



THINKING ALLOWED

# The processing of multi-word expressions: A research agenda for the next 10 years

Anna Siyanova-Chanturia<sup>1</sup>  and Suhad Sonbul<sup>2</sup> 

<sup>1</sup>School of Linguistics and Applied Language Studies, Te Herenga Waka – Victoria University of Wellington, Wellington, New Zealand and <sup>2</sup>English Language Institute, Umm Al-Qura University, Makkah, Saudi Arabia

**Corresponding author:** Anna Siyanova-Chanturia; Email: [Anna.siyanova@vuw.ac.nz](mailto:Anna.siyanova@vuw.ac.nz)

## Abstract

Multi-word expressions (MWEs) are fixed, conventional strings of language (e.g. idioms, collocations, binomials, proverbs) which have been found to be widespread in language use. Research has shown that MWEs exhibit an online processing advantage over control phrases by first language (L1) and second language (L2) speakers. While this line of research has helped us better understand the nature of MWEs and factors that may influence their processing in real time, there remain several gaps that future research should focus on. In this piece, we focus on four main topics related to the online processing of MWEs: (1) comprehension of MWEs by L1 and L2 speakers, (2) production of MWEs by L1 and L2 speakers, (3) the processing of modified MWEs by L1 and L2 speakers, and (4) the processing of MWEs by L1 children. Under each topic, we propose nine research tasks that will further advance our understanding of MWE processing in real time. We conclude with relevance of MWE processing research to L2 teaching and learning.

**Keywords:** multi-word expressions; language processing; comprehension; production; first/second language

## 1. Introduction

Recent years have seen growing interest in multi-word expressions (MWEs) and factors affecting their online processing. MWEs are ‘strings of letters, words, sounds, or other elements, contiguous or noncontiguous [...] that necessarily enjoy a degree of conventionality or familiarity among (typical) speakers of a language community or group, and that hold a strong relationship in communicating meaning’ (Siyanova-Chanturia & Pellicer-Sánchez, 2019, p. 5). They are a large and diverse set of expressions that includes idioms (e.g. *spill the beans*), collocations (e.g. *cup of tea*), lexical bundles (e.g. *despite the fact that*), binomials (e.g. *fish and chips*), proverbs (e.g. *early bird catches the worm*), and other sequences. While MWEs are rather heterogeneous, they nonetheless share several key properties – familiarity, frequency, and predictability (Siyanova-Chanturia & Van Lancker Sidtis, 2019).

MWEs are conventional strings of language, typically highly familiar to proficient language users. Most (but not all) MWEs enjoy high frequency of occurrence, as attested by corpora counts. Interestingly, even if a MWE is low frequency, such as some idioms (e.g. *raining cats and dogs*), it is still perceived as a highly conventional expression. MWEs are also highly predictable, meaning that incomplete expressions can easily be completed with the most expected word/s (e.g. *excruciating* → *pain*, *on the other* → *hand*, *it’s never too* → *late*). Individual components within MWEs have also been shown to be strongly associated. For example, *excruciating pain* is not only a highly familiar, fixed expression; it is also an expression in which Word 2 is very likely to follow Word 1 (and Word

1 is very likely to precede Word 2), as determined by such association measures as t-score, mutual information, and Delta P (e.g. see Siyanova-Chanturia & Spina, 2020).

Research into MWE online processing has by and large focused on two modalities – comprehension (e.g. reading and listening) and production (e.g. speaking). These two modalities are also focused on in the present piece. The online processing refers to language comprehension and production happening in ‘real time’, that is, under time pressure with no preparation or revision possible.

One of the key theoretical questions that has guided MWE processing research is whether or not language users are sensitive to phrase frequency manipulations during language comprehension and production. Phrase frequency effects have, in particular, been used to test the models of language acquisition, processing, and use (e.g. see Siyanova-Chanturia et al., 2011). According to one such model – the words-and-rules approach – the lexicon (comprising memorized forms) and grammar (comprising grammar rules) are distinct (Pinker, 1999; Pinker & Ullman, 2002). In line with the words-and-rules approach, frequency effects should be found in the processing of memorized forms (e.g. words), but should not be found in the processing of compositional strings of language (that many MWEs are). On the contrary, according to usage-based models (Bybee, 1998; Goldberg, 2006; Tomasello, 2003), all linguistic information, at the morpheme, word, phrase, or sentence level, literal and figurative, compositional and non-compositional, is represented and processed in a comparable way, and should be similarly affected by frequency manipulations. Thus, language should be viewed as a statistical accumulation of discrete experiences (Bod, 2006; Bybee, 2006). In line with this view, the frequency with which a linguistic exemplar occurs in language is deemed a determining factor in how this exemplar will be represented and processed (accessed and retrieved, Bod, 2006). Corpus frequencies can thus be used as a proxy for exposure to or experience with the language; while the inclusion of language proficiency and type of exposure to second language (L2, e.g. age of first exposure, time spent learning L2, time spent in L2 country) can help determine how frequency may interact with proficiency and experience.

The research accumulated so far has shown that both first language (L1) speakers and L2 learners are uniquely attuned to frequency distributions at various levels of granularity: both at the word and phrase level. However, much of the research available to date has targeted adult L1 speakers, with L2 speakers and children, in particular, still being under-represented in this line of enquiry. Crucially, we believe it is time to move beyond simple phrase frequency effects that manifest themselves in quantitatively faster processing of frequent versus infrequent phrasal information. MWEs come in many shapes and sizes, with non-adjacent or modified MWEs being one such manifestation of phrasal diversity. So far, few studies have considered how the language processor deals with the modification of familiar and predictable sequences. Another domain that is still very much under-represented is the production of MWEs. While MWE comprehension has received sustained attention, production studies are still scarce, with virtually no studies to date probing naturalistically elicited (rather than experimentally manipulated) production of MWEs in children or L2 speakers. Lastly, the vast majority of MWE processing studies have to date employed behavioural measures, such as reaction times and eye movements (in comprehension studies), or elicited production (in production studies). These methodologies can attest to quantitatively faster processing of MWEs relative to novel language. However, they are silent as to possible qualitative differences (e.g. lexical vs syntactic processing). Electroencephalography (EEG) is one methodology that can be used to probe the nature of the cognitive processes involved in MWE comprehension. Key properties of MWEs, such as frequency and predictability, lend themselves particularly well to the adoption of this methodology, which has been shown to be highly sensitive to the manipulations of these properties.

Capitalizing on the research already available and the identified gaps, the current agenda has as its aim to focus on nine specific tasks within the following four areas of enquiry: 1. the application of

ERPs in the study of MWE comprehension, 2. the production of MWEs, 3. the processing of modified MWEs, and 4. MWE processing in children. It is important to note that the present piece is not intended as an extensive or exhaustive review of the MWE processing literature, with many pertinent studies inevitably omitted. Rather, it should be viewed along the lines of ‘What do we currently know?’ and ‘What additional insights can be gained and how?’

## 2. MWE comprehension in L1 and L2 speakers

Comprehension of MWEs in L1 and L2 adult speakers has received a substantial amount of attention in the literature. A key question that has been asked is whether L1 speakers and L2 learners are sensitive to phrase frequency manipulations during language comprehension (reading, in particular). Compared to L1 speakers, L2 learners would have had far less exposure to L2 MWEs, and so their sensitivity to phrase frequency effects might not be as pronounced. In one of the earliest such studies, Siyanova-Chanturia et al. (2011) used eye movements to examine the comprehension of binomials (e.g. *bride and groom*) and their reversed forms (e.g. *groom and bride*) by L1 and L2 speakers of English. In both participant groups, phrase frequency was found to predict the reading speed of target sequences, although the effect was more robust in L1 speakers. This finding has since been replicated in a multitude of comprehension studies using a range of methodologies, such as reaction times and eye movements, and employing a variety of MWE types (for a review see Siyanova-Chanturia & Van Lancker Sidtis, 2019). A large body of evidence has now accumulated attesting to quantitatively faster processing of MWEs (e.g. binomials, collocations, lexical bundles, etc.) compared to novel strings of language.

