The Effect of Input Modes and Number of Exposures on the Learning of L2 Binomials

Sara Alotaibi[1], Ana Pellicer-Sánchez[2], and Kathy Conklin[3]

[1]University of Jeddah, [2]University College London, [3]University of Nottingham

Abstract

Despite the importance of mastering different types of formulaic sequences in a second language, little is known about the relative effect of different input modes on their acquisition. This study explores the learning of a particular type of formulaic language (binomials) in three input modes (reading-only, listening-only, and reading-while-listening) at different frequencies of exposure (2, 4, 5 and 6 occurrences). Arabic learners of English were presented with three stories, each in a different mode, that contained novel binomials (e.g., *wires and pipes*) and existing binomials (e.g., *brother and sister*). Two post-tests (multiple-choice and familiarity ratings) assessed learners' knowledge of the binomials. Results showed that reading-only and reading-while-listening led to better performance on the tasks than listening-only. Frequency of exposure had an effect on the perceived familiarity of binomials.

Keywords: incidental learning, formulaic language, reading, reading-while-listening, listening.

Introduction

Multi-word sequences (MWSs) – also referred to as 'formulaic language' (Wray, 2002) – are recurring patterns consisting of multiple words. MWSs are generally identified and/or defined based on how frequently the combination of words occurs in corpora and whether the co-occurrence of words is random or not (e.g., Biber, Johansson, Leech, Conrad & Finegan, 1999). MWSs are ubiquitous in the English language and are therefore a major component of language competence. Although estimates have varied, research has indicated that formulaic sequences are widespread in both spoken (e.g., Biber et al., 1999) and written discourse (Erman & Warren, 2000) and that they serve a crucial function in language use, especially when their pragmatic value is considered (e.g., Conklin & Schmitt, 2008). The importance of acquiring formulaic sequences for achieving high levels of competence has led to a surge in research exploring the conditions that facilitate their acquisition. Many studies have shown that formulaic sequences can be learnt from explicit, focused instruction (e.g., Laufer & Girsai, 2008; Webb & Kagimoto, 2009). However, the large amount of formulaic language that learners need to acquire, together with the often-limited classroom time, points to the need to explore methods that could be exploited in- and out-side of the classroom (Pellicer-Sánchez, 2017). Recent studies have suggested that formulaic sequences can be learnt incidentally from different input modes: reading (Pellicer-Sánchez, 2017; Vilkaite, 2017), reading-while-listening (Webb, Newton, & Chang, 2013; Webb & Chang, 2020), listening (Webb & Chang, 2020), and viewing (Puimège & Peters, 2019, 2020). Importantly, with the exception of Puimège and Peters (2019, 2020), who looked at a range of formulaic sequences, the majority of these studies have focused on the acquisition of collocations. The effectiveness of different input modes on the incidental acquisition of other types of formulaic sequences needs to be explored more fully.

Binomials are one type of formulaic language, and although many binomials are more frequent than idioms, they have been studied far less (Siyanova-Chanturia, Conklin, & Van Heuven, 2011). Binomials consist of two lexical items of the same lexical class joined by a coordinating conjunction and generally have a conventional word order that may be determined by semantic, orthographical, social, phonological and frequency-based factors (e.g., male before female as in *men and women*; Benor & Levy, 2006). They can be reversible, that is, the words can change position (e.g., *salt and pepper*) or irreversible (i.e., the meaning is anomalous when reversed, as in *hit and run*; Malkiel, 1959). Binomials vary on a continuum from transparent to opaque; the meaning of transparent binomials is directly derived from the two content words (e.g., *tall and short*), while the meaning of opaque binomials is not evident from the content words (e.g., *by and large*; Farghal & Jabber, 1995). For transparent binomials, when the individual words are already known to learners, developing the form-meaning connection should not pose a challenge. What might be more difficult is learning the conventional or preferred word order (it is *salt and pepper* not *pepper and salt*). The present study examined whether L2 learners could incidentally acquire new (transparent) binomials in three different input modes (reading-only, listening-only and reading-while-listening) and compared performance on novel binomials to existing binomials.

Background

Incidental Learning of Individual Words and Formulaic Sequences

In language acquisition, learning new words without the intention of doing so is referred to as incidental vocabulary acquisition (van Zeeland & Schmitt, 2013). It occurs when vocabulary learning is the 'by-product' of another activity like reading (Huckin & Coady, 1999). Learning new words is described as incidental when learners are *not* forewarned that a

vocabulary test will follow (Hulstijn, 2001). Incidental learning occurs through mere exposure to new words through various types of input (e.g., written and/or spoken) when learners focus on either understanding a story or reading/listening for pleasure. Research on incidental vocabulary learning has shown that new lexical items can be acquired from reading (e.g., Horst, 2005; Malone, 2018; Pellicer-Sánchez & Schmitt, 2010; Vu & Peters, 2020; Zahar, Cobb, & Spada, 2001), reading-while-listening (e.g., Malone, 2018; Vu & Peters, 2020; Webb et al., 2013), viewing (e.g., Peters & Webb, 2018; Rodgers & Webb, 2019) and, although with generally smaller gains, from listening (e.g., Vidal 2003; van Zeeland & Schmitt, 2013).

A few studies have looked at the incidental learning of formulaic sequences from reading. Szudarski (2012) compared EFL learners' acquisition of verb-noun collocations in two conditions: reading-plus treatment and reading-only. Participants in the reading-plus treatment group read stories that contained target collocations and completed explicit exercises focused on collocational patterns, while those in the reading-only group were simply asked to read the same stories. The results showed that the collocational knowledge of those in the reading-plus treatment group was significantly better than that of learners in the reading-only group and that this was true on both the productive and the receptive tests. In contrast, there was no significant difference between the reading-only group and the no-treatment control group, which led the author to conclude that reading-only did not contribute much to the learners' knowledge. Similarly, Szudarski and Carter (2014) investigated L2 learners' acquisition of infrequent verb-noun (e.g., *take a swipe*) and adjective-noun (e.g., *quick retort*) collocations in two conditions: reading-only and reading with target words underlined. They found that reading a story with underlined collocations lead to significant gains in their form recall and recognition, but that reading-only did not.

Pellicer-Sánchez (2017) examined the incidental acquisition of collocational knowledge when learners encountered adjective-pseudoword collocations (e.g., *dangerous bancel* for *dangerous criminal*) while reading. She found that exposure to the input led to collocational gains in terms of form recall and form recognition at a rate similar to that of learning the form and meaning of a single word. Vilkaitė (2017) looked at the incidental acquisition of both adjacent verb-noun collocations (i.e., the components directly follow each other, as in *to spend time*) and nonadjacent verb-noun collocations (i.e., intervening words between the two components, as in *to spend a lot of time*) from reading. The results showed that learners had equivalent gains for adjacent and nonadjacent collocations at the recognition level, while the learning gains were negligible at the recall level for both types.

Despite aural input playing a major role in overall language development (Vandergrift, 1999), research on learning from L2 listening is relatively scarce and limited to the learning of single words. Vidal (2003) found that listening to lectures containing target words resulted in vocabulary learning and that learners with a larger vocabulary size gained greater vocabulary knowledge. Van Zeeland and Schmitt (2013) showed that listening to passages containing target items that were repeated 3, 7 and 11 times resulted in gains in different aspects of vocabulary knowledge, with very small gains on a meaning-recall test compared to those on form- and grammar-recognition tests.

Due to the ephemeral nature of listening, it might be difficult for L2 learners to recognize words in aural input and learn new lexical items from it. It is, however, thought that exposure to written plus spoken input may help low-proficiency learners to develop their 'auditory discrimination skills' (Vandergrift, 2007). Reading-while-listening allows learners to follow the written words as they listen to the pronunciation. Conklin, Alotaibi, Vilkaitė, and Pellicer-Sánchez (2020) used eye movement measures to explore how reading patterns were impacted by the simultaneous presentation of auditory input. They examined whether

fixations on words in a reading-while-listening mode were aligned with the audio. In other words, whether readers fixated on the word *tea* when they heard it. The overall results showed that L2 speakers were ahead about 60.7% of the time, behind 7.1 % of the time, and aligned with the audio only 21.4 % of the time. Thus, reading was mostly unaligned with the audio. It is unclear how the auditory and written input are integrated when reading and listening are not aligned and, therefore, how reading-while-listening might benefit learners. The authors speculated that exposure to written text provides listeners with visual cues for the upcoming spoken word, hence helping (1) speed word identification and (2) segment individual words from the continuous speech stream and (3) match the spoken form with the written form.

Research with single words has shown that reading-while-listening is an effective source of L2 vocabulary growth. Chang (2009) showed that reading-while-listening to stories resulted in better performance than listening-only in terms of vocabulary learning and comprehension. Reading-while-listening has also been shown to be effective for the acquisition of formulaic sequences. Webb et al. (2013) investigated the incidental acquisition of verb-noun collocations (e.g., *blow your nose*) when learners simultaneously read and listened to a modified graded reader. Overall, the results confirmed that reading-while-listening led to increased collocational knowledge at the level of form recall and recognition.

