

## ORIGINAL ARTICLE

# Production of ambiguous idioms in English: A reading aloud study

Anna Siyanova-Chanturia<sup>1</sup>  | Phoebe M.S. Lin<sup>2</sup>

<sup>1</sup>Victoria University of Wellington, New Zealand

<sup>2</sup>The Hong Kong Polytechnic University, Hong Kong

## Correspondence

Anna Siyanova-Chanturia, School of Linguistics and Applied Language Studies, Victoria University of Wellington, Kelburn Parade, Wellington 6012, New Zealand.  
Email: anna.siyanova@vuw.ac.nz

In a reading-aloud experiment, we investigated the on-line processing of English idioms from a production perspective. The stimuli were ambiguous idioms used figuratively and literally, and matched novel control phrases. The analysis of the articulatory durations showed a processing advantage for idioms over controls. Further, we found that figurative meanings were articulated somewhat faster than their literal counterparts. The results suggest that the processing advantage for idioms over control phrases, previously reported in comprehension studies, is also present during their production. Unlike the comprehension idiom literature, however, the two idiom meanings might be processed differently during reading aloud. The study concludes with directions for future research, and a case is made for why this line of research is important for the field of applied linguistics.

## KEYWORDS

ambiguous idioms, frequency, multi-word expressions, novel language, production

本研究利用朗读(reading aloud)方法,探究英语习语的在线产生及加工优势。受试者朗读一系列英语语篇,每篇均出现有歧义的习语,或在频率及长度方面与其相配的生成字符串。录音分析结果显示,习语相比生成字符串,前者有明显加工优势。此外,在语境下朗读带喻义及字面义双解的习语,喻义用法朗读时需时较字面义用法者短。本研究为英语习语的加工优势这个课题带来崭新的实验证明,英语习语的加工优势不仅体现在言语感知(perception)方面,还体现在言语产生(production)方面。可是,有别于习语感知研究文献,本研究发现习语带喻义用法及其字面义用法在加工时的差异。本文提出习语加工优势研究的未来发展方向,以及此类研究对应用语言学的重要性。

## 关键词

习语歧义, 多词序列, 生成字符串, 言语产生, 词频频率

## 1 | INTRODUCTION

Frequently occurring word combinations, such as idioms and collocations, have been the focus of research in various linguistic disciplines for many years. Corpus linguists, for example, have long noted that natural corpora contain hundreds of thousands of word combinations whose constituents co-occur together “more often than is explicable in terms of random distribution” (Hoey, 2005: 5). The pervasiveness of frequently occurring word combinations in corpora has been taken to suggest that humans have the ability to “store” large numbers of frequent semi-preconstructed phrases (Sinclair, 1987). It has further been argued that such word combinations form “chunks” in long-term memory (Ellis, 2001). What these propositions imply is that frequent and familiar chunks of language, collectively referred to in the literature as multi-word expressions (MWEs), should be processed – comprehended and produced – differently from less frequent sequences.

Indeed, studies have shown that, due to their frequency, familiarity, and hence predictability (Siyanova-Chanturia & Martinez, 2015), MWEs are comprehended differently from novel language. A number of recent reading studies with idioms, collocations, binomials, phrasal verbs, and other frequent chunks have demonstrated that MWEs are characterized by faster reading times, as well as fewer and shorter fixations in eye-tracking studies than matched novel strings of language (e.g., Arnon & Snider, 2010; Conklin & Schmitt, 2008; Edmonds, 2014; Matlock & Heredia, 2002; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Sosa & MacFarlane, 2002; Underwood, Schmitt, & Galpin, 2004, and others). MWEs have also been found to be judged faster than novel language in the studies that used on-line grammaticality and frequency judgements (e.g., Jiang & Nekrasova, 2007; Siyanova & Schmitt, 2008).

## 2 | PRODUCTION OF MWEs

By and large, the studies mentioned above have focused on the comprehension of recurrent phrases during silent reading. Relatively little is known about the processing of MWEs from a production perspective. Although the existing evidence is still rather scarce, it, nevertheless, suggests that frequent phrases are produced (i.e., articulated) faster than their less frequent novel counterparts. For example, Bybee and Scheibman (1999) and Bell et al. (2003) demonstrated that words were more likely to be phonetically reduced when they appeared in frequent phrases (e.g., *I don't know, middle of the*). Bybee (2000) found that in recurrent word pairs, the boundary between the two words was akin to that between word-internal segments. A processing advantage during the production of frequent versus infrequent sequences has also been observed in young children. Bannard and Matthews (2008) found that two- and three-year old children's phrase production was modulated by the phrase frequency; higher frequency phrases were articulated more quickly than lower frequency ones. More recently, the studies looking at the production of n-grams varying in size, syntactic structure and the level of abstractness (e.g., Arnon & Cohen Priva, 2014; Janssen & Barber, 2012; Tremblay & Tucker, 2011) have reaffirmed the earlier findings reported in the literature, as well as provided a richer account of the frequency effects involved in the production of units above the word level.

