Sara D. Beck & Andrea Weber

2022

Unlocking the Key to L2 Idiomatic Processing: Non-native Listeners' Idiomatic Processing is not Immediately Affected by the Idiomatic Key

in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),

Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data

pp. 97–112

Tübingen: University of Tübingen

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

In R. Hörnig, S. von Wietersheim, A. Konietzko & S. Featherston (eds.) (2022), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*. Tübingen: University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

Unlocking the Key to L2 Idiomatic Processing: Non-native Listeners' Idiomatic Processing is not Immediately Affected by the Idiomatic Key

Sara D. Beck & Andrea Weber
University of Tübingen
sara.beck@uni-tuebingen.de

1 Introduction

One of the defining factors of idioms, setting them apart from other types of figurative language, is their formulaic nature (see e.g., Wray, 2002). While idioms do allow for a certain amount of flexibility in syntactic structure (e.g., Kyriacou et al., 2019; Nunberg et al., 1994), the constituents are generally fixed, and there is a preferred canonical form (e.g., McGlone et al., 1994). The consequence of this fixed nature is that idioms which have been frequently encountered can also be highly predictable, reducing processing costs. Research done on native (L1) speakers has confirmed this advantage by comparing idioms to similar novel language and finding evidence of eased processing in a number of ways including, but not limited to, faster reaction times in semantic judgement tasks (Tabossi et al., 2009), faster reading times (Conklin & Schmitt, 2008), and fewer and shorter fixations on idioms during reading (Underwood et al., 2004). In addition to these faster recognition processes, native speakers are able to access the idiomatic meaning even before being exposed to the final constituent of the idiom in cases where an idiom has been shown to be highly predictable (e.g., Cacciari & Tabossi, 1988; Rommers et al., 2013; Tabossi & Zardon, 1993). What is unclear, however, is the extent to which such advantages also apply to highly proficient non-native speakers. Though it has been established that non-native speakers, like native speakers, have online access to both figurative and literal-constituent meaning in idioms at idiom offset (e.g., Beck & Weber, 2016a), research has yet to clearly establish the boundaries of this access to meaning and whether some of the same processing advantages apply. In particular, there is some evidence that idiomatic processing is slower overall for L2 speakers compared to L1 speakers (e.g., Beck & Weber, 2016a), and that idioms do not provide L2 speakers with the same processing advantages over novel phrases that have been observed for L1 speakers (e.g., Siyanova-Chanturia et al., 2011). If this is the case, idiom predictability may not play a role in L2 idiom processing. In the studies presented here, we look at the aspect of predictability, and whether or not non-native speakers are able to access the idiomatic meaning of highly-predictable idioms before the offset of the final idiomatic constituent.

Prior to the study of predictability in idioms, idiomatic processing was assumed to be a different and separate process from literal language processing. Following this assumption, processing idiomatic phrases was generally viewed in a step-wise manner. In the traditions of standard pragmatic views of language processing (e.g., Searle, 1979), a literal-first approach requires that literal meaning is processed first, and only when this meaning is rejected or in the presence of exceptional circumstances (i.e., biasing context) is figurative language directly retrieved (e.g., Bobrow & Bell, 1973). Conversely, a figurative-first approach suggests that the figurative meanings of idioms must be obligatorily retrieved immediately before literal

meanings are activated, again based on a meaning and integration mismatch (e.g., Gibbs, 1980; Schweigert & Moates, 1988). A third possibility following the two-step approach is that figurative and literal processing occur simultaneously. Because figurative retrieval has been found to be faster than literal computation, this hypothesis predicts faster access to figurative meanings (e.g., Swinney & Cutler, 1979). However, it is not the case that any of these step-wise procedures explain the variable timing of access to figurative meaning as it occurs in every situation, so no single hypothesis following these strict approaches can account for the variation found in the timing of access to meaning between idioms (e.g., Gibbs et al., 1989; Libben & Titone, 2008; Titone & Connine, 1994; Titone & Libben, 2014).

One source of such variation between idioms is predictability. The seminal study on predictability of idioms was undertaken by Cacciari and Tabossi (1988) in which the notion of the idiomatic key was derived. The idiomatic key is the point at which a sufficient portion of the phrase is processed, and key content words have been encountered that clearly point to a figurative interpretation. This position of the key is usually determined in norming tests. Based on a series of cross-modal priming studies in Italian, the authors found that an idiom's figurative and literal constituent meanings were available at different times depending on the point at which an idiom became recognizable as such (e.g., *to be in seventh heaven* is recognized after "seventh", and *go to the devil* after "devil"). In a lexical decision task, the priming of words related to the figurative meaning of an idiom and a literal constituent were compared with unrelated control words (e.g., figurative: HAPPY, literal: SAINT, control: UMBRELLA). If an idiom was recognizable by most native speakers before the final constituent (based on cloze-probability testing for idiomatic completions), only the figurative meaning and not the literal meaning was available at the offset of the idiom. Conversely, if the idiom was not predictably completed as an idiom before the final constituent, then the literal constituent meaning was activated while figuratively related words were not. An additional study confirmed that both meanings were available in the latter case when tested 300ms after the offset of the idiom. These results formed the basis for the authors' Configuration Hypothesis. Under this theory of processing, idioms are made up of configurations of words, and processing is necessarily literal until the occurrence of the recognition point within the idiom is reached, known as the idiomatic key. Once the idiomatic key has been encountered, or enough information has been gathered to recognize the string as idiomatic, literal processing ceases, and access to figurative meaning occurs. The concept of the idiomatic key crucially implies that processing of idioms occurs in a probabilistic, or predictability-based manner. Though, since meaning activation was always tested following the idiomatic key, lexical models of processing could not be entirely excluded based on their findings.

Tabossi and Zardon (1993) followed up on this hypothesis by testing in Italian whether the idiomatic meaning was available prior to the idiomatic key or not. In two cross-modal priming studies, the authors first tested predictable idioms with two content words following the verb, and the first content word was the idiomatic key (e.g., *Finally, Silvio had succeeded in setting his mind at rest*). The target words were presented at three points: following the verb (e.g., *setting*), the first content word (e.g., *mind*), and the second content word (e.g., *rest*). As predicted by the Configuration Hypothesis, figuratively related words showed priming effects following the first and second content word, with no differences between these two placements. In a second study, unpredictable idioms were chosen and tested (e.g., *In the end, the man hit the nail on the head*) following the same experimental design. For these idioms, the meaning was only available after the second content word, confirming the importance of the idiomatic key in processing.

