Ruth Kessler & Sara D. Beck

2022

L1 and L2 Learners Keep Their Eyes on the Prize: Eye-tracking Evidence during Idiom Recognition

in

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

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

pp. 23–43

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

L1 and L2 Learners Keep Their Eyes on the Prize: Eye-tracking Evidence during Idiom Recognition

Ruth Kessler & Sara D. Beck

University of Leipzig & University of Tübingen

ruth.kessler@uni-leipzig.de & sara.beck@uni-tuebingen.de

1 Introduction

Idioms are both formulaic sequences as well as figurative expressions, identities which both present their own implications and possible challenges for language processing. Although idiomatic meaning appears to challenge compositional theories of language processing, for adult native speakers of a language, the use and comprehension of idioms is a breeze. In fact, research shows that not only are idioms like *to let the cat out of the bag* and *to have butterflies in one's stomach* understood figuratively by native speakers online (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988), but there may even be an advantage in speed of comprehension when compared with comparable novel expressions (e.g., Gibbs, 1980; McGlone et al., 1994; Swinney & Cutler, 1979). This advantage might stem from predictive properties based on their formulaic nature, in which idiomatic strings are recognized early during comprehension. However, the same speed advantages may also stem from a direct mapping of the form to the idiomatic or figurative meaning, bypassing compositional processes and associated word-meaning activation processes. The figurative meaning of the idiom (e.g., *to reveal a secret* as figurative meaning of the idiom *to let the cat out of the bag*) is thus associated with a fixed, word-like form rather than building the meaning of an idiom online from its individual constituents (i.e., CAT and BAG). In spite of this lack of apparent contribution of the literal words to the idiomatic meaning, it seems that, in many instances, native speakers still activate individual literal constituents (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988; Titone & Connine, 1994) and even literal phrasal meaning (e.g., Beck & Weber, 2019; Holsinger & Kaiser, 2013) without hindering the speed of access to the idiomatic meaning.

However, processing advantages in idioms are not to be taken for granted, particularly when comparing first (L1) and second (L2) language learners. For language users to understand an idiom, they must first recognize the phrase as an idiom, have been exposed to the figurative meaning of the idiom, and must ignore the possible literal interpretation achieved through traditional meaning computation. Additionally, they must do so quickly if predictive processes are to offer processing advantages. However, language learners have less experience with idiomatic expressions, which might negatively influence idiom recognition and speed of processing. Furthermore, the role of literal meaning might differ both between adult native speakers as well as between L1 and L2 learners of a language. Language learners may activate literal meaning differently than experienced users of a language based on language experience. L1 and L2 learner groups may also differ because of different underlying mechanisms of acquisition (Arnon & Christiansen, 2017). For instance, children may acquire formulaic sequences as units, and literal word meanings might be less prominent in processing of those sequences, while adult L2 learners are more likely to combine words within formulaic sequences and rely more on literal constituent meaning.

The current study addresses both the issue of prediction and literal activation. Specifically, the current study asks to what extent L1 adults predictive processes associated with formulaic

sequences also apply to language learners, and what processes occur following idiom recognition, particularly concerning literal meaning activation. In order to address these questions, we will first review these processes in native speakers and then attend to the current research in language learners, specifically proficient non-native (L2) speakers and child (L1) speakers.

1.1 Predictive Processing in Idioms

The processing advantages of idioms compared to similar novel phrases are well-documented (Conklin & Schmitt, 2008; Gibbs, 1980; McGlone et al., 1994; Siyanova-Chanturia et al., 2011; Swinney & Cutler, 1979). Some theories of idiomatic processing account for this processing advantage by suggesting that a holistic access to figurative phrase meaning is either partially, in the case of hybrid theories (e.g., Cacciari & Tabossi, 1988; Sprenger et al., 2006), or even fully responsible for fast idiomatic processing (e.g., Swinney & Cutler, 1979). Alternatively, processing advantages are often also attributed to familiarity with the conventionalized form of idioms (e.g., Tabossi et al., 2009). Providing evidence for the latter instance, reading studies using eye-tracking techniques have consistently shown processing benefits for idioms as well as other formulaic sequences. Underwood et al. (2004) used a reading paradigm that measured fixations on individual words within formulaic sequences, including idioms embedded in unambiguous contexts. They found that participants fixated less while reading all types of formulaic sequences compared to novel phrases. They also found fewer fixations on terminal words in formulaic sequences, which indicates prediction of these words based on the previous elements of the sequence. This result was confirmed by Siyanova-Chanturia et al. (2011), who found that idioms (e.g., *at the end of the day*), used in both literal and figurative contexts, incited fewer and shorter fixations than similar novel phrases (e.g., *at the end of the war*). However, they did not find any differences between idiomatic and literal readings of the idioms, providing even further evidence that the processing benefit associated with idioms may indeed relate to the conventionalized form of the phrase.

Additionally, idioms pose an interesting case because of their distinct idiomatic and phrasal meanings. These meaning differences are a common testing ground for predictive processes that occur during listening. There is evidence of qualitative differences in idiomatic processing compared to other types of formulaic language. Carrol and Conklin (2020) compared processing in idioms, collocations, and binomials and found that specific idiomatic properties such as phrasal frequency, familiarity, and decomposability play an important role in the speed of processing and that familiarity with the phrase might explain shorter reading times associated with idioms, unlike other types of formulaic language. Additionally, ERP studies conducted by Vespignani et al. (2010) and by Canal et al. (2017) provide further evidence for such differences by comparing literal and figurative processing. The former study found that prediction within idioms is qualitatively different from predictive processing in novel, literal language based on the timing and distribution of N400 and P300 effects. The latter study added to this finding with evidence of effects in sentence contexts that biased a figurative interpretation of the expression as early as the first constituent of the idiom. This suggests that prediction and possible integration processes of the figurative meaning occur immediately at recognition of the idiom, which may be sooner than in other types of formulaic language. Together, the reasons for processing advantages of idioms are still in need of further investigation.