The above studies have all looked at highly frequent MWEs. Less frequent MWEs have also been investigated from a production perspective. Hallin and Van Lancker Sidtis (2017) explored the prosodic characteristics of highly familiar, low frequency Swedish proverbs spoken by ten adults and five children. These participants produced seven proverbs (e.g., *bättre sent än aldrig* meaning “better late than never”) and matched control sentences (e.g., *bättre kallt än ljummet* meaning “better cold than lukewarm”). All adult participants reported 100% recognition of the proverbs, while for children the figure was around 69%. Thus, the items were deemed highly familiar. In terms of the frequency, all proverbs were very low frequency items (0–0.3 occurrences per million words in a reference corpus). The analyses revealed that the proverbs exhibited less stressed tonal patterns and faster articulation rate than novel sentences. In

line with earlier research (e.g., Bell et al., 2003; Bybee & Scheibman, 1999), Hallin and Van Lancker Sidtis (2017) concluded that proverbs are phonetically distinct structures.

## 2.1 | Production of idioms

With regard to what are arguably some of the most prototypical of all MWEs – idioms (e.g., Van Lancker Sidtis, 2012), only a handful of studies have focused on the mechanisms involved in their on-line production. In particular, one kind of expressions has been looked at – ambiguous idioms. Ambiguous idioms (also known as ‘ditropically’ ambiguous idioms) are interesting exemplars to investigate because they can be used both figuratively and literally. For example, *twist someone’s arm* can be used to mean either “to pressure or force someone to do something they do not want to do” or “to physically hurt someone”. On the contrary, the idiom *be on cloud nine* can only mean “to be extremely happy”; that is, it can only be used figuratively. Because of the availability of two distinct meanings, ambiguous idioms have been extensively studied in the linguistic and psycholinguistic literature and various accounts have been proposed as to how the two meanings should be processed in relation to each other (e.g., Bobrow & Bell, 1973; Cacciari & Tabossi, 1988; Swinney & Cutler, 1979). The two idiom renderings are identical in form; they share the same constituent words and are, therefore, perfectly matched for individual word frequency and length (the factors known to affect word recognition, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). This makes them ideal candidates for the investigation of frequent figurative versus infrequent literal phrasal processing.

In what is, probably, one of the earliest idiom studies, Van Lancker, Canter, and Terbeek (1981) carried out acoustic and phonetic analyses on a range of literal and idiomatic utterances that were produced (read aloud) by five speakers. These authors found that literal renderings of ambiguous idioms were characterized by longer durations and longer and more numerous pauses than their figurative equivalents. Van Lancker et al. (1981) concluded that various prosodic cues, such as sentence duration, word duration and pausing marked utterances as either literal or figurative. The idea that the two meanings available in ambiguous idioms have distinct articulatory characteristics was later confirmed in Van Lancker Sidtis (2003) and Yang, Ahn, and Van Lancker Sidtis (2015). In Van Lancker Sidtis (2003), participants judged the nature (i.e., whether the intended meaning was figurative or literal) of pre-recorded ambiguous idioms in English. It was found that a range of cues helped the judges successfully differentiate between idioms used figuratively and idioms used literally, suggesting that the two idiom uses were phonetically distinct. In a similar study, Yang et al. (2015) investigated listeners’ ability to distinguish between literal and figurative instances of ambiguous idioms in Korean. These authors found that native speakers of Korean were successful in the identification of the intended meaning. The cues that helped the listeners were, among others, greater intensity and shorter durations observed for figurative meanings compared to literal ones.

The findings of the three production studies described above, Van Lancker et al. (1981), Van Lancker Sidtis (2003), and Yang et al. (2015), all point to the conclusion that when presented in isolation, figurative and literal renderings of ambiguous idioms are processed according to different characteristics. One of the key findings commonly reported is faster articulation times for figurative meanings compared to literal ones. Interestingly, this goes against the results reported in comparable comprehension studies. For example, using the eye-tracking paradigm, Siyanova-Chanturia, Conklin, & Schmitt (2011) investigated the comprehension of idioms embedded in story contexts. These authors looked at figurative meanings of ambiguous idioms (e.g., *at the end of the day* – ‘eventually’), literal meanings of ambiguous idioms (e.g., *at the end of the day* – ‘in the evening’), and closely matched novel phrases (e.g., *at the end of the war*). As expected, native speakers of English read the idioms more quickly than novel sequences. Importantly, unlike the differences in articulation rates reported in the production literature, figurative and literal meanings were read at a similar pace. No differences were observed between the two idiom renderings either in the full idiom analysis, or in the recognition point analysis where the utterances were looked at before and after the recognition point (a recognition point is the point at which the idiom becomes recognisable as a conventional phrase; for example, for the idiom *on the other hand* the recognition point is the word *other*). The analyses of Siyanova-Chanturia, Conklin, & Schmitt (2011) showed similar processing patterns for figurative and literal interpretations in ambiguous idioms.

