

Familiarity and Decomposability Modulate the Prosodic Realization of Figuratively vs. Literally Intended Idioms During Natural Speech Production

Kyle Lovseth^{1,2}, Lauren de la Parra^{1,2}, Michael Wagner^{2,3}, Debra A. Titone^{1,2}

¹Department of Psychology, McGill University
1205 avenue Dr. Penfield, Montreal, Canada

²CRLMB – Centre for Research on Language, Mind and Brain
3640 rue de la Montagne, Montreal, Canada

³Department of Linguistics, McGill University
1085 avenue Dr. Penfield, Montreal, Canada

kyle.lovseth@mail.mcgill.ca

lauren.delaparra@mail.mcgill.ca

chael@mcgill.ca

dtitone@psych.mcgill.ca

Abstract. *Idioms are fixed expressions whose figurative meanings are not usually derived from a literal interpretation of their component words. When people produce literally plausible idioms, such as “kick the bucket”, their literal and figurative productions can be acoustically distinguished by prosodic cues such as word duration and fundamental frequency. To the extent that speakers produce these cues reliably bears on whether idioms are mentally represented as holistic units or independent constituent words. In this study, we investigated whether linguistic attributes known to track the lexicalization of idioms (e.g., familiarity and decomposability), affect the prosodic realization of figuratively and literally intended idioms in natural speech. We recorded 14 native-English speakers while they naturally read aloud idioms in figurative or literal biasing contexts. The results showed that both familiarity and decomposability affected the prosodic realization of idioms intended figuratively versus those intended literally. This suggests that prosodic cues are differentially utilized in speech production for idioms that are more likely to be lexicalized (i.e., those that are highly familiar and semantically nondecomposable) versus idioms that are less likely to be lexicalized (i.e., those that are low familiarity and semantically decomposable).*

1. Introduction

Prosody, also known as the melody of speech, is a level of linguistic representation at which acoustic-phonetic properties of a speech signal vary independently of uttered lexical items (Wagner & Watson, 2010). The perceptual experience of prosody can manifest itself in many ways, which include changes in the duration and fundamental frequency (F_0) of an utterance. Among its many roles, the presence of prosodic cues during speech is crucial for resolving interpretive or semantic ambiguities resulting from nuances in syntactic structure (Cutler, Dahan, & van Donselaar, 1997). For example, consider the phrase, *tap the frog with the flower*. However, whether similar prosodic cues are also used to signal how a phrase is represented in the mental lexicon is an open question. In an attempt to answer this question, researchers have turned to fixed expressions such as idioms, which have a unique ability to be represented in memory both as a lexical unit, and as individual words.

Idioms are classically defined as expressions whose figurative meanings cannot be derived from a traditional compositional analysis of their component words (Fraser, 1970). For example, the figurative meaning of the idiom *kick the bucket* is ‘to die suddenly.’ If one were to compositionally analyze the expression (i.e., build its meaning semantically and syntactically on-demand), they would arrive at a markedly different interpretation (i.e., to literally kick a bucket). This has led to the suggestion that idioms’ figurative meanings are not derived through the same compositional analysis that gives rise to their literal meanings. Rather, they are stored and directly retrieved from memory as holistic units (Libben & Titone, 2008).

With respect to idiom production, some prior work has focused on identifying prosodic differences between idioms spoken in their figurative or literal forms (Van Lancker, Canter, & Terbeek, 1981; Belanger, Baum, & Titone, 2009). For example, Van Lancker et al. found that idioms intended literally had longer spoken durations, as well more frequent rapid changes in F_0 compared to idioms intended figuratively. They concluded that prosody for figurative productions served to fuse constituent words into a single utterance, whereas prosody for literal productions served to individuate the constituent words of idioms. However, differences in temporal and spectral cues were only observed when participants were explicitly instructed to read the idioms with either a figuratively or literally intended meaning. No discernible differences were detected when participants simply read idioms naturally. There are two possible explanations for this result. First, it is possible that idioms varied in ways that diluted the observation of prosodic effects during natural speech production, such as idiom familiarity (Libben & Titone, 2008). Second, methodological aspects of the study may have minimized the likelihood of observing prosodic effects during natural speech. For example, speakers encountered idioms in contexts that biased their figurative and literal interpretations simultaneously. Also, the idioms produced varied in their syntactic structure and in length, which may have increased variability with respect to how they were realized prosodically.