However, it is often argued that MWE processing is associated not only with quantitatively faster processing, relative to novel language, but also with easier semantic integration (i.e. MWEs are easier to process and integrate due to their high familiarity and predictability; e.g. see Vespignani et al., 2010). Electroencephalography (EEG) is one methodology that has allowed researchers to more directly probe the nature of MWE processing, above and beyond the speed of processing. EEG is the recording of electrical activity produced by the brain, while event-related brain potentials (ERPs) are EEG responses time-locked to a particular stimulus and averaged over trials (Van Petten & Kutas, 1991). Not only can ERPs tell us *WHEN* something happened, but they can also reveal the *NATURE* of the cognitive processes involved (Kutas & Van Petten, 1994). Event-related brain potentials are represented by a series of positive and negative waves, which are associated with different ERP components. Two components, in particular, have been linked to MWE processing – the P300 (a positive wave peaking around 300 ms post-stimulus) and the N400 (a negative wave peaking around 400 ms post-stimulus).

In a study investigating the comprehension of idiomatic expressions in L1 Italian, Vespignani et al. (2010) found that the final word within an idiom elicited an increased positivity around 300 ms after stimulus onset. The authors interpreted this finding as evidence for categorical template matching – upon encountering a highly conventional expression, the brain uniquely predicts, in cognitive terms, the ‘right’ continuation. For example, *you can’t judge a book by its cover*, while *excruciating* evokes *pain*, and so on. When the expectation is met, the P300 effect is elicited. Siyanova-Chanturia et al. (2017) used ERPs to examine the processing of literal English binomial expressions (e.g. *knife and fork*) in L1 speakers. Similar to Vespignani et al. (2010), the final constituent within MWEs elicited larger positivity around 300 ms after the stimulus onset compared to the control condition, which was interpreted as evidence for categorical template matching.

Studies with L1 speakers have also linked MWE processing to the N400 component, specifically, a decreased negativity around 400 ms after stimulus onset. In Strandburg et al. (1993), smaller N400 amplitudes were observed on frequent idiomatic phrases (L1 English) compared to novel ones, while in Laurent et al. (2006) N400 amplitudes were smaller on the last word of conventional metaphors than on the last word of novel metaphors. Taken together, larger P300 amplitudes and reduced N400

amplitudes on MWEs relative to novel sequences suggest easier semantic integration for routinized, conventional language.

The ERP studies so far conducted have overwhelmingly probed the processing of figurative MWEs (e.g. idioms and metaphors), and most have targeted L1 speakers. It has been shown that figurative MWEs are processed differently from novel ones (for a review, see Siyanova-Chanturia et al., 2019). Virtually nothing is known about the neural correlates involved in the comprehension of literal, compositional MWEs in L2 speakers. Compositional MWEs are distinct from their figurative counterparts, as there is no dissociation between the meanings of the components and the meaning of the phrase. However, it remains unclear whether different neural mechanisms underlie the processing of predictable compositional phrases relative to novel ones. The studies with L1 speakers point to the involvement of the P300 and N400 components (e.g. Siyanova-Chanturia et al., 2017). Are the same components involved in L2 learners' processing of literal, compositional MWEs, and does L2 proficiency play a role?

### 2.1. Research task 1: What can event-related brain potentials reveal about compositional MWE processing out of context in L1 speakers and L2 learners?

We propose an extension and replication of Siyanova-Chanturia et al. (2017) with L2 speakers (using L1 speakers as a baseline). In Experiment 1a, L1 and L2 speakers will read three types of phrases:

Frequent English binomial expressions: *knife and fork*

Infrequent equally strongly associated phrases: *spoon and fork*

Semantic violations: *theme and fork*

In total, 120 matched triplets will be used. These items will be presented out of sentence context using rapid serial visual presentation (RSVP, wherein phrases or sentences are presented quickly one word at a time). To ensure participants read the sequences for comprehension, a Go-noGo task will be used on the filler items (e.g. filler items can be animal words requiring participants to press a button [Go response]; no animal words will be used in the stimuli [noGo response, i.e. no response]). ERPs will be compared on the last word – *fork* – which will be the same across all conditions (i.e. ERP response will be measured on the exact same word within all conditions). In Experiment 1b, the same L1 and L2 speakers will read the same stimuli as in Experiment 1a. However, these will be presented without the conjunction 'and'. L1 speakers would have had large amounts of experience with binomials to recognize them as highly fixed, conventional expressions. We thus expect the last word within the binomial (Experiment 1a) to elicit larger P300s and smaller N400s compared to the same word within infrequent strongly associated phrases. However, we expect the P300 and N400 effects to disappear in the absence of the conjunction 'and' (Experiment 1b), since the mental template will no longer match sequences presented.

With respect to L2 learners, we expect L2 proficiency to play a role. As noted by Hahne (2001, p. 252), 'proficiency level in L2 might be the most important variable'. Ibáñez et al. (2010, 2011), used ERPs to investigate L1 and L2 speakers' processing of non-literal sequences. Higher proficiency L2 users exhibited responses comparable to those found in L1 speakers; while lower proficiency L2 learners differed both from their higher proficiency L2 counterparts and L1 speakers. To measure L2 proficiency, several measures can be used, such as: LexTale (Lemhöfer & Broersma, 2012), the Oxford Placement Test (Oxford University Press, 2001), the four self-reported skills (reading, writing, speaking, listening), as well as time learning L2 and time spent in a L2 country. We expect more proficient speakers to show the pattern of results comparable to L1 speakers (the P300/N400 complex), while less proficient L2 users are unlikely to be sensitive to the differences between frequent and strongly associated items (*knife and fork*) versus infrequent yet equally strongly associated ones (*spoon and fork*). While most items can be borrowed from Siyanova-Chanturia et al. (2017), care should be taken

not to use items that may not be familiar to L2 speakers. The two experiments will allow replication of the L1 findings of Siyanova-Chanturia et al. (2017) and extend them to L2 learners.

## 2.2. Research task 2: What can event-related brain potentials reveal about compositional MWE processing in context in bilingual individuals?

Much of MWE processing research has been done with L1 and L2 speakers of English. Other languages have received far less attention. It would thus be instructive to consider non-English languages. Below, we propose a series of experiments with Spanish-Basque bilinguals. It is important to note that a similar design and comparable stimuli can be adapted to a variety of other bilingual contexts. The Basque Country provides researchers with a unique linguistic environment wherein two very different languages are spoken – Spanish and Basque. Basque is taught in all schools. However, while in some schools, half of the subjects are taught in Spanish and the other half in Basque, in others, *all subjects* are taught in Basque, and Spanish is taught as a L2. This implies that there are both Spanish-Basque (L1 Spanish, L2 Basque) and Basque-Spanish (L1 Basque, L2 Spanish) bilinguals.

In Experiment 1a/1b, the experiment items will be verb + noun collocations and novel phrases embedded in context. Half of the collocations will be similar in Spanish and Basque (can be translated word-by-word):

Collocation: *Es necesario tomar una decisión junto con nuestros amigos* ('It is important to take a decision together with our friends')

Novel phrase: *Es necesario tomar una cerveza junto con nuestros amigos* ('It is important to have [lit. take] a beer together with our friends')

In Basque, *tomar una decisión* (lit. 'take a decision') is *erabaki bat hartu* (lit. 'take a decision').

The other half of collocations will be different in that they exist in both languages but cannot be translated word-by-word:

Collocation: *Es necesario prestar atención a estos problemas* ('It is necessary to pay attention to these problems')

Novel phrase: *Es necesario prestar obediencia a las leyes* ('It is necessary to obey [lit. render obedience] these laws')

In Basque, *prestar atención* (lit. 'render attention') is *arreta jarri* (lit. 'pay attention').

In Experiment 2a/2b, noun + adj. collocations and novel phrases will be embedded in context. Half of the collocations will be similar in Spanish and Basque (can be translated word-by-word):

Collocation: *Necesito comprar chocolate negro para hacer la tarta* ('I need to buy dark chocolate to make the cake')

Novel phrase: *Necesito comprar chocolate bueno para hacer la tarta* ('I need to buy good chocolate to make the cake')

In Basque, *chocolate negro* (lit. 'black chocolate') is *txokolate beltza* (lit. 'black chocolate').

The other half of collocations will be different in that they exist in both languages but cannot be translated word-by-word:

Collocation: *Me gusta el vino tinto con la carne* ('I like red wine with meat')

Novel phrase: *Me gusta el vino bueno con la carne* ('I like good wine with meat')

In Basque, *vino tinto* (lit. 'red wine') is *ardo beltza* (lit. 'black wine').