The findings of Cacciari and Tabossi (1988) and Tabossi and Zardon (1993) for Italian suggest that step-wise models of idiomatic processing are insufficient in describing L1 idiom processing. Rather, hybrid models that better describe the flexibility in meaning access native speakers display are necessary. The Configuration Hypothesis was later extended by Titone and Connine (1994) to suggest that literal meaning need not be terminated upon activation. Rather,

the idiom representation is linked with the composite literal meaning, therefore still allowing for activation of the literal constituents and even phrases as well as competition in meanings under certain conditions. This adjustment is in tune with more recent research on idioms as well as other types of formulaic language in which access to the whole does not preclude meaning access to the literal constituents' meaning (e.g., Beck & Weber, 2016a) and structure (e.g., Peterson et al., 2001).

The flexibility assumed of native speaker idiomatic processing, however, is not necessarily indicative of non-native idiomatic processing. Instead, a fundamental difference in L1 and L2 processing is often assumed, namely, that non-native speakers will always first process literal meaning by default. Cieślicka (2006) conducted a cross-modal priming study in English on non-native Polish listeners in which figuratively or a literally related targets were presented either at the offset of the phrase or at the offset of the penultimate word in the phrase (e.g., the targets FORGIVE and AXE were presented either after *hatchet* or after *the* in '*George wanted to bury the hatchet soon after Susan left*'). Overall, the results indicated more priming for literally related targets and only marginal results for figurative targets. There were also no significant differences in figurative priming between the two positions. These results were interpreted to suggest that literal processing indeed has priority over figurative processing and suggested that the salience of literal language (see e.g., Giora, 1997) may be at the root of this difference. Note, however, that the figurative targets were abstract words and concepts in this study, while all literal targets were concrete objects (e.g., for the idiom *bury the hatchet*, literal target: AXE, figurative target: FORGIVE). It is conceivable that this lexical difference between items could have influenced the results.

Beck and Weber (2016a) followed up on these claims with another cross-modal priming study in English. In this study, both a native and a highly proficient non-native group of speakers (German L1) participated in the lexical decision task (again with figuratively and literally related targets), in which idioms were varied only systematically in their translatability from English to German (e.g., *lend one's ear* is translatable word-for-word into the same German idiom, while the translation of *kick the bucket* does not result in an idiom in German). Targets were presented at the offset of sentences (e.g., directly following *bucket*), which coincided with the last idiom constituent (e.g., '*His uncle kicked the bucket'*; figurative target: DIE, literal target: PAIL, unrelated controls: ZOO, BOAT). The authors found significant priming effects in both types of targets compared to their controls for both native and non-native speakers. Based on the similar behavior of L1 and L2 participants, the authors suggest that highly proficient non-native processing is indeed similar to native processing. This result was also confirmed in a later but similarly designed cross-modal priming study conducted by van Ginkel and Dijkstra with Dutch learners of English (2019).

Additional studies have compared native and non-native processing using idioms as well as other formulaic language compared to novel phrases and have found more mixed results. Using a self-paced reading paradigm, Conklin and Schmitt (2008) found reading advantages for idioms (e.g., *hit the nail on the head*) compared to novel phrases (e.g., *hit his head on the nail*) in both English L1 and L2 readers with various L1 backgrounds, suggesting that both groups are sensitive to the processing advantages of formulaic phrases. On the other hand, Underwood et al. (2004) as well as Siyanova-Chanturia et al. (2011) investigated such phrasal advantages using eye-tracking in reading and found that while natives fixated less often and for shorter periods of time on final elements in idiomatic or formulaic sequences compared to novel phrases, non-natives did not reliably show the same effects. Rather, Siyanova-Chanturia et al. (2011) concluded that idioms were processed in the same manner as novel phrases for L2 readers in contrast to L1 readers. These studies suggest that while access to meaning may be available online, L2 speakers may not be reliably sensitive to the same predictive-based advantages seen in L1 speakers. Though, more direct measures are still lacking.

In the current study, we asked if an idiom's predictability can also impact figurative priming for non-native listeners as it has been shown to do so in native listeners (e.g., Cacciari

& Tabossi, 1988; Tabossi & Zardon, 1993). English was the target language, and the highly-proficient L2 participants were native speakers of German. In particular, when considering the Configuration Hypothesis and the idea of the idiomatic key, is figurative meaning available as early as the idiomatic key? If we expect L2 processing to be comparable to L1 processing, the predictability of an idiom and therefore the placement of the idiomatic key should impact the availability of figurative meaning in online processing. However, if predictability does not play the same role for L2 speakers as it does for L1 speakers, we do not expect idiom's predictability to reliably impact meaning access. We conducted two cross-modal priming studies in order to investigate this question.

2 Experiment 1

2.1 Method

In a cross-modal priming study, German L2 listeners heard English sentences with short, non-biasing contexts containing idioms. In a lexical decision task, response times to target words related to the figurative meaning or a literal constituent meaning in the idiom were taken and were compared to matched unrelated control words in order to measure lexical priming. The idioms varied in predictability, that is the placement of the idiomatic key (early-key idioms and late-key idioms), and target words were presented at the offset of the idiomatic key (i.e., during the sentence for early-key idioms, and at the end of the sentence for late-key idioms). If L2 processing is sensitive to predictability in the same manner as L1 processing, we expected figurative priming for both types of idioms following the idiomatic key. If L2 priming is not sensitive to predictability, we expected to find priming only in the presentation of the idiomatic key at the end of the sentence (following late-key idioms), and not earlier. We expected literal priming to occur regardless of idiom type, based on the strength of activation in previous L2 studies (e.g., Beck & Weber, 2016a, 2016c; Cieślicka, 2006).