Some studies have compared the effectiveness of different sources of input. Brown, Waring, and Donkaewbua (2008) explored incidental vocabulary acquisition of single words in three input modes (reading-only, listening-only and reading-while-listening) and the effect of number of exposures. Results showed that new words were learned in all three input modes, but the most learning occurred from reading-while-listening and the least in listening-only. Reading-while-listening and reading-only led to similar gains. Vidal (2011) compared

incidental learning from academic reading and listening and found that both immediate and delayed vocabulary gains were higher for reading than listening, especially for low-proficiency students. Further, Malone (2018) found that a reading-while-listening group had an advantage over a reading-only group in a form-recognition task. In a recent study, however, Vu and Peters (2020) did not find a difference in single word gains between reading-while-listening and reading-only. Similarly, Feng and Webb (2020) compared the acquisition of single words from reading, listening and viewing and found similar gains in the three modes.

To the authors' knowledge, only one study has compared the effectiveness of different input modes on the acquisition of formulaic sequences. Webb and Chang (2020) compared the effect of reading, listening, and reading-while-listening with EFL students using graded readers. Findings showed that all three input modes contributed to the incidental learning of L2 multi-word combinations (e.g., *unpack bag*), with the reading-while-listening condition resulting in the highest immediate post-test gains, followed by both the listening-only and the reading-only conditions. Unlike the findings of Brown et al. (2008) and Vidal (2011), these results showed that listening-only had an effect on learning that was similar to that of reading-only, suggesting that listening may play a more important role in the learning of formulaic sequences than for single words. This advantage is interpreted by the authors as pointing to the important role of intonation, which may have contributed to acquiring the prosodic form of the sequences. As Lin (2018) argues, the prosodic cues in spoken input can (1) guide the segmentation of input hence promoting the acquisition of L2 formulaic sequences and (2) potentially facilitate how they are processed and stored in the brain.

Overall, research suggests that successful incidental learning of new lexical items can take place from all three input modes, but that reading-while-listening seems to have an advantage over the other modes. However, there is contradictory evidence about the benefit

of reading-while-listening relative to reading-only, and of reading-only relative to listening-only. Webb and Chang (2020) argue that these differences may be due to the input modes having differing effects on the incidental learning of single versus multi-word units. However, there is very limited empirical research to support such a claim. Crucially, we need evidence about a much wider range of MWSs.

As a final point, it is important to note that in the literature on incidental vocabulary learning, most studies have employed a battery of tests to measure learners' knowledge of new words at different levels of mastery. For example, in Brown et al.'s (2008) study, the three learning conditions (reading-only, listening-only, and reading-while-listening) were each followed by a test. It might be argued that after the first test, participants would expect a test after the other input conditions. This makes it difficult to draw concrete conclusions about *incidental* learning. Indeed, this is a common problem in incidental learning studies: once assessment has been carried out, any subsequent learning may no longer be incidental as participants might try to learn (i.e., intentionally learn) in case there are further tests. While keeping this caveat in mind, in line with the much of the literature, learning in the current study is referred to as incidental, as items were not explicitly taught nor were participants told that they would be tested.

The Effect of Number of Exposures on Incidental Vocabulary Learning

It is widely acknowledged that new lexical items are learned gradually and incrementally through repeated encounters in context. Research on single words has demonstrated a positive effect of number of exposures on learners' incidental gains from reading (e.g., Brown et al., 2008; Chen & Truscott, 2010; Vidal, 2011; Waring & Takaki, 2003; Webb, 2007; Zahar, Cobb, & Spada, 2001). However, in the listening modality, studies have demonstrated either no or marginal effects for one or repeated exposures (e.g., Brown et al.,

2008; van Zeeland & Schmitt, 2013; Vidal, 2011). More recently, Feng and Webb (2020) found that frequency of exposure was not related to incidental vocabulary gains in any of the conditions they examined (i.e., reading, listening, and viewing), calling for more research in the area.

Very little evidence exists about the effect of frequency of exposure on the acquisition of formulaic language, and it is somewhat contradictory. Pellicer-Sánchez (2017) found that differences in the number of exposures to collocations (4 and 8 times) in reading had no significant effect on their acquisition. This could be due to the narrow range in frequencies of the target items. In Szudarki and Carter's (2014) study, there were greater gains in terms of form recognition of collocations with 6 encounters and form recall with 12 encounters. Similar to single-word research, this suggests that different levels of exposure are needed for form recognition and recall of MWSs.

In reading-while-listening, Webb et al. (2013) found that repeated exposure (1, 5, 10 and 15 times) had a positive effect on the learning of collocations at both the receptive and productive levels. In particular, 5 encounters to target items led to significant learning and 15 encounters contributed to significantly greater gains than the other frequencies. Similarly, Webb and Chang (2020) found a significant positive correlation ($r = .61$) between number of exposures (between 1 and 16) and vocabulary gains for collocations. Importantly, this was modulated by input mode, with a significant positive correlation between number of exposures and vocabulary gains in reading-while-listening but only moderate and non-significant correlations in reading-only and listening-only.

A recent meta-analysis by Uchihara, Webb, and Yanagisawa (2019) provides evidence for the complex relationship between number of exposures and learning. Uchihara and colleagues examined studies reporting the correlation coefficient between number of exposures and single word learning. Their goal was to explore the overall relationship

between frequency of encounters and incidental vocabulary learning, as well as whether and to what extent other variables might account for the variability in repetition effects. Results showed that there was a medium effect ($r = .34$) of repetition on incidental vocabulary learning. There was significant variation in the size of frequency effects across the studies, which was explained by a range of variables: learner variables (e.g., age, vocabulary knowledge); methodological differences (e.g., nonword use, whether or not participants were informed of an upcoming test, test format); and treatment (spacing, mode of input, visual aids, engagement, range in number of exposures). Particularly relevant for the present investigation is the finding of the association between input mode and repetition. Although there were no significant differences between four input modes (reading-only, listening-only, reading-while-listening, and viewing), repetition appeared more beneficial in reading-only ($r = .41$) and listening-only ($r = .39$) than in reading-while-listening ($r = .28$) and viewing ($r = .22$). The authors concluded that, although number of exposures is an important predictor of incidental vocabulary learning, it is one of many factors affecting vocabulary learning in meaning-focused input. This provides a rationale for the current study, which explores the impact of and relationship between input mode and number of exposures in the incidental learning of MWSs.

As pointed to by Uchihara et al. (2019), the seemingly conflicting findings across studies could be explained by methodological differences: for instance, the aspects of knowledge measured, test formats, and the length of time between encountering the items and the testing phase. More research is needed to gain a better understanding of the role of number of exposures and how it interacts with factors such as input mode. Importantly, the focus of previous MWS research has been on the acquisition of collocations. While this is a frequent and important type of MWS, the multifarious nature of formulaic sequences means that we cannot necessarily generalise findings from one type of MWSs to others. A notable

exception to the focus on collocations is the research by Puimège and Peters (2019, 2020) that looked at the acquisition of different types of formulaic sequences (e.g., compounds, idioms, phrasal verbs, collocations, similes, binomials) from watching English language television. Further, an eye-tracking study on non-idiomatic reversible binomials by Siyanova-Chanturia et al. (2011) found that binomials in their more frequent 'forward' form (e.g., *bride and groom*) were read more quickly than their less frequent 'reversed' form (e.g., *groom and bride*) by L1 speakers and more proficient L2 speakers.

Studies have mainly looked at gains in learners' knowledge of form and meaning (at receptive and productive levels of mastery) using multiple-choice and translation tests. The use of other measures would allow us to explore gains at potentially earlier and less explicit stages of learning. For example, familiarity has been shown to be related to the comprehension of formulaic language (e.g., Aljabri, 2013; Nippold & Taylor, 2002) and is therefore an important aspect to examine when investigating the learning of new formulaic sequences. While a number of psycholinguistic studies have used familiarity ratings as a criterion for item selection (e.g., Carrol & Conklin, 2019) and to provide evidence supporting the role of word familiarity in word/phrases recognition (e.g., Isobe, 2011), to our best knowledge, there has not been research on vocabulary acquisition that used familiarity ratings as a measure to assess whether target items were familiar after exposure.

To address the gaps in the literature, the present study examined the effect of three input modalities (reading-only, listening-only, and reading-while-listening) on the incidental acquisition of an under-researched type of MWS, *binomials*. The study focused on transparent, compositional binomials where the constituents are known to the learners and the lexical aspect to be learnt is the association among the two components and the order in which those components appear (e.g., *fish and chips* not *chips and fish*). The study aimed to identify whether L2 learners can acquire knowledge of the correct order of the binomials

incidentally from various input modes. The role of number of exposures was also examined. In order to assess knowledge of binomials, different measures were used: a multiple-choice form recognition task and a familiarity rating task. The following research questions were addressed:

1. Do L2 learners acquire new binomials (wires and pipes) from meaning-focused input?

2. Do the different input modes (reading-only, listening-only or reading-while-listening) lead to different performance on post-tests?