More recently, Belanger, Baum, and Titone (2009) investigated prosodic cues for idioms that were shorter and had a more consistent syntactic structure than those used by Van Lancker et al. (1981). Similar to Van Lancker et al., they used the same method of presenting idioms in figurative and literal contexts simultaneously, and instructed participants to explicitly contrast the figurative and literal meanings

prosodically. Belanger et al. did not manipulate idiom familiarity, however, they did investigate the effect of decomposability on the prosodic realization of idioms. Decomposability can be defined as the degree to which the figurative and literal interpretations of an idiom overlap in meaning. Accordingly, decomposable idioms have a high degree of overlap between their figurative and literal meanings (e.g., *missed the mark*), while nondecomposable idioms have little or no overlap between the two meanings (e.g., *kick the bucket*). Although early work suggests that increased compositionality aids comprehension (e.g., Gibbs, Nayak, & Cutting, 1989), recent work suggests that compositional processes play a limited role in appreciating an idiom's figurative meaning, especially if idioms are highly familiar (Libben & Titone, 2008). Whether a similar limited role of decomposability holds for language production is an open question.

To investigate compositionality effects during idiom production, Belanger et al. examined prosodic effects for idioms in which the noun, verb, or both contributed semantically to the figurative meaning. Their participants consisted of healthy older adults and neurologically damaged patients matched in age. We focus here only on the results for the healthy older adult group. In contrast with Van Lanker et al. (1981), phrase-final nouns for figuratively intended utterances had longer durations than those of literally intended utterances, although there was no difference between the two in *F0*. Moreover, differences in duration between verbs and nouns were greatest for idioms whose nouns solely contributed semantically to the figurative meaning. Although their results differ from Van Lancker et al., Belanger et al. similarly concluded that significant prosodic differences exist between utterances intended figuratively versus literally. Furthermore, these differences manifest themselves variably at different levels of decomposability. While they did not observe any differences in *F0* between idioms intended figuratively or literally, it is possible that a younger adult population would show differences. It is also possible that the results would differ when idioms were produced naturally, that is, when speakers were not explicitly instructed to differentiate figurative and literal interpretations prosodically.

We investigated whether young adult speakers naturally used prosodic cues to differentiate between the figurative and literal representations of idiomatic expressions during speech production. We were particularly interested in whether differences between the two interpretations would arise even if we did not explicitly inform participants that the phrases could be interpreted in multiple ways. Like Belanger et al., our idioms had a consistent *verb + determiner + noun* syntactic structure. However, our idioms varied continuously in familiarity, ranging from low to high familiar idioms. Similarly, to test for compositional effects, we used continuous measures of verb relatedness, noun relatedness, and global decomposability for each idiom. Unlike previous studies, participants only produced a given idiom in either its figurative or literal context, but not both. Moreover, participants were instructed to produce sentences containing idioms naturally, with no explicit instruction regarding prosody.

Consistent with previous studies, we predicted that speakers would prosodically distinguish between the figurative and literal productions of idioms. Based on Belanger et al. (2009), we did not anticipate any differences between figurative and literal productions for either duration or *F0* of the verb. Rather, we predicted that prosodic differences for idioms intended figuratively versus literally would occur specifically for the noun. With respect to temporal prosodic cues, we predicted that

nouns for highly lexicalized idioms (i.e., high familiarity, low decomposability) intended figuratively would have shorter durations than nouns for the same idioms intended literally (e.g., Van Lancker et al., 1981). However, less lexicalized idioms (i.e., low familiarity, high decomposability) would not. F_0 variability across verbs and nouns could differ to the extent that idioms were highly lexicalized.

2. Method

2.1. Participants

Participants were 14 (6 Male, 8 Female) healthy, native English-speaking individuals from the Montreal community. They were compensated \$10/hour for their participation. The age range of participants was 20-44, while the median age was 22.