2.1.1 Participants

A total of 40 non-native speakers of English (German L1) took part in the study, and 38 were included in the analysis. One participant was excluded from the analysis for not meeting the participation requirements (L1 only German), and one was excluded based on data loss. All participants (30 female, average age of 23.71, *SD* = 3.32) received financial compensation for participation and were recruited at the University of Tübingen, primarily from the department of English. Participants were all advanced users of English (many of them students of English) and reported an average of 10 years of formal English instruction and rated their English skills at 5.8 on a 7-point scale (1 corresponds to very poor and 7 native-like). Participants from this pool have on previous occasions shown high Lex-TALE scores (Lemhöfer & Broersma, 2012). A total of 5 participants were left-handed, and none reported any visual or hearing impairments.

2.1.2 Materials

The experiment consisted of 36 target and 104 filler trials. Target trials consisted of short, neutral sentences ending in idioms (e.g., I think she *got up on the wrong side of the bed*.). Sentence contexts (e.g., *I think she*…) were chosen by the authors, did not contain any obvious cues to the interpretation of the upcoming idiomatic phrase, and thus neither clearly biased towards a figurative or literal interpretation. Idioms were taken from the German Database of Idiom Norms (DIN; Beck & Weber, 2016b). Twenty-nine of the 36 idioms had also been used in Beck and Weber (2016a). Idioms were selected that had been rated highly for familiarity of encounter (mean = 5.2, *SD* = 0.81, on a scale from 1 to 7) and familiarity with the meaning (mean = 5.8, *SD* = 0.56, on a scale from 1 to 7) in the DIN database, and mid-range on the scale

for literality (i.e., idioms potential for literal interpretation) to not bias towards a literal or figurative interpretation of the idiomatic phrase itself.

Idioms varied systematically in their predictability, that is in the position of the idiomatic key. Idioms were divided into two categories with either an early idiomatic key or a late idiomatic key based on the L1 cloze-probability testing done in the DIN database (Beck & Weber, 2016b). Highly predictable idioms that were completed by at least 70 % of participants correctly at any word before the final constituent word were classified as early-key idioms, and idioms with less than 70 % correct completions before the final constituents were classified as late-key idioms. Early-key idioms had an average correct completion rate of 85 % (*SD* = 10) at the point identified as the idiomatic key, whereas late-key idioms had an average of only 23 % (*SD* = 20), and groups significantly differed from one another (t = -11.56, df = 25.62, $p < .001$). For an example of each idiom type, see Table 1. Notably, idioms classified as early-key idioms align with highly predictable idioms, and those classified as late-key idioms as unpredictable.

Table 1. Example items. The Symbol | is placed after the idiomatic key. (F) is the figurative target, and (L) is the literal target

Key	Idiom	Correct Completions	Targets	Controls
Early	get up on the wrong | side of the bed	86 %	MISTAKE (L) MOODY (F)	PLANET BOTTLE
Late	Pop the question |	18 %	ANSWER (L) PROPOSE (F)	ORANGE PROBLEM

Four target words associated with each idiom were either literally related to the idiomatic key (or last content word before the idiomatic key), figuratively related to the idiom, or a respective control. For the late-key idioms taken from Beck and Weber (2016a), the same targets were used. For all other idioms and targets, literal targets were taken from established associative norms (Nelson et al., 1998), and figurative targets were developed by two native speakers as there is no similar database for figuratively related meanings known to these authors. The literal targets were either the top associate or taken from the top three associates in order to balance the lists for frequency and orthographic complexity. The figurative targets taken from Beck and Weber (2016a) were, however, pre-tested to check that the target was associated with the figurative meaning. For the early-key idiom *get up on the wrong side of the bed*, the literal target MISTAKE was chosen as it corresponds to the key "wrong" and the control target PLANET was matched for orthographic complexity (syllable structure and length) and frequency according to the CELEX database (Max Planck Institute for Psycholinguistics, 2001). Likewise, the figurative target MOODY relates to the overall idiomatic meaning, to have a bad start to the day or wake up in a bad mood, and the matched control was BOTTLE. For a late-key idiom such as *pop the question*, the literal target ANSWER corresponded to the final constituent "question" and the figurative target PROPOSE still corresponded to the overall meaning of the idiom, to propose marriage, both with matching controls ORANGE and PROBLEM, respectively.

In order to make as few changes as possible to the data from Beck and Weber (2016a), only targets for new idioms or literal targets that were related to words earlier than the final constituent were added or changed, and there were slight differences in list-frequencies based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009). Individual t-tests confirmed that figuratively related targets were overall more frequent than their controls (means = 164.472, 131.611 respectively, t = 2.716, $p < 0.01$) and the literal control targets (mean = 130.722, t = 2.879, $p < 0.01$). Since lexical frequency might therefore disadvantage figurative targets, frequency counts were included in the final models in the statistical analysis.

Half (18) of the target trials consisted of early-key idioms, and the other half of late-key idioms. The 104 filler trials were the same ones used in Beck and Weber (2016a), consisting of 84 non-idiomatic trails and 20 idiomatic trials. These idiomatic filler trials were followed by non-words, helping ensure a balance of word and non-word targets for all trial types. Overall, half of the targets were words, and half non-words, and of the word-targets some filler targets also had meanings related to the literal meanings of a constituent word in the filler sentence. The experiment was divided into four counter-balanced lists so that each participant heard each idiom only once. Only target items varied between lists.

The sentences used in all trials were recorded by a male speaker of American English in the recording studio of the LingTüLab (in the English Department of The University of Tübingen). The experiment began with four practice trials followed by the 140 pseudo-randomized experimental trials. Participants were allowed a self-directed concentration break after trial 70.

2.1.3 Procedure

Participants were tested individually in single rooms in the LingTüLab (University of Tübingen). Individuals were instructed that they would be hearing English sentences and would be presented with either an English word or non-word during or at the end of the auditory sentence. They had to decide if a word or non-word was presented. Participants were told to listen carefully and respond as quickly, but accurately as possible by pushing the red or green button on a Cedrus Button-box. The green "YES" button was always pushed with the dominant hand, while the red "NO" button was pushed by the non-dominant hand. Participants were advised that careful listening was important, even while making lexical decisions, as they would be asked questions about what was heard following the lexical decision task.