3. Does number of exposures (2, 4, 5 and 6) affect the incidental learning of binomials?

Methodology

Participants

Thirty-nine female students studying at a university in Saudi Arabia took part in this study (BA level = 34; MA level = 5). For practical reasons, participants were recruited from two classes: 19 participants were majoring in English and 20 were non-English majors learning English at the English Language Institute. Participants were asked to complete a language background questionnaire in which they self-rated their proficiency in speaking, listening, reading and writing, as well as how often they used English in various contexts: speaking (with family, friends etc.), writing (to family and friends etc.), reading (things in English for academic purposes, work, etc.) and watching/hearing/listening (to TV, radio in English etc.). This should help establish that the participants' exposure to English was similar and that they did not have any particular difficulties with reading and listening. The purpose of the self-reporting measure was to ensure that none of the participants had difficulties with reading

and listening (i.e., ratings below 4), which might influence their ability to understand the written/spoken texts. Participants were also asked to complete a shortened version of the VST (modified from Nation & Beglar, 2007), which will be referred to as a vocabulary knowledge test (VKT). The VKT consisted of 20 items, of which two words were randomly selected from the first 10 bands. A maximum score of 20 could be achieved, which would roughly correspond to a vocabulary size of 10,000 words (Carrol & Conklin, 2014, 2017; Carrol, Conklin, & Gyllstad, 2016). This vocabulary measure simply provided an estimate of participants' vocabulary knowledge and was included in the analysis. Data from five participants, whose shortened VKT indicated a vocabulary of less than 4000 word families, were discarded. Finally, data from two participants were not included in the analysis because they did not complete the experimental procedure. This resulted in 32 participants being included in the analyses (see Table 1 for a summary of their characteristics).

Table 1. Summary of participants' mean age, years studying English, self-rating on a 7-point scale (1 = very low to 7 = native-like) of speaking, auditory comprehension, reading and writing, use of English language in everyday life (0% English to 100% English), as well as their score on the shortened version of the VKT (maximum score of 20) with Standard deviations in parentheses.

	Age	Yrs. studying English	Speaking	Auditory comprehension	Reading	Writing	Use of English	VKT
English majors	20.4	8.73	5.4	6.3	6.00	5.73	52.1%	14.44
	(1.1)	(1.5)	(.76)	(1.1)	(.99)	(1.3)	(13.2)	(1.2)
n=18								
Non-English majors	19.3	7.66	3.8	4.44	4.53	3.55	44.4%	11.9
	(2.2)	(1.9)	(1.7)	(1.05)	(1.6)	(1.95)	(20.7)	(1.8)
n=14								
Both	19.9	8.2	4.6	5.4	5.3	4.7	48.3%	13.1
n=32	(1.6)	(1.7)	(1.2)	(1.1)	(1.3)	(1.7)	16.9	(1.5)

Binomials and Passages

The stimuli consisted of a set of novel and existing binomials. By using a set of novel binomials (e.g., *wires and pipes*), we could examine learning without having to pre-test knowledge. The inclusion of existing binomials allowed us to determine whether repeated exposures influence performance for novel binomials, such that they begin to behave like existing binomials. The nine existing and 24 novel (invented) binomials were taken from a study by Conklin and Carrol (2020) on adult L1 speaker acquisition of lexical (word order) patterns. Existing binomials were all 'noun and noun', were highly frequent, and had a highly conventionalized order (forward form '*son and daughter*' significantly greater than reversed form *'daughter and son'*, $t(11) = 9.23$, $p < .001$). To ensure that the novel binomials were 'plausible' phrases in English, they occurred at least once in either order in the British National Corpus (BNC, 2007) but not more than 11 times. In Siyanova-Chanturia et al. (2011) binomial phrases had a mean of 247.3 occurrences in the BNC (per 100 million words) while their reversed forms had a mean of 27.4 occurrences, with the latter being considered novel phrases. Based on this, 11 occurrences in the BNC seemed to be a reasonable upper limit for novel items. It is important to note that the items did not exist in English as formulaic items (which was established in Conklin & Carrol, 2020). A native speaker of Arabic judged that the novel items were not formulaic in Arabic. Thus, the novel items should have been unknown to participants. Further, while some binomials can be idiomatic (e.g., *bread and butter*, meaning 'basic needs'), all of the items in the study (existing and novel) were: (1) transparent (i.e., have only a literal meaning), and (2) reversible (i.e., the meaning does not change if the word order of expression is reversed '*wind and rain*' versus *'rain and wind'*). The list of target items is presented in Appendix 1.

The novel binomials were characterized by a lack of a clear word order preference (forward and reversed forms *not* significantly different ($t(24) = 0.91$, $p = .37$). Twenty of the novel binomials were 'noun and noun' as in *spoons and bowls* and four were 'verb and verb' as in *write and phone*.[1] The words in each novel binomial were matched for length (in characters and syllables) and frequency. In terms of frequency, raw frequencies were converted to Zipf values. This measure of word frequency is independent of the corpus size and has been suggested to lead to a better understanding of the word frequency effect (Van Heuven, Mandera, Keuleers & Brysbaert, 2014).[2] To assess the association strength of the content words of both existing and novel binomials, the Edinburgh Associative Thesaurus was used. Table 2 summarizes the characteristics of the stimuli.

---

[1] The novel items did not violate the word order constraints considered in Morgan and Levy (2016).

[2] The Zipf scale goes from 1 to 6 or 7 and can be interpreted as follows: values of 3 and less are low-frequency words, and values of 4 or more are high-frequency words (Van Heuven, Mandera, Keuleers & Brysbaert, 2014). The calculation of Zipf values is straightforward, as it equals log10 (frequency per billion words) or log10 (frequency per million words) + 3. As such, words with a frequency of 1 per 100 million words get a Zipf value of 1, words with a frequency of 1 per 10 million words get a Zipf value of 2, and words with a frequency of 1 per million words get a Zipf value of 3, and so on.

Table 2. Item characteristics for all existing and novel binomials. Mean values are provided with standard deviation in brackets and range underneath. Zipf frequencies are on a scale from 1 to 7; association strength is measured on a scale from 0 to 100; word length is measured in characters/letters and syllables.

	Zipf	Ratio	W1 Length (chars)	W1 length (syll)	W1 Zipf	W2 Length (chars)	W2 length (syll)	W2 Zipf	Association
Existing Binomial	3.29 (0.44) 2.26-3.71	18.53 (15.19) 4.12-46.91	4.11 (1.27) 3-7	1.11 (0.33) 1-2	5.15 (0.48) 4.51-6.24	5.11 (1.45) 3-8	1.44 (0.53) 1-2	4.91 (0.44) 4.00-5.56	0.38 (0.29) 0-0.81
Novel Binomial	1.62 (0.23) 1.3-2.1	1.47 (1.27) 1-5	5.46 (1.22) 4-8	1.21 (0.41) 1-2	4.76 (0.37) 4.02-5.30	5.29 (1.19) 4-8	1.33 (0.41) 1-2	4.69 (0.41) 4.03-5.79	0.01 (0.27) 0 - 0.12

The three passages that were about daily life were taken from a study by Conklin and Carrol (2020) on adult L1 speaker acquisition and adapted such that each contained three existing binomials and eight novel ones (see Appendix 2 for a sample text). The three existing binomials appeared twice in a text and the eight novel ones occurred at different frequency levels (two at each frequency level: 2, 4, 5 and 6). All of the words that made up the texts were high frequency words (belonging to the first 3000 most frequent word families in English), and therefore were likely to be known by the participants. No comprehension difficulties were informally reported by any of the participants, either during or after the treatment.

The mean length of the passages/stories was 1443 words (*Max* = 1494, *Min* = 1402). The stories were divided into 8-9 paragraphs, with a mean length of 167 words per paragraph. Each passage was presented in one of the three modes in a counterbalanced design (see Table 3 below). For listening, the passages were recorded by an L1 speaker of British English at a

normal speech rate (i.e., 3.5 words per second, with a typical speech rate being 3.7 words per second for L1 speakers of English; Goldman-Eisler, 1961), and were all around six minutes long.

Post-tests

Two tests were created to assess different aspects of learners' knowledge of the existing and novel binomials presented in the passage: 1. a multiple-choice test and 2. familiarity ratings. For both the multiple-choice test and the familiarity ratings two categories of 'filler' binomials were created to avoid making the target binomials overly salient: 'filler existing binomials' and 'filler novel binomials'. The filler existing binomials were from Siyanova-Chanturia et al. (2011) and the filler novel ones were created by the researchers. Care was taken to ensure that the filler existing binomials were of low frequency (Mean frequency = 1.5 per million words in the BNC, *SD* = .53). The filler novel binomials had a frequency of 0 in the BNC. Both types of filler items were matched with the target items for part of speech, length, and frequency. The multiple-choice test and familiarity ratings after each passage contained six filler existing binomials and ten filler novel binomials.