While well-established general speed and processing advantages in idioms indirectly indicate predictive mechanisms for final words as result of their familiar and conventionalized form, only few studies have looked directly at prediction within idioms. Cloze-probability tests show that native speakers can produce predictions for constituent words within idioms. In fact, most native speakers predict the final word *BAG*, when encountering the phrase *let the cat out of the …* (Beck & Weber, 2016b; Van Lancker Sidtis et al., 2015), and some idioms allow for prediction very early in the phrase. An online study by Kessler et al. (2020) confirms these processes and found evidence for predictive processing both with eye-tracking and ERPs. In their eye-tracking study, native speakers listened to a non-biasing sentence ending with a predictable

idiomatic phrase missing its final constituent. Participants saw the correct final constituent of the idiom, a literal distractor, and two unrelated distractors. Fixation rates on the correct completion were significantly higher than to all other options early during listening. In their ERP data, in line with the results of Rommers et al. (2013), Kessler and colleagues found that when the final constituents of these predictable idioms were replaced with incorrect and unexpected words (e.g., let the cat out of the BASKET/STOMACH), N400 effects occurred, yielding higher amplitudes for incorrect completions compared to the correct completion. These results are in line with previous work showing the sensitivity of the N400 to predictive properties, such as lexical violations like unexpected words (Federmeier & Kutas, 1999). Thus, regardless of the root of idiomatic processing advantages, adult native speakers consistently show evidence of prediction during idiomatic processing.

1.1.1 Predictive Idiomatic Processing in Language Learners

Most research on predictive processing of idioms in language learners has been conducted with proficient non-native speakers. However, the results of these studies appear to be mixed. In several studies, the same advantages for formulaic sequences or idiomatic phrases compared to novel phrases, such as faster reading times and fewer fixations during the phrase and on terminal words, have also been present when compared directly to native adult speakers (Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2011; Underwood et al., 2004). Conklin and Schmitt (2008) tested both figurative and literal uses of idioms in a self-paced reading study on L1 and L2 readers and found a speed advantage for formulaic phrases compared to novel phrases in both reader groups. In direct contrast to these findings, Siyanova-Chanturia et al. (2011) found differences between L1 and L2 readers in eye-movements depending on the phrase-type; unlike in their L1 reader group, idiomatic readings did not produce any speed or fixation differences from novel phrases. Furthermore, idiomatic readings produced a processing cost for L2 readers not present in L1 readers. However, the results of a study conducted by Siyanova-Chanturia et al. (2011) suggests that the inconsistencies found between studies may be due to experience with the language. Namely, in their study on frequency in binomial processing (e.g., *bride and groom* vs. *groom and bride*), phrasal frequency interacted with proficiency in a manner suggesting that only known, memorized forms can be subject to phrasal frequency effects and this experience is necessary for processing effects to be visible in L2 readers.

Beyond the limited L2 research on the possible advantages of idiomatic and formulaic phrases, there is little research directly on L2 prediction in idiomatic processing. It is clear, however, that L2 learners are also able to predict during processing, though these processes may be more limited than in L1 speakers (see Kaan, 2014 for an overview). For example, there is growing evidence that L2 speakers use contextual information to form predictions of upcoming words differently from L1 speakers in literal language (e.g., Dussias et al., 2013; Ito et al., 2017). For instance, Ito et al. (2017) conducted an ERP study to test prediction of form and meaning in highly predictable sentence-endings (e.g., *The student is going to borrow a…book*.) on Spanish-English bilinguals. While the authors found reduced N400 amplitudes for words semantically related to predicted words, this reduction was not dependent on predictability. However, as also suggested by the reading and eye-tracking studies discussed previously, proficiency may modulate prediction abilities rather than distinguishing them from native processing. More studies looking at prediction in idioms are needed in order to compare L1 and L2 predictive processes in idiom comprehension. To our knowledge, there are not yet comparable eye-tracking or ERP studies on prediction within idioms in the L2.

For L1 language learners, there is some evidence that children show a processing advantage for idioms over novel expressions. In a study by Qualls et al. (2003), 10-year-old children read correct idioms with high, moderate, or low familiarity. These were compared to control phrases that were manipulated idiom forms in which either the initial or the final word was replaced with an unrelated word (e.g., *put/stomp their heads together; go around in circles/trouble*). For each sentence, children had to judge whether it was an idiom or not. Children showed shorter

response latencies for idiomatic than for control expressions but only when the final word was substituted. Furthermore, children showed shorter response times for highly familiar idioms reflecting the L1 adult data showing that higher familiarity leads to faster processing (Carrol & Conklin, 2020). Thus, similar to second language learners, children might show a stronger processing advantage if they are familiar with the canonical form of the idiom. Taken together, idioms also seem to have a privileged status for children, but direct evidence for predictive processing within idioms is missing.

L1 language learners are also able to perform predictions about upcoming words in literal language (e.g., Mani & Huettig, 2012, 2014). Mani and Huettig (2014) conducted a visual world experiment in which children listened to sentences like *the boy eats the big cake* while at the same time seeing pictures of eatable and non-eatable objects on the screen. Children showed more fixations towards eatable objects shortly after hearing the verb *eat* (Mani & Huettig, 2014). In another experiment, Mahler and Chenery (2018) looked at naming latency. Children were presented auditorily with high and low cloze sentences (*The dog buried the bone/stick*) and had to repeat the final word. They were faster at repeating words in high (*bone*) compared to low cloze words (*stick*) suggesting a benefit for predictable items. Still, more evidence is needed to explain predictive processing mechanisms in children, but overall research supports general predictive abilities.

1.2 Literal Activation in Idiomatic Constituents

After recognizing an idiomatic phrase, ongoing literal computation and therefore literal word meaning activation seems unnecessary. However, experiments have repeatedly shown that online literal activation occurs in adult native speakers by finding semantic spreading activation from idiom constituents to semantically related words (Beck & Weber, 2016a; Holsinger, 2013; Kessler et al., 2020; Rabanus et al., 2008; Smolka et al., 2007; Sprenger et al., 2006; van Ginkel & Dijkstra, 2019). In addition to consistent semantic priming effects for literal constituents of idioms (e.g. kick the bucket primes PAIL, Beck & Weber, 2016a; Cacciari & Tabossi, 1988), eye-tracking also provides further evidence of such activation. For example, Holsinger (2013) tested literal constituent meaning activation in a visual world paradigm. While listening to idioms (*kick the bucket*), participants saw four words on the screen: one related to the figurative meaning (DEATH), one literally related to an idiom constituent (FOOT) and two unrelated distractors (TRIANGLE, ANIMAL). Despite increased fixations towards the figuratively related word, participants also showed more fixations towards literally related words compared to unrelated distractors. Thus, eye-tracking also shows evidence of literal constituent meaning activation.