Once it was clear that participants understood the task, the experiment began. Auditory stimuli were presented at a clear volume on closed headphones. Target words appeared on a computer screen with a black background and white text in size 20 font. Audio files played to completion for each trial, but targets were presented at different times during the trials. The targets were presented immediately at the offset of the idiomatic key and varied for filler trials. For late-key idioms, the presentation of the target coincided with the final word of the idiom; for early-key idioms, the presentation of the target appeared immediately at the offset of the word identified as the idiomatic key. If a button was pressed still during the presentation of the auditory sentence, then the next trial began 2000ms after the end of the auditory stimulus. If the button press occurred after the end of the auditory stimulus, then the next trial began 2000ms later. If no button was pressed, the next stimulus began after 3000ms (ensuring at least 2000ms of silence).

Following the lexical decision task, participants completed a short comprehension task and a language background questionnaire. Both tasks were completed on forms using Adobe Professional. For the comprehension task, participants saw a list of 60 sentences, and had to decide whether they had heard the sentences listed by checking "YES" or "NO." Half of the sentences on the test had been heard in the experiment, and half were new (see Beck & Weber, 2016a; Cieślicka, 2006). The language background questionnaire collected basic information from the participants about their language skills and confirmed their status as L1 German and L2 English speakers.

2.1.4 Results

The analysis of reaction times included only trials with correct responses to targets. On average, participants responded correctly 98 % of all target items. Based on these performance levels, high proficiency in English can be assumed. Additionally, outliers were identified and removed from the analysis if they were either outside of 2 standard deviations from the mean reaction

time per participant considering the main experimental factors of *figurativeness*, *relatedness*, and *key-timing* or they exceeded 1500ms (a total of 2.6 % of the data).

Table 1. Mean reaction times (ms)

	Early Key				**Late Key**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**
Related	660	-1,641	716	-1,546	592	-1,834	584	-1,861
Unrelated	708	-1,568	692	-1,571	595	-1,784	576	-1,855

R (R Core Team, 2013) and lme4 (Bates, Maechler et al., 2015) were used to perform a linear mixed effects analysis of the relationships between the factors of *figurativeness*, *relatedness*, and *key-timing* on the inverse reaction times. The inverse reaction times were used, as this transformation created the most normal distribution based on the Box-Cox transformation test and a visual confirmation of the data (Venables & Ripley, 2002). The mean reaction times and inverse reaction times, measured from the offset of the idiomatic key, for each condition are displayed in Table 2. The mean reaction times are shown in Figure 1, with bars representing the standard error of the mean.

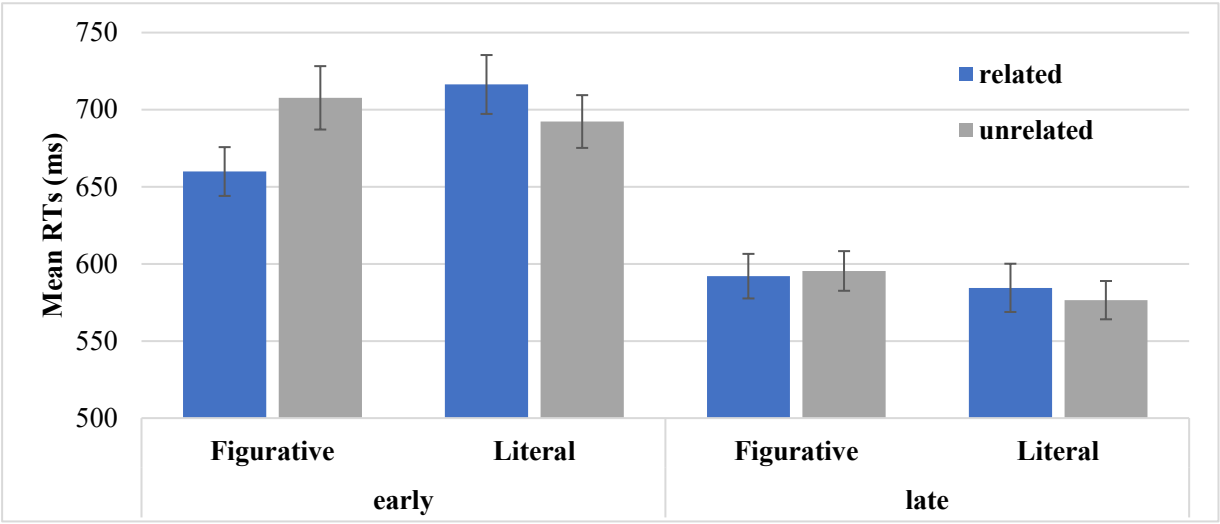

Figure 1. Experiment 1 mean RTs (ms) with standard error bars. Grouped by key-position (early vs. late from left to right), figurativeness (figurative vs. literal from left to right), and relatedness (related vs. unrelated)

LMER models with inverse RTs as the dependent variable and fixed factors, centered around 0, were coded and included as follows: *figurativeness* (figurative: 0.5 and literal: -0.5), *relatedness* (related: 0.5 and unrelated: -0.5), and *key-timing* (late: 0.5 and early: -0.5). Additional numeric fixed factors included after centering were *order* (order of experimental trial), *frequency* (target frequency per million), *length* (target length in letters), *L2 meaningfulness* (ratings on the meaningfulness of the idiom, on a scale from 1-7), *L2 familiarity* (ratings on the subjective frequency of encounter of the idiom, on a scale from 1-7), *L2 literality* (ratings on the literal interpretability of the idiom, on a scale from 1-7), and *proficiency* (self-scored proficiency ratings on reading, writing, speaking, and listening, on a scale from 1 [not at all proficient] to 5 [native-like proficiency] averaged across all four ratings). The idiomatic ratings included were taken from the DIN database as they have been shown to affect access to meaning in cross-modal priming experiments (e.g., Beck & Weber, 2019; Titone & Connine, 1994; Titone & Libben, 2014). Participants and items were included in the models as random factors with random slopes. A maximally justified random effects structure was determined by step-wise selection and model comparison (see e.g., Bates, Kliegl et al., 2015) using

RePsychLing (Baayen et al., 2015) that included random slopes for *figurativeness*, *relatedness*, and *key-timing*. Backward step-wise selection was used to eliminate the additional factors not relevant to the theoretical questions of this study where model comparison showed that they did not contribute to a better fit of the model. The values of the final model are shown in Table 3.