### 3 | THE PRESENT STUDY

The above review suggests a wealth of studies that have looked at the comprehension of various kinds of MWEs, fewer studies on the production (articulation) of MWEs, and a lack of production studies that focus specifically on ambiguous idioms. In addition, in the studies reviewed above (e.g., Van Lancker et al., 1981; Van Lancker Sidtis, 2003; Yang et al., 2015), the idioms were presented in isolation. In other words, no wider context was provided. This might have contributed to the unnaturalness of the task, since human interactions tend to happen in meaningful contexts. What remains an answered question is how figurative and literal meanings, embedded in larger contexts, are processed during production. The goal of the present investigation was to address this gap in the literature. To this aim, a group of native speakers of English was recorded while reading passages that contained one of the following: an idiom used figuratively, the same idiom used literally, or a control phrase. Our overarching research questions were:

1. Do idioms exhibit a processing advantage over matched novel phrases during reading aloud?
2. Do figurative interpretations of ambiguous idioms exhibit a processing advantage over their literal counterparts during reading aloud?

#### 3.1 | Method

##### 3.1.1 | Participants

Sixty-six native speakers of English took part in the study. All participants were full-time undergraduate students at a large public research university in the UK. They received (linguistics) course credit for their participation, which was voluntary. These participants were chosen because they can be deemed representative of young adult native speakers of British English. The research was conducted in accordance with the standard ethical procedures of the university.

##### 3.1.2 | Experimental materials

The experimental materials were borrowed from Siyanova-Chanturia, Conklin, & Schmitt (2011). They were 21 idioms used figuratively (*at the end of the day* – ‘eventually’), the same 21 idioms used literally (*at the end of the day* – ‘in the evening’), and 21 matched, novel control phrases (*at the end of the war*). The idioms used in the study were ambiguous idioms; that is, they could be used figuratively as well as literally. Control phrases were created such that they were as close to the idiom (in the form) as possible, and their constituent words were matched with the constituents of the corresponding idiom as closely as possible (e.g., *at the end of the day* vs. *at the end of the war*). The frequency of the novel phrase was much lower than that of the idiom, because, by definition, it was a newly created, fully compositional literal phrase (see Siyanova-Chanturia, Conklin, & Schmitt (2011) for a detailed description of the materials and norming procedures).

The original study by Siyanova-Chanturia, Conklin, & Schmitt (2011) employed a reading task. Given that the present study focused on the production (i.e., articulation) of idioms and control phrases, we checked the target items for possible differences in phoneme duration. To this aim, the Lancaster/IBM Spoken English Corpus (Knowles, Williams, & Taylor, 1996) was used. *T*-tests indicated that there were no significant differences between the idioms and controls in terms of phoneme duration ( $t = 0.13$ ,  $p = 0.89$ ), or the number of syllables ( $t = 0.09$ ,  $p = 0.93$ ).

The frequency information was obtained from the British National Corpus (BNC). A British corpus was used because the original (as well as the current) study was carried out in the UK. In addition, at the time the original study was conducted, the BNC was the best and largest corpus available. While some of the target idioms are highly frequent (e.g., *on the other hand*, 5,311 occurrences in the BNC), others are relatively infrequent (e.g., *can't judge a book by its cover*, 11 occurrences in the BNC). In the original study, it was argued that idiom frequencies cannot be directly compared with those of single words. That is, while 11 occurrences in the BNC might be very low for a word, it is, nevertheless, rather frequent for a longer expression, such as *can't judge a book by its cover*. Still, given a number of

idioms of relatively low frequency that were used as stimuli (see Appendix A), we deemed it important to provide additional measures of their 'formulaicity'. In the original study, Siyanova-Chanturia, Conklin, & Schmitt (2011) provided only one statistical measure – frequency. Given a wide range of frequencies of the target idioms (11–5,311 in the BNC; all frequencies borrowed from the original study), we also calculated other frequently used in corpus linguistics statistical measures – t-score and mutual information (MI).

A t-score has been widely used for collocation identification in corpus linguistics. It tells us how probable or improbable it is that a certain constellation of words will occur (Manning & Schütze, 1999). However, t-score correlates strongly with frequency. A measure that does not correlate strongly with frequency and which may thus be a more accurate measure of association strength to be used with lower frequency items is MI. MI tells us how strongly associated the words are and how likely it is that they will co-occur together. According to Manning and Schütze (1999), MI is particularly good for identifying 'interesting collocations'. A t-score of  $\geq 2$  and a MI score of  $\geq 3$  are often cited as a threshold for identifying significant collocations (e.g., Durrant & Doherty, 2010; Hunston, 2002; Stubbs, 1995).