With regard to literal word meaning activation during processing of idioms that are recognized early within the phrase (i.e., idioms in which constituent words are highly predictable based on preceding constituent words) evidence is mixed. In a production study, Sprenger et al. (2006) found that speakers preparing to produce an idiom-final word (e.g., English translation: *Jan walked against the lamp*) produce words semantically related to the respective idiom-final word (*candle*) faster than unrelated words. In contrast, there is some evidence that literal constituents are not always activated during the prediction of idiom-final constituents. Rommers et al. (2013) conducted a semantic expectancy paradigm using ERP in which the final word of Dutch idioms in biasing contexts was highly predictable, but the idiom was presented either with its correct completion or replaced by either a word semantically related to the correct completion or an unrelated word (e.g., English translation: *After many transactions the careless scammer eventually walked against the LAMP/CANDLE/FISH yesterday*.). Neither the N400 nor the P600 effect showed sensitivity towards semantic relatedness, which suggested that literal activation of the final constituent did not occur. However, a recent study by Kessler et al. (2020) with native German speakers using a similar paradigm challenges this finding when testing auditorily instead of visually presented idioms in neutral contexts (e.g., English translation: *Hannes let the cat out of the BAG/BASKET/ARM*). While the authors did not find semantic

activation for word-by-word reading, they found evidence of semantic expectancy within idioms during listening – as indexed by reduced N400 amplitudes for related compared to unrelated words. In another experiment, Kessler et al. (2020) adapted the paradigm used by Holsinger (2013) and tested semantic expectancy in a visual world paradigm using the same stimuli as in the ERP experiment. Participants listened to idioms missing the final, highly predictable constituent word while viewing four words on the screen (correct completion, related distractor and two unrelated distractors). Increased fixations towards the correct completion around 460 ms prior to the offset of the acoustic stimulus indicated prediction of the correct idiom form. Simultaneously, there were more fixations to related than to unrelated distractors. Together, these mixed results indicate that also design-related differences might play a role. Nevertheless, one can conclude that literal word meaning is often co-activated during predictive processing after recognition of the idiom.

1.2.1 Literal Constituent Activation in Language Learners

For language learners, it is possible that literal meaning has a different status than for native adult speakers. And because of possibly different acquisition mechanisms involved (Arnon & Christiansen, 2017), L1 and L2 language learners might also differ in their reliance on individual constituents and thus, their literal word meaning activation. Although research with children has focused on the understanding and interpretation of figurative expressions (e.g., Bernicot et al., 2007), the topic of literal meaning activation in idioms has not yet been explored. However, this topic has been widely addressed in L2 learners.

For non-native speakers, a possible difference between L1 and L2 language users in literal activation may be the status of literal language as highly salient in comparison to figurative language as a whole (e.g., Cieślicka, 2006; Giora, 1997). This line of research suggests that not only does literal activation occur even when a figurative meaning is likely (i.e., in a biasing context), but activation of literal word and even phrasal meaning is obligatory for idioms as a general processing strategy (i.e., for *kick the bucket*, both the word BUCKET and the literal phrasal meaning of striking a pail with one's foot are activated automatically). In L2 cross-modal priming studies, both Cieślicka (2006) and later Beck and Weber (2016a) found that the literal meaning of idiom-final words was activated (e.g., *kick the bucket* activated PAIL), and these priming effects appeared to be stronger than the priming of the figurative meaning (e.g., DIE). However, Cieślicka (2006) did not include native speakers as a control group in her study. In contrast, Beck and Weber (2016a), and an additional study also employing both speaker groups by van Ginkel and Dijkstra (2019) showed that there were no differences between L1 and L2 activation of literal meaning. Beck and Weber (2019, 2020) also looked into the special status of literal meaning in L1 and L2 speakers on a phrasal level in two self-paced reading studies. Readings of idiomatic sentences were biased literally or figuratively, and sentence endings either followed this expectation or went against it. While L1 readers (Beck & Weber, 2020) showed evidence that context influenced literal and figurative readings, L2 readers (Beck & Weber, 2019) showed evidence only on figurative readings. However, there was no evidence that literal meaning is obligatorily computed in L2 speakers compared to L1 speakers, so differences in literal processing on the phrasal level also remain unsubstantiated. Overall, while L2 speakers do activate literal constituent meaning during processing, it is unclear whether literal constituent words play more importance in L2 compared to L1 processing.

1.3 Research Questions

We conducted two visual world eye-tracking experiments with printed words (cf. Huettig et al., 2011) using L1 and L2 language learners to address the questions:

(1) Do L1 and L2 learners predict words within an idiom during listening?

(2) Are literal words activated during this process, and if so, when?

By looking at both L1 and L2 language learners of German, we aim to fill in largely unexplored gaps on idiom processing in the progression of language development and look at processing differences that might result from different acquisition mechanisms.

In the current study, we use the visual world paradigm with printed words to answer our research questions. The visual paradigm is suitable to answer both research questions because in adults and children it is sensitive towards predictive mechanisms (Altmann & Kamide, 1999, 2007; Kamide et al., 2003; Mani & Huettig, 2012) as well as semantic competition in literal expressions (Cortés-Monter et al., 2017; Huettig & Altmann, 2005). Furthermore, this paradigm has been previously used to study lexical access in idioms (Holsinger, 2013; Kessler et al., 2020). Literal constituent meaning activation can be measured by looking at co-activation of semantically related words. Here, we used the experimental design of Kessler et al. (2020): participants listened to neutral sentences containing idioms without their final word (e.g., German: *Hannes ließ die Katze aus dem …*/English translation: Hannes let the cat out of the…). We chose idioms in which the final word was highly predictable. At the same time participants viewed words displayed on the screen (German/English translation) that were (a) correct completions of the idiom (*SACK*/BAG), (b) distractors semantically related to the correct completion (*KORB*/BASKET), and (c) two unrelated distractors (*BAUCH*/STOMACH*, ARM*/ARM).

In answering question (1), we hypothesized that prediction will be seen in many and early fixations towards the correct completion that is not presented auditorily. If evidence of an idiomatic processing advantage for both L1 and L2 listeners is present in prediction, then we should find looks to the correct completion before or around the offset of the final auditorily presented word. Concerning question (2), increased fixations towards related compared to unrelated distractors will index effects of semantic competition and thus, literal constituent meaning activation. We expect the results of previous priming experiments to extend to the current experiment, and L2 participants are expected to look more towards related compared to unrelated distractors. If there is a literal priority for L2 listeners, the amount of these fixations may be greater than in the L1 listeners groups examined in Kessler et al. (2020). As there is no evidence to the contrary for L1 language learners, we expect that they, too, will activate the literal meaning of idioms. Though, if only L2 learners show activation of the literal meaning, this would imply that L1 and L2 learners store and access idioms differently.

2 Experiment 1: L1 Language Learners

2.1 Methods

2.1.1 Participants

Twenty-six 7th graders (17 female; mean age = 12;8 years) that we recruited from local schools participated in the experiment. L1 learners from this age group were chosen based on the great probability of exposure to the idioms included in the study while still early enough in language development to consider them learners. Parents gave informed consent prior to the experiment. All participants were native, monolingual speakers of German and had no history of hearing disorders and normal or to normal corrected vision. As compensation, the children received a voucher for a local toy store. The design of the study was approved by the local ethical committee (reference number: 2016/1027/22).