Table 2. Experiment 1 final LMER model

Fixed Effects and Controls	Effect Size	SE	*t*-Value	Pr(<|*t*|)
Intercept	-1.690	0.0495	-34.165	<2e-16***
Figurativeness (coded)	0.006	0.0271	0.218	0.828
Relatedness (coded)	-0.007	0.0276	-0.237	0.813
Key-Timing (coded)	-0.265	0.0355	-7.465	<1.3e-10***
Frequency	-0.070	0.0144	-4.848	3.63e-6***
L2 Meaningfulness	-0.097	0.0352	-2.760	0.007**
L2 Familiarity	0.082	0.0333	2.467	0.015*
L2 Literality	0.028	0.0159	1.732	0.086.
Proficiency	-0.115	0.0488	-2.364	0.024*
Figurativeness* Relatedness (coded)	-0.066	0.0532	-1.235	0.219
Figurativeness* Key-timing (coded)	0.056	0.0536	1.054	0.294
Relatedness* Key-timing (coded)	0.013	0.0534	0.251	0.802
Figurativeness*Relatedness* Key-timing (coded)	0.072	0.1068	0.672	0.503

Note: *.p < .10 *p < .05 **p < .01 ***p < .001*

Of the three factors under investigation in this study, only *key-timing* shows a significant main effect (β = -0.26, t = -7.46, p < .001); all targets appearing in the late-key position (i.e., sentence offset) were responded to more quickly than those in the early-key position. Additional fixed factors also affected reaction times. *Frequency* (β = -0.07, t = 4.84, p < .001) affected reaction times in that more frequent targets were reacted to more quickly. The idiomatic factors *L2 familiarity* (β = 0.08, t = 2.46, p < .05), *L2 meaningfulness* (β = -0.09, t = -2.76, p < .01), and, marginally, *L2 literality* (β = 0.02, t = 1.73, p = .08) all improved the model fit, but did not impact the interpretation of the main results. Familiarity with the meaning of the idiom decreased reaction times, while subjective frequency of encounter increased reaction times. While the latter result may seem surprising, the ratings were collected from another group of non-native speakers and reflect only how often L2 speakers thought they had come across the idiom. The literal interpretability of an idiom also slowed down reaction times where an idiom was more literally interpretable. Though this result is only marginal, it reflects other results commonly found in idiomatic processing literature (e.g., Titone & Connine, 1994). Finally, a significant effect of *proficiency* (β = 0.08, t = 0.03, p < .05) confirms that an increased self-rated English proficiency decreased reaction times.

2.2 Discussion

The analysis shows that facilitatory priming was neither observed for literally or figuratively related targets nor was there an interaction of these factors with the timing of the idiomatic key. While it was unclear whether or not priming would be found for figuratively related targets in the early-key condition, it is surprising that no priming was found in the late-key condition. Both Beck and Weber (2016a) as well as van Ginkel and Dijkstra (2019) found figurative and literal priming following the offset of the final idiomatic constituent, and while Cieślicka (2006)

did not find figurative priming in this position, literal priming was also present in her study. Although different items were used from the latter two studies, the current study re-used many of the experimental items from Beck and Weber (2016a), so it would be surprising that these idioms and targets already shown to reproduce similar results (e.g., Beck & Weber, 2016c) would be the cause of the null result in the current study.

One possible reason for the lack of figurative priming in the current study is that previous studies did not consider the factor of predictability. While this has been shown to affect L1 processing (e.g., Titone & Connine, 1994), it has largely been ignored in many L2 studies. A closer look at the items from Beck and Weber (2016a) shows that the idioms included in the original study were indeed both predictable and non-predictable idioms. Additionally, neither van Ginkel and Dijkstra (2019) nor Cieślicka (2006) included such norms in their data, and it appears that here, too, there is possible variation. While all of the studies mentioned tested either after the offset or immediately following the penultimate word in the idiom, results indicated the presence of priming that should be comparable to the late-key idioms in the current study. However, one result of such predictability variability is that the high-predictability of some idioms may cause an overall increase in priming for figuratively related targets at the offset (or even penultimate) position, and a main effect of priming for figuratively related targets may be found though there is variation between individual idioms. This possibility is supported by the variability in item-based data in previous studies (e.g., Beck & Weber, 2019). However, this factor should only affect the figurative targets, and we should still have expected to see literal priming in the late-key targets.

Another possible explanation for the lack of facilitatory priming not just in figuratively related targets, but overall, is that the task demands were too great on L2 listeners. Unlike in previous experiments, the current experiment asked the participants to make fast, online decisions both during listening and immediately at the offset to the lexical items presented at targets. Additionally, in the instructions, the participants were asked to listen carefully, as there would be questions about the sentences heard at the end of the experiment. While it is common practice to use such instructions and the goal is generally to ensure that participants actively listened, the result might have been that the demands of listening may have impeded their ability to respond to the targets or overshadowed priming effects (e.g., Cutler & Clifton, 1999). Slow reaction times displayed by L2 listeners when directly compared to L1 listeners in previous and similar cross-modal priming studies (e.g., Beck & Weber, 2016a; van Ginkel & Dijkstra, 2019) suggests that these listening demands can be exaggerated for L2 listeners in a cross-modal priming task using idioms.

In order to test the latter theory, a second experiment was conducted in which task demands were decreased for listeners. First, the possibility of semantic activation for both literally and figuratively related targets was increased by displaying targets 400ms after the offset of the idiomatic key (see e.g., Titone & Libben, 2014) rather than directly at the offset. Second, rather than playing all sentences to completion, listeners only heard the sentences until the offset of the idiomatic key. While this results in sentence fragments for the early-key idioms (and some filler sentences), participants' attention should be equally and wholly focused on the lexical decision task for both idiom types. If, after these changes, facilitatory priming can be found for the late-key idioms as in previous experiments, we assume that the demands of the experimental task influenced the results, and the results for both early- and late-key idioms can better be interpreted in terms of the placement of the idiomatic key.