All collocations will be embedded in context. Experiment 1a and 2a will be run in Spanish with L1 Spanish speakers and L2 Spanish speakers (Basque-Spanish bilinguals). The rationale is that where

collocations are the same in Spanish and Basque (*tomar una decisión* vs *erabaki bat hartu*), both participant groups will develop similar expectations regarding the upcoming information after seeing the first word of the collocation, evidenced by larger P300s and smaller N400s relative to novel phrases. On the other hand, where the two collocations are different (*prestar atención* vs *arreta jarri*), the effect should be stronger for L1 Spanish speakers than L2 Spanish speakers (Basque-Spanish bilinguals) because, in the mother tongue of the latter, these collocations are different from the Spanish ones.

Experiment 1b and 2b will be run in Basque with L1 Basque speakers and L2 Basque speakers (Spanish-Basque bilinguals). The rationale is that where collocations are the same (*txokolate beltza* vs *chocolate negro*), both groups will develop similar expectations regarding the upcoming information after seeing the first word of the collocation, evidenced by the P300/N400 complex relative to novel phrases. On the contrary, where the two types of collocations are different (*vino tinto* vs *ardo beltza*), the effect should be stronger for L1 Basque speakers than L2 Basque speakers (Spanish-Basque bilinguals) because, in the mother tongue of the latter, these collocations are different from the Basque ones.

In the above experiments, the final word within MWEs will be manipulated by means of substitution with another word, matched in key lexical properties, to form a novel phrase. The critical comparison will be the last word of a highly familiar phrase (high cloze probability) versus the last word of a novel phrase (low cloze probability). Cloze probability is the probability of a phrase being completed correctly when only the beginning is provided. ERPs will be measured on the phrase-final word. Across all experiments, the stimuli will be presented using rapid serial visual presentation (RSVP). Familiarity, predictability, and literalness of the selected MWEs will be pretested. Where target stimuli are embedded in a sentence context, comprehension questions will be used to ensure participants are reading for comprehension. Where the stimuli are presented in isolation, a Go-noGO task will be used on the filler items. We expect that the manipulations in phrase frequency and the predictability of the phrase-final word will result in the P300/N400 complex.

The above experiments will have important implications for our understanding of the neural correlates involved in MWE processing. Research has showed that the P300 may be observed within the N400 time range. Thus, the latencies and peaks of the two components may overlap. It is debatable whether an increased positivity around 300 ms followed by a reduced negativity around 400 ms elicited by predictable stimuli is the manifestation of two distinct components, or whether they are manifestations of one component (e.g. Molinaro & Carreiras, 2010). The proposed experiments will provide empirical evidence in support of the view that these effects may in fact reflect two distinct components: one elicited by expected stimuli (P300), and the other indicative of a more general semantic processing (N400).

According to Vespiñani et al. (2010), the words in an idiom are anticipated differently from the way a literal string is anticipated, because the comprehension of literal language proceeds *incrementally* and *compositionally*. The proposed experiments will show that anticipatory mechanisms are at work not because one string of words is figurative, while the other one is literal, but because one is highly predictable, and the other one is not. It will be demonstrated that anticipatory mechanisms involved in the processing of two comparably compositional literal phrases might still differ. This is because one phrase is a novel word combination that is indeed activated incrementally by integrating each piece of semantic information, while the other one is a MWE, where initial word/s preactivate the upcoming configuration.

Finally, the proposed experiments will offer neurophysiological support to the usage-based theories. These theories predict distinct processing for frequent/predictable phrases over infrequent ones, because all linguistic units are thought to exist in networks. This is why initial word/s within a MWE are likely to activate final one/s. Due to their fixedness and familiarity, MWEs are believed to be stored in long-term memory. By making use of a relatively abundant resource (long-term memory),



the brain may compensate for a relative lack in working memory. If compositional MWEs are processed differently from novel phrases, as evidenced by the P300/N400 complex for the former, this will provide further empirical evidence to support the usage-based approaches to language acquisition, processing, and use.

### 3. MWE production in L1 and L2 speakers

Research into MWE processing with L1 and L2 speakers has mainly focused on MWE comprehension, in particular, through tasks that involve reading (self-paced reading, eye-tracking, EEG). Relatively few studies have so far examined the processing of L1 MWEs from the productive perspective. Fewer still have looked at L2 speakers' production of MWEs.

Early L1 studies by Van Lancker and colleagues (Van Lancker & Canter, 1981; Van Lancker et al., 1981) examined whether language users are able to utilize prosodic cues to be able to identify the intended meaning of an ambiguous idiom (with both a literal and figurative meaning) from pre-recorded sentences (e.g. *It was the usual procedure about once a week to wash their dirty linen in public.*) where the MWE can either be interpreted figuratively or literally. Results showed that when the speakers were instructed to intentionally convey the required meaning when recording the sentences (Experiment 3 and Experiment 4), the listeners were successful in identifying that meaning. In a follow-up study, Van Lancker et al. (1981) identified those prosodic cues: utterance duration, pause length, and pitch contours. In a more recent study, Yang et al. (2015) used a similar design but focused on L1 Korean. In this study, the recordings were intended to be as natural as possible (avoiding the artificial nature of recordings used in previous research). It was found that Korean speakers could successfully identify the intended meaning, and that the clues were similar to L1 English (although certain differences were found to exist). In Van Lancker (2003), the author compared L1 and L2 recognition of prosodic cues. Results showed that both L1 and proficient L2 speakers were able to identify the intended meaning based on such cues. However, less proficient L2 speakers were not successful in this task. These results suggest that L2 speakers can be less sensitive to prosodic cues than L1 speakers when recognizing the MWE meaning. It should be noted that although the study is pioneering in that it included L2 speakers, it did not employ any objective measure of L2 proficiency. Rather, subjective judgement of exposure (in daily communication versus in classroom context) was used as a categorical variable.

While the studies reviewed above focused on the processing of figurative MWEs and prosodic cues that disambiguate the intended meaning, Siyanova-Chanturia and Janssen (2018) looked at L1/L2 production of literal MWEs, such as binomial expressions (e.g. *fish and chips*) borrowed from Siyanova-Chanturia et al. (2011). In a phrase-elicitation task, L1 speakers articulated the target binomial expressions in their typical order as well as their reversed forms (varying in frequency) out of context. Results showed that L1 speakers showed sensitivity (shorter articulatory durations) to the frequency of binomials regardless of their type (typical or reversed) (see Janssen et al., 2012, for similar evidence from L1 Spanish and L1 French). L2 speakers, however, were not sensitive either to phrasal frequency or type. These findings are contrasted with those of Siyanova-Chanturia et al.'s (2011), where both L1 and the more proficient L2 speakers exhibited a processing advantage for more frequent configurations, pointing to the differences between L1 and L2 speakers when recognizing and producing MWEs.

#### 3.1. Research task 3: Does L2 proficiency modulate sensitivity to the prosodic cues of figurative MWEs?

As noted above, only one study compared L1 and L2 processing of ambiguous MWEs (Van Lancker, 2003). However, this study is limited in several ways. First, the carrier sentences were recorded in an

artificial manner whereby the speakers were asked to purposefully stress the figurative/literal sense. Second, the L2 participants were categorized based on their exposure (naturalistic versus formal) rather than through employing an objective measure which would have yielded more reliable results. Third, the study did not control for congruency, as participants came from a variety of L1 backgrounds. Previous research on MWEs (e.g. Carrol et al., 2016; Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2011) has shown that congruent MWEs (those with a direct equivalent in the L1) are processed faster than incongruent ones (those that do not have a direct translation equivalent in the L1).

Thus, to expand this line of research, we suggest that a study is designed whereby idioms are carefully selected to be incongruent (based on the L2 participants' background). This should ensure that any potential L1 interference is partialled out. Thus, if the target population is L1 Arabic–L2 English speakers, for example, then an idiom *to wash their dirty linen in public* should not be targeted as a direct equivalent exists in Arabic. However, an idiom such as *spill the beans* would be suitable as its direct translation does not exist in Arabic. Transparency should also be controlled for as previous research has shown that it may have a modulating effect in the processing of MWEs (see Gyllstad & Wolter, 2016). The target idioms should be normed with a group of L1 raters using a scale from 1 = transparent to 7 = opaque, with the average score included as a covariate in the analysis. Then, the carrier sentences should be recorded in a natural manner (akin to Yang et al., 2015). This should allow for more naturalistic data in comparison to previous research by Van Lancker and colleagues (see Siyanova-Chanturia & Van Lancker Sidtis, 2019).