2.1.2 Materials

Experimental items from Kessler et al. (2020) were used in the current study. These items consisted of twenty familiar German idioms (see Appendix) embedded in sentences using the following structure (see Table 1): (i) a person carrying out the action of the sentence, (ii) an idiom fragment that originated from a German idiom (e.g., *die Katze aus dem Sack lassen*, English translation: *to let the cat out of the bag*) and (iii) the final target word of the idiom (which was

not presented auditorily). Pre-tests with adult native speakers of German ensured a mean familiarity rating of 5.7 on a scale from 1 (not familiar) to 7 (highly familiar). Sentences were recorded using a native speaker of German, and the final target word was cut from the recording. Participants heard each idiom fragment once, during which four visual words were presented on a computer screen. The words presented visually were categorized as: (1) Correct Completion: correct completion of the idiomatic phrase (e.g., *SACK*/BAG), (2) Related Distractor: semantically related word of the Correct Completion (e.g., *KORB*/BASKET), (3&4) Unrelated Distractor 1 and Unrelated Distractor 2: semantically unrelated to the Correct Completion (e.g., *BAUCH*/STOMACH, *ARM*/ARM). Unrelated 1 and Unrelated 2 words were word pairs taken from Correct Completions and Related Distractors used from other idiomatic sentences in the experiment. For example, the word pair *BAUCH* (*STOMACH*) & *ARM* (*ARM*) was used as correct completion and related distractor for the idiom fragment *Isabell hatte Schmetterlinge im…* (*Isabell had butterflies in the…*), and as Unrelated Distractor 1 and Unrelated Distractor 2 for the idiom fragment *Hannes ließ die Katze aus dem…* (*Hannes let the cat out of the…*). All four presented words on the screen matched the grammatical gender expected from the preceding sentence context and did not overlap phonologically.

Table 1. German example sentence for types 1-4 with English equivalent

(i) Person	(ii) Idiom Fragment	(iii) Target words			
		(1) Correct	(2) Related	(3) Unrelated 1	(4) Unrelated 2
Hannes	ließ die Katze aus dem	Sack	Korb	Bauch	Arm
Hannes	*let the cat out of the*	*bag*	*basket*	*stomach*	*arm*

Semantic relatedness between correct and related words was confirmed by computing semantic similarities in the German LSA Space using the R package LSAfun (Günther et al., 2015). In general, similarities in a latent semantic analysis (LSA) relate to lexical priming effects (Günther et al., 2016). In the present experiment, the semantic similarity between correct and related words was significantly higher than between correct words and both unrelated words (Wilcoxon signed rank test: unrelated1 $Z = 189$, $p < .001$; unrelated2 $Z = 185$, $p = .002$). There was no difference in semantic similarity between correct words and both unrelated words ($Z = 75$, $p = .28$). Target words were visually presented in white font (Arial, font size 28) on a grey background (see Figure 1), and the position of displayed words was counterbalanced across items and participants. The inter-trial interval was 1500 ms, and the presentation of a fixation cross for 500 ms marked the beginning of a new trial. Next, the set of four words were displayed on the screen, and the presentation of the audio stimuli began 2150 ms later. The order of the trials was randomized.

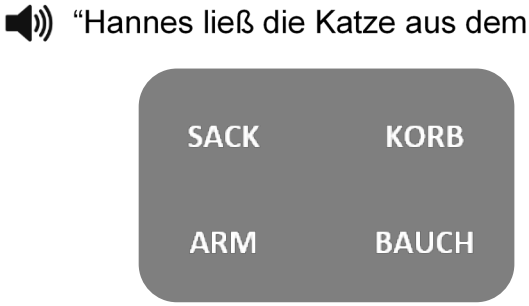

Figure 1. Example for visual display and auditory stimuli used in the experiment

2.1.3 Procedure

The experiment was conducted either at local schools or in the lab of the University in a single session. Participants were given both written and oral instructions for the experimental task. They were instructed using a cover story in which they teach an alien correct idiomatic expressions. After listening to each idiom fragment via headphones, participants should decide which of the visually presented words was the correct completion for the idiomatic expression by saying their choice out loud. Participant responses were not recorded, but noted by the experimenter. After their oral response, participants were instructed to press a button in order to continue to the next trial. Prior to the experimental task, a 5-point eye calibration was carried out for each participant followed by a practice block consisting of five trials. For the eye-tracking task, participants were asked to sit in a relaxed position and fixate on the screen. Fixations were recorded using a portable Tobii eye-tracker with a sampling size of 60 Hz. The eye-tracking experiment took around 20 minutes in total including instructions, eye-calibration and the experimental task, which lasted about 10 minutes on its own.

2.1.4 Analysis

In order to analyze fixations on the four visually presented words, we divided the screen into four areas of interest (AOIs). The time window of the analysis was aligned to the offset of each audio stimulus (offset = 0 ms). Only items with correct responses were included in the analysis (i.e. trials in which the participants chose the correct completion of the idiom). We assume that correct answers indicate knowledge of the idiom forms. Values of the two unrelated distractors were averaged. We used the fixation data to conduct two analyses to determine 1) the point at which participants reliably looked to the correct response, and 2) the time-course of fixations to each response type. In our first analysis, we conducted running t-tests ($p < .01$) comparing fixations towards correct completions and unrelated distractors at succeeding measurement points (every 16.67 ms resulting from 60 Hz sampling rate of the eye-tracker). These tests were used to determine whether anticipation occurred by identifying a point of divergence between fixations on correct completions compared to unrelated distractors. We excluded fixations to related distractors from this comparison because we assumed that the result could be biased due to possible semantic spreading activation. In the second analysis, we conducted a Growth Curve Analysis (GCA) with orthogonal polynomials (Mirman et al., 2008) to compare the amount and time-course of fixations towards Distractor Types (related and unrelated). By using orthogonal polynomials, growth curve models capture non-linear changes over time. Therefore, they are well-suited to model the time-course of fixations in visual world paradigms. In growth curve models, polynomial terms reflect different changes in the curve: the intercept term relates to the average magnitude of the curve, the linear term relates to the slope of the curve, the quadratic term relates to in inflection around the center of the curve, and cubic and quartic terms relate to inflections at the extremities of the curve.

As the starting point of the GCA time window, we chose the start of the anticipation, that is the point at which fixation proportions between correct completions and unrelated distractors diverge (first analysis). As length of the GCA time window, we chose 1200 ms.