3 Experiment 2

3.1 Method

Experiment 2 followed the same method as Experiment 1, with the exception of the timing of the target presentation and the cut-off of the auditory stimuli immediately after the offset of the idiomatic key.

3.1.1 Participants

A total of 41 L2 speakers of English (German L1) took part in the study who did not participate in Experiment 1. One participant was excluded from the analysis for not meeting the participation requirements. All participants (27 female, average age of 23.56, SD = 3.43) received financial compensation for participation and were recruited as in Experiment 1. All were advanced users of English and reported an average of 11 years of formal English instruction and rated their English skills at 5.7 on a 7-point scale (1 corresponds to very poor and 7 native-like). One participant was left-handed, and none reported any visual or hearing impairments.

3.1.2 Materials

The same materials were used as in Experiment 1.

3.1.3 Procedure

The same laboratory space, task, and order of procedure was used as in Experiment 1. Crucially, there were two differences from Experiment 1:

1. The presentation of all targets occurred 400ms after the offset of the idiomatic key rather than directly at the offset.
2. The presentation of the auditory stimuli sentence stopped immediately after the offset of the idiomatic key. For late-key idioms, there was no change from Experiment 1 during listening as the key occurred at the end of the sentence. For early-key idioms, the full idiom was not presented in the stimuli.

As in Experiment 1, following a button-press and 2000ms or a maximum of 3000ms with no response, the next trial began.

3.1.4 Results

The analysis followed the same order and procedure as in Experiment 1. Again, analyzed reaction times included only trials with correct responses to targets. On average, participants responded correctly to 99 % of all target items. Additionally, outliers were removed from the analysis if they were either outside of 2 standard deviations from the mean reaction time of the sample considering the main experimental factors of *figurativeness*, *relatedness*, and *key-timing* or they were faster than 300ms or exceeded 1500ms (1.8 % of the data).

Table 4. Mean reaction times (ms) and inverse reaction times

	Early				**Late**			
	Figurative		**Literal**		**Figurative**		**Literal**	
	RT	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**	**RT**	**inverse RT**
Related	641	-1.657	590	-1.782	611	-1.722	593	-1.771
Unrelated	623	-1.696	607	-1.728	635	-1.660	613	-1.716

A linear mixed effects analysis of the relationships between the factors of *figurativeness*, *relatedness*, and *key-timing* was performed on the transformed reaction times, again the inverse reaction times as determined by the Box-Cox transformation test and a visual confirmation of the data (Venables & Ripley, 2002). The mean reaction times and inverse reaction times, measured from the offset of the idiomatic key, for each condition are displayed in Table 4. The mean reaction times are graphed in Figure 2, with bars representing the standard error.

LMER models with inverse RTs as the dependent variable and fixed factors, centered around 0, were coded and included as follows: *figurativeness* (figurative: 0.5 and literal: -0.5), *relatedness* (related: 0.5 and unrelated: -0.5), and *key-timing* (late: 0.5 and early: -0.5). Additional numeric fixed factors included after centering were *order*, *frequency*, *length*, *L2 meaningfulness*, *L2 familiarity*, *L2 literality*, and *proficiency*. Participants and items were included in the models as random factors with random slopes. A maximally justified random effects structure was determined by step-wise selection and model comparison using RePsychLing (Baayen et al., 2015) that included a random slope for *relatedness*. Backward step-wise selection was used to eliminate the additional factors not relevant to the theoretical questions of this study where model comparison showed that they did not contribute to a better fit of the model. The values of the final model are shown in Table 5.

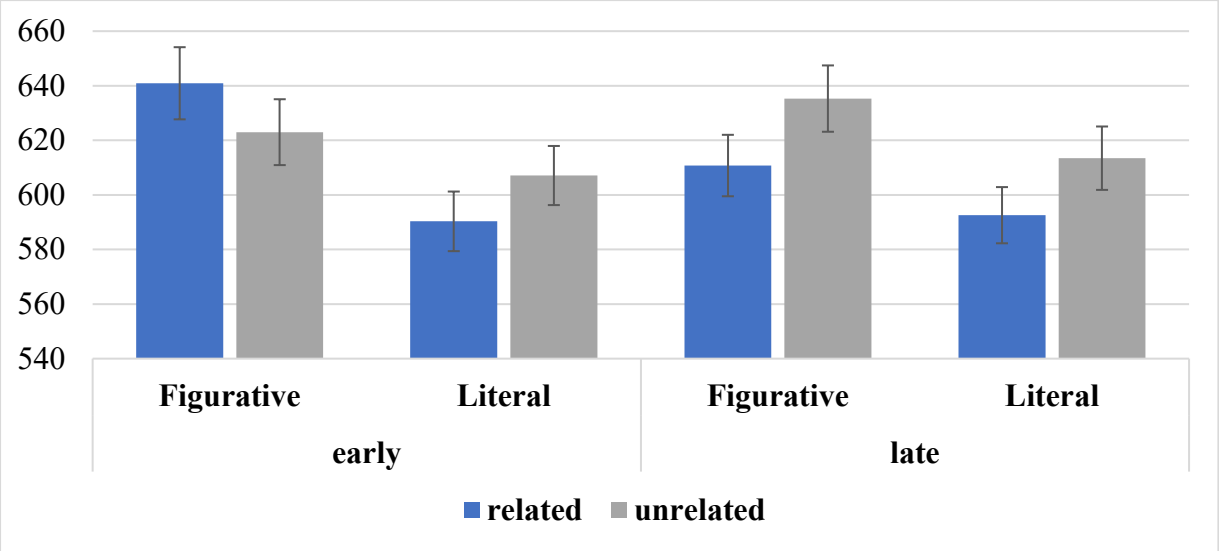

Figure 1. Experiment 2 mean RTs (ms) with standard error bars. Grouped by key-position (early vs. late from left to right), figurativeness (figurative vs. literal from left to right), and relatedness (related vs. unrelated)