In the multiple-choice test, participants' recognition of the form of the binomials was measured. Participants were presented with a list of three existing binomials and eight novel binomials. Each item included three options: the correct form of a binomial, the reversed form and a *neither* option to reduce guessing.[3] It is worth noting that this test does not reflect

---

[3] When the multiple-choice test was administered, the participants were told that in each test item they were going to see three options. The first two options presented two possible orders for the same phrase, and the third option was 'neither'. They had to choose one of the first two options if they felt that one of the orders was preferred in English and sounded more natural, whereas the 'neither' option should be selected if they felt that both orders were equally natural in English and there was no preferred order. The novel filler items had no preferred word order and therefore in those cases the 'neither' option was the correct response. Thus, the multiple-choice test offered three options, which means that chance level, or guessing, is 33.33%.

the standard format of a multiple-choice test used in L2 vocabulary research. Instead, it is designed to measure learners' ability to recognize the correct word order of binomials by asking them to choose between options that vary only in the word order, which is likely very challenging (see Appendix 3).

The familiarity ratings asked participants to rate how familiar they were with the existing and novel binomials on a five-point scale (1= I have never heard/used this phrase, 2 = I've very rarely heard/used this phrase, 3 = I've rarely heard/used this phrase, 4 = I've frequently heard/used this phrase to 5 = I have very frequently very frequently heard/used this phrase).[4] This test included both the existing and novel target binomials that had been encountered in the just seen and/or heard passage as well as the filler existing and novel binomials that had not been encountered in the text (see Appendix 3).

Procedure

The study was carried out in the participants' English class. They provided their informed consent for taking part in the research. Participants were given the three passages, each one in a different mode, and were told that they would answer some comprehension questions (see Appendix 4) about them. A within participant, counterbalanced design was adopted. Thus, a participant read three texts, each in a different input mode. The three passages were presented in the different modes to different but an equal number of participants. Table 3 demonstrates the counterbalance design. Participants were instructed to read as naturally as possible for comprehension (no announcement was made about the posttests). After each passage, participants were given three true-false statements to (1) ensure that they read/listened to the

---

[4] Learning binomials involves associating two words in memory as well as gaining knowledge about their word order, e.g., the words *wires* and *pipes* occur together (collocate) and have a particular word order (*wires and pipes* not *pipes and wires*). The familiarity task may tap into the development of associations between the constituent words and/or knowledge about their canonical word order.

text and (2) check their overall understanding (questions did not contain the target binomials), followed by the two post-tests for the items presented in that passage. The tests were administered and collected after each passage to prevent participants going back to previous tests. The medium of test administration was paper and pencil and participants reading was self-paced in the reading-only condition. The scoring of the multiple-choice task involved giving 1 point for a correct response and 0 points if the response was incorrect or the '*neither*' option was selected. The vocabulary test and language background questionnaire were completed at the end of the session. The entire procedure took approximately 50 minutes.

Table 3. Summary of the counterbalance design

Input Mode	Reading	Listening	Reading-while-listening
Group 1	Passage 1	Passage 2	Passage 3
Group 2	Passage 2	Passage 3	Passage 1
Group 3	Passage 3	Passage 1	Passage 2

Analysis

To assess whether participants read attentively, the reading comprehension questions were inspected. The mean scores showed good comprehension of the passages and independent sample *t*-tests showed no significant differences between the input modes (67.5% reading-only, 71.7% listening-only, and 66.2% reading-while listening): reading-only vs. listening-only $t(31) = -1.71$, $p = .097$, reading-only vs. reading-while-listening $t(31)= .37$, $p = 71$, listening-only vs. reading-while-listening $t(31) = 1.54$, $p = 1.33$. An independent sample *t*-test showed significant differences between participant groups (74.2% correct for English majors and 62.7% for non-English majors), $t(30) = 4.29$, $p = .002$.

The analyses were carried out with the Statistical Package for the Social Sciences (SPSS) software. The data from the two tasks were analysed using Generalised Estimating Equations (GEEs, see Zeger & Liang, 1986), which is an extension of generalized linear models (Owusu-Darko, Kwasi Adu, & Frempong, 2014). While other types of analyses, including generalised linear mixed models (GLMMs), could arguably have been used, GEEs provided the best means of examining the data, as they can be used for a repeated-measures design and accommodate a wide range of different data types (continuous, categorical, etc.).

GEE is based on the number of observations (the number of participants multiplied by the number of items) rather than on the test scores, as in an AN(C)OVA (Peters, 2016). This means that the combination of values of the specified variables uniquely define subjects within the dataset. For example, the combination of 'Participant, Test Type and Score defines, for each case, a particular score (correct or incorrect) on a particular test type for a particular participant. In GEEs the odds ratio (= expβ, or exponential parameter estimate) provides a measure of effect size and expresses the relative chance of an event happening under the different conditions analysed. The odds ratio predicts the change from one level to the next in a factor variable relative to the lowest or highest level.[5] It should be noted that a GEE model does not allow for *pairwise comparisons* of categorical data; hence, only the odds ratio is reported in the analyses of familiarity-ratings tasks.

---

[5] The odds ratios reflect the relative chance of an event of interest happening. If odds ratios < 1, it means that the event or condition is less likely to occur; and if odds ratios > 1, it means that the event is more likely to occur. If the ratio =1 (e.g., when outcome is the same in the compared groups), it implies that there is no difference between the compared conditions. In the current study, 'input mode' has three levels (reading-only, listening-only, reading-while-listening), hence the comparisons are between: (1) the odds of answering correctly when learners are exposed to binomials in reading-only to the odds of answering correctly when they are exposed to binomials in reading-while-listening (reference category) and (2) the odds of answering correctly when they are exposed to binomials in reading-only to the odds of answering correctly when they are exposed to binomials in reading-while-listening. For example, if we're comparing learners' performance in X to Y & odd ratio of X for Y is >1, it means that performance in X is better than Y.

Results

Multiple-choice test

Table 4 presents accuracy in the multiple-choice test. The first set of GEEs explored

performance in this task and included the factors: input mode (reading-only, listening-only,

and reading-while-listening); number of exposures (2x, 4x, 5x, and 6x) as within-subject

factors; and vocabulary knowledge (measured by the shortened VKT) as a covariate; as well

as exploring the two-way interactions between them. We included the Group variable

(English majors vs. non-English majors) in the initial analyses, but since there were no

differences in vocabulary learning between the two groups it was excluded from the analyses.

With a binary dependent variable in each case (correct vs. incorrect), the model employed

a *binomial* distribution and *logit* link. In these analyses (Tables 5 and 6), only the novel

binomials were considered, as they had an exposure manipulation, while the existing

binomials did not. In a second set of analyses, we compared performance on the existing and

a subset of the novel binomials to look at performance on the novel binomials relative to ones

that are likely to have been encountered before.

Table 4. Percent correct responses, means with standard deviations in parentheses, and maximum possible score per item type and condition for the existing and novel binomials in the multiple-choice task at each level of exposure in the different input modes (reading-only, listening-only, reading-while-listening) .

		Existing	**Novel**				**Novel**
		2x	2x	4x	5x	6x	**Total**
Reading-only	Percentage	72%	45%	48.5%	48.5%	55.5%	**49.38%**
	Mean (SD)	2.16 (0.78)	0.90 (0.65)	0.97 (0.67)	0.97 (0.72)	1.09 (0.64)	**3.93 (0.67)**
	Max	3	2	2	2	2	**8**
Listening-only	Percentage	59.3%	28.9%	34%	33%	47%	**35.73%**
	Mean (SD)	1.78 (0.94)	0.58 (0.61)	0.65 (0.64)	0.66 (0.69)	0.94 (0.56)	**2.83 (0.63)**
	Max	3	2	2	2	2	**8**
Reading-while-listening	Percentage	67.7%	53%	63%	61%	50%	**56.75%**
	Mean (SD)	2.03 (0.99)	1.06 (0.74)	1.25 (0.66)	1.22 (0.69)	1 (0.56)	**4.53 (0.66)**
	Max	3	2	2	2	2	**8**
Total	Percentage	66.33%	42.3%	48.5%	47.5%	50.83%	**47.28%**
	Mean (SD)	5.97 (1.63)	2.62 (1.03)	2.87 (1.17)	2.85 (1.30)	3.03 (1.85)	**11.29 (3.41)**
	Max	9	6	6	6	6	**24**

Table 5. Generalized Estimating Equations analysis on the multiple-choice test scores: Test of model effects (input mode and gains)

	Multiple-choice test		
Source	Wald Chi-square	Df	**Sig**
Input mode	25.9	2	**.001**
Number of exposures	6.3	3	**.09**
Input mode*Number of exposures	12.2	6	**.06**
Vocab test	.002	1	**.96**

Table 6. Generalized Estimating Equations analysis on the multiple-choice test scores: Parameter estimates (input mode and gains)

Parameter	B	SE	95% Wald confidence Interval		Wald Chi-square	*P*	Expβ
			Lower	Upper			
Multiple-choice test							
(Intercept)	.83	.48	-.87	1.03	.03	.86	**1.1**
Reading-only	.13	.27	-.40	.67	.23	.63	**1.1**
Listening-only	-.19	.28	-.75	.37	.45	.50	**.83**
Reading-while-listening	0a						**1**

Note. a Set to zero because this parameter is redundant, β: regression coefficient

The model for the multiple-choice test revealed a significant main effect of input mode on participants' performance ($\chi^2(2) = 25.9$, $p < .001$). The odds of correct answers in listening-only did not differ from those in reading-while-listening ($\exp^{-.19} = .83$). Further, the odds of correct answers in reading-only did not significantly differ from those in reading-while-listening ($\exp^{.13} = 1.1$). The post-hoc, pairwise comparisons showed that listening-only led to lower scores than both reading-only ($p = .01$); and reading-while-listening ($p < .001$). There was no significant difference between reading-only and reading-while-listening ($p = .23$). The number of exposures did not predict performance ($\chi^2(3) = 6.32$, $p = .09$). The interaction between input mode and number of exposures was not significant ($\chi^2(6) = 12.22$, $p = .06$), indicating a lack of difference in performance between input modes at the different levels of exposure. The effect of vocabulary knowledge was not significant ($\chi^2(1) = .002$, $p = .96$).