Most importantly, participants should be recruited to represent different L2 proficiency levels. An objective measure of proficiency (e.g. Oxford Placement Test, see Research task 1 above for more options) should be administered to allow a reliable estimation of proficiency in English. The results of the test can either be used to categorize the participants into two separate groups based on proficiency level ('higher level' versus 'lower level') or, preferably, the score can be included as a raw continuous predictor in the analysis. Either way, the effect of proficiency can be included as a main predictor of decision accuracy (0/1) and as an interacting variable with sentence type (figurative versus literal).

Another potential expansion of the study involves the addition of a reaction time measure as a dependent variable. Previous research examining speakers' ability to identify the intended meaning of an idiom through prosodic cues has mainly used a 0/1 decision accuracy measure (intended meaning correctly identified or not). Through measuring response latency, it will be possible to examine how much time it took the participants to make a decision and whether this is relevant to the availability of cues in the recorded sentence.

It should be noted that in such a study, a group of L1 speakers is needed as a baseline for comparison. Thus, the study can follow the two-experiment design with Experiment 1 focusing on L1 speakers of English and Experiment 2 focusing on L2 speakers with proficiency as a modulating factor.

### 3.2. Research task 4: How much exposure is needed to allow L2 speakers to cross the threshold of sensitivity to MWEs from the receptive to the production level?

We have seen that Siyanova-Chanturia and Janssen (2018) provided evidence for a dissociation between receptive and productive processing of MWEs. This was done through comparing the results of their phrase-elicitation task (productive processing) to those of Siyanova-Chanturia et al. (2011) who employed eye-tracking as a measure of receptive processing. It was found that L2 speakers showed no sensitivity to the typical binomial configurations during production, although the more proficient speakers did show such an advantage during recognition. This is not surprising given the general finding in vocabulary research that recognition precedes production (Nation, 2013). One notable gap in L2 vocabulary research according to Schmitt (2019) is examining how knowledge develops from receptive to productive mastery. Addressing this question, according to Schmitt



(2019), requires a design where lexical items are known partially (at the receptive level), followed by exposure to enhance this partial knowledge. Then, a measure of productive mastery can be administered to evaluate whether productive mastery has developed or not. A pretest–post-test design might be most suitable for this study.

First, participants take a measure of receptive processing and productive processing of binomials (pretests). The receptive measure can be a self-paced reading task, and the productive measure can follow Siyanova-Chanturia and Janssen's (2018) phrase-elicitation task. The study can use the same materials as Siyanova-Chanturia et al. (2011) with typical and reversed forms. A proficiency measure can also be included to control for its potential effect on processing. The next step is exposing the learners to reading texts purposefully designed to be seeded with the target items (treatment). Frequency of exposure can be varied (three versus nine, see Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). Then, a post-test can be administered (same receptive and productive measures) to examine which frequency condition led learners to cross the 'recall threshold' (post-tests).

The analysis can include both test time (pretest versus post-test) and item type (typical versus reversed) and the interaction between them, in addition to proficiency as a covariate. Based on a comparison between typical and reversed forms under both testing sessions, it can be identified whether participants have developed sensitivity to binomial type in the pretest in receptive processing (advantage in a self-paced reading task) and productive processing (advantage in a phrase-elicitation task) and whether this has improved in the post-test. It is hypothesized that the L2 participants may show receptive advantage in the pretest but no productive advantage at this stage. It would be interesting to see if any of the two treatment conditions can lead to the development of productive advantage in the post-test.

#### 4. Processing of modified MWEs

Research on the processing of different types of MWEs has generally shown that L1 speakers process typical MWEs of various types faster than novel, control phrases; in the L2, however, this advantage is modulated by several factors (including congruency, transparency, and proficiency, see El-Dakhs et al., 2024). While this line of research has indeed expanded our understanding of how L1 and L2 speakers process MWEs, the items targeted have almost exhaustively been adjacent MWEs such as the verb-noun collocation *provide information* or the phrasal verb *find out*. In natural language, however, the constituents of a MWE are not always adjacent, and MWEs might be modified in various ways. For example, the collocation *provide information* can have an intervening word (*provide some information*) or several words (*provide some useful information*). The verb and particle in *find out something* are adjacent but the object can be moved to separate the two components *find something out*. Another form of modification in MWEs concerns passivation. The passive construction *information is provided* is a modified version of the active form *provide information*. Further, MWEs can be modified through varying the morphological forms of the components words (*provide information*, *provides information*, and *provided information*).

Research on L1 processing of modified MWEs is rather scarce (see Molinaro et al., 2013). In a pioneering study, Vilkaitė (2016) examined how L1 English speakers process adjacent and non-adjacent verb-noun collocations (*provide information* versus *provide some of the information*) in comparison to matched control phrases (*compare information* versus *compare some of the information*). The results showed a processing advantage both for adjacent and non-adjacent collocations over control phrases, although the facilitative effect on the last word (i.e. *information*) was smaller for non-adjacent collocations.

Research on the processing of modified collocations in L2 is even scarcer. Vilkaitė and Schmitt (2019) addressed the question of how non-adjacent collocations are processed by L2 English speakers. The study employed the same design as Vilkaitė (2016) and used the same stimuli. The L2 speakers exhibited a processing advantage for collocations over novel phrases only when these were adjacent.

However, with an intervening modifier, the MWE advantage diminished, pointing to an ‘adjacency effect’ in L2 collocation processing. However, the study is limited in several ways: (1) not examining the congruency effect as speakers came from a variety of L1 backgrounds, (2) not manipulating L2 proficiency levels as all participants were advanced L2 users, and (3) not controlling for the type of the inserted material which ranged from frequent to novel expressions.

As indicated above, modification is not only about insertion but can also be related to morphological changes. This is most evident in agglutinative languages such as Finnish and Turkish and fusional languages such as Arabic and Spanish. In agglutinative languages morphemes are ‘glued’ together to make up complex words. Durrant (2013) gave the Turkish example *olabileceğini* which is a single word composed of several morphemes: one root ‘ol’ (meaning ‘be’) and four suffixes which give the meaning of possibility, subordination, possession, and accusation. Fusional languages are also complex in structure but ‘fuse’ the meaning into one suffix. For example, the word *تكتسب* *taktasib* meaning *she gains* in Arabic has the prefix ‘ta’ which represents third person singular, present tense, and feminine subject.

Research examining the processing of MWEs in agglutinative and fusional languages can be interesting as some MWEs exhibit rich morphological variations. However, to the best of our knowledge, only one study compared the processing of MWEs in an agglutinative language (i.e. Turkish) with English (Öksüz et al., 2024, Study 2). These authors compared reaction times (in a decontextualized timed acceptability judgement task) of L1 Turkish and L1 English speakers in their respective languages as they processed equivalent adjective-noun collocations. Both groups showed sensitivity to phrasal frequency (higher frequency led to shorter reading times), but the effect was stronger for English, suggesting that speakers of morphologically complex languages (Turkish) may be less sensitive to phrase frequency than speakers of English.

#### 4.1. Research task 5: Which factors modulate the L2 adjacency effect when processing collocations?

Vilkaite and Schmitt (2019) were the first to demonstrate an L2 adjacency effect. However, they did not control for several factors that might have influenced (or at least modulated) the processing of adjacent/non-adjacent collocations. These include congruency, proficiency, transparency, and variation in the frequency of the intervening phrases. It would not be possible to control for all such factors in one study. These factors can instead be manipulated in a series of studies which may help in teasing the various factors apart. This line of research can use a variety of measures including reaction times, eye-tracking, and ERPs.

First, the potential effect of congruency can be examined through carefully selecting items to represent a homogeneous L1 population (e.g. Arabic). Half of the target collocations can be congruent with a direct translation equivalent in the participants’ L1, while the other half can be incongruent with no direct translation. Each target collocation should be paired with a control, and an intervening phrase is added to create the non-adjacent collocation. Thus, in total four item pairs should be generated:

- Congruent adjacent vs. control phrase
- Congruent non-adjacent vs. control phrase
- Incongruent adjacent vs. control phrase
- Incongruent non-adjacent vs. control phrase

Given the robust congruency effect on L2 collocation processing, the L2 adjacency effect may only be evident for incongruent collocations. When collocations are congruent, however, both adjacent and non-adjacent collocations may exhibit an advantage over control pairs.