2.2 Results

In 91.72 % of the trials, participants responded correctly (average responses to related distractors: 7.16 %; unrelated distractors 1: 1.55 %, unrelated distractors 2: 0.58 %). In our first analysis, a running t-test showed that fixations on correct completions and unrelated distractors diverged around 128 ms before the stimulus offset and therefore indicate anticipation. The overall fixation pattern is displayed in Figure 2, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, we modeled fixation patterns for related and

unrelated distractors with third-order orthogonal polynomials. The modeled data are also depicted in Figure 2, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 2.

Table 2. Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated)

Term	Estimate	Standard Error	t	*p*
Intercept	-0.02	0.00	-4.78	< .001
Linear	0.02	0.03	0.80	.42
Quadratic	0.03	0.00	6.35	< .001
Cubic	-0.01	0.00	-2.93	< .001

We find a significant effect on the intercept that indicates more fixations on related than unrelated distractors. The quadratic term indicates that the curves for both distractors differ in their central inflection. As visualized in Figure 2, Panel (B), this relates to a lower peak for unrelated distractors than for related distractors. The effect on the cubic term suggests that towards the extremities of the model inflections for both distractors differ. This effect captures that at the extremities both curves converge, as seen in Figure 2, Panel (B). Together, the present results indicate that while participants generally fixate more on related distractors overall than on unrelated distractors, this difference is largest over central parts of the model.

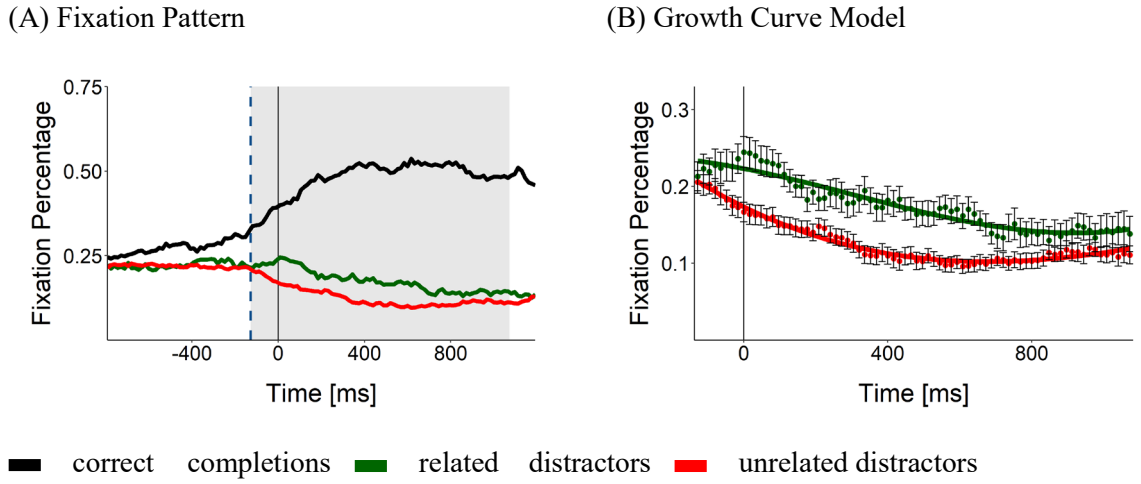

Figure 2. (A) Fixation patterns for L1 Children: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line)

2.3 Discussion

In Experiment 1, we tested idiom processing in first language learners (7th graders). They showed relatively good performance on idiom recognition as displayed by correct responses in more than 90 % of the trials. This recognition of the idiom-final word was also observable in predictive fixations of the correct completion. Although the onset of predictive fixations could be accelerated by the task of finding the correct completion of the idiom fragment, it nevertheless shows an increasing fixation on these correct complements without being presented within the idiomatic expression. Although not transferable to natural language processing, it supports the idea that knowledge of idioms is reflected in implicit measures in thirteen-year-old children. Thus, children not only show a processing benefit for the recognition of highly familiar idioms

(Qualls et al., 2003), but also fast online prediction of idiom-final constituents. This is in line with other studies showing the abilities of children to predict upcoming words in sentences (e.g., Mani & Huettig, 2012, 2014). However, unlike previous studies, the current study examines highly predictable idioms rather than using semantic cues for prediction (e.g., EAT increases looks to CAKE), and indicates an additional type of prediction in L1 children. We interpret these results as positive evidence for our first research question concerning prediction during listening, as seen via early anticipation of the idiom-final word.

Regarding literal constituent meaning activation, children indeed show more fixations on related than unrelated distractors, and we interpret this as positive evidence for literal activation in L1 child listeners. This effect seems to be strongest shortly after anticipation of the correct idiom-final word.

3 Experiment 2: L2 Language Learners

3.1 Methods

3.1.1 Participants

We recruited 33 non-native speakers of German (English L1, 12 female; mean age = 25.7 years) from the University of Tübingen for the experiment. Participants were living in Germany at the time of the experiment and were asked only to participate if they had at least a B2 level of proficiency (Common European framework of references for languages, 2001). Of the 33 participants, only 26 were included in the analysis because five participants fell below the criteria for appropriate language proficiency based on task performance (correct responses < 40 %), LexTale score (scores < 50/100), and the language background questionnaire (years of study < 2, average self-rated proficiency < 3 on a scale of 1-7). The 26 participants included in the analysis had an average of 5.05 (*SD* = 2.34) years of German instruction, an average LexTale score of 68.62 out of 100 (*SD* = 9.39), and an average self-reported proficiency (averaged across reading, writing, speaking, and listening) of 5.01 (*SD* = 0.94) on a scale of 1-7. Participants had no history of hearing disorders and normal or to normal corrected vision.

3.1.2 Materials

The same materials from Experiment 1 were used in Experiment 2.

3.1.3 Procedure

The procedure was the same as in Experiment 1 with the addition of a LexTale vocabulary test (Lemhöfer & Broersma, 2012) prior to the eye-tracking task and a language background questionnaire at the end of the experimental session. These additional tasks were used as measures of language proficiency. The LexTale is a lexical decision task used to test vocabulary knowledge in non-native speakers (Lemhöfer & Broersma, 2012). While this test is not a direct assessment of language proficiency, it has been shown to correlate with other measures of proficiency in L2 leaners (Beck, 2020). The experiment lasted about 30 minutes.

3.1.4 Analysis

The analyses were conducted following the same procedure as Experiment 1. Again, only items with correct responses were included in the analysis, and fixations were used to determine 1) the point at which participants reliably looked to the correct response via t-tests, and 2) the time-course of fixations to each response type via Growth Curve Analysis (GCA) (Mirman et al., 2008).