There was no effect of number of exposures. However, differences between 2, 4, 5, and 6 exposures are expected to be relatively small and are unlikely to be picked up when all of the levels are considered together. To further explore the effect of number of exposures, we compared performance on the minimum and maximum number of exposures (two vs. six) for each mode, using a paired sample t-test. Results revealed that items repeated six times were better learned than those repeated two times after listening-only ($t(30) = 2.9$, $p = .01$). There were no significant differences between items repeated six and two times after reading-only ($t(28) = 1.1$, $p = .28$), or reading-while-listening ($t(31) = -.22$, $p = .83$).

In sum, reading-only and reading-while-listening had a similar effect on participants' knowledge of binomials. Further, there was no main effect of number of exposures on participants' performance on the multiple-choice task. However, when comparing the minimum and maximum number of exposures, results showed a significant difference between two and six exposures in the listening-only mode. Finally, participants' vocabulary

knowledge, as measured by the shortened VKT, did not modulate their performance on the task. While the analysis thus far looked at performance on the novel binomials across the number of exposures, another important comparison involves that of the novel binomials to the existing ones (which were only presented twice). In this subsequent analysis, we explored whether: 1) the novel binomials that were presented twice and the existing binomials that were also presented twice demonstrate differences in accuracy; and 2) whether the novel binomials presented six times (the maximum number of exposures) differ in accuracy from existing binomials. The analyses of performance on the multiple-choice task included the factors: input modes (reading-only, listening-only, and reading-while-listening), binomial type (existing 2x, novel 2x, novel 6x) as within-subject factors, vocabulary knowledge (VKT) as a covariate, and the two-way interactions between them. Similar to the previous analysis, the model employed a *binomial* distribution and *logit* link function. Table 4 presents the percentage of correct responses in the multiple-choice test.

The model in Table 7 revealed a main effect of binomial type ($\chi^2(2) = 34.76$, $p < .001$). From Table 8, we see that the odds of correct answers for novel binomials repeated six times were lower than existing binomials repeated twice ($\exp^{-.77} = .47$). The odds of correct answers for novel binomials repeated twice were lower than existing binomials that also appeared twice ($\exp^{-1.08} = .33$). The post-hoc, pairwise comparisons indicate that existing binomials occurring twice yielded better performance than novel binomials repeated twice ($p < .001$); and novel binomials repeated six times ($p = .003$). It also showed that performance on novel binomials seen six times was significantly better than those seen twice ($p = .02$). Furthermore, results revealed a main effect of input mode ($\chi^2(2) = 8.71$, $p = .01$). The odds of correct answers in the listening-only were not significantly different from those in the reading-while-listening

$(\exp^{-.53} = .59)$.[6] The odds of correct answers in the reading-while-listening were not significantly different from those in the reading-only $(\exp^{-.30} = .74)$. The post-hoc, pairwise comparisons revealed that performance in listening-only was worse than in both reading-only $(p = .01)$ and reading-while-listening $(p = .01)$. There was no difference between reading-only and reading-while-listening $(p = .76)$. The

interaction between binomial type and input mode was not significant $(p = .22)$, indicating that the type of target items (novel or existing phrases) did not modulate performance across the input modes. Vocabulary knowledge contributed significantly to the model $(\chi^2(1) = 5.28, p = .02)$: learners with larger vocabulary knowledge performed better on the multiple-choice test.

Table 7. Generalized Estimating Equations analysis on the multiple-choice test scores: Test of model effects (binomial type and gains)

	Multiple-choice test		
Source	Wald Chi-square	Df	**Sig**
Binomial Type	34.76	2	**.001**
Input Mode	8.71	2	**.013**
Binomial Type * Input Mode	5.70	4	**0.22**
Vocab Test	5.28	1	**.02**

---

[6] As an example, this result reflects the overall main effect of 'input mode' only, not the interaction between input mode and binomial type. It compares the odds of correct answers for all types of target items in the listening-only vs. reading-while-listening modes.

Table 8. Generalized Estimating Equations analysis on the multiple-choice test scores: Parameter estimates (binomial type and gains)

Parameter	B	SE	95% Wald confidence Interval		Wald Chi-square	P	Expβ
			Lower	Upper			
Multiple-choice test							
(Intercept)	-.25	.35	-.44	.94	.51	.48	**1.28**
Novel (x6)	-.77	.39	-.15	.01	3.94	.04	**.47**
Novel (x2)	-1.08	.33	-1.7	-.44	10.75	.001	**.33**
Existing (x2)	0a						**1**
Reading-only	-.30	.33	-.95	.35	.82	.36	**.74**
Listening-only	-.53	.30	-1.1	-.07	3.03	.08	**.59**
Reading-while-listening	0a						**1**

Note. a Set to zero because this parameter is redundant, β: regression coefficient

Taken together, these results indicate that participants were more likely to recognize the 'correct' order of existing binomials than novel binomials. Exposure played a role in the recognition of novel binomials; seeing a new phrase six times significantly increased participants' ability to recognize its correct order relative to seeing it twice. Even after having seen the novel binomials six times, knowledge of the correct order of the binomials was significantly lower than existing binomials. Learners' vocabulary knowledge appeared to modulate their performance on the multiple-choice task.

It is possible that responses in the multiple-choice task reflected guessing. In the test, participants were presented with two possible orders for the same phrase (e.g., *fish and chips*, *chips and fish*), and a third 'neither' option. They were instructed to choose the option that sounded more natural or 'neither' if neither sounded better. It could be argued that participants would simply ignore the 'neither' option, which would mean that chance-level or

guessing was 50% (instead of 33.33% if there were three legitimate options). In an attempt to ascertain whether guessing was responsible for the pattern of results, we first examined whether the 'neither' option was indeed a legitimate option and whether participants selected it for the filler novel items (i.e., for which the 'neither' option was the correct option). The data revealed that the 'neither' option for the filler novel items was selected in 13.43% of the cases, while for the target novel items it was selected in 11.48% of the cases. This shows that the 'neither' option was indeed selected by participants. However, the small percentage of 'neither' responses still leaves open the possibility that participants were simply guessing when choosing one of the two other options. Learners' performance on the target novel items had a mean of 47.28, indicating that learning was above chance (if we consider chance to be 33.33% rather than 50%). Finally, we also examined whether there was a difference between the target novel items and filler novel items. Paired sample t-tests showed that there was a significant difference in the scores for the target novel items and filler novel items, $t(31) = 7.74$, $p < .000$. The fact that there was a difference in performance between target items (that were exposed in the reading texts) and filler novel items (those that were not exposed to) indicates that target items were learnt from the treatment.

Familiarity Ratings

Another goal of this study was to determine whether number of exposures and input modes affected participants' familiarity level with novel binomials. Given that the dependent variable was categorical (familiarity ratings), we used an ordinal logistic GEE model, with *Multinomial* distribution, *Cumulative logit* function and *Independent* structure. The following factors were included: input mode (reading-only, listening-only, and reading-while-listening) and number of exposures (2x, 4x, 5x, and 6x) as within-subject factors; vocabulary knowledge (measured by the vocabulary test) as a covariate; and the two-way interactions

between them. Table 9 shows participants' familiarity ratings for existing and novel binomials.

Table 9. Familiarity ratings on a 5-point scale (with 1 = I have never heard/used this phrase and 5 = I have very frequently heard/used this phrase) with standard deviations in parentheses for novel binomials at each level of exposure across the different input modes (reading-only, listening-only, reading-while-listening)

Input Mode	Existing	Novel			
	2x	2x	4x	5x	**6x**
Reading-only	3.91 (0.94)	2.68 (1.05)	3.28 (0.85)	3.25 (0.94)	**3.87** **(0.81)**
Listening-only	3.52 (0.85)	2.59 (0.85)	2.85 (1.00)	3.11 (1.08)	**3.53** **(1.01)**
Reading-while-listening	3.68 (0.97)	2.87 (0.93)	3.32 (1.01)	2.88 (1.05)	**3.73** **(0.87)**

The results in Table 10 revealed a significant main effect of input mode on familiarity, ($\chi^2(2)$ = 25.11, $p < .001$). Based on the odds ratio, presented in Table 11, we see that familiarity ratings were lower in listening-only than in reading-while-listening ($\exp^{-1.07}$ = .34). In contrast, familiarity ratings in reading-only did not significantly differ from those in reading-while-listening ($\exp^{.24}$ = .78). The model also revealed a significant main effect of number of exposures on familiarity scores, ($\chi^2(3)$ = 238.19, $p < .001$). Familiarity ratings were higher for novel binomials seen six times than all other repetitions: two exposures ($\exp^{-1.35}$ = .26), four exposures ($\exp^{-1.35}$ = .26), and five exposures ($\exp^{-1.39}$ = .25). The interaction between input mode and number of exposures was significant ($\chi^2(6)$ = 156.85, $p < .001$). This indicates that number of exposures to novel items varied depending on the input mode.

