 2011; Lemhöfer et al., 2008; Van Wijnendaele & Brysbaert, 2002; Whitford & Titone, 2012), and has driven different theoretical accounts about L2 lexicon integration. One of the most accepted explanations for this phenomenon is the "lexical entrenchment account": the effortful discrimination between a low frequency word and its neighbors is due to L2 words being more weakly represented in the lexicon. Although this weakness is conceptualized differently by the various versions of this account (e.g., L2 reduced resting levels, lower L2 lexical precision), it is argued to be the result of a generally lower proficiency in L2. The "lexical entrenchment account" would, in principle, be as applicable to multiword phrases as it is to single words. If this were the case, we would expect stronger phrase frequency effects in non-natives (relative to natives). However, this exaggerated frequency effect in non-natives was not observed in our phrasal-decision task or in prior studies using two-word combinations. One possibility is that there is a larger multiword frequency effect in L2 speakers, but it was not detected because of the cutoff point between high and low items. That is, it is possible that the mean frequency difference between high and low frequency items is much smaller in the studies using multiword combinations than in studies using single words. This is the case of our study: in our phrasal-decision task, the mean frequency difference (per million words) between the high and low frequency variants was 16.33 and 2.76 for phrases in the high and low frequency bin (see original study by Arnon &

Snider, 2010). In contrast, the difference between high and low frequency words (in the LDT) was clearly higher (103.85). Nevertheless, whether or not this is the (only) reason why non-natives' typically show stronger frequency effects (relative to natives) in single words but not in phrases is unclear and calls for further research.[3]

With regards to our secondary question, our data suggests that immersion settings – which prompt more implicit learning – do not lead to increased sensitivity to multiword frequency information relative to classroom settings – where learning is typically more explicit. Instead, our results suggest that non-native speakers in both settings show native-like sensitivity to the distributional properties of multiword phrases. This finding goes in line with the results of prior research showing that immersion exposure is not necessary to develop native-like cognitive mechanisms, at least for some aspects of L2 processing (e.g., the ability to use prosodic cues in syntactic parsing, Nickels et al., 2013). Note, however, that our study and also prior studies comparing learning settings cannot rule out that classroom learners may also have been exposed to the L2 in immersion-like contexts as well. Non-natives may be exposed to immersion-like contexts through a variety of out-of-class activities (e.g., interacting with foreign work mates, watching films in original version, reading leisure books, etc.), as well as through certain classroom materials (movies, songs, etc.). What we can conclude is that living in a native context is not necessary to develop sensitivity to multiword frequency. However, whether this is because such sensitivity can arise through both implicit and explicit learning or because classroom learners are also exposed to immersion-like contexts is a matter for further research.

In conclusion, the results of this study suggest that acquiring a second language after childhood does not result in a lack of sensitivity to distributional information at the phrase level. Like native speakers, L2 learners are sensitive to the frequency of four-word compositional phrases across the continuum, even at upper-intermediate levels of proficiency. Moreover, these effects were found to a similar extent regardless of current exposure setting (classroom vs. immersion). These findings highlight additional parallels between the lexical processes of native and non-native speakers. Further research is needed to understand when such sensitivity develops, whether it is found also in lower proficiency levels, and the degree to which it is dependent on implicit vs. explicit learning mechanisms.

Acknowledgement

We thank the Boston Spanish group for the help with participants' recruitment.

Funding

This work was supported by the "Agència de Gestió d'Ajuts Universitaris i de Recerca" (Catalan Government) [GRNC-2014SGR1210, SGR 2009-1521]; the Spanish Government [PSI2011-23033, CONSOLIDER-INGENIO 2010 CSD2007-00048]; and the European Research Council under 7th European Community Framework (FP7/2007-2013) [613465-AThEME]. MH was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework (FP7/2007-2013), and by the postdoctoral fellowship "Formación Posdoctoral" within the Spanish Government program "Programa Estatal de Promoción del Talento y su Empleabilidad del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016".