2.2. Stimuli

We selected idioms from Libben and Titone (2008), and used normative measures from this study to assess familiarity, global decomposability, verb relatedness, and noun relatedness for 32 literally plausible idioms. On a 1 (unfamiliar) to 5 (familiar) Likert scale, familiarity ratings ranged from 2.57 to 4.73 (mean = 3.81). On a 0 (nondecomposable) to 1 (decomposable) proportionate scale, global decomposability proportions ranged from 0.15 to 0.98 (mean = 0.52). On a 1 (unrelated) to 5 (related) Likert scale, verb relatedness ratings ranged from 0.80 to 4.63 (mean = 2.81), while noun relatedness ratings ranged from 0.54 to 4.83 (mean = 2.52).

For each idiom, two contexts were created. Thus, for the idiom, *spill the beans*, one context led to a figurative production (“Jill could not keep a secret. She *spilled the beans* a minute ago.”), while the other context led to a literal production (“Margaret was making supper. She *spilled the beans* a minute ago.”). Literal control phrases, whose constituent verbs and nouns matched their idiom counterparts in length and frequency, were also created and embedded in a literal context. For example, the literal control for *spill the beans* was *water the plants* (“Andrea worried about the drought. She *watered the plants* a minute ago.”).

2.3. Procedure

Each participant saw any given idiom in either its figurative, literal, or non-idiomatic control form presented on a computer screen. However, each participant did not see an idiom in more than one given form. Thus, one third of the stimuli produced by participants were figurative, one-third literal, and one-third non-idiomatic controls. Stimuli were presented in random order, and participants were instructed to read them silently. Participants were then instructed to read the sentence out loud as naturally as possible. Their speech was recorded by a Logitech USB headset microphone.

2.4. Acoustic analyses

Each sound file was manually checked, and trials were discarded if participants said something other than what was written on the computer screen, if the recording

failed, or if the speech was dysfluent. This resulted in 97 out of 512 trials being discarded. Next, target idiomatic expressions were acoustically extracted from the rest of the sentence. We then conducted a forced-alignment using HTK software, which gave us segment-by-segment and word-by-word boundaries. Finally, we ran a script in PRAAT to extract duration and F_0 information for each word the idioms.

3. Results

We constructed separate linear mixed effects models with noun/verb ratios for duration and F_0 as dependent variables (DVs). Following Belanger et al., we used ratios as DVs to control for within- subject variability and gender differences for duration and F_0 , respectively. Each DV was log-transformed to adjust for distributional skew. For each DV, three different models were generated, which included a different measure of decomposability (global decomposability, verb relatedness, or noun relatedness). Thus, six models were generated in total. Each model included subjects and items as random effects. We included all possible 2-way interactions between condition (Figurative vs. Literal), familiarity, and noun relatedness, as well as their lower order fixed effects. Finally, for models with F_0 as a DV, we included Gender (Male vs. Female) as a fixed effect, and for models with duration as a DV, we included a noun/verb ratio of phoneme length as a fixed effect to account for additional variance. The models that included noun relatedness as a measure of decomposability yielded the most interesting results, so we focus on those here.

3.1. Duration

Across all idioms, nouns had longer durations than verbs, thus nearly all noun/verb duration ratios were greater than one. There was a significant interaction between condition and familiarity ($t = 2.29$, $p = .02$). As shown in Figure 1a, at low levels of familiarity there was a greater difference between nouns and verbs for figuratively produced utterances compared to literally produced utterances. However, at high levels of familiarity the direction of the effect reversed, and a greater difference between nouns and verbs now existed for utterances that were intended literally versus figuratively. Thus speakers produced nouns of low familiar idioms with relatively longer durations when intended figuratively compared to when intended literally.

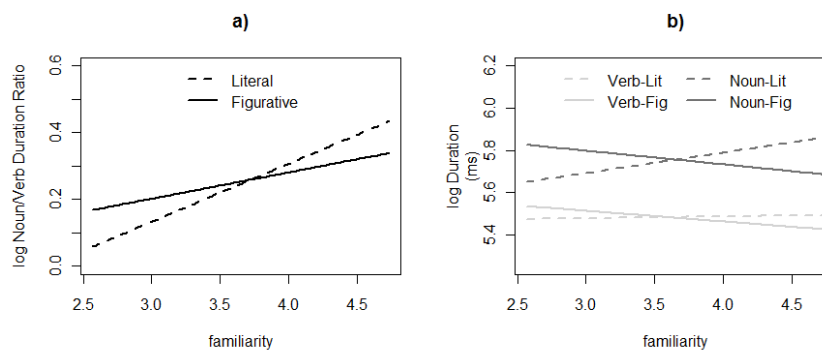


Figure 1. Condition x Familiarity interaction for log-transformed duration ratios (**1a**) and log duration for both nouns and verbs (**1b**).