3.2 Results

On average, participants responded in 64.81 % of the trials with the correct completion of the idiom (average responses to related distractors: 17.88 %; unrelated distractors 1: 11.15 %, unrelated distractors 2: 6.15 %). In our first analysis, a running t-test revealed a reliable difference in fixations towards correct completions and unrelated distractors starting around 48 ms prior to the stimulus offset. The overall fixation pattern is displayed in Figure 3, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, differences between fixation patterns for related and unrelated distractors were again modeled using a Growth Curve Analysis with third-order orthogonal polynomials because by visual inspection of the time-course, two bends were observable in the curve. The modeled data are also depicted in Figure 3, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 3.

Table 3. Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated)

Term	Estimate	Standard Error	t	*p*
Intercept	-0.02	0.00	-4.16	< .001
Linear	0.04	0.03	1.06	.29
Quadratic	0.05	0.01	8.39	< .001
Cubic	-0.07	0.01	-11.37	< .001

There was a significant effect of Distractor Type on all terms but the linear term (see Table 3). The main effect of Distractor Type indicates overall more fixations towards related than unrelated distractors. The quadratic term reflects that the curves for both distractors differ in their inflection around the center. Together with the depiction of the model in Figure 3, Panel (B), this indicates that the curve for related distractors peaks higher than for unrelated distractors. The effect on the cubic term reflects differences at the extremities of the curve.

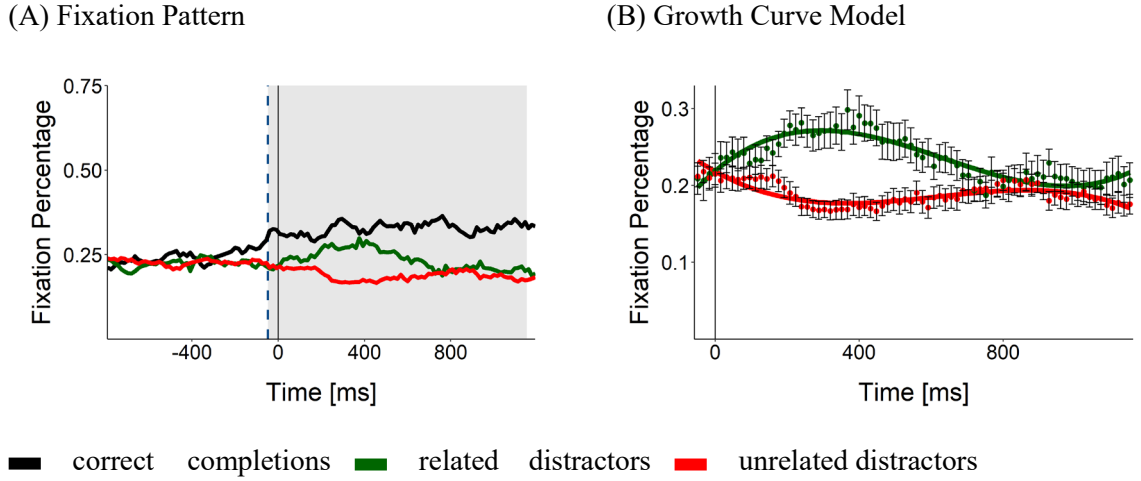

Figure 3. (A) Fixation patterns for L2 Adults: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line)

3.3 Discussion

In Experiment 2, we tested idiom processing in second language learners of German (English L1). Their performance, averaging about 65 %, reflected moderate figurative language proficiency, which was lower than the proficiency of the L1 learner group in Experiment 1. However, when idioms were recognized, there was also evidence of prediction. Although looks to the correct completions only occurred 48 ms before the offset of the idiom fragment, the programming of a saccade following a critical word takes approximately 200 ms (Saslow, 1967), a delay which suggests that recognition occurred prior to the anticipation point of 48 ms identified here. These results are in line with studies suggesting that there may be processing advantages associated with idioms, even for non-native learners (Conklin & Schmitt, 2008; Underwood et al., 2004). We interpret these results as positive evidence for our first research question, namely that L2 learners exhibit predictive processing of idioms during listening.

Concerning literal constituent meaning activation, more fixations on related distractors compared to unrelated distractors suggest that, like L1 learners, L2 learners activated the literal meaning of constituent words. This pattern of results is consistent with literature suggesting strong activation of literal constituents in L2 idiom processing (e.g., Beck & Weber, 2016a; Cieślicka, 2006), especially considering that this literal activation continues to increase even after correct recognition of the final constituent. While this does not provide evidence of an L2 literal priority directly (e.g., Cieślicka, 2006), it is further confirmation of the importance of literal word meaning even in figurative phrases. It should also be noted that activation occurs even without the auditory cue of the final word, suggesting that this activation is also predictive in nature and strengthens our conclusions on prediction in L2 listening. Overall, these results indicate that L2 learners also display patterns of predictive processing and semantic activation early on in idiom processing.

4 General Discussion

Overall, the results from Experiments 1 and 2 show evidence that both groups of L1 and L2 language learners are able to predict idiom-final constituents during listening in addition to activating the literal meaning of these constituents. In the following, we will give tentative explanations for possible differences related to language proficiency when comparing both groups to native adult speakers. Regarding the task itself (i.e. identifying the correct completion of each idiom), we obtained graded effects in accuracy and timing of eye movements. In a previous study, adult native speakers recognized idioms' correct final word at ceiling level with 99 % correct responses (Kessler et al., 2020). In the present study using the same materials and procedures, L1 language learners recognized 91 % of the idiom completions correctly, while L2 language learners only recognized 65 % correctly. Similarly, overall looks to the correct response were higher for the L1 language learners than for the L2 language learners (see Figures 2 and 3, Panels A). Furthermore, adult native speakers directed their gazes earlier to the correct completion (approx. 460 ms before its onset) than L1 language learners (approx. 130 ms) and L2 language learners (approx. 50 ms), respectively. Graded accuracy in fixating and recognizing the correct idiom completions might simply reflect that adult native speakers had the most experience with the language, and L1 children had more experience than our L2 learners did. Differences in proficiency and therefore experience and familiarity with respective idioms may impact stored frequency information (Lee et al., 2013; Qualls et al., 2003; Siyanova-Chanturia et al., 2011; Tabossi et al., 2009) resulting in differences in the efficiency of predictions based on the same phrases.

Like native adults (Kessler et al., 2020), language learners in the present study appeared to activate literal meanings of predicted idiom-final constituents. All three groups showed more fixations towards related distractors than towards unrelated distractors. While predicting the correct completion *bag* of the idiom onset *to let the cat out of the ...,* participants' fixations were biased towards *BASKET* (being a semantic associate of *bag).* They appeared to predict

single constituents, which made spreading activation to semantic associates of these single constituents possible. Also similar to the timing obtained for fixations to the correct completions, the timing of fixations towards related distractors (relative to the anticipation of the correct completion) seems graded between the three groups. Again, this outcome might relate to differences in proficiency.