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111 (4),* 1036–1060. doi:10.1037/0033-295X.111.4.1036

Arnon, I., & Clark, E. V. (2011). When ‘*on your feet*’ is better than ‘*feet*’: Children’s word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7, 107–129. doi:10.1080/15475441.2010.505489

Arnon, I., & Cohen Priva, U. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech, 56 (3),* 349–371. doi:10.1177/0023830913484891

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., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

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

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

Batterink, L., & Neville, H. (2013). Implicit and Explicit Second Language Training Recruit Common Neural Mechanisms for Syntactic Processing. *Journal of Cognitive Neurosci*ence, 25 (6), 936–951. doi:10.1162/jocn_a_00354

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). Longman grammar of spoken and written English. MIT Press.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175–198. doi:10.1017/S0272263109090275

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51(13)*, 2492–2511. doi:10.1016/j.neuropsychologia.2013.09.004

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

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157–205. doi:10.1016/S0364-0213(99)00003-8

Cieri, C., Miller, D., & Walker, K. (2004). The fisher corpus: A resource for the next generations of speech-to-text. In *Proceedings of LREC 2004*.

Clashen, H., & Muysken, P. (1986). The availability of universal grammar to adult and child learners - a study of the acquisition of German and word order. *Second Language Research, 2*, 93–119. doi:10.1177/026765838600200201

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29 (1)*, 72–89. doi:10.1093/applin/amm022

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2012). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The*

Quarterly Journal of Experimental Psychology, 66 (5), 843–863. doi:10.1080/17470218.2012.720994

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25,* 104–123. doi:10.1016/j.newideapsych.2007.02.002

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26 (2),* 163–188. doi:10.1177/0267658309349431

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency-effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15,* 850–855. doi:10.3758/PBR.15.4.850

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188. doi:10.1017.S0272263102002024

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27,* 164–194. doi:10.1093/applin/aml015

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *41*, 375–396. doi:10.1002/j.1545-7249.2008.tb00137.x

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582. doi:10.1111/j.1551-6709.2009.01023.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116–124. doi:10.3758/BF03195503

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, *22*, 829–834. doi:10.1177/0956797611409589

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition, 8,* 449–456. doi: 10.3758/BF03213418

Gibbs, R. W. (1992). What do idioms really mean? *Journal of Memory and Language, 31*, 485–506. doi:10.1016/0749-596X(92)90025-S

Gibbs, R. W. (1993). Why idioms are not dead metaphors. In C. Cacciari & P. Tabossi (Eds.), *Idioms:Processing, structure and interpretation* (pp. 57–77). Hillsdale: Lawrence Erlbaum.

Gibbs, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition, 20,* 243–259. doi:10.1016/0010-0277(85)90010-1

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language, 28,* 576–93. doi:10.1016/0749-596X(89)90014-4

Godfrey, J. J., Holliman, E. C., McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *IEEE ICASSP* (pp. 517–520). doi:10.1109/ICASSP.1992.225858

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

Gollan, T. H., Montoya, R. I., Cera, C. M., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58,* 787–814. doi:10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140,* 186–209. doi:10.1037/a0022256

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes–no vocabulary test: Correction for guessing and response style. *Language Testing, 19,* 227–245. doi:10.1191/0265532202lt229oa

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44,* 325–343. doi:10.3758/s13428-011-0146-0

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwisterlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*, 12–31. doi:10.1037/0278-7393.34.1.12

MacWhinney, B. (2008). A Unified Model. In P. Robinson & N. Ellis (Eds.), Handbook of Cognitive Linguistics and Second Language Acquisition. Mahwah, NJ: Lawrence Erlbaum Associates.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14 (*8*),* 348–356. doi:10.1016/j.tics.2010.06.002

Morgan-Short K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and Implicit Second Language Training Differentially Affect the achievement of

Native-like Brain Activation Patterns. *Journal of cognitive neuroscience, 24 (4),* 933–947. doi:10.1162/jocn_a_00119

Nickels, S., Opitz, B., & Steinhauer, K. (2013). ERPs show that classroom-instructed late second language learnersrely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters, 557,* 107–111. doi:10.1016/j.neulet.2013.10.019

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50,* 417–528. doi: 10.1111/0023-8333.00136

Paradis, M. (2004). A Neurolinguistic Theory of Bilingualism. Amsterdam: John Benjamins.

Patil, U., Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2016). A computational evaluation of sentence processing deficits in aphasia. *Cognitive Science, 40 (1),* 5–50. doi:10.1111/cogs.12250

Pinker, S. (1991). Rules of language. *Science, 253,* 530–535. doi:10.1126/science.1857983

Pinker, S. (1999). Words-and-rules: The ingredients of language. Basic Books.