Another potential extension concerns the manipulation of L2 proficiency through targeting different groups of L2 speakers from beginner to advanced. An objective measure of proficiency can be

employed and be included in the analysis as a modulating factor. One might argue that the adjacency effect may be clearer for lower-level L2 participants than those with higher proficiency. If the design suggested above is adopted (with congruent and incongruent collocations), the researchers can look at the interaction between adjacency, congruency, and proficiency (see Sonbul & El-Dakhs, 2020, for evidence of a modulating effect of proficiency on congruency).

Then, the frequency of the intervening phrase can be altered to represent different frequency levels. For instance, the two components of the collocation *provide information* can be separated by a highly frequent phrase (*provide additional information*, COCA frequency = 228), a phrase that is less frequent (*provide vital information*, COCA frequency = 30), or an infrequent one (*provide concrete information*, COCA frequency = 5). Our hypothesis is that the adjacency effect may be more evident in collocations with less frequent intervening phrases than more frequent ones.

#### 4.2. Research task 6: Do L1/L2 speakers of agglutinative and fusional languages show sensitivity to the frequency of modified MWEs?

Researchers can also examine how L1 and L2 users of agglutinative and fusional languages process frequent MWEs. Variation in MWEs in this case concerns morphological changes in the form of the constituent words. Let's take the collocation 'yaktisab althiqa' (*gain confidence*) in Arabic as an example. To keep things simple, let's focus on changing the form of the verb only as follows:

'yaktisab althiqa' (*he gains confidence*)  
 'tabktisab althiqa' (*she gains confidence*)  
 'ikhtasaba althiqa' (*he gained confidence*)  
 'ikhtasabat althiqa' (*she gained confidence*)

A control pair can be devised for each item above:

'yaktisab alamal' (*he gains hope*)  
 'tabktisab alamal' (*she gains hope*)  
 'ikhtasaba alamal' (*he gained hope*)  
 'ikhtasabat alamal' (*she gained hope*)

First, research can look at how L1 Arabic speakers process these variations and whether their processing reflects differences in the frequency of verb forms and phrasal forms. This type of research can present items out of context in an acceptability judgement task (akin to Öksüz et al., 2024) or in a reading/listening task to more closely reflect natural language use (e.g. using self-paced reading or eye-tracking). It is assumed that lexical and phrasal frequency effects may be stronger when items are presented in isolation than when they are inserted in context (see El-Dakhs et al., 2024).

In addition to L1 speakers of a fusional language like Arabic, researchers can recruit a group of L2 Arabic speakers to examine their processing of modified MWEs. This line of research would be interesting as Arabic (along with other agglutinative and fusional languages) is often assumed to be difficult to learn given its complex morphological systems. Hence, unlike L2 users of morphologically transparent languages like English, L2 Arabic speakers might struggle to perceive the various forms of the collocation 'yaktisab althiqa' as representing the same collocation. This may be particularly the case for beginners who could be 'blind' to such subtle alterations in the form of a MWE.

Another important factor to consider when examining the processing of L2 MWEs in agglutinative and fusional languages is L1 typology. For example, an L1 Spanish speaker learning Arabic as an L2 (both languages being fusional) might quite quickly appreciate the Arabic morphological system and thus show sensitivity to various forms of a given MWE (see the example above) over their control counterparts. On the other hand, an L1 English/L2 Arabic speaker might struggle with the complex

Arabic word structure and might require a lot more exposure in order to develop the MWE processing advantage, especially for less frequent variations.

## 5. MWE processing in children

While MWE processing studies with adults are many and varied (see above), only a handful of studies have so far considered MWE processing in young children (preschool and primary school age). This is ironic, considering that some of the pioneering and formative works in the area of usage-based linguistics are in fact grounded in the field of L1 acquisition (Goldberg, 2006; Peters, 1983; Tomasello, 2003), with MWEs being attributed a pivotal role in L1 learning.<sup>1</sup>

The studies that have contributed to our understanding of the key role played by MWEs in L1 acquisition have traditionally drawn on data obtained via naturalistic observations. Experimental, laboratory-based studies probing the role of multi-word information in L1 children's language processing are limited, with most being in the domain of comprehension. Using eye movements, Jiang et al. (2020, Study 1) exposed 2-year old children to two-word sequences made of a prime and a noun (e.g. *pretty dress*, *pretty cow*), with two pictures appearing on the screen, one depicting the noun (*dress*) and the other depicting the distractor (*cow*). Children consistently looked more quickly at noun pictures for frequent two-word sequences (*pretty dress*) than for novel sequences (*pretty cow*). The authors concluded that not only are children sensitive to frequency distributions of two-word sequences from a very young age, but that this sensitivity also facilitates utterance processing, allowing children to focus on the more novel aspects of what is being said (also see Skarabela et al., 2021).

While the above studies have tapped into phrase frequency effects in young children, researchers have also considered older, primary school-aged children. In one of the first such studies, Jiang et al. (2020) recorded the eye movements of third- (8-year-olds) and fourth-graders (9-year-olds, all L1 Mandarin) as they read MWEs of different frequencies embedded in sentence contexts. These authors operationalized phrase frequency as a dichotomous variable (i.e. phrase type: collocation vs control) as well as a continuous variable. Adults, used as a baseline, read higher frequency phrases faster than lower frequency phrases across all areas of interest and eye-tracking measures, early and late. Importantly, fourth-graders showed a somewhat similar pattern of results, although the phrase frequency effect was confined to a late measure only. On the contrary, third-graders did not show any sensitivity to phrase frequency manipulations, in any of the areas of interest or eye-tracking measures analyzed. The studies that have since followed have further confirmed children's sensitivity to linguistic information at different levels of granularity, focusing, in particular, on forming predictions (e.g. Abu-Zhaya et al., 2022; Kessler & Friedrich, 2022).

Much of MWE processing research has been conducted with immediately adjacent, or unmodified, MWEs. Although one of the properties of MWEs is their relative fixedness, many MWEs can in fact be modified (e.g. *in the hands of* → *in the capable hands of*). While the processing of modified MWEs in adults has been explored to some extent from a variety of perspectives and using different methodologies (see the section 'Processing of modified MWEs' above), how such sequences are processed by children, who are still in the process of acquiring their L1, has so far received almost no attention. To the best of our knowledge, Jiang and Siyanova-Chanturia (2024) is the only published study to have examined this issue.

Using a self-paced reading paradigm, Jiang and Siyanova-Chanturia (2024) examined how third-graders (9-year-olds), sixth-graders (12-year-olds), and adults (L1 Chinese) processed three types of stimuli (all presented in Chinese): adjacent collocations (e.g. *protect animals*), non-adjacent collocation with two characters inserted (e.g. *protect these animals*), and non-adjacent collocations with four characters inserted (e.g. *protect the small animals around here*) relative to their respective controls (e.g. *know animals*, *know these animals*, *know the small animals around here*). Note: in Chinese, the underlined words in the above examples all appear in-between the two content words *protect*

and *animals*, rendering the collocation non-adjacent). Experimental items were embedded in sentence context and were presented one portion at a time, via a self-paced reading task. All age groups read adjacent collocations and collocations with two characters inserted faster than their respective controls. However, only adults and sixth-graders, but not third-graders, read collocations with four characters inserted faster than their controls. The magnitude of the effect was greatest in the adjacent conditions and smallest (but still significant) in the four-character insertion condition. Further, the processing advantage for collocations over controls – and hence sensitivity to phrase frequency manipulations – increased with age, a finding that is in line with Jiang et al. (2020). The findings of Jiang and Siyanova-Chanturia (2024) are important as they extend and enrich those of earlier studies, showing that L1 children are sensitive to linguistic exemplars that really and truly are of different shapes and sizes (adjacent and non-adjacent).