Based on the combined frequency information, t-score and MI scores (Appendix A), we can argue that all the target idioms, irrespective of frequency, are prototypical idiomatic expressions (i.e., recurrent, familiar and predictable).

It is important to note that a BNC search conducted in the original study showed that all 21 idioms appeared figuratively more frequently than they did literally (83% of figurative uses versus 17% of literal uses). That is, if the idiom *at the end of the day* is found in the BNC  $n$  number of times, most of these times, it is used in its figurative rather than literal sense. Thus, we can argue that in ambiguous idioms, figurative senses are generally more frequent than literal meanings.

As was mentioned above, Siyanova-Chanturia, Conklin, & Schmitt (2011) conducted two analyses of the figurative and literal senses. First, they looked at the reading of the entire sequence. Second, they looked at the processing of the two uses before and after the idiom's recognition point. In a series of cloze tests with a group of native speakers, these authors identified a recognition point for 14 out of the 21 target idioms. The remaining seven idioms were not found to have a reliable recognition point. Their rationale behind this analysis was to find out whether the (silent) reading of the figurative versus literal meaning might differ before and after the recognition point. In the present investigation, we too wanted to look at the articulation times for the entire string, as well as before and after the point at which a phrase is recognized as routinized. To this aim, recognition point locations were also borrowed from Siyanova-Chanturia, Conklin, & Schmitt (2011).

Finally, the three conditions – idioms used figuratively, idioms used literally and matched controls – were embedded in story contexts (see Appendix B for an example). The story contexts used in the present study were identical to those used in the original study.

### 3.1.3 | Procedure and task

The stories were presented across three presentation lists. Thus, each list contained exactly 21 items: seven idioms used figuratively, seven idioms used literally, and seven novel phrases. Each participant saw only one version of each phrase. The story order was randomized. Participants were seated in a quiet room; they were provided with the instructions and were asked to complete a short practice session. They were asked to read the stories aloud, one by one in their most natural way, while being recorded with a high quality digital recorder. They were asked to avoid coughing, sneezing, and getting distracted as much as possible.

Because the aim of this study was to investigate articulatory aspects in terms of durations, associated with figurative and literal idiom uses, it was necessary to provide sufficient story contexts, in order to support either the figurative or literal meaning. Asking participants to read aloud the target sequences embedded in story contexts was thus deemed the best way to get participants produce the idioms (but see a discussion of this approach below).

## 3.2 | Analysis and results

We coded how long, in milliseconds, it took each participant to articulate the sequence. We measured the duration of each sequence from the onset of the first word of the sequence to the end of the last word of the sequence. To this

aim, Praat software was used. Having obtained such measures, we compared the articulatory durations of idiom figurative and literal uses, and novel phrases. One articulatory duration was obtained per each sequence.

Four participants were excluded from the analysis due to unnecessary pauses, excessive coughing, sneezing, or other noise. The analysis thus included the data from 62 participants. Twenty-one articulatory durations were obtained for each participant. Articulatory durations  $\pm 3SD$  were deleted prior to the analysis, resulting in 1.8% of lost data (evenly distributed across the conditions). The analyses were performed in two ways: the full idiom analysis and the recognition point analysis. The Greenhouse–Geisser procedure was applied on every within-subject effect where the sphericity assumption was violated.

3.2.1 | Full idiom analysis

In total, 21 figurative and 21 literal uses, and 21 control phrases were included in this analysis. Articulatory durations were analysed using repeated measures ANOVA treating participants ( $F_1$ ) and items ( $F_2$ ) as random variables. The dependent variable was mean total articulatory duration (Table 1).

A main effect of Phrase Type was found significant in the analysis by participants and items (Table 2). Planned comparisons showed that figurative idiom senses were articulated significantly faster than novel phrases. Literal idiom readings, on the other hand, did not differ significantly from novel phrases. Finally, no significant differences were found between the two idiom readings in the analysis by participants. The analysis by items showed that figurative phrases were articulated somewhat faster than literal ones. This difference, however, was found to be approaching significance ( $p = 0.07$ ).

3.2.2 | Recognition point analysis

The control phrases were not included in the recognition point analysis because, being novel word combinations, they do not have a recognition point. Thus, only the 14 idioms for which a recognition point was identified (see above) and their literal equivalents were included in this analysis.

The mean articulatory durations are given in Table 3, while the statistical comparisons are summarized in Table 4. Before the recognition point in the analysis by participants, we observed no processing differences between figurative versus literal idiom interpretations. However, the analysis by items suggested a small but significant processing advantage for figurative idiom uses. After the recognition point, no differences between the two idiom senses were found in the analysis by participants or items.

TABLE 1    Articulatory durations (in ms) for the three conditions

Full idiom analysis		
Figurative	Literal	Control
879 (149*)	923 (160)	952 (159)

\*Standard Deviation in parentheses (SD).