We further deconstructed the interaction depicted in Figure 1a by constructing separate models with duration as the DV (instead of noun/verb duration ratios) for verbs and nouns separately. The results from this analysis are depicted in Figure 1b. There were no differences between figurative and literal productions of verbs at any level of familiarity. However, there was a significant interaction between condition and familiarity for nouns ($t = 4.63, p < .001$). Consistent with the ratio results, at low levels of familiarity, nouns in figuratively produced idioms were *longer* in duration than their literally produced counterparts. In contrast, at high levels of familiarity, nouns in figuratively produced idioms were *shorter* in duration than literally produced idioms.

3.2. F0 Variability

Across all idioms, F_0 for verbs was higher than F_0 for nouns, thus nearly all noun/verb F_0 ratios were less than one. There was a significant interaction between condition and familiarity ($t = 2.01, p = .04$). Note that because F_0 ratios were less than one, and they were log-transformed, greater differences between nouns and verbs in F_0 would be represented as increasingly negative numbers. Thus, as ratios reduced in negativity, and approached zero, differences between nouns and verbs were reduced. With that in mind, Figure 2a illustrates that at low levels of familiarity, there was a reduced spectral difference between nouns and verbs for figuratively produced utterances compared to literally produced utterances. However, at high levels of familiarity there was an opposite effect, in that spectral differences between nouns and verbs were now higher for figurative productions, as compared to literal productions.

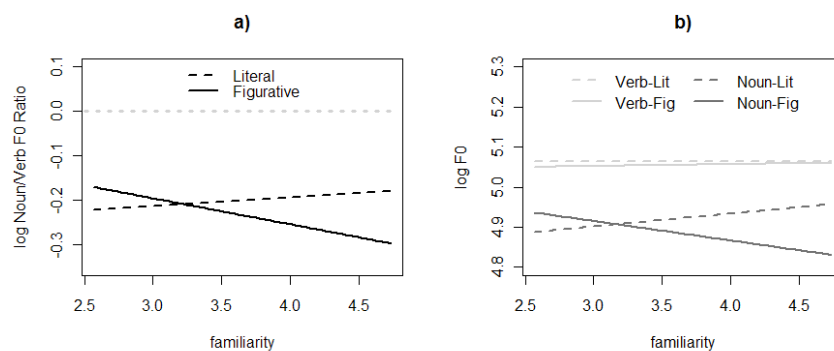


Figure 2. Condition x Familiarity interaction for log-transformed F_0 ratios (**2a**) and log F_0 for both nouns and verbs (**2b**).

Deconstructing the interaction depicted in Figure 2a for both verb and noun F_0 , we can see in Figure 2b that there was a significant interaction between condition and familiarity for nouns ($t = 2.07, p = .04$) but not for verbs. At low levels of familiarity, nouns from idioms intended figuratively were spoken with a *higher* F_0 than nouns from idioms intended literally. In contrast, at high levels of familiarity, nouns from idioms intended figuratively were spoken with a significantly *lower* F_0 than nouns from idioms intended literally.

Familiarity was not the only predictor to influence F_0 variability. There was also a significant interaction between condition and noun relatedness ($t = -2.68, p = .008$). Figure 3a illustrates that at low levels of noun relatedness, spectral differences between

nouns and verbs from idioms intended figuratively were significantly *increased* compared to those from idioms intended literally. However, at high levels of noun relatedness spectral differences between nouns and verbs from idioms intended figuratively were significantly *decreased* compared to idioms intended literally.