Looking at the difference between input modes at each exposure, familiarity ratings for items repeated twice were lower in listening-only than in reading-while-listening ($\exp^{-.95} = 2.59$). Familiarity ratings for items repeated five times were higher in listening-only than in reading-while-listening ($\exp^{1.43} = 4.19$) and in reading-only than in reading-while-listening ($\exp^{.51} = 1.66$). The main effect of vocabulary knowledge was significant ($\chi^2(1) = 4.57$, $p = .03$). Greater vocabulary knowledge led to higher familiarity ratings.

Table 10. Generalized Estimating Equations analysis on the familiarity rating task: Test of model effects (input mode and familiarity ratings)

Source	Wald Chi-square	Df	Sig
Input mode	25.11	2	**.001**
Number of exposures	238.19	3	**.001**
Input mode*number of exposures	156.85	6	**.001**
Vocab test	4.57	1	**.03**

Table 11. Generalized Estimating Equations analysis on the familiarity rating task: Parameter estimates (input mode and familiarity ratings)

Parameter	B	SE	95% Wald confidence Interval		Wald Chi-square	P	Expβ
			Upper	Lower			
2x	-1.35	.11	-1.58	-1.13	139.57	.000	**.26**
4x	-1.35	.13	-1.60	-1.09	109.43	.000	**.26**
5x	-1.39	.12	-1.63	-1.17	144.98	.000	**.25**
6x	0a						**1**
RO	.24	.14	-.52	.03	2.99	.08	**.78**
LO	-1.07	.11	-1.29	.84	87.83	.000	**.34**
RL	0a						**1**
2x*Reading-only	-.17	.15	-.47	.13	1.25	.26	**.84**
2x*Listening-only	-.95	.14	.67	1.2	43.19	.000	**2.59**
2x*Reading-while-listening	0a						**1**
4x*Reading-only	-.19	.29	-.75	.37	.44	.51	**1.45**
4x*Listening-only	-.24	.32	-.83.	.39	.52	.47	**1.49**
4x*Reading-while-listening	0a						**1**
5x*Reading-only	.51	.16	.19	.83	9.94	.002	**1.66**
5x*Listening-only	1.43	.17	1.09	1.77	68.22	.000	**4.19**
5x*Reading-while-listening	0a						**1**
6x*Reading-only	0a						**1**
6x*Listening-only	0a						**1**
6x*Reading-while-listening	0a						**1**

Note. a Set to zero because this parameter is redundant, β: regression coefficient

These results showed that participants developed a certain level of familiarity with the novel binomials in the three input modes, with reading-while-listening having an advantage over the listening-only mode. Number of exposures also appeared to have an effect, with greater familiarity appearing at four exposures. This analysis informed us about learners' familiarity with novel binomials only. In order to investigate differences between familiarity ratings to the different types of binomials, another GEE (ordinal model) was carried out that included the factors: input modes (reading-only, listening-only, reading-while-listening), binomial type (existing 2x, novel 2x, and novel 6x) as within-subject variables; vocabulary knowledge (VKT) as a covariate; and the interaction between them. Similar to the previous analysis, the model employed a *Multinomial* distribution, *Cumulative logit* function and *Independent* structure. Results (see Table 12) revealed a main effect of binomial type ($\chi^2(2) = 76.76$, < .001). Based on the odds ratio (see Table 13), there was no significant difference in familiarity ratings between novel binomials seen six times and existing binomials ($\exp^{-.08} = .93$). In contrast, familiarity ratings for novel binomials seen twice were lower than existing binomials ($\exp^{-1.42} = .24$). In addition, there was a main effect of input mode ($\chi^2(2) = 16.14$, $p < .001$). Familiarity ratings were lower in listening-only than in reading-only ($\exp^{-.49} = .61$). Also, familiarity ratings in reading-while-listening did not differ from those in reading-only ($\exp^{-.28} = .75$). The interaction between binomial type and input mode was not significant, $p = .69$. This shows that learners' familiarity with binomials (novel or existing) did not depend on the input mode. Vocabulary knowledge did not contribute significantly to the model, $p = .92$. In sum, participants demonstrated a higher familiarity level with existing binomials over novel binomials seen twice. However, with greater exposure, the novel binomials seen six times became as familiar as existing ones. This indicates that repeated exposure to new phrases in a meaningful context can promote lexical development.

Table 12. Generalized Estimating Equations analysis on the familiarity rating task: Test of model effects (binomial type and familiarity ratings)

Source	Wald Chi-square	Df	Sig
Binomial type	76.76	2	**.000**
Input mode	16.14	2	**.000**
Input mode*binomial type	2.25	4	**.69**
Vocab Test	.010	1	**.92**

Table 13. Generalized Estimating Equations analysis on the familiarity rating task: Parameter estimates (binomial type and familiarity ratings)

Parameter	B	SE	95% Wald confidence Interval		Wald Chi-square	P	Expβ
			Lower	Upper			
Novel (6x)	-.075	.17	-.41	.26	.19	.66	**.93**
Novel (2x)	-1.42	.27	-1.93	-.89	28.45	.000	**.24**
Existing (2x)	0a						**1**
Reading-while-listening	-.28	.27	-.81	.24	1.13	.28	**.75**
Listening-only	-.49	.18	-86	-.13	7.00	.008	**.61**
Reading-only	0a						**1**

Note. a Set to zero because this parameter is redundant, β: regression coefficient

Discussion

The purpose of this study was to investigate the effect of different input modes and number of exposures on the incidental learning of binomials. It also explored an aspect of formulaic language knowledge that previous studies on vocabulary acquisition have not investigated (i.e., the development of learners' familiarity with target items, using familiarity ratings).

In response to our first research question about the effectiveness of different input modes, our results indicate that learners can incidentally develop receptive knowledge of the 'correct' order of novel binomials (as measured by the multiple-choice test) and develop familiarity with them from all three input modes. Notably, in response to the second research question, there were generally no differences in performance in the reading-only and the reading-while-listening modes, and they both lead to better performance than the listening-only mode. This supports previous research on single words (e.g., Brown et al.'s, 2008; Vidal, 2011; Vu & Peters 2020) and points to an interesting similarity between the learning of single words and MWSs. However, the current findings seem to contradict those of Webb and Chang's (2020) indicating that reading-while-listening led to best performance for MWSs, while listening-only and reading-only yielded similar results. They also appear to contrast with the findings of Malone (2018) who found that simultaneous input modalities (i.e., reading-while-listening) led to higher learning outcomes in the acquisition of single words. The relatively small sample size in the current study could explain the lack of difference in performance between reading-while-listening and reading-only modes, and hence the contradictory findings between the current study and previous research (e.g., Webb & Chang, 2020; Malone, 2018). In addition, the difference in results could point to different input modes being more/less beneficial for the learning of different types of MWSs. Listening-only may be more beneficial when learning collocations than binomials. While prosodic cues in spoken input are thought to play a role in the development of formulaic sequences (e.g., Vidal, 2011; Webb & Chang, 2020), it is unclear whether they are equally important for the different types of formulaic language. Future research will need to look at prosodic cues across the types of formulaic sequences and how they might benefit learners. Another possible explanation for the differing results may be related to the use of different tasks. In their post-test, Webb and Chang presented the collocations auditorily (i.e.,

collocations were presented aurally, and participants had to write their meanings on a corresponding test sheet), while in the current study they were presented in a 'reading' mode. This could have disadvantaged items that were learned orally. The lack of a clear advantage for reading-while-listening in the current study may imply that its advantages are limited to tasks involving aural input. There is evidence suggesting a test-modality congruency effect (Sydorenko, 2010; Jelani & Boers, 2018). Sydorenko (2010) found that a video with audio group performed better on an aural than on a written recognition test, and the reverse pattern was found for a video with caption group. Jelani and Boers (2018) also found that a group watching video clips with captions outperformed a group watching uncaptioned clips on a written test, but not on a test presented in aural format.