TABLE 2    Analyses of variance and planned comparisons. Full idiom analysis

	Df	F <sub>1</sub>	p	Partial $\eta^2$	Df	F <sub>2</sub>	p	Partial $\eta^2$
Phrase type	2,122	3.30	0.04	0.05	2.40	3.44	0.04	0.15
	Df	t <sub>1</sub>	p	Cohen's d	Df	t <sub>2</sub>	p	Cohen's d
Figurative vs. control	61	2.93	0.005	0.48	20	2.54	0.02	0.18
Literal vs. control	61	0.97	0.34	0.18	20	0.84	0.41	0.06
Figurative vs. literal	61	1.42	0.16	0.29	20	1.90	0.07	0.12

Numbers in bold shows the significant result.

**TABLE 3** Articulatory durations (in ms) for the two idiom uses before and after the recognition point

Recognition point analysis			
Before		After	
Figurative	Literal	Figurative	Literal
459 (144*)	500 (158)	446 (55)	464 (67)

\*Standard Deviation in parentheses (SD).

**TABLE 4** Planned comparisons for the recognition point analysis

	Df	t <sub>1</sub>	p	Cohen's d	Df	t <sub>2</sub>	p	Cohen's d
<b>Before</b>								
Figurative vs. literal	61	1.29	0.20	0.27	13	2.14	<b>0.05</b>	0.15
<b>After</b>								
Figurative vs. literal	61	1.58	0.12	0.28	13	1.36	0.19	0.08

Number in bold shows the significant result.

## 4 | DISCUSSION

In the present investigation, we looked at native English speaker processing of familiar idioms used figurative and literally, as well as infrequent novel phrases presented in a story context. The rationale for the study was to establish whether the processing advantage for recurrent phrases over infrequent strings found in a wealth of comprehension studies equally extends to their production. Specifically, our aims were, first, to compare the production of frequent idioms and infrequent control strings, and, second, to compare articulatory durations for the two idiom interpretations (figurative vs. literal). In addition, we wanted to compare the processing of the two idiom renderings before and after the recognition point, as has previously been done in the literature (e.g., Siyanova-Chanturia, Conklin, & Schmitt (2011); Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010).

Two findings emerged from the present study. First, as was shown in the full idiom analysis, figurative idiom uses were articulated significantly faster than control phrases. This finding is in line with that of Siyanova-Chanturia, Conklin, & Schmitt (2011), who, in an eye-tracking experiment using the same stimuli, observed shorter and fewer fixations on idioms relative to novel language. The processing advantage for idioms observed in the present production study is comparable to faster reading times for other types of MWEs reported in comprehension studies, such as frequent chunks (e.g., Arnon & Snider, 2010), binomials (e.g., Siyanova-Chanturia, Conklin, & Schmitt (2011), phrasal verbs (e.g., Matlock & Heredia, 2002), frequent collocations (e.g., Sosa & MacFarlane, 2002), and other kinds of MWEs. This finding is also in line with the production studies, which report on phonetic reduction and faster articulation times for a range of MWEs (e.g., Bell et al., 2003; Bybee, 2000; Bybee & Scheibman, 1999; Hallin & Van Lancker Sidtis, 2017), and further lends support to the results of recent studies looking at n-gram frequency effects in elicited and spontaneous speech (e.g., Arnon & Cohen Priva, 2014; Janssen & Barber, 2012; Tremblay & Tucker, 2011).

Second, we found that the more frequent figurative and the less frequent literal idiom uses might not be processed at a similar speed, as has been shown in comprehension studies with native English speakers. Conklin and Schmitt (2008) and Siyanova-Chanturia, Conklin, & Schmitt (2011) found no differences in reading times, and fixation durations and fixation counts for the two idiom interpretations. The findings of the present investigation, however, tentatively suggest that this might be different in production. First, in the full idiom analysis, the literal interpretation was not read significantly faster than the control. This implies that the production of the low frequency literal meaning of a frequent idiom is not unlike that of the low frequency novel control phrase. Second, the recognition-point analysis showed that literal idiom uses might, in fact, be articulated somewhat slower than their figurative counterparts. This fits in well with the fact that figurative uses are more frequent than literal ones, as suggested by the frequencies



obtained from the BNC. It needs to be pointed out, however, that this finding is only suggestive, as the effect was found to be small and significant only in the analysis by items. Because this effect was not significant in the analysis by participants, we cannot make unequivocal claims about possible differences between figurative and literal meaning processing during reading aloud.

Nevertheless, our results do seem to offer some support to those of Van Lancker Sidtis (2003), who investigated the prosodic cues of pre-recorded English idioms. She found that such cues helped native English speakers distinguish between idioms used figuratively and literally, suggesting that the two idiom interpretations are phonetically distinct. Similarly, in Van Lancker et al. (1981) and Yang et al. (2015), figurative idiom uses were characterized by shorter durations and shorter phrase-internal pauses compared to literal uses. These findings led the authors to conclude that “the physical shapes” of literal and figurative phrases reflected their linguistic forms and functions (Van Lancker et al., 1981: 334).