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

Prince, A., & Pinker, S. (1988). Rules and connections in human language. *Trends in Neurosciences, 11,* 195–202. doi:10.1016/0166-2236(88)90122-1

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

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011a). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27(2),* 251–272. doi:10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011b). 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. doi:10.1037/a0022531

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In S. Gries & D. Divjak (Eds.), *Frequency effects in language*. Berlin, Germany: Mouton de Gruyter.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18,* 419–437. doi:10.1017/S1366728914000674

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63 (1),* 121–159. doi:10.1111/j.1467-9922.2012.00730.x

Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour, 18,* 523–534. doi:10.1016/S0022-5371(79)90284-6

Szudarski, P., & Conklin, K. (2014). Short-and Long-Term Effects of Rote Rehearsal on ESL Learners' Processing of L2 Collocations. *TESOL Quarterly, 48 (4),* 833–842. doi:10.1002/tesq.201

Tabossi, P., Fanari, R., & Wolf, K. (2008). Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34,* 313–327. doi:10.1037/0278-7393.34.2.313

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition,* 37, 529–540. doi:10.3758/MC.37.4.529

Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 1126–1138. doi:10.1037/0278-7393.20.5.1126

Tremblay, A., & Baayen, 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: The Continuum International Publishing Group.

Tremblay, A., Derwing, B., Libben, G., & Westbury, G. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*, 569–613. doi:10.1111/j.1467-9922.2010.00622.x

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92 (1–2),* 231–270. doi:10.1016/j.cognition.2003.10.008

Ullman, M. T., & Walenski, M. (2005). Reply: Moving past the past tense. *Brain and Language, 93,* 248–252. doi:10.1016/j.bandl.2004.10.004

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In: Schmitt N (ed.) *Formulaic Sequences*. Amsterdam: John Benjamins, 153–72.

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. doi:10.1044/jshr.2403.330

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 616–627. doi:10.1037/0096-1523.28.3.616

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review, 19,* 73–80. doi:10.3758/s13423-011-0179-5

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition, 35 (3),* 451-482. doi:10.1017/S0272263113000107

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

Appendix A. Materials used in the Phrasal-Decision Task, with the frequency per million words.

1.	A lot of places A lot of days	10.45 0.55	14.	Out of the house Out of the game	9.75 0.7
2.	A lot of work A lot of years	14.7 1.9	15.	Where do you live Where do you work	44.8 2.6
3.	All over the place All over the city	21.45 0.65	16.	A lot of rain A lot of blood	4.65 0.2
4.	Don't have to worry Don't have to wait	15.3 1.4	17.	Don't have any money Don't have any place	2.35 0.25
5.	Don't know how much Don't know how many	12.8 7.8	18.	Going to come back Going to come down	1.35 0.4
6.	Go to the doctor Go to the beach	16.7 5.65	19.	Have to be careful Have to be quiet	5.9 0.15
7.	How do you feel How do you do	29.6 4.95	20.	I have a sister I have a game	4.9 0.1
8.	I don't know why I don't know who	35.15 7	21.	I have to pay I have to play	1.8 0.1
9.	I have a lot I have a little	26.45 8.95	22.	I want to say I want to sit	3.6 0.2
10.	I have to say I have to see	15.4 0.95	23.	It was really funny It was really big	2.65 0.15
11.	I want to go I want to know	9.1 2.95	24.	Out of the car Out of the box	2 0.2
12.	It's kind of hard It's kind of funny	13.3 7.2	25	We have to wait We have to leave	1.65 0.25
13.	On the other hand On the other end	27.15 3.95	26	You like to read You like to try	1.55 0.1

Table 1. *Mean (SD) age and educational attainment broken by Language group (native, immersion exposure, and classroom exposure).*

	Native	Immersion exposure	Classroom exposure
Age (years)	26.85 (4.06)	28.74 (5.1)	21.12 (1.24)
Educational attainment	5.89 (1.5)	6.48 (1.53)	5.4 (0.82)

Educational attainment values range from 1 to 8 (1 = Less than High School, 2 = High School, 3 = Professional Training, 4 = Some College, 6 = Some Graduate School, 7 = Masters, 8 = PhD).