As discussed earlier, much of the L2 vocabulary research suggests that reading-while-listening leads to greater gains than either reading- or listening-only. However, the results of the current study do not generally show a difference in performance in the reading-while-listening and the reading-only modes. This may be related to narration pace in the reading-while-listening, which may not have been aligned with the participants' reading speed. While the reading pace was intended to be natural, it is possible that (some of) our participants were unable to follow the narration. Research on children's reading comprehension from reading-while-listening has demonstrated that some readers do not benefit from combining auditory input with reading because the narration rate is either slower or faster than the readers' reading rate (e.g., McMahon, 1983). As Chang (2009) notes, the reading-while-listening mode will be more beneficial if learners have a reading speed that matches the speech rate in the auditory input. In addition, similar to what Vu and Peters (2020) suggested, it may have been that some participants had negative perceptions about reading and listening at the same time, which could have influenced their performance. Notably, the lack of difference between

reading-only and reading-while-listening in the present study supports recent findings with single words (Feng & Webb, 2020; Vu & Peters, 2020).

While it may be tempting to compare performance on incidental vocabulary learning across studies, any such endeavor should be treated with caution as the results are based on very different methodologies. While noting this, we consider our findings in relation to others in the field. Our study yielded the following scores on the multiple-choice test: 49.38% for reading-only; 35.73% for listening-only; and 56.75% for reading-while-listening. This is fairly similar to the percentages reported by Brown et al. (2008) for single-words in a multiple-choice task: 45% for reading-only; 29% for listening-only; and 56.91% for reading-while-listening. However, it contrasts with the findings of Webb and Chang (2020), which reveal lower gains: immediate 12% / delayed 23% for reading-only; immediate 16% / delayed 26% listening-only; and immediate 28% / 57% for the reading-while-listening.

It is important to note that in the current study, while we have talked about the results in terms of incidental learning, the frequent occurrence of binomials in the texts and the fact that learners completed tasks after each session, could have increased the salience of the items leading to more intentional learning. Further, demonstrating lexical gains, in the current study, is considerably different than in other studies in the literature. The binomials in this study were transparent (the meaning is the sum of the parts) and 'learning' amounts to demonstrating command over the word order (form); thus, it is *fish and chips* not *chips and fish* for the existing binomials and *plates and glasses* not *glasses and plates* for the novel ones. On the multiple-choice task learners had to choose from three options – 'forward' form (e.g., *plates and glasses*), the reversed from (e.g., *glasses and plates*), and a 'neither' option. Importantly, performance on the task implies that it is possible for L2 learners to develop knowledge of novel binomial word order.

Concerning the effect of binomial type (novel versus existing), results indicated that learners were better on existing binomials (e.g., *brother and sister*) than novel ones (e.g., *wires and pipes*). Given the fact that the EFL students in the current study have been exposed to English for a number of years (approximately 5 years studying English), these results are promising; they indicate that the students have had exposure to English MWSs, which is reflected in their performance.

With regard to the effect of number of exposures on the incidental learning of binomials in the multiple-choice task, encountering novel binomials an increasing number of times did not have an impact on overall performance in any of the modes. This finding aligns with both Pellicer-Sánchez' (2017) and Szudarski's (2012) studies, which showed that number of exposures did not affect incidental collocation learning from reading. However, this finding contradicts several studies on single words (Brown et al., 2008; Vidal, 2011; Webb, 2007) and studies on collocations (Webb et al., 2013; Webb & Chang, 2020). The lack of an effect of exposure in the present study is likely due to frequency levels that are minimally different (2, 4, 5 and 6 exposures). A significant role of number of exposures on incidental learning of MWSs has been found in studies having a higher number of exposures and wider range (e.g., Webb et al.'s, 2013; Webb & Chang, 2020). Another possible explanation is that more items at each frequency level are needed in order to find a significant difference. Further, as has been speculated in the literature (van Zeeland & Schmitt, 2013; Vidal 2003, 2011), number of exposures may be more effective if the encounters are spread over a greater period of time, as in a naturalistic learning environment, and not a short experimental context.

However, a significant effect of frequency of exposures emerged in the familiarity rating task. Seeing a novel phrase multiple times had a positive effect on its familiarity. Learners rated the existing binomials as being more familiar than novel binomials that were

encountered twice. However, they showed similar levels of familiarity with the existing binomials and novel binomials that were encountered six times in the text. This suggests that repeated exposure incrementally builds knowledge of new MWSs that regularly occur in a classroom 'story' context. It is worth noting that we seem to see earlier development of binomial knowledge in a familiarity-rating task than in form recognition task – one that shows a clear sensitivity to frequency of exposure. Notably, an effect of exposure for familiarity ratings aligns with that finding in the recognition task showing that there was an effect of exposure when comparing novel binomials that occurred twice to those that occurred six times. It may be that to find an effect of exposure on recognition tasks, the exposures need to be sufficiently different from each other, whereas familiarity ratings are more sensitive to smaller differences in number of exposures.

There are a few limitations of the current research that are important to mention. This study only attempted to gather data on L2 learners' immediate knowledge of binomials. However, as recognized by Waring and Takaki (2003), using immediate posttests to assess learners' knowledge, when new lexical items are still fresh in the mind, may result in higher gains than might be found at a delay. Therefore, future research should assess whether learners retain this knowledge over time. In addition, as discussed earlier, the items in the post-tests were presented in a written format. While participants in the listening-only condition had only been exposed to binomials in the listening modality, the post-test presented the binomials in a 'reading' mode, which could have impacted performance. Presenting the items auditorily in the post-tests may have revealed differences in performance between reading-only and reading-while listening that were not captured by the tests in the present study. Other limitations relate more specifically to the multiple-choice test format. As discussed earlier, while we checked that the learning reported in the multiple-choice test was above chance, the participants' guessing could have affected scores. Future research could

also include a control group with participants only completing the tests and receiving no treatment, in order to further control for the possibility of guessing.

Conclusion

This study demonstrates that incidental learning is not limited to the type of lexical items examined thus far in the literature but can be extended to other types of formulaic sequences, namely binomials. All three input modes – reading-only, listening-only, and reading-while-listening – contributed to binomial learning. Notably, both reading-while-listening and reading-only led to better performance than listening-only, while no differences were found between reading-while-listening and reading-only. In addition, the results of this study provide some indication for the role of number of exposures on binomials acquisition, but indicate that at an initial stage, frequency might have a clearer effect at the level of familiarity than at the level of form recognition. Six exposures were not enough to develop knowledge of the correct order of binomials at a level comparable to existing binomials, while they were enough to develop a level of familiarity similar to that of existing items. More work needs to be done to explore frequency and its relation to other factors in order to gain a more accurate picture of its role in incidental learning across a range of MWSs.

References

Aljabri, S. S. (2013). EFL students' judgments of English idiom familiarity and transparency. *Journal of Language Teaching and Research, 4*(4), 662-66.

Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, *82*(2), 233-27.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Biber, D., Johansson, J., Leech, G., Conrad S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Carrol, G. & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784-797.

Carrol, G. & Conklin, K. (2017). Cross language priming extends to formulaic units: evidence from eye-tracking suggests that this idea "has legs". *Bilingualism: Language and Cognition, 20*(2), 299-317.

Carrol, G., Conklin, K. & Gyllstad, H. (2016). Found in translation: The influence of L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition, 38*(3), 403-443.

Chang, A. C-S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652-663.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693-713.

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257-276.

Conklin, K., & Carrol, G. (2020). Words go together like 'bread and butter': The rapid, automatic learning of lexical patterns in natural reading contexts. *Applied Linguistics*. Online First, 1-23.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and non-native speakers? *Applied Linguistics, 29*(1), 72-89.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse, 20*(1), 29-62.

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing. Which mode of input is most effective? *Studies in Second Language Acquisition, 42*(3), 499-52.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review, 61*(3), 355-382.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A Review. *Studies in Second Language Acquisition, 21*(2), 181-193.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258-86). Cambridge, England: Cambridge University Press.

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

Jelani, N. A. M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video. *ITL-International Journal of Applied Linguistics*, 169(1), 169-190.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua 8,* 113-160.McMahon, M. L. (1983). Development of reading-while-listening skills in the primary grades. *Reading Research Quarterly, 19*(1), 38-52.

Malone, J. (2018). Incidental vocabulary learning in SLA. *Studies in Second Language Acquisition, 40*(3), 651-675.

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition, 157,* 382-402.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nippold, M. A., & Taylor, C. L. (2002). Judgments of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech, Language, and Hearing Research, 45*(2), 384-391.

Owusu-Darko, I., Kwasi Adu, I., & Frempong N. K. (2014). Application of generalized estimating equation (GEE) model on students' academic performance. *Applied Mathematical Sciences, 68*(8), 3359-3374.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22(*1), 31-55.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113-138.

Puimège, E., & Peters, E. (2019a). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47,* 424-438.

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language acquisition, 42*(3), 525-549.

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through watching television. *ITL-International Journal of Applied* Linguistics, *171*(2),191-220.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 776-791.

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, *14*(2), 50–73.

Szudarski, P. (2012). Effects of meaning- and form- focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching and Research, 1*(2), 3-37.

Szudarski, P., & Carter, R. (2014). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245-265.

The British National Corpus (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal, 53*(1), 168-176.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

Vidal, K. (2003). Academic listening: a source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56-89.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258.