It is noteworthy that the difference between the two idiom meanings emerged before, rather than after the recognition point, contrary to what was expected by Siyanova-Chanturia, Conklin, & Schmitt (2011). Indeed, it would be logical to predict that, if the two meanings are processed differently, the difference should become evident after the reader has reached the recognition point. However, our findings, albeit only speculative, suggest that this prediction might not hold true in production. Further research will be able to determine whether the processing difference between the more frequent figurative rendering and its less frequent literal counterpart is more likely to manifest itself before or after the recognition point.

#### 4.1 | Reading aloud: Interpreting the results

A few words need to be said about the task involved in the experiment. It has been pointed out that reading aloud is only half as fast as natural silent reading and thus may not share the same processes (Field, 2004). Buswell's (1929) classic study that investigated the eye-voice span (EVS) – the time that elapses between fixating a word and reading it out loud – showed that in skilled readers, the average span was around two words. These findings suggest that the viewing of a word (the identification stage) precedes its articulation (the articulation stage). Balota and Chumbley (1985) argue that speech realization is rather time-consuming, such that the articulation of a word initiates only after the eyes have moved to the following word/s. Indeed, Laubrock and Bohn (2008) found that more than half a second intervened between the viewing of a word and its eventual articulation. What this implies is that when a single word (presented in context) is under investigation, its production time reflects mostly the articulation stage of on-line processing. Given that the stimuli in the present study were long multi-word sequences, the above does not hold true. That is, the obtained reading aloud time for a given idiom was cumulative (i.e., it was made of durations for *all* idiom constituents) and thus reflected *both* the identification and articulation stages for the words within the idiom. Based on the two-word span findings reported in the EVS studies, we can argue that a reader identified word *n* of an idiom, while preparing to articulate word *n-1*, while articulating word *n-2*. All of these processes, we believe, are reflected in each duration obtained for a given idiom.

Based on the EVS findings, researchers have argued that reading aloud might not be able to provide same insights into the on-line processes of reading as silent reading (Field, 2004), which might explain the differences between the current study and that of Siyanova-Chanturia, Conklin, & Schmitt (2011). Critically, our findings point to the conclusion that future studies using elicited (rather than spontaneous) speech should be able to disentangle the processes involved in the identification and articulation stages. In order to separate the recognition and production stages, Arnon and Cohen Priva's (2013) participants were first visually presented with a *n*-gram; they were then asked to produce it once it disappeared from the screen. Because one of the aims of this study was to look at ambiguous idiom production in meaningful story contexts, it was not possible to adopt a similar design. And, given that literal uses of ambiguous idioms are rare, it was not possible to extract such instances from a corpus of spontaneous speech (as was done in Arnon & Cohen Priva, 2014). Future work will be able to clarify the present findings and to find better ways of pinning down the processes involved in the production of ambiguous idioms in context.



We interpret our findings in line with those of Hallin and Van Lancker Sidtis (2017), Lin (2010a) and others who have reported systematic differences in the prosodic characteristics for conventional and novel language. The differences in the speech rates observed for idioms versus novel language indicate faster processing for the former; furthermore, these differences suggest that a prosodic pattern might in fact be stored together with the form of the figurative expression (e.g., Hallin & Van Lancker Sidtis, 2017; also, see Calhoun & Schweitzer, 2012; Schweitzer et al., 2015). More broadly, the findings of the present study support a model of language which argues that frequent and familiar phrasal units differ from newly constructed propositional word sequences in how they are acquired, processed and used (e.g., Van Lancker Sidtis, 2012; Wray & Perkins, 2000).

## 5 | IMPLICATIONS FOR THE FIELD OF APPLIED LINGUISTICS

Finally, it is important to consider some of the implications of this line of research for the field of applied linguistics. Specifically, why does it matter whether or not MWEs are characterized by distinct phonology when compared to novel language? There are a number of noteworthy reasons.

First, researchers have long called for a greater focus on the prosody of MWEs within applied linguistics. As Lin (2012) points out, although we are beginning to learn more about MWEs, their learning and use, we still know surprisingly little about the phonology of MWEs. A more thorough investigation of the phonetic cues associated with MWEs is needed to be able to better understand and verify formulaicity of word combinations, and, more importantly, to help us identify the mechanisms by which we learn and remember language (e.g., Lin, 2012). Indeed, the tenet that a range of phonetic cues (e.g., speed of articulation, intonation, tone, stress and rhythm) are key to how languages are learned has long been recognized in the literature on first language acquisition (e.g., Peters, 1977). Admittedly, the present study has looked at only one manifestation of MWE phonology – articulatory durations – and thus has barely scratched the surface of the multitude of factors affecting MWE prosody. A lot still remains to be addressed in future research. For example, to date, there have not been any studies looking at the phonology of MWEs in a second language. The studies reviewed in this paper, albeit limited, do point to important phonological differences between MWEs and novel language in a first language. How the production of MWEs versus novel language might manifest itself in a second language remains virtually an uncharted territory. Because second language acquisition is distinct from first language learning in terms of the conditions and outcomes, we might expect to observe both qualitative and quantitative differences between MWE production in native and non-native speakers. We might also expect the articulatory pattern to vary as a function of such variables as second language proficiency and age of acquisition. The research into MWE phonology is still very much in its infancy, but it may truly be “the missing piece of the jigsaw in our understanding of formulaicity” (Lin, 2012: 345).