### 5.1. Research task 7: How do young children comprehend modified MWEs?

There are several ways in which MWE processing research can extend the limited body of evidence with L1 children. First, we suggest replicating the findings of Jiang and Siyanova-Chanturia (2024). It is unclear why third-graders in Jiang et al. (2020) did not show sensitivity to phrase frequency manipulations in adjacent collocations, despite the use of eye movements (which is arguably a much more powerful method than a self-paced reading paradigm), while in Jiang and Siyanova-Chanturia (2024) the same age group read both adjacent as well as non-adjacent collocations faster than their respective controls. We propose a conceptual replication of Jiang and Siyanova-Chanturia (2024) using comparable experimental items and age groups, but a different L1 (namely, English). Similar to the original study, three groups of participants will be used: 9-year-olds, 12-year-olds, and a baseline adult group (all L1 English speakers). The following experimental items will be used:

Adjacent collocations and their controls: *draw a painting* vs. *view a painting*

Short modification of collocations and controls: *draw a beautiful painting* vs. *view a beautiful painting*

Long modification of collocations and controls: *draw a beautiful large painting* vs. *view a beautiful large painting*

Similar to Jiang and Siyanova-Chanturia (2024), this will allow researchers to probe the role of two independent variables of interest on reading times: age (and hence L1 proficiency) and modification length (short/one word inserted vs long/two words inserted). This research can be conducted using two methodologies: a self-paced reading paradigm and/or eye movements. The dependent variables will thus be reading times on the target collocation (in self-paced reading) or different eye movement measures specific to the areas of interest (e.g. first pass reading time, total reading time, fixation count; see Durrant et al., 2022, for an overview of different online methodologies in the context of vocabulary processing research).

Future research should also seek to use a wider range of participants' ages. At what age do L1 children become sensitive to phrase frequency distributions of modified and non-modified MWEs? Based on the finding of just one study (Jiang & Siyanova-Chanturia, 2024), it may be around the age of nine with fewer elements intervening, or around 12 with more elements intervening in the middle of the MWE. This conjecture, however, requires further experimentation. We thus propose a study with a wider age range that has previously been probed, specifically the following four groups: 7- to 8-year-olds, 9- to 10-year-olds, 11- to 12-year-olds, and a baseline adult group. The experimental items can be similar to those used in Jiang and Siyanova-Chanturia (2024) – literal collocations with different insertion lengths. For example, if conducted in Chinese, experimental items can be borrowed in their entirety from Jiang and Siyanova-Chanturia (2024), since these items were



specifically created with L1 children in mind, and hence they do not contain low frequency lexical items or concepts unknown to primary school children. If conducted in English, the items may be adapted from Vilkaitė (2016). However, it should be noted that Vilkaitė's (2016) study is somewhat underpowered (40 items per each of the four conditions were used, 160 experimental items in total) and hence additional items may need to be included in the study (for comparison, 48 items per each of the six conditions, 288 experimental items in total, were used in Jiang & Siyanova-Chanturia, 2024). Second, the items in Vilkaitė (2016) were created for an experiment with adult participants. Thus, some items may need to be revised to render them suitable for the use with young children.

A different type of frequent and literal MWEs may be used. Binomials, in particular, are suitable candidates. Many binomials are literal sequences made of frequent and familiar lexical items (e.g. *fish and chips*), and hence should pose no comprehension difficulties to primary-school children. Figurative MWEs (e.g. idioms and proverbs) may not yet be well familiar to young children, as they tend to be less frequent in language than other types of MWEs. Thus, literal MWEs, such as collocations and binomials may be better suited for use with children. Importantly, binomials easily allow for modification by means of insertion (e.g. *fish and chips* → *fish and yummy chips* → *fish and yummy golden chips*). Control items would also need to be created and matched with target binomials (e.g. *fish and crisps* → *fish and yummy crisps* → *fish and yummy golden crisps*). To make the control items as natural as possible, it may not be possible to closely match them on the properties of the target items known to affect processing, such as lexical frequency and length. However, this should not pose any threat to experimental design, as these variables can be included as covariates in the linear mixed effect modelling, now the go-to analysis in lexical processing research and the field of L2 acquisition, more broadly.

Similar to Jiang and Siyanova-Chanturia (2024), target items may be presented using a self-paced reading paradigm, wherein items are presented either word-by-word, or portion-by-portion (Jiang & Siyanova-Chanturia, 2024 used the latter). However, the use of a more powerful and sensitive methodology, such as eye movements, is likely to shed additional insights on the role of modification. Eye movements can show what was fixated and for how long, what was skipped altogether, or read multiple times. Skipping rates, in particular, can be highly informative, suggesting decreased cognitive load and ease of processing (e.g. Rayner et al., 2011). For example, the final word within MWEs is uniquely predictable (e.g. *excruciating ... pain, fish and ... chips*) and may be skipped altogether. The final word within MWEs has also been shown to elicit fewer and shorter fixations than the same word in novel sequences (e.g. Jiang et al., 2020). Further, the use of eye-tracking will allow researchers to tap into the early versus late stage of phrasal processing, as well as explore different areas of interest (e.g. the entire phrase vs. the final word, as well as the spillover region, i.e. the region immediately following the final word). The use of eye movements will thus allow for an extremely detailed picture to emerge as to the processes involved in adjacent versus non-adjacent MWE comprehension in children.

## 5.2. Research task 8: Are young children sensitive to phrase frequency distributions in elicited language production?

In the introduction above, we noted that online language processing broadly encompasses two modalities: comprehension (researched via reading or listening) and production (researched via speaking). Much of MWE processing research, however, has focused on the former. Production studies with adults are few (see the section 'MWE production in L1 and L2 speakers' above), and with children, there are fewer still.

In what was arguably the first study examining MWE processing in children, Bannard and Matthews (2008) used a repetition task to probe children's sensitivity to phrase frequency in production. Bannard and Matthews (2008) examined the accuracy and speed with which 2- and 3-year-old



children articulated four-word MWEs that varied in frequency (e.g. *a lot of noise* vs *a lot of juice*). The phrases within each pair were controlled for the final word frequency (e.g. *noise* vs *juice*), the final bigram frequency (e.g. *of noise* vs *of juice*), as well as the length of the final word in syllables. This was done to ascertain that any effect found could be attributed to the role of phrase frequency, rather than constituent frequency. Two-year-olds were more likely to repeat a MWE correctly if it was a higher rather than lower frequency, while 3-year-olds articulated more frequent phrases faster than less frequent ones. The authors concluded that children as young as two possess ‘complementary representations at different levels of granularity’ (Bannard & Matthews, 2008, p. 246). What this means is that young children, just like adults, are highly attuned to the frequency with which linguistic exemplars – both at the word and phrase level – occur in their input.

To extend this line of research, we propose an elicited production approach with two age groups: 8- and 9-year-old children, and 10- and 11-year-old children. The children will perform a reading task, hence younger children not yet able to read may not be suitable. A phrase-elicitation task used in Siyanova-Chanturia and Janssen (2018) and Arnon and Cohen Priva (2013) can be adopted. In a phrase-elicitation task, participants first silently read a phrase that briefly appears on the screen in front of them (this is a comprehension stage). Once the phrase has disappeared from the screen, participants articulate it out loud in their most natural way (this is a production stage) while being recorded. Articulatory durations are extracted and calculated by subtracting the onset from the offset time for each trial. In terms of stimulus type, we suggest using a wide range of phrases that vary in frequency. That is, rather than binning items into specific conditions (e.g. collocations vs controls), we suggest treating phrase frequency as a continuum, and including items along the entire frequency continuum. To ensure items of various frequencies are used (very high, high, mid, low, very low, etc.), a pool of potential target items can be extracted from a large representative corpus (e.g. COCA for English) using specific syntactic queries (e.g. adjective + noun, verb + noun, noun + noun, noun + preposition, etc., *blue skies*, *do homework*, *train ride*). Given the age of the participants, it is also important to use phrases made of high frequency individual lexical items. While phrase frequency should vary, individual word frequency should be consistently high. Given the study’s focus on probing children’s sensitivity to phrase frequency distributions, we suggest using literal rather than figurative items, as they are more likely to be familiar to children of this age (e.g. the figurative meanings of *red herring*, *top drawer*, and *cut corners* are likely to be unfamiliar to the target age group, despite these phrases potentially meeting other selection criteria).