Vilkaitė, L. (2017). Incidental acquisition of collocations: Effects of adjacency and prior vocabulary knowledge. *International Journal of Applied Linguistics, 168*(2), 248-277.

Vu, D. V., & Peters, E. (2020). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*. https://doi.org/10.1177/0033688220911485

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly, 43*(1), 55–77.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91-120.

Webb, S., & Chang, A. C-S. (2020). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*,1-22. doi:10.1017/S0272263120000297.

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

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541-572.

Zeger, S., & Liang, K. (1986). Longitudinal data analysis for discrete and continuous

outcomes. *Biometrics, 42*(1), 121-130.

Appendix 1: List of Target Items

Passage	Existing binomials	Novel binomials
Passage 1	- mum and dad (2x) - knives and forks (2x) - tea and coffee (2x)	- write and phone (2x) - bottles and tins (2x) - spoons and bowls (4x) - wires and pipes (4x) - plates and glasses (5x) - shelves and drawers (5x) - clean and polish (6x) - bags and coats (6x)
Passage 2	- son and daughter (2x) - time and money (2x) - wind and rain (2x)	- kitchen and bedrooms (2x) - potatoes and beans (2x) - stone and wood (4x) - chickens and rabbits (4x) - goats and pigs (5x) - walls and fences (5x) - eggs and milk (6x) - farms and houses (6x)
Passage 3	- brother and sister (2x) - boys and girls (2x) - food and drink (2x)	- paint and glue (2x) - boat and train (2x) - cards and gifts (4x) - tennis and football (4x) - stories and jokes (5x) - games and music (5X) - walk and swim (6x) - chat and laugh (6x)

Appendix 2: A Sample Text

We've recently moved house so I've been very busy trying to get the new place into shape. It needs a bit of work, and in particular I think a lot of the wires and pipes in the garage are going to need looking at. When we first went in, the floor was flooded, and after a bit of investigation we realised that a whole load of leaves had blocked the guttering and caused the problem. The rest of the house is generally ok I think, but I'd really like to get some new shelves and drawers for the kitchen as the old ones were well past their best. I'm also fairly sure that quite a lot of our plates and glasses got broken during the move, so we'll probably have to stock up on a lot of things before we can cook our first meal. I found a box with spoons and bowls in it so I guess we can live on soup for a while if we need to! I'll maybe need to clean and polish a few things before we use them since the move was a little on the dusty side, but we should be ok for now. One thing we do have plenty of is tea and coffee since I know how useless I am without it!

One of the main reasons we've moved is to be nearer to my mum and dad. I'm aware that they're slowing down quite a lot now, so it'll be nice to spend time with them while I can. I used to write and phone a lot but it's not the same as being within walking distance. I called round to see them this morning actually as my mother said that she had some knives and forks that we could have, so that's one less thing that I'll need to buy. One thing we certainly don't need is any more bags and coats since my wife has brought boxes of the things with her in the move. She's not a big fan of shopping in general, so I guess I should count myself lucky, but I was rather hoping that the move would be a good opportunity to get rid of a few things like that.

When I arrived to see my parents, my father was in the lounge. He told me that my mother had been keen to clean and polish everything before I arrived, so when I went into the kitchen, I was presented with quite a sight. There were plates and glasses everywhere and they were all sparkling, so I certainly didn't need to worry about buying more. The shelves and drawers were all empty and it looked like everything had been piled on the table in the middle of the room. I thought that my wife was bad with her bags and coats but that was nothing compared to what I was seeing here! There were also bottles and tins of all kinds piled up all over the floor – I was amazed at what I was seeing.

My mother stopped what she was doing and smiled as I came in. She explained that she wanted to make sure that we had everything we needed in the new house, so she had dug out everything she could think of. She said that she had been to the supermarket and stocked up on plenty of things, including tea and coffee since she knew how much we went through as a family. She had also been collecting various items and had already boxed up some spoons and bowls for us (the one thing I knew we didn't need!) and had been busy all morning trying to clean and polish everything so that it was ready for me to take away. I asked where all of this had come from and she said that it was just what had accumulated over the years. The knives and forks were very good quality, so she was very keen for me to have them.

I came back from my visit to see mum and dad with several boxes full of stuff. After everything that I'd said to my wife about the bags and coats she was going to tease me about this, I was sure! But she was very pleased to learn that we now had plates and glasses for the kitchen, and even more pleased to learn that she wouldn't have to clean and polish them since my mother had already done so. We wanted to unpack the boxes, but since the shelves and drawers were yet to be replaced, we didn't really have anywhere to put the contents, so we decided that it was as good an excuse as any to head to the big shopping park that was located a few miles away. The bottles and tins were all going to go straight into the garage, so I unpacked those, but I made sure that the wires and pipes were kept clear, ready for when the men came to look at them. I also thought that I must make sure to clear the leaves out of the gutters to make sure that the garage didn't flood again in future. I had a rather wicked thought that we could maybe use the bags and coats as a barrier, but I didn't think I'd mention it to my wife!

We got in the car and drove out to the shopping centre. I was glad that we didn't need plates and glasses anymore, as it meant that we could think of much more interesting things to spend our money on. We had never been people with particularly expensive tastes – as far as I was concerned having nice things to look after was just annoying, so not having fancy jewellery or things like that just meant fewer things to have to clean and polish on a regular basis! It was nice to know that we wouldn't have to spend money on things for the kitchen since I had a nasty feeling that the wires and pipes in the garage were going to end up costing us quite a bit to fix, so saving a bit of money here was a big bonus.

We walked around for quite a while just browsing some of the shops. I knew that we no longer needed spoons and bowls, but my wife found a set that she absolutely fell in love with,

so I decided to buy it for her as a moving in gift. They had quite a striking pattern which would match our colour scheme, so I was happy to get them for her. I did insist that along with the plates and glasses from my mother, we would now definitely need shelves and drawers though, or else she would have nowhere to put them, so we asked if there was a shop that might sell such things in the shopping centre. A helpful young lady told us that there was, so we set off to find it.

We passed right by a shop selling bags and coats and I was a bit concerned that my wife would insist on going in, but she didn't seem to notice. We went on to find the shop that the girl had mentioned. It was right at the far end of the shopping centre, so it took a few minutes to get there, but when we did, we found an amazing selection awaiting us. It took us a while to browse through everything but eventually we found something we could agree on. The shelves and drawers we picked out were not cheap, but they would certainly be sturdy and would last us a long time. While I was there, I also asked if the shop dealt with wires and pipes to see whether they could send someone to look at our garage. They said they could, so I booked a man to come out later in the week, although I must make sure I had time to clear out the leaves before he came.

We headed home very satisfied with our day's shopping. I was particularly pleased that we had gotten our kitchen sorted out, and that we had acquired so much stuff from my parents to fill the cupboards! And the fact that we didn't need to clean and polish any of it ourselves was a bonus, as was the fact that we'd gone shopping and bags and coats hadn't even entered my wife's mind! She was happy with her spoons and bowls so that would probably keep her happy for a while, but getting them did remind me that I would need to buy white paint for the kitchen walls. As we drove home, I thought about how happy we were going to be in our new home, and I decided that I really must remember to write and phone to tell all of our friends our new address.

Appendix 3: Sample Tests

Multiple-choice Test

You will see a set of three-word phrases. Some of these are natural phrases in English and some are not.

Indicate, which you think is more natural, or if you don't think either more is natural than the other then

indicate "neither". For example:

For number 1. I would circle "high and low" because it sounds more natural to me.

1. (high and low)	low and high	Neither

2. north and west	west and north	(Neither)

3. street and dog	dog and street	neither
4. basket and table	table and basket	neither
5. mind and body	body and mind	neither

Familiarity Ratings Task

In this task, you will see a set of three-word phrases. You are asked to rate your familiarity with each

phrase by circling a number from 1 to 5. For example:

In the first example "black and white", I would circle "5" because I have heard this phrase a lot of times.

	I've never heard/used this phrase	*I've very rarely heard/used this phrase*	*I've rarely heard/used this phrase*	*I've frequently heard/used this phrase*	*I've very frequently heard/used this phrase*
Example **black and white**	1	2	3	4	5 (○)
street and dog	1	2	3	4	5
table and basket	1	2	3	4	5
mind and body	1	2	3	4	5
shelves and drawers	1	2	3	4	5

Appendix 4

A Sample of the Comprehensions Questions

Are the following statements true (√) or false (×)?

    1- The kitchen doesn't need any tools for cooking.   (   )

    2- They decided to move out because they did not like their house.   (   )

    3- They didn't buy anything from the shopping centre.   (   )

Sara Alotaibi

University of Jeddah

Ismat Ibn Malik, As Salamah District, Jeddah 23436 6290 (Saudi Arabia)

e-mail: salotaibi@uj.edu.sa

Ana Pellicer-Sánchez

University College London

20 Bedford Way, London WC1H 0AL (United Kingdom)

e-mail: a.pellicer-sanchez@ucl.ac.uk

Kathy Conklin

University of Nottingham

Trent Building, University Park, Nottingham NG7 2RD (United Kingdom)

e-mail: Kathy.conklin@nottingham.ac.uk