Table 2. S*elf-report information on English acquisition, exposure, and proficiency of the two non-native groups of participants (immersion exposure, and classroom exposure).*

	Immersion exposure		Classroom exposure		
	Average (SD)	Range	Average (SD)	Range	p-value[+]
English acquisition (years)					
Age start acquisition	10.26 (2.92)	3-14	7.4 (2.52)	3-12	.0001
Age became fluent	20.7 (5.84)	13-32	16.16 (2.03)	13-22	.001
Current Exposure to English					
Family (0-10)	1.56 (2.64)	0-10	0.36 (0.7)	0-3	.033
Friends (0-10)	6.93 (2.56)	1-10	3.88 (2.11)	0-8	.0001
Reading (0-10)	7.89 (1.95)	3-10	6.44 (2.71)	2-10	.031
Classroom inst. (0-10)	5.93 (4.17)	0-10	6.64 (3.08)	0-10	.49
Self inst. (0-10)	3.3 (3.71)	0-10	4.32 (2.88)	0-10	.27
Watching TV (0-10)	6.52 (3.03)	1-10	6.68 (3.22)	0-10	.77
Radio/music (0-10)	6.19 (2.39)	2-10	6.8 (2.68)	1-10	.44
TOTAL (0-70)	38.63 (11.22)	23-70	34.68 (11.16)	12-56	.21
Months in an English-speaking country	40.93 (47.08)	3-205	2.44 (2.2)	0-12[*]	.0001
LexTALE score[§]	0.59 (0.18)	0.25-0.93	0.52 (0.13)	0.25-0.8	.08
Self-rated English proficiency (ILR)[ζ]					
Speaking	78.7% (15.66)	46-100	78.68% (12.96%)	44-95	.99
Reading	87.74% (11.86)	62-100	87.68% (12.8%)	62-100	.82
Listening	79.67% (18.03)	40-100	80.64% (12.81%)	47-100	.95

[*] No participant in the classroom exposure group have lived more than three consecutive weeks in an English-speaking country; the number of months reported here is the sum of all the periods abroad.

[+]p-values are the result of t-tests comparing the immersion exposure group versus the classroom exposure group.

[§] Score range: 0-1, higher scores indicate better performance (I_{SDT} scoring method; Huibregtse, Admiraal, & Meara, 2002).

[ζ] We computed the percentage of "yes" answers in each language domain (39, 21, and 30 statements in the speaking, reading, and listening domains, respectively).

Table 3. *Mean (SD) frequency and logarithmic (Log) frequency as well as number (N) of syllables, phonemes, letters, and senses of the words used in the two conditions (High frequency and Low frequency) of the LDT.*

	High frequency	Low frequency	*t*	*p-value*
Word frequency (CELEX)	109.3 (133.99)	5.45 (2.85)	6.17	.0001
Log frequency (CELEX)	2.2 (0.27)	0.77 (0.18)	19.48	.0001
N. Syllables	1.15 (0.37)	1.15 (0.37)	< 1	.999
N. Phonemes	3.4 (0.6)	3.6 (0.82)	< 1	.38
N. Letters	4.5 (1)	4.3 (0.87)	< 1	.5
N. Senses	8.8 (4.24)	5.25 (3.51)	2.89	.01

Table 4. *Accuracy rates (% correct) per frequency condition (high and low) and pseudo-words in each group of participants in the Single Word Lexical Decision Task.*

	Native	Immersion exposure	Classroom exposure
High frequency	95.74	96.3	98.8
Low frequency	88.89	87.59	86.6
Pseudo-words	93.15	87.63	86.64

Table 5. *Accuracy rates (%) per phrases and fillers in each group of participants in the Phrasal-Decision Task.*

	Native	Immersion exposure	Classroom exposure
Phrases	98.6	96.2	94
Fillers	90.3	86	85.2

Table 6: *Mixed-effect regression. Significance obtained using the lmerTest package in R. Formula: log(RT.clean) ~ log(chunk.freq) * group + firstBlock + dmdx.order + nchar + log(unigram4) + bigram3.r + trigram2.r + (1 + log(chunk.freq) | subject) + (1 | UniqueItem).*

Fixed Effects	Coef.	SE	*t*	*p-value*
Intercept	6.01	.01	49.78	< .001
Log(phrase-freq)	-.03	0.1	-3.06	< .001
Group-immersion	.06	.05	1.06	> .3
Group-classroom	.02	.05	4.1	< .001
Block-second	-.06	.007	-9.06	< .001
Trial-num	.003	.004	2.43	< .05
Num-letters	.05	.005	10.14	< .001
Log(unigram4)	.01	.009	1.96	= .055
Log(bigram3)	.007	.01	0.61	> 0.5
Log(trigram2)	.02	.01	1.44	> 0.1
Log(phrase-freq) X group-immersion	.001	.005	0.35	> 0.7
Log(phrase-freq) X group-classroom	-.002	.005	-.38	> 0.7