### 5.3. Research task 9: Are young children sensitive to phrase frequency distributions in natural language production?

Much of evidence available to date is strictly limited to the repetition of MWEs following a prompt (see above). In this paradigm, articulatory durations are elicited via a controlled repetition task, with no meaningful sentential contexts employed. On the contrary, some of the pioneering production studies with L1 adults drew on naturalistic data, extracted from large spoken corpora. Bell et al. (2003) and Bybee and Scheibman (1999) demonstrated that words were more likely to be phonetically reduced when they appeared within highly predictable contexts, such as common phrasal configurations (e.g. *I don’t know*, *middle of the*); while Bybee (2000) showed that boundaries between words within MWEs were akin to those between word-internal segments. More recently, Arnon and Cohen Priva (2013) further confirmed that articulatory durations are likely to be reduced in higher frequency phrases compared to lower frequency ones. Production studies based on naturalistically elicited spoken data have been limited to adult populations largely due to availability of suitably annotated spoken corpora, such as the *Buckeye Speech Corpus* (Pitt et al., 2007) and the *Switchboard Corpus* (Godfrey et al., 1992). These corpora, consisting of naturalistically elicited, conversational speech, are orthographically transcribed and phonetically annotated for various features, making it possible to extract articulatory durations of n-grams of different lengths (bigrams, trigrams, etc.).

Thus, we propose a line of research that will use a spoken corpus of children's productions – to confirm and extend the findings of the laboratory-based studies. Some of the important questions that can be answered using this approach are as follows: Are children's articulatory durations likely to be reduced for higher frequency phrases compared to lower frequency ones? Are words more likely to be phonetically reduced when they appear within frequent sequences than when they appear within novel sequences? How does children's pausing behaviour (if any) compare within frequent phrases versus infrequent ones?

To answer these questions, the CHILDES (MacWhinney, 2000) database can be used. The CHILDES database contains a large collection of transcripts of spontaneous interactions and conversations between young children and their caregivers. The proposed study will explore the effect of multi-word frequencies on phonetic durations in spontaneous speech using the CHILDES corpus. The use of spontaneous, rather than elicited, children's speech will allow researchers to ensure the analysis focuses on sequences produced in meaningful context and with natural prosody (hardly possible during an elicitation task in a laboratory setting). Given the naturalistic and hence unpredictable nature of the corpus, it is not possible to focus on one specific type of MWEs – there may simply not be enough instances of collocations, idioms, binomials, or other types of MWEs used spontaneously. Similar to Arnon and Cohen Priva (2013), specific syntactic structures can be explored, for example, two three-word syntactic structures: 1. subject–auxiliary–verb sequences (e.g. *everybody was trying*), and 2. verb–determiner–noun sequences (e.g. *saw the boy*). Given the age of children who contribute to CHILDES (mostly under 5 years old), more common syntactic structures are advisable to increase the likelihood of them having been extensively heard and produced by children (e.g. verb–determiner–noun sequences would be good candidates). Children's age (in months) can be treated as a continuous variable and included in the analysis as a covariate, to see whether children's articulatory durations (i.e. how long, in milliseconds, children took to articulate a phrase) vary not only as a function of frequency but also age. Frequency counts of the extracted target sequences can be obtained both from the CHILDES corpus (specifically, from the caregivers' part, which reflects the amount of input received) as well as from an L1 reference corpus (e.g. COCA for English). Of note is that CHILDES contains linguistic data specific to learners of different L1s; thus, the study can be extended to other languages. The procedures specific to item identification in a spoken corpus and articulatory duration calculations are described in detail in Arnon and Cohen Priva (2013) and Arnon and Cohen Priva (2014), and can be adapted in future research.

#### 5.4. What relevance does MWE processing research have to L2 teaching and learning?

To conclude this research agenda, it would be instructive and, indeed, prudent, given the readership of *Language Teaching*, to briefly address the following question: What broader relevance does MWE processing research have to L2 teaching and learning?

Psycholinguistic studies have shown that MWEs are accessed, retrieved, and produced faster and more easily compared to novel language. In other words, MWEs are associated with a processing advantage, and, as such, they require relatively little cognitive effort. While this idea goes back to the 1980s at the very least (e.g. see the seminal yet still current work by Pawley & Syder, 1983), most of actual empirical research dates back to much later, post-2010. The observed processing advantage for MWEs, while being of huge theoretical importance, also has non-trivial implications for 'real' language use and for L2 pedagogy. Drawing on a rich repertoire of MWEs in language comprehension and production allows L2 learners to focus on the content, that is, the novel aspects of what is being communicated. Given the sheer number of MWEs in natural language (estimates suggest that there are as many MWEs as there are single words), the *cumulative* processing advantage associated with MWE use can be immense. It has long been argued by L2 researchers that speech fluency may lie in the control of MWEs, and that the process of 'chunking' reduces the amount of planning, processing, and encoding needed for language comprehension and production (Wood, 2002). If a speaker can

pull MWEs readily from memory, fluency can be enhanced (Wood, 2002, 2010). It has even been shown that L2 learners who use MWEs are perceived as more fluent and more proficient than those who do not (Boers et al., 2006). Ironically, many of these early conjectures were based largely on researchers' intuition or subjective perceptions of fluency. Empirical evidence in support of these claims appeared later and has come almost entirely from psycholinguistically oriented research. This is a prime example of how psycholinguistics can inform and advance pedagogy, offering the means and a testing ground for major pedagogical claims about the workings of the human brain.

Psycholinguistic research into MWEs has also demonstrated that words are not learnt in isolation; rather, words form and exist in relationships or networks. One word in the mental lexicon activates another that it is commonly used with (e.g. *excruciating* → *pain*, *run* → *a marathon*). Empirical evidence strongly points to the conclusion that MWEs, rather than single words, are the essential building blocks of language learning. Correspondingly, the focus of L2 teaching and learning should be on MWEs, not on single words. Despite this tenet, it has been shown that, much unlike child L1 learners, L2 learners (and L2 pedagogy, overall) have a strong tendency to focus on words rather than multi-word information.

Lastly, it is worth noting that MWEs are an intrinsically psycholinguistic phenomenon. Whatever line of enquiry one adopts – pedagogical, corpus, computational, and so on – the psycholinguistic nature of MWEs and the core features that define them – frequency, familiarity, and predictability – must necessarily influence and inform the research in question. Relatedly, research into L2 learning and teaching is quickly becoming highly inter- and multidisciplinary. The field of L2 learning has expanded dramatically, and in the process, it has adapted and adopted new methodologies, most notably those from the neighbouring field of psycholinguistics (e.g. reaction time measures, eye-tracking, ERPs). These online measures have revolutionized the way in which much of L2 lexical research is done and have quickly established themselves as the go-to methods in the field (e.g. Godfroid, 2019; Pellicer-Sánchez & Siyanova-Chanturia, 2018). With the present contribution, we hope to continue to pave the way for further cross-fertilization between the more pedagogically oriented linguistic enquiry and research probing mental processing.

## Note

<sup>1</sup> The section on MWE processing in children deals only with L1 children. No study to date has probed L2 children's sensitivity to phrase frequency manipulations during language processing, likely due to the following reasons. First, young children are still in the process of acquiring their L1. Thus any observed patterns in the processing of MWEs vs. infrequent sequences may be due to: 1) incomplete L1 acquisition (L1 acquisition being still in progress), 2) L2 proficiency, or 3) both. Second, finding a large enough pool of child L2 learners (of comparable L2 proficiency) and child L1 baseline (with all children being of a similar age) would be difficult. Thus, this section deals with MWE processing in L1, but not L2, children. Future research may address this gap.