Table 1. Experiment 2 final LMER model

Fixed Effects and Controls	**Effect Size**	**SE**	***t*-Value**	**Pr(<|*t*|)**
Intercept	6.394	0.021	305.049	2e-16***
Figurativeness (coded)	0.040	0.016	2.476	0.015*
Relatedness (coded)	-0.009	0.015	-0.611	0.542
Key-Timing (coded)	0.005	0.015	0.342	0.733
Frequency	-003.1	0.008	-3.915	1.45e-04***
Figurativeness* Relatedness (coded)	0.027	0.031	0.883	0.379
Figurativeness*Key-timing (coded)	-0.009	0.031	-0.283	0.777
Relatedness*Key-timing (coded)	-0.022	0.031	-0.698	0.486
Figurativeness*Relatedness* Key-timing (coded)	-0.056	0.062	-0.903	0.368

Note: *.p < .10 *p < .05 **p < .01 ***p < .001*

Even though the results descriptively align with the predicted patterns (see Figure 2), only the factor figurativeness showed a main effect (β = 0.04, t = 2.47, p < .05) suggesting that all figurative targets were responded to more slowly than literal targets (regardless of relatedness). Additionally, *frequency* (β = -0.03, t = -3.91, p < .001) showed the effect of faster reaction times for more frequent targets. No other factors or interactions reached significance.

3.2 Discussion

As in Experiment 1, no facilitatory priming was found for literally or figuratively related targets in either the early-key or late-key positions. Thus, in addressing the concerns of task demands in Experiment 1, Experiment 2 does not deliver a clear answer. Though, the effect of *key-timing* is no longer significant in Experiment 2. While the reaction times graphed in Figure 2 do present the expected patterns as well as those seen in previous experiments (e.g., Beck & Weber, 2016a, 2016c; Cieślicka, 2006; van Ginkel & Dijkstra, 2019), they fail to reach significance. Since facilitatory priming was neither reliably found in early-key nor late-key positions, the latter of which has been consistently found for L2 listeners, the results should not be interpreted as negative evidence of the effect of predictability. Possible issues will be discussed in the general discussion.

4 General discussion and conclusion

In addressing the question of whether predictability can impact L2 listeners' access to meaning, particularly considering the placement of the idiomatic key, both experiments provide inconclusive results. While the mean reaction times in Experiment 2 generally reflect expected patterns of responses, unlike in Experiment 1, these patterns are not reliably corroborated in the statistical analyses. Experiment 2 is therefore also inconclusive in whether task demands may have impacted reaction times in Experiment 1. However, as the main effect of *key-timing* was not significant in Experiment 2 compared to Experiment 1, it is still a possibility. Thus, a closer look at possible issues with the experimental methods and/or materials must be taken into consideration.

One possible issue is the (un)reliability of native speaker idiom predictability measures. The idioms for the current study were taken from the DIN database, a database with both L1 and L2 ratings on over 300 idioms. While the results are generally similar to other databases that have been released (e.g., Titone & Connine, 1994b), marked differences in the same idioms between studies suggest that participants in the ratings experiments may not be accurate or unbiased in their judgements. Inaccuracies in ratings are particularly crucial in determining the placement of the idiomatic key. For example, the idiom *kick the bucket* was rated as highly predictable (70 % correct completions after "the"), a result that is surprising given the broad and general nature of "kick" as a verb. These results may be skewed by the bias of a large number of idioms in the ratings tasks. Where a large number of idioms are present in experimental tasks, both L1 (e.g., Bobrow & Bell, 1973) and L2 (e.g., Beck & Weber, 2016b) studies have shown that figurative completions or interpretations become more frequent or the strength of these activations can increase. Thus, the reliability of such cloze-probability tasks with a high ratio of idiomatic to non-idiomatic items cannot be taken for granted.

Another issue with the predictability scores is that they were taken from L1 speaker judgements. Cieślicka (2006) also addressed the issue of using L1 predictability for L2 listeners, noting that a best practice for such studies would be to include predictability measures from non-native speakers rather than native speakers. However, in order to accurately represent the knowledge variability in an L2 group in a particular experiment, the same speaker group should be used to determine predictability from the experiment, presenting a problematic experimental situation in which either participants must be exposed to the idioms before the experiment and risk a change in behavior, or the participants must see the idioms afterwards, which would

impact their responses. A more plausible compromise for further studies would be cloze-probability testing in a similar group of L2 speakers as the one tested, much like the L2 familiarity measures used in Beck and Weber (2016b). However, the same concerns remain that subjective familiarity with the idiom will greatly impact individual responses. Furthermore, the tendency for L2 speakers to use literal language preferentially in offline and production tasks may prevent any idioms from appearing as predictable by L2 listeners (Irujo, 1993).

Word frequency also had a marked impact on the data. In both experiments, frequency was a highly significant factor for reaction times. The word lists were originally controlled for frequency based on Beck and Weber (2016a) using the online CELEX database (Max Planck Institute for Psycholinguistics, 2001) for consistency. However, due to a large number of missing frequencies, the words were later re-calculated based on the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009) for inclusion in the LMER models. The differences in lists were discussed in the materials section, and, importantly, there were list differences between the figuratively related and unrelated control word lists. In particular, figuratively related targets were more frequent than the two control target groups. While the inclusion of this measure in the regression model should help identify the effects of the factors under consideration, the strength of frequency as a predictor suggests that the materials were not adequately controlled for considering the speaker group at hand, and this may provide some indication as to why figurative priming that appears in a visual inspection of the late-key data does not reach significance.