Table 7: *Mixed-effect regression only with immersion learners. Significance obtained using the lmerTest package in R. Formula: log(RT.clean) ~ log(chunk.freq) * timeUS + firstBlock + dmdx.order + nchar + log(unigram4) + bigram3.r + trigram2.r + (1 + log(chunk.freq) | subject) + (1 | UniqueItem).*

Fixed Effects	Coef.	SE	*t*	*p-value*
Intercept	6.01	.01	45.75	< .001
Log(phrase-freq)	-.03	.01	-2.68	< .001
timeUS	-.02	.01	-2.5	< .05
Block-second	-.05	.01	-4.08	< .001
Trial-num	.004	.002	0.99	> .03
Num-letters	.06	.005	11.13	< .001
Log(unigram4)	.01	.009	1.56	> .01
Log(bigram3)	-.002	.01	-0.2	> 0.8
Log(trigram2)	.02	.01	1.57	> 0.1
Log(phrase-freq) X timeUS	-.001	.01	0.12	> 0.9

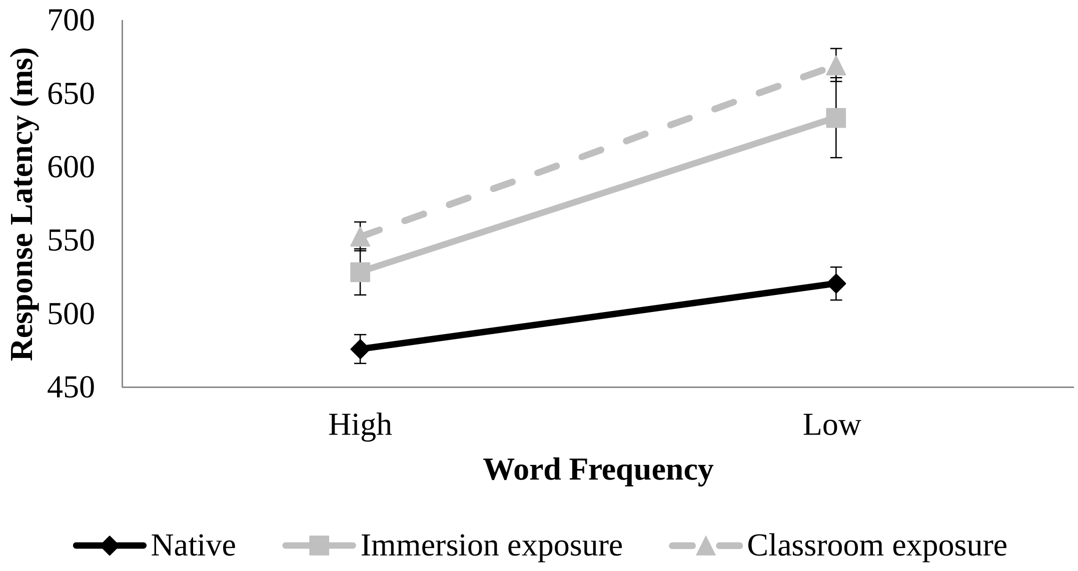

Figure 1. Mean response latencies as a function of word frequency (high vs. low), and Group of Participants (Native, Immersion exposure, and Classroom exposure) in the Single Word Lexical Decision Task. For the sake of visualization, mean RTs are shown in milliseconds instead of log(RT).

Footnotes

[1] This does not mean, however, that multiword phrases are stored in the lexicon or treated as unanalyzed wholes. Indeed, language users are also sensitive to the frequency of the smaller parts of multiword phrases, indicating that they have an internal structure.

[2] Participants' L1 varied much more in the immersion group than in the classroom group. Nevertheless, only six participants in the immersion group had an L1 with a non-Latin script. All results reported in this article remain the same when those six participants were removed from the analyses.

[3] An alternative explanation for the larger word frequency effects in non-natives is that of language competition effects, which is based on the two languages of a bilingual being simultaneously activated by the conceptual system (for a detailed rationale of this account, see Diependaele et al., 2012). This forces bilinguals to discriminate a given word from both its L1 and L2 neighbors. Considering that (a) any lexical discrimination process is particularly costly for low frequency words with high frequency neighbors, and that (b) L1 words are assumed to be of higher frequency than L2 words, the larger frequency effect in L2 is the result of the stronger competition from L1 words. As with the "lexical entrenchment account", the language competition account would also predict larger multiword frequency effects in non-native speakers, which were not found.