When comparing L1 adults (Kessler et al., 2020) and L1 children, differences in timing for increased fixations towards related distractors relative to increased fixations towards correct completions might be explained by age-related differences in experience and familiarity with respective idioms. Higher exposure and familiarity might relate to stronger multi-word representations of these expressions. Early fixations towards related distractors in adults might indicate automatic semantic spreading activation (for further discussion see, Kessler et al., 2020). In turn, late fixations towards related distractors in children might indicate weak multi-word representations. In this case, delayed lexical activation of the idiom-final constituent might be due to more flexible representations that accept related idiom variants, and less confidence with the conventionalized form of the idiomatic phrase. Although not the case for all used idiomatic expressions, some related distractors also yield literally plausible continuations of the idiom fragment which might explain increased fixations towards these words. This explanation could also support weak multi-word representations because this effect occurs after the correct idiom has been recognized as such. Alternatively, late lexical activation might suggest delayed spreading activation in language learners compared to proficient native adults in general.

When comparing L1 and L2 speakers of German, looks towards related distractors also revealed qualitative differences in processing. While looks to semantically related distractors linearly decreased in time in L1 learners (see Figure 2, Panel B) and in L1 adults (Kessler et al., 2020), they initially increased for L2 learners (see Figure 3, Panel B). In terms of qualitative differences, it has been suggested that literal meaning plays a more important role in L2 idiom processing compared to L1 idiom processing (e.g., Cieślicka, 2006; Giora, 1997), though evidence has been mixed (e.g., Beck & Weber, 2016a; van Ginkel & Dijkstra, 2019). Responsible for the present result may indeed be an underlying learning mechanism that is more additive in nature for L2 learners compared to L1 learners (Arnon & Christiansen, 2017), which might account for a stronger reliance on constituent meanings. Following to this line of argumentation, increased looks to semantically related distractors in L2 might reflect obligatory activation of literal meaning rather than a more shallow activation that may be displayed in L1 groups (e.g., Peterson et al., 2001). Accordingly, constituent meanings within idioms play a different role in online processing.

Alternatively, however, the present outcome might also still be a result of graded language proficiency. Participants of the least proficient group (L2 learners) might either be less confident in their choice of the idiom-final word and/or weaker in inhibiting automatic activation of semantic associates. Furthermore, increased looks to the related distractors may be a result of second-guessing their given responses as these choices sometimes reflect a likely response (e.g., a cat may be let out of a BAG/*SACK* or a BASKET/*KORB*). In the case of the latter possibility, the visual presence of both the correct final constituent as well as its semantic associate may cause a greater challenge for non-native speakers, who are less skilled at inhibiting related but inappropriate information (e.g., Gernsbacher & Faust, 1991). As the final idiomatic constituent as well as its related distractor were available on the screen during the listening process, the current study is limited in its ability to clearly tease apart those alternative conclusions.

In conclusion, our data shows evidence that L1 and L2 language learners are able to predict during listening and activate the meaning of literal constituents. Whereas prediction indicates that both sets of learners interact with the idiom as a unit, literal constituent activation indicates that the parts of the unit remain available. Differences in timing between these groups and adult native speakers (Kessler et al., 2020) suggest that proficiency plays a role in both the timing of anticipation during listening as well as the pattern of literal activation. Additionally, in line with growing evidence that L1 and L2 idiomatic processing are similar (e.g., Beck & Weber, 2016a,

2019; Conklin & Schmitt, 2008; Tabossi et al., 2009; van Ginkel & Dijkstra, 2019) we propose that the differences found between our L1 and L2 learners here may be an interaction between proficiency and L2-specific challenges rather than fundamental differences in storage and processing. Future studies must include more varied proficiencies in learner groups in order to better understand how prediction and activation evolves and use additional methods such as ERP (e.g., Kessler et al., 2020) in which literal activation can be more clearly separated from other possibilities such as spreading activation by excluding target words from the experimental trials.

Acknowledgements

We would like to thank Stacie Boswell, Claudia K. Friedrich, Birte Herter, Kerstin Jendrysik, Thanh Lan Truong, Andrea Weber, Miriam Wild. We also warmly thank all participants. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science, 9*(3), 621-636.

Beck, S. D. (2020). *Native and Non-native Idiom Processing: Same Difference.* (Doctoral dissertation), University Tübingen, Tübingen. Available from TOBIAS-lib.

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*, 1305.

Beck, S. D., & Weber, A. (2016b). *English-German Database of Idiom Norms (DIN)*. Retrieved from: https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/englisches-seminar/sections/english-linguistics/lehrstuhl-prof-dr-andrea-weber/staff/sara-d-beck/din/

Beck, S. D., & Weber, A. (2019). Context matters, figuratively, for L2 readers: Evidence from self-paced reading. *Proceedings of the 11th International Conference on the Mental Lexicon, 1*, e057.

Beck, S. D., & Weber, A. (2020). Context and Literality in Idiom Processing: Evidence from Self-paced Reading. *Journal of Psycholinguistic Research, 49*, 837-863.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics, 39*(12), 2115-2132.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*(6), 668-683.

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(6), 928-943.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 63*(1), 95-122.

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22*(2), 115-144.

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.

Cortés-Monter, D. R., Angulo-Chavira, A. Q., & Arias-Trejo, N. (2017). Differences between skilled and less-skilled young readers in the retrieval of semantic, phonological, and shape information. *Journal of Research in Reading, 40*, S170-S189.

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching and assessment.* Cambridge: Cambridge University Press.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35*(2), 353-387.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 245-262.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition, 8*(2), 149-156.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*(3), 183-206.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47*(4), 930-944.

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology, 69*(4), 626-653.

Holsinger, E. (2013). Representing idioms: syntactic and contextual effects on idiom processing. *Language and Speech, 56*(3), 373-394.

Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 866-878.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23-B32.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica, 137*(2), 151-171.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(4), 635-652.

Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism, 4*(2), 257-282.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

Kessler, R., Weber, A., & Friedrich, C. K. (2021). Activation of literal word meanings in idioms: Evidence from eye-tracking and ERP experiments. *Language and Speech 64*(3), 594-624.

Lee, E. K., Hsin-Yi Lu, D., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition, 16*(4), 761-775.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325-343.

Mahler, N. A., & Chenery, H. J. (2018). A developmental perspective on processing semantic context: Preliminary evidence from sentential auditory word repetition in school-aged children. *Journal of Psycholinguistic Research, 48*(1), 81-105.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843-847.