## References

- Abu-Zhaya, R., Arnon, I., & Borovsky, A. (2022). Do children use multi-word information in real-time sentence comprehension? *Cognitive Science*, 46(3), e13111. <https://doi.org/10.1111/cogs.13111>
- 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. <https://doi.org/10.1177/0023830913484891>
- Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent phrases. *The Mental Lexicon*, 9(3), 377–400. <https://doi.org/10.1075/ml.9.3.01arn>
- 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(3), 241–248. <https://doi.org/10.1111/j.1467-9280.2008.02075.x>
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001–1024. <https://doi.org/10.1121/1.1534836>
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, 23(3), 291–320. <https://doi.org/10.1515/tlr.2006.012>

- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245–261. <https://doi.org/10.1191/1362168806lr1950a>
- Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society*, 34, 421–435. <https://doi.org/10.1093/acprof:oso/9780195301571.003.0013>
- Bybee, J. (2000). The phonology of the lexicon. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). CSLI Publications.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733. <https://doi.org/10.1353/lan.2006.0186>
- Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, 37(4), 575–596. <https://doi.org/10.1515/ling.37.4.575>
- Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, 38(3), 403–443. <https://doi.org/10.1017/s0272263115000492>
- Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9(1), 1–38. <https://doi.org/10.1515/cllt-2013-0009>
- Durrant, P., Siyanova-Chanturia, A., Sonbul, S., & Kremmel, B. (2022). *Research methods in vocabulary studies*. John Benjamins. <https://doi.org/10.1075/rmal.2>
- El-Dakhs, D. A. S., Sonbul, S., & Masrai, A. (2024). An eye-tracking study on the processing of L2 collocations: The effect of congruency, proficiency, and transparency. *Journal of Psycholinguistic Research*, 53(2), 30. <https://doi.org/10.1007/s10936-024-10064-x>
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, 517–520. <https://doi.org/10.1109/icassp.1992.225858>
- Godfroid, A. (2019). Sensitive measures of vocabulary knowledge and processing: Expanding nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). Routledge. <https://doi.org/10.4324/9780429291586-28>
- Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press. <https://doi.org/10.2298/jfi0662368i>
- Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, 66(2), 296–323. <https://doi.org/10.1111/lang.12143>
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30(3), 251–266. <https://doi.org/10.1023/a:1010490917575>
- Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters*, 471(1), 48–52. <https://doi.org/10.1016/j.neulet.2010.01.009>
- Ibáñez, A., Toro, P., Cornejo, C., Hurquina, H., Manes, F., Weisbrod, M., & Schröder, J. (2011). High contextual sensitivity of metaphorical expressions and gesture blending: A video event-related potential design. *Psychiatry Research: Neuroimaging*, 191(1), 68–75. <https://doi.org/10.1016/j.neulet.2010.01.009>
- Janssen, N., Barber, H. A., & Rodriguez-Fornells, A. (2012). Phrase frequency effects in language production. *PloS One*, 7(3), e33202. <https://doi.org/10.1371/journal.pone.0033202>
- Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). Multi-word expression processing in children: An eye-tracking study. *Applied Psycholinguistics*, 41(4), 901–931. <https://doi.org/10.1017/s0142716420000296>
- Jiang, S., & Siyanova-Chanturia, A. (2024). Does the processing advantage of formulaic language persist in its nonadjacent forms? Evidence from Chinese collocation processing in children. *First Language*, 44(5), 533–558. <https://doi.org/10.1177/01427237231201580>
- Kessler, R., & Friedrich, C. (2022). Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children. *Language, Cognition and Neuroscience*, 37(8), 1040–1061. <https://doi.org/10.1080/23273798.2022.2035781>
- Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). Academic Press. <https://doi.org/10.1016/b978-012369374-7/50018-3>
- Laurent, J., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, 1068(1), 151–160. <https://doi.org/10.1016/j.brainres.2005.10.076>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. (3rd ed.). Lawrence Erlbaum Associates.
- Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, 28(6), 762–788. <https://doi.org/10.1080/01690965.2012.665465>
- Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, 83(3), 176–190. <https://doi.org/10.1016/j.biopsycho.2009.12.006>

- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139858656>
- Öksüz, D., Brezina, V., Monaghan, P., & Rebuschat, P. (2024). Individual word and phrase frequency effects in collocational processing: Evidence from typologically different languages, English and Turkish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(8), 1287–1314. <https://doi.org/10.1037/xlm0001324>
- Oxford University Press. (2001). Quick Placement Test. Paper and pen test. Oxford University Press.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards, and R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman.
- Pellicer-Sánchez, A., & Siyanova-Chanturia, A. (2018). Eye movements in vocabulary research. *ITL – International Journal of Applied Linguistics*, 169(1), 5–29. <https://doi.org/10.1075/itl.00004.pel>
- Peters, A. (1983). *The units of language*. Cambridge University Press.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. HarperCollins. <https://doi.org/10.5070/l4102005018>
- Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463. [https://doi.org/10.1016/s1364-6613\(02\)01990-3](https://doi.org/10.1016/s1364-6613(02)01990-3)
- Pitt, M., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech (2nd release)*. Department of Psychology, Ohio State University. <https://doi.org/10.21437/interspeech.2007-336>
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 514–528. <https://doi.org/10.1037/a0020990>
- Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, 52(2), 261–274. <https://doi.org/10.1017/s0261444819000053>
- Siyanova-Chanturia, A., & Spina, S. (2020). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, 70(2). <https://doi.org/10.1016/j.bandl.2017.10.004>
- Siyanova-Chanturia, A., Canal, P., & Heredia, R. R. (2019). Event-related potentials in monolingual and bilingual non-literal language processing. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 508–529). Wiley–Blackwell. <https://doi.org/10.1002/9781119387725.ch25>
- Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & Van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, 175, 111–122. <https://doi.org/10.1016/j.bandl.2017.10.004>
- Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. (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), 776–784. <https://doi.org/10.1037/a0022531>
- Siyanova-Chanturia, A., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(12), 2009–2018. <https://doi.org/10.1037/xlm0000562>
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia, and A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language* (pp. 1–15). Routledge. [ISBN: 9781138345409]
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2019). *Understanding formulaic language: A second language acquisition perspective*. Routledge. <https://doi.org/10.4324/9781315206615-1>
- Siyanova-Chanturia, A., & Van Lancker Sittis, D. (2019). What on-line processing tells us about formulaic language. In A. Siyanova-Chanturia, & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 38–61). Routledge. <https://doi.org/10.4324/9781315206615-3>
- Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). ‘Clap your hands’ or ‘take your hands’? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, 211, 104612. <https://doi.org/10.1016/j.cognition.2021.104612>
- Sonbul, S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, 41(5), 1197–1222. <https://doi.org/10.1017/s014271642000051x>
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63(1), 121–159. <https://doi.org/10.1111/j.1467-9922.2012.00730.x>
- Strandburg, R., Marsh, J., Brown, W., Asarnow, R., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: Linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia*, 31, 413–434. [https://doi.org/10.1016/0028-3932\(93\)90058-8](https://doi.org/10.1016/0028-3932(93)90058-8)
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press. <https://doi.org/10.2307/j.ctv26070v8>
- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt. *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>
- Van Lancker, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics*, 24(1), 45–57. <https://doi.org/10.1017/s0142716403000031>



- Van Lancker, D., & Canter, G. J. (1981). Idiomatic versus literal interpretations of ditropically ambiguous sentences. *Journal of Speech, Language, and Hearing Research*, 24(1), 64–69. <https://doi.org/10.1044/jshr.2401.64>
- Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech, Language, and Hearing Research*, 24(3), 330–335. <https://doi.org/10.1044/jshr.2403.330>
- Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context in open-and closed-class words. *Memory and Cognition*, 19, 95–112. <https://doi.org/10.3758/bf03198500>
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8), 1682–1700. <https://doi.org/10.1162/jocn.2009.21293>
- Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1632–1642. <https://doi.org/10.1037/xlm0000259>
- Vilkaitė, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, 40(2), 329–354. <https://doi.org/10.1093/applin/amx030>
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32(4), 430–449. <https://doi.org/10.1093/applin/amr011>
- Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal*, 20(1), 1–15. <https://doi.org/10.18806/tesl.v20i1.935>
- Wood, D. (2010). *Formulaic language and second language speech fluency*. Bloomsbury Publishing. <https://doi.org/10.5040/9781474212069>
- Yang, S. Y., Ahn, J. S., & Van Lancker Sidtis, D. (2015). The perceptual and acoustic characteristics of Korean idiomatic and literal sentences. *Speech, Language and Hearing*, 18(3), 166–178. <https://doi.org/10.1179/2050572814y.0000000061>

**Anna Siyanova-Chanturia** is Associate Professor in Applied Linguistics at Te Herenga Waka – Victoria University of Wellington, New Zealand. Anna's research interests include second language acquisition and bilingualism, psycholinguistics, vocabulary, dyslexia, and quantitative research methods.

**Suhad Sonbul** is Associate Professor of Applied Linguistics at Umm Al-Qura University, Saudi Arabia. Her research interests include vocabulary teaching/learning, formulaic language, and psycholinguistic measures. Her work has appeared in *Language Learning*, *Language Teaching Research*, *Bilingualism: Language and Cognition*, and *Applied Psycholinguistics*. She is a member of the editorial boards of several flagship journals, including *System* and *Studies in Second Language Acquisition*.