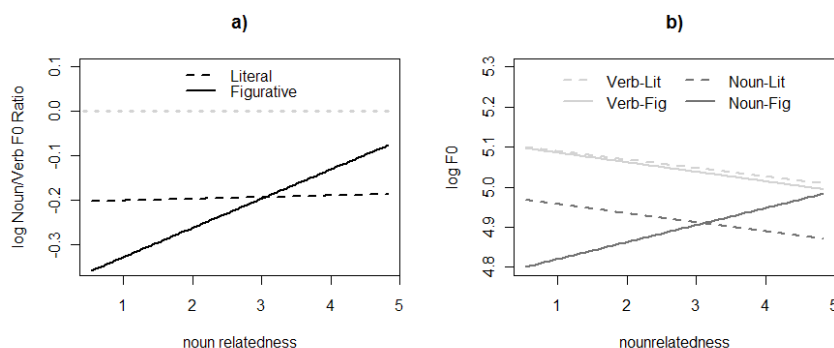


Figure 3. Condition x Noun Relatedness interaction for log-transformed F_0 ratios (**3a**) and log F_0 for both nouns and verbs (**3b**).

Deconstructing the interaction depicted in Figure 3a for both verb and noun F_0 , Figure 3b shows that there was a significant interaction between condition and familiarity for nouns ($t = -2.90$, $p = .004$), but not verbs. At low levels of noun relatedness, nouns from idioms intended figuratively were spoken with a *lower* F_0 than nouns from idioms intended literally. In contrast, at high levels of noun relatedness, nouns from idioms intended figuratively were spoken with a significantly *higher* F_0 than nouns from idioms intended literally.

4. Discussion

We found that the prosodic realization of idioms spoken with a figurative intent differed from that of idioms spoken with a literal intent as a function of both idiom familiarity and semantic decomposability. As expected, these differences only manifested themselves on the nouns of idioms, which could reflect the importance of phrase-final words in marking idioms prosodically (Belanger et al., 2009). Of note, we observed these effects in a relatively natural language production task in which participants produced each idiom in only one context. Thus, in contrast with prior work, the task used here did not explicitly call attention to prosodic differences across interpretations. Previous studies failed to find prosodic differences under similar implicit conditions, rather differences were only found when both contexts were presented simultaneously.

In Figure 1b, we showed that when highly familiar idioms were produced with figurative intent, durations for nouns were significantly reduced compared to when the same idioms were produced with literal intent. Recall, that as familiarity for idioms increases, the likelihood that they will be lexically represented and produced as a single phrasal unit also increases. Thus, our findings are consistent with Van Lancker et al. (1981), and solidify the notion that when the figurative forms of idioms are represented holistically, the less time it takes to produce them.

We also observed that speakers produced greater differences in F_0 between nouns and verbs for idioms that were most likely to be lexicalized (i.e., high familiarity and low decomposability), compared to idioms that were less likely to be lexicalized (i.e., low familiarity and high decomposability). For example, in Figure 2a, when idioms were highly familiar, the difference in F_0 between nouns and verbs was greater when idioms were produced with a figurative intent, as opposed to a literal intent. However, when idioms were less familiar, the extent to which F_0 varied for idioms produced with figurative versus literal intent was also reduced (i.e., smaller differences in F_0 between nouns and verbs). A similar pattern was also observed for the degree of noun relatedness. However, because increased noun relatedness is potentially associated with reduced idiom lexicalization, prosodic changes in F_0 were highest at *low* levels of noun relatedness (see Figure 3a).

Although preliminary, the results of the present study suggest that prosodic cues are differentially utilized during speech production for idioms that are likely to be lexicalized versus idioms that are less likely to be lexicalized. Unresolved in this report is how other potential prosodic cues varied for these idioms. For example, it is possible that differences in the frequency or duration of pauses between the individual words of idioms intended figuratively versus literally may have also existed. As well, it would be important to compare the acoustic patterns of idioms with those of the non-idiomatic control sentences. While we continue to explore these and other possibilities in the data, taken together, these preliminary results presented here suggest that naturally produced prosody can signal how multiword expressions are represented in the mental lexicon.

5. References

- Belanger, N., Baum, S. R., & Titone, D. A. (2009). Use of prosodic cues in the production of idiomatic and literal sentences by individuals with right- and left-hemisphere damage. *Brain & Language*, *110*, 38-42.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A review. *Language & Speech*, *40*, 141-201.
- Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of Language*, *6*, 22-42.
- Gibbs, R.W., Nayak, N. P., & Cutting, J. C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory & Language*, *28*, 576-593.
- Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*, 1103-1121.
- Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech & Hearing Research*, *24*, 330-335.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language & Cognitive Processes*, *25*, 905-945.