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of Experimental Child Psychology, 126*, 264-279.

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes, 17*(2), 167-190.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475-494.

Peterson, R., Burgess, C., Dell, G., & Eberhard, K. (2001). Dissociation Between Syntactic and Semantic Processing During Idiom Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1223-1237.

Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders, 36*(4), 245-261.

Rabanus, S., Smolka, E., Streb, J., & Rösler, F. (2008). Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik, 36*(1), 27-47.

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.

Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America, 57*(8), 1024-1029.

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. (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.

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing verbs in German idioms: Evidence against the Configuration Hypothesis. *Metaphor and Symbol, 22*(3), 213-231.

Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*(2), 161-184.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18*(5), 523-534.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition, 37*(4), 529-540.

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.

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). Amsterdam: 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, 23*(1), 131-147.

Van Lancker Sidtis, D., Cameron, K., Bridges, K., & Sidtis, J. J. (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand, 2*, 39-48.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22*(8), 1682-1700.

Item	Idiom Fragment	Correct Completion	Related Distractor	Unrelated Distractor 1	Unrelated Distractor 2	Familiarity, Scale from 1 (not familiar) to 7 (highly familiar)[1]		Relation of final word to figurative meanings, Scale from 1 (not related) to 7 (highly related)[2]		Cloze Probability for correct idiomatic completion[3]
						Mean	Standard Error	Mean	Standard Error	Percentage
1	Julia rutschte das Herz in die	Hose	Jacke	Welt	Venus	6.05	0.31	3	0.34	100%
	Julia slid the heart in the	*pants*	*coat*	*world*	*venus*					
2	Lena setzte alle Hebel in	Bewegung	Sprünge	Löwen	Hasen	6.35	0.21	4.92	0.33	88%
	Lena put all levers in	*move*	*jumps*	*lion*	*bunny*					
3	Marie stand Gabriel Rede und	Antwort	Frage	Seele	Gefühle	5.65	0.36	5.72	0.32	100%
	Marie stood Gabriel speech and	*answer*	*question*	*mind*	*emotions*					
4	Hannah schlug sich die Zeit um die	Ohren	Augen	Bewegung	Sprünge	5.35	0.44	2.17	0.37	100%
	Hannah hit herself the time around the	*ears*	*eyes*	*move*	*jumps*					
5	Sofia brachte die Aufgaben unter Dach und	Fach	Schrank	Ohren	Augen	5.6	0.37	3.08	0.36	100%
	Sofia brought the duties under roof and	*shelf*	*cupboard*	*ears*	*eyes*					

6	Annika war Balsam für die	Seele	Gefühle	Hose	Jacke	5.95	0.20	4.96	0.35	94%
	Annika was balm for the	*mind*	*emotions*	*pants*	*coat*					
7	Amelie hatte einen Frosch im	Hals	Rücken	Speck	Käse	6	0.29	5.88	0.29	100%
	Amelie had a frog in the	*throat*	*back*	*bacon*	*cheese*					
8	Isabell hatte Schmetterlinge im	Bauch	Arm	Wasser	Regen	6.5	0.24	4.96	0.34	100%
	Isabell had butterflies in the	*stomach*	*arm*	*water*	*rain*					
9	Jasmin lebte wie die Made im	Speck	Käse	Kopf	Bart	4.4	0.41	3.12	0.36	94%
	Jasmin lived like the grub in the	*bacon*	*cheese*	*head*	*beard*					
10	Melina ließ die Kirche im	Dorf	Feld	Hals	Rücken	5.15	0.32	2.36	0.29	94%
	Melina left the church in the	*village*	*field*	*throat*	*back*					
11	Helena fiel ein Stein vom	Herzen	Magen	Dorf	Feld	6.8	0.12	4.96	0.31	100%
	Helena fell a stone from the	*heart*	*stomach*	*village*	*field*					
12	Nora hatte ihr Herz am rechten	Fleck	Platz	Sack	Korb	6.35	0.20	2.92	0.35	100%
	Nora had her heart on the right	*spot*	*place*	*bag*	*basket*					0%
13	Nico erblickte das Licht der	Welt	Venus	Antwort	Frage	5.75	0.32	5.68	0.32	88%
	Nico beheld the light of the	*world*	*venus*	*answer*	*question*					
14	Robin legte für Natalie die Hand ins	Feuer	Holz	Fach	Schrank	6.45	0.15	3	0.32	100%

	Robin put for Natalie the hand in the	*fire*	*wood*	*shelf*	*cupboard*					
15	Linus verlor den Boden unter den	Füßen	Händen	Rollen	Kugeln	6.35	0.21	3.84	0.38	100%
	Linus lost the floor under the	*feet*	*hands*	*rolling*	*rolling* (syn.)					
16	Jannis fiel die Decke auf den	Kopf	Bart	Fleck	Platz	6.45	0.15	3.6	0.37	100%
	Jannis fell the ceiling on the	*head*	*beard*	*spot*	*place*					
17	Julian hielt den Kopf über	Wasser	Regen	Füßen	Händen	5.1	0.27	3.42	0.37	100%
	Julian held the head above	*water*	*rain*	*feet*	*hands*					
18	Moritz begab sich in die Höhle des	Löwen	Hasen	Herzen	Magen	5.45	0.37	3.44	0.4	100%
	Moritz repaired himself to the hole of the	*lion*	*bunny*	*heart*	*stomach*					
19	Timo brachte den Stein ins	Rollen	Kugeln	Feuer	Holz	5.2	0.32	4.4	0.33	100%
	Timo brought the stone in the	*rolling*	*rolling* (syn.)	*fire*	*wood*					
20	Hannes ließ die Katze aus dem	Sack	Korb	Bauch	Arm	5.75	0.31	2.76	0.31	94%
	Hannes let the cat out of the	*bag*	*basket*	*stomach*	*arm*					

[1] Pre-test Familiarity: Participants (N=20, adult native speakers of German) were asked to rate familiarity for each idiom on a scale from 1 (not familiar) to 7 (highly familiar). Numbers indicate means and standard errors for each idiom.
[2] Pre-test Relatedness of final constituent to figurative phrase meaning. Participants (N=25, adult native speakers of German) were asked to rate how much the final constituent relates to the figurative phrase meaning of each idiom on a scale from 1 (not related) to 7 (highly related). Numbers indicate means and standard errors for each idiom.
[3] Pre-test Cloze Probability: Participants (N = 17, adult native speakers of German) were given written forms of idiom fragments missing the final word. Their task was to fill in the missing blank space with a word they thought would complete the fragment best. Numbers indicate percentages of idiomatically correct responses.