Second, the results with native speakers of English presented here as well as the findings reported in earlier studies have important implications for second language pedagogy. If MWEs have a unique prosody compared to literal language, then prosody should clearly be part of a MWE teaching syllabus. As suggested by Lin (2010b), teachers should consider drawing learners' attention specifically to the fact that MWEs are pronounced in a certain manner, one that is distinct from the way in which novel, propositional language is articulated. As noted by Ashby (2006), any deviation from the conventional pronunciation pattern of idioms will result in the loss of the original idiomatic meaning. We believe that awareness-raising activities coupled with speech fluency training with MWEs as practice material can positively influence the comprehensibility and fluency of second language learners' speech (e.g., Dechert, 1983; Raupach, 1984). Although currently little empirical evidence exists, we can hypothesise that there should be a link between second language learners' comprehensibility and perceived fluency, and their delivery of MWEs. If learners' articulation of MWEs deviates from native speaker norms, the end result is likely to be low comprehensibility and poor fluency. How exactly such deviations may affect second language learners' speech, and what consequences there might be for a native speaker listener, remains an important unanswered question.

Lastly, raising second language learners' awareness of the prosody of MWEs will help learners disambiguate figurative and literal uses of idioms and other instances of non-literal language, such as jokes and irony. Human language is inevitably ambiguous. While native speakers will have had many years of experience to be able to successfully disambiguate *He's skating on thin ice* or *Yeah, right* as figurative or literal instances based purely on the phonological cues available, lower proficiency speakers are likely to miss these cues, which may result in a breakdown in communication. Again, including prosody in a MWE teaching syllabus will benefit second and foreign language learners.

In conclusion, in this study investigating the on-line production of frequent figurative meanings, infrequent literal meanings and matched novel phrases, we observed a general effect of phrasal frequency on phonetic duration: higher frequency phrases were characterized by shorter durations; correspondingly, lower frequency phrases exhibited longer durations. In addition, our findings have added to a limited body of research into the processing of ambiguous idioms from a production perspective, and offered some support to the view according to which figurative and literal interpretations might be processed according to different principles.

## REFERENCES

- 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.
- Arnon, I., & Cohen Priva, U. (2014). The changing effect of word and multiword frequency on phonetic duration for highly frequent phrases. *The Mental Lexicon*, 9(3), 377–400.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82.
- Ashby, M. (2006). Prosody and idioms in English. *Journal of Pragmatics*, 38(10), 1580–1597.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24(1), 89–106.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133(2), 283–316.
- 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.
- 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.
- Bobrow, S., & Bell, S. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, 1(3), 343–46.
- Buswell, G. T. (1929). The relationship between eye-perception and voice response in reading. *Journal of Educational Psychology*, 12(4), 217–227.
- Bybee, J. (2000). The phonology of the lexicon. In M. Barlow, & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. 65–85). Stanford, CA: CSLI Publications.
- Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics*, 37(4), 575–596.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27(6), 668–83.
- Calhoun, S., & Schweitzer, S. (2012). Can intonation contours be lexicalised? Implications for discourse meanings. In G. Elordieta & P. Prieto (eds.) *Prosody and Meaning* (Interface Explorations 15, pp. 271–328), Berlin: De Gruyter Mouton.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89.
- Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch, & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp. 175–195). London: Longman.
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2), 125–155.
- Edmonds, A. (2014). Conventional expressions. Investigating pragmatics and processing. *Studies in Second Language Acquisition*, 36(1), 69–99.
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge: Cambridge University Press.
- Field, J. (2004). *Psycholinguistics: The key concepts*. London: Taylor and Francis.