Critically, however, the concept of the idiomatic key and its role in idiomatic processing should be looked at with more scrutiny. With consideration of the criticisms of the identification of the idiomatic key are taken to heart, the challenge of correctly identifying the key remains unanswered. Whether or not such a task (i.e., identifying a true idiomatic key) is possible suggests that the concept of the idiomatic key may need reconsideration, as should studies attempting to identify predictability in idioms. Namely, previous authors have defined this as the point at which an idiom can be identified as such (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993), though, more recent research has suggested that access to idiomatic meaning is incremental (see Cacciari, 2014 for an overview) and subject to a variable number of factors during processing (e.g., Libben & Titone, 2008). For example, the presence of context is one factor that can also aid in the speed of idiomatic recognition (e.g., Beck & Weber, submitted, 2019; Colombo, 1993, 1998). If the idiomatic key represents the point at which enough information accumulates for recognition, then we should recognize the idiomatic key as a concept more fluid than a single word within an idiom. The idiomatic key may vary for an individual in the same idiom given a variety of contexts, but it also may vary greatly between individuals. Whereas an L2 speaker with advanced English experience and knowledge of L1 idioms may recognize an idiom very early during listening, another less-experienced or simply otherwise-experienced L2 listener may not recognize the idiom until its completion, or even at all. While this does not negate the importance of studies using predictability as a measure in idiom processing studies, as it has been shown to affect processing (e.g., Cacciari & Tabossi, 1988; Tabossi & Zardon, 1993; Titone & Connine, 1994), it's important to consider predictability and the idea of the idiomatic key in combination with other factors that may influence idiom processing and access to figurative meaning.

Overall, no evidence was found for early access to idiomatic meaning in L2 listeners based on a high-predictability or an early idiomatic key. However, the inconclusive results also should not be used as negative evidence. Rather, further investigation using highly-predictable idioms based on the target listener group and careful experimental controls may help further investigate the issue. In particular, later presentation of the target (i.e., one word later in early-key idioms) may be another experimental variation worth exploring. Considering L2 ratings for predictability may also add worthwhile information during item selection. Finally, the use of more sensitive measures such as eye-tracking (see e.g., Underwood et al., 2004) or EEG (e.g.,

Rommers et al., 2013) may help to re-examine this effect where reaction times to a lexical decision task may be insufficient.

Acknowledgements

We wish to thank Ida Sanders for her work on this project as a student assistant and M.A. student. Additional acknowledgements for data collection go to Stacie Boswell, Sabrina Bucher, Johanna Geyer, and Eliane Leberer.

References

Baayen, R.H., Bates, D., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from Psychology and Linguistics experiments* (Version R package version 0.0.4) [Computer software]. (2015).

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. ArXiv Preprint. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker S., (2015). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-8) [Computer software]. (2015). R Package. http://CRAN.R-project.org/package=lme4

Beck, S. D., & Weber, A. (2020). Context and Literality in Idiom Processing: Evidence from Self-paced Reading. *Journal of Psycholinguistic Research*, *49*, 837-863. https://doi.org/10.1007/s10936-020-09719-2

Beck, S. D., & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The Role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, *7*, 1350. https://doi.org/10.3389/fpsyg.2016.01350

Beck, S. D., & Weber, A. (2016b). English-German Database of Idiom Norms. https://hdl.handle.net/11022/0000-0007-EB26-6

Beck, S. D., & Weber, A. (2016c). L2 Idiom processing: Figurative attunement in highly idiomatic contexts. In A. Papafragou, D. Grodner, D. Mirman, & J. C. & Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* Cognitive Science Society.

Beck, S. D., & Weber, A. (2019). Context matters, figuratively, for L2 readers: Evidence from self-paced reading. In *Proceedings of The 11th International Conference on the Mental Lexicon.* University of Alberta Libraries. https://doi.org/10.7939/R3-MNCB-C270

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343–346. https://doi.org/10.3758/BF03198118

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, *9*(2), 267–293. https://doi.org/10.1075/ml.9.2.05cac

Cacciari, C., & Tabossi, P. (1988). The Comprehension of Idioms. *Journal of Memory and Language*, *27*, 668–683. https://psycnet.apa.org/doi/10.1016/0749-596X(88)90014-9

Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

Colombo, L. (1993). The Comprehension of Ambiguous Idioms in Context. In C. Cacciari & P. Tabossi (Eds.), Idioms: Processing, Structure, and Interpretation (pp. 163–200). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Colombo, L. (1998). Role of context in the comprehension of ambiguous Italian idioms. In D. Hillert (Ed.), Syntax and Semantics: Vol. 31. Sentence processing: A cross-linguistic perspective (pp. 405–425). New York: Academic Press.

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. https://doi.org/10.1093/applin/amm022

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford Univ. Press.

Gibbs, R. W. Jr. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. https://doi.org/10.3758/BF03213418

Gibbs, R. W. Jr., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593. https://doi.org/10.1016/0749-596X(89)90014-4

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Kyriacou, M., Conklin, K., & Thompson, D. (2019). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, 23830919847691. https://doi.org/10.1177/0023830919847691

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

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Max Planck Institute for Psycholinguistics (2001). *WebCelex [Computer software]*. http://celex.mpi.nl/

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, *17*(2), 167–190. https://doi.org/10.1080/01638539409544865

Nelson, D. L., McEvoy, C. L., Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms [Computer software]*. http://www.usf.edu/FreeAssociation/

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491–538. https://doi.org/10.1353/lan.1994.0007

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(5), 1223–1237. https://doi.org/10.1037/0278-7393.27.5.1223

R Core Team (2013). *R: A language and environment for statistical computing* (Version 3.4.3) [Computer software]. R Foundation for statistical computing. Vienna, Austria. http://www.R-project.org/

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776. https://doi.org/10.1162/jocn_a_00337

Schweigert, W. A., & Moates, D. R. (1988). Familiar idiom comprehension. *Journal of Psycholinguistic Research*, *17*(4), 281–296. https://doi.org/10.1007/BF01067198

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

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*, *29*(2), 72–89. https://doi.org/10.1177/0267658310382068

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

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540. https://doi.org/10.3758/MC.37.4.529

Tabossi, P., & Zardon, F. (1993). The Activation of idiomatic meaning in spoken language comprehension. In C. Cacciari & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation* (pp. 145–162). Lawrence Erlbaum Associates, Publishers.

Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1126–1138. https://doi.org/10.1037//0278-7393.20.5.1126

Titone, D. A., & Libben, M. R. (2014). Time-dependent effects on decomposability, familiarity and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon*, *9*(3), 473–496. https://doi.org/10.1075/ml.9.3.05tit

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. 155–172). John Benjamins.

Van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, *3*, 1–17. https://doi.org/10.1017/S1366728918001219

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). *Statistics and computing*. Springer.

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