- Hallin, A. E., & Van Lancker Sidtis, D. (2017). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics*, 38(1), 68–89.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PloS One*, 7(3), e33202. <https://doi.org/10.1371/journal.pone.0033202>
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433–445.
- Knowles, G., Williams, B., & Taylor, L. (1996). *A corpus of formal British English speech: the Lancaster/IBM Spoken English Corpus*. London: Longman.
- Laubrock, J., & Bohn, C. (2008). Broadening the scope of eye-movement research in reading: Oral reading and proof reading. *International Journal of Psychology*, 43(3), 190.
- Lin, P. M. S. (2010a). The phonology of formulaic sequences: A review. In D. Wood (Ed.), *Perspectives on the formulaic language: Acquisition and communication* (pp. 174–193). London: Continuum.
- Lin, P. M. S. (2010b). *The prosody of formulaic sequences in spontaneous speech*. The University of Nottingham: Unpublished doctoral dissertation.
- Lin, P. M. S. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, 33(3), 342–347.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Matlock, T., & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. Heredia, & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 251–274). Amsterdam: Elsevier Press.
- Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language*, 53(3), 560–573.
- Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Tübingen: Gunter Narr.
- Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalized storage of intonation. *Speech Communication* 66, 65–81.
- Sinclair, J. M. (1987). Collocation: a progress report. In R. Steele, & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday*, Vol. 2 (pp. 319–331). Amsterdam: John Benjamins.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429–258.
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251–272.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776–784.
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, 36(5), 549–569.
- Sosa, A., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language*, 83(2), 236–248.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of Language*, 2(1), 1–33.
- Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, 18(5), 523–534.
- Tremblay, A., & Tucker, B. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2), 302–324.
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 153–172). Amsterdam: John Benjamins.
- Van Lancker, D., Canter, G., & Terbeek, D. (1981). Disambiguation of Ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24(3), 330–335.
- Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The handbook of the neurophysiology of language* (pp. 342–367). Oxford: Blackwell.
- Van Lancker Sidtis, D. (2003). Auditory recognition of idioms by first and second speakers of English. *Applied Psycholinguistics*, 24, 45–57.

- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8), 1682–1700.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20(1), 1–28.
- Yang, S., Ahn, J. S., & Van Lancker Sidtis, D. (2015). The perceptual and acoustic characteristics of Koreana idiomatic and literal sentences. *Speech, Language and Hearing*, 18(3), 166–178.

**How to cite this article:** Siyanova-Chanturia A, Lin PMS. Production of ambiguous idioms in English: A reading aloud study. *Int J Appl Linguist*. 2017;1–13. <https://doi.org/10.1111/ijal.12183>

## APPENDIX A

### A1 | Idioms used in the study, their frequency, t-score, and MI score

	Idiom	Idiom frequency <sup>a</sup>	t-score	MI
1	(a) breath of fresh air	89	9.4	25.6
2	(a) piece of cake	70	8.4	13.2
3	add fuel to the fire	14	3.7	27.6
4	(as) cold as ice	24	4.9	13.0
5	at the end of the day	760	27.6	25.5
6	kill two birds with one stone	36	6.0	45.8
7	leave a bad taste in your mouth	13	3.6	37.6
8	left in the dark	17	4.1	11.3
9	cut a long story short	39	6.2	32.8
10	not my cup of tea	19	4.3	26.6
11	on the other hand	5,311	72.9	17.9
12	pain in the neck	36	6.0	16.4
13	put your foot down	112	10.6	24.4
14	ring a bell	75	8.7	14.0
15	see which way the wind is blowing	23	4.8	30.3
16	sick and tired	58	7.6	13.7
17	(the) other side of the coin	63	7.9	25.8
18	tie the knot	48	6.9	15.4
19	twist my arm	36	6.0	17.9
20	under your nose	104	10.2	15.0
21	(you) cannot judge a book by its cover	11	3.3	37.4

<sup>a</sup>Frequencies, obtained from the British National Corpus, are given per 100 million words.

## APPENDIX B

Examples of stories used in the study (the materials were borrowed from Siyanova-Chanturia, Conklin, & Schmitt (2011). In the reading aloud experiment, the target sequences were not underlined.

### B1 | Figurative

I had my younger brother and my sister-in-law over for dinner yesterday. They both have their degrees from Cambridge, whereas most of the people they work with have theirs from less well-known overseas and British universities. Personally, I think you can have the highest degree from the best university in the world, but at the end of the day it's your contribution to the society that matters, and not the name of the university you went to at all. Sadly, they didn't agree with me.

### B2 | Literal

After my second year at university, I moved house. When I started packing, I realized that I had a lot more stuff than I had when I moved in as a first-year student. The house I was moving to was next door to the house I was moving from, which was very handy. However, I still had to carry most of my stuff in small boxes from my old room to the new one. I had to make at least 50 trips so at the end of the day I was absolutely exhausted. I'm hoping to stay at this house for at least another two years. I really don't want to move any more.

### B3 | Novel

One of my granddads was an army officer for most of his life. Despite being an army guy, he's always been a very humane and kind person. He is also a very artistic and creative person. For example, one of his hobbies is writing poetry. He's a retired man now who served in Vietnam and who's been through many things in his life, so he's got plenty of things to write about. I know that at the end of the war he went on to teach students at the Military Academy. That was something he found particularly challenging but also rewarding in many respects.