MDPI

Article

Not Batting an Eye: Figurative Meanings of L2 Idioms Do Not Interfere with Literal Uses

Marianna Kyriacou * and **Franziska Köder**

Center for Multilingualism in Society across the Lifespan (MultiLing), University of Oslo, 0313 Oslo, Norway; franziska.koder@iln.uio.no
* Correspondence: marianna.kyriacou@iln.uio.no

Abstract: Encountering idioms (*hit the sack* = "go to bed") in a second language (L2) often results in a literal-first understanding ("literally hit a sack"). The figurative meaning is retrieved later, subject to idiom familiarity and L2 proficiency, and typically at a processing cost. Intriguingly recent findings report the overextension of idiom use in inappropriate contexts by advanced L2 users, with greater L2 proficiency somewhat mitigating this effect. In this study, we tested the tenability of this finding by comparing eye-movement patterns for idioms used literally, vs. literal control phrases (*hit the dirt*) in an eye-tracking-while-reading paradigm. We hypothesised that if idiom overextension holds, processing delays should be observed for idioms, as the (over)activated but contextually irrelevant figurative meanings would cause interference. In contrast, unambiguous control phrases should be faster to process. The results demonstrated undifferentiated processing for idioms used literally and control phrases across measures, with L2 proficiency affecting both similarly. Therefore, the findings do not support the hypothesis that advanced L2 users overextend idiom use in inappropriate contexts, nor that L2 proficiency modulates this tendency. The results are also discussed in light of potential pitfalls pertaining to idiom priming under typical experimental settings.

Keywords: idioms; second language (L2); eye-tracking reading; literal meaning; literal idiom uses

1. Introduction

A substantial part of everyday communication is expressed with prefabricated formulaic sequences, such as speech formulae (*how are you*), collocations (*strong coffee*), phrasal verbs (*give up*), and idioms (*throw in the towel*; Erman and Warren 2000; Foster 2001). Acquiring these multi-word expressions in a second language (L2) is, therefore, crucial for effective communication, and indicative of the highest stages of L2 mastery (Wray and Perkins 2000). Nevertheless, mastering formulaic language is a long and laborious endeavour (Schmitt 2012), with some types of formulaic sequences being harder to acquire than others. Idioms, for instance, can be particularly challenging to learn. These expressions are considerably less frequent than other types of formulaic sequences (Moon 1998), thus minimising opportunities for implicit learning and the formation of strong representations (entrenchment) in memory (Conklin 2019; Ellis 2002; Pulido 2022). Idioms are further susceptible to negative cross-linguistic transfer from the first language (L1) when L2 equivalents do not match in terms of both meaning and form (Pulido 2021; Pulido and Dussias 2020). Finally, idioms are notoriously known for their often idiosyncratic, non-compositional nature, since the meanings of the individual component words contribute little, if anything, to the figurative meaning (*throw + in + the towel* ≠ "to give up"), and many have a plausible literal interpretation which may compete with the figurative for activation (e.g., "to literally throw a towel in [a physical place]").

Several studies have investigated to what extent the processing of idioms differs in L1 and L2 users of a language. Most previous research suggests that the literal meaning

of idioms is particularly prominent in L2 processing (literal bias), while the figurative meaning is only accessed later during processing (e.g., Carrol and Conklin 2017; Cieślicka and Heredia 2017; Senaldi and Titone 2022; Siyanova-Chanturia et al. 2011). Intriguingly, however, a recent study by Milburn et al. (2021) reported that advanced L2 users not only overcame this literal bias, but also started overextending idiom use in inappropriate contexts, suggesting an overactivation of figurative meanings. Specifically, using a visual world paradigm, the researchers observed activation of idiom-related targets in contexts where they were not warranted. Since this is a surprising, and so far, single finding, potentially confounded with some methodological limitations in Milburn et al.'s (2021) study design and analysis, we tried to replicate these results using eye-tracking while reading. In our study, advanced L2 users read sentences containing idioms used in their literal sense (i.e., literally *throw in the towel*) or matched, nonidiomatic phrases (e.g., *throw in the coin*). Overall, the results speak against idiom overextension and cast some light on how L2 users process figurative sequences in the absence of a figurative bias.

1.1. Differences between L1 and L2 Idiom Processing

A number of models hold that L2 idioms are more likely to undergo a full compositional analysis, because the literal meaning is more prominent (*Literal Salience Model*; Cieślicka 2006) or because accessing the individual meanings of an idiom's component words (*throw*, *in*, *the*, *towel*) is obligatory (*Idiom Diffusion Model of Second Languages*; Liontas 2002), and/or more accessible to L2 users (*Model of Dual L2 Idiom Representation*; Abel 2003). While literal compositional analysis also features in the current models of L1 idiom processing, this terminates as soon as an idiom is recognised, subject to phrasal predictability (*Configuration Hypothesis*; Cacciari and Tabossi 1988), as well as familiarity, frequency, literalness, and meaning transparency (*Multidetermined Model*; Libben and Titone 2008; Titone and Libben 2014). Familiarity is of particular importance as it can influence perceptions of transparency (i.e., how guessable the meaning of an idiom is based on its component words alone) and may override frequency effects when a phrase is rare in language use but very well-known to a speaker. With regard to transparency, findings show that both L1 and L2 users tend to classify idioms as more transparent when they are highly familiar to them, and once familiarity is accounted for, both L1 and L2 users show similar patterns in the way they interpret figurative phrases (Carrol et al. 2018). Further findings suggest that L1 and L2 intuitions on idiom transparency are less reliant on language experience, and that L1 judgements on transparency can predict L2 idiom familiarity (Hubers et al. 2020). Relative weights and intercorrelations aside, these phrasal properties essentially contribute to the entrenchment of idioms in the L1. For example, highly familiar idioms are more deeply entrenched in memory (i.e., they have stronger representations), and as a result, they are recognised sooner than less familiar ones. Once the recognition threshold has been reached, L1 idioms along with their figurative meaning can be retrieved directly from memory (lexical route), negating the need for further compositional analysis.

It is argued that because L1 idioms can (at some point) benefit from the lexical route, they often enjoy a processing advantage relative to literal phrases whose only available route is the compositional (Conklin and Schmitt 2008; Rommers et al. 2013; Senaldi et al. 2022; Tabossi et al. 2009; Vespignani et al. 2010). Eye-tracking reading studies, for instance, report faster reading times and fewer regressions for idioms (*spill the beans*) compared to control phrases (*spill the chips*), as well as increased skipping rates for idiom final words as these tend to be more predictable than final words in nonidiomatic phrases (Carrol and Conklin 2020; Underwood et al. 2004). What is more, this processing advantage in the L1 may be sustained regardless of which idiom meaning is intended in the context. For instance, native speakers of English have demonstrated faster reading times for English idioms used figuratively (*at the end of the day* = "eventually") as well as literally (*at the end of the day* = "in the evening") compared to literal control phrases (*at the end of the war*; Siyanova-Chanturia et al. 2011).

In the same study, L2 users were found to process both idioms and control phrases at a similar speed, although idioms used figuratively were processed more slowly than

idioms used literally (Siyanova-Chanturia et al. 2011), an effect replicated in other studies (e.g., Carrol and Conklin 2017; Cieślicka and Heredia 2017). This implies that L2 idioms are processed like literal phrases, and, as such, they do not enjoy a similar processing advantage as L1 idioms. Moreover, accessing the figurative (as opposed to the literal) meaning of an L2 idiom incurs additional processing costs. A more recent eye-tracking study examined how French–English bilinguals processed L2 English idioms and control phrases containing a language switch in French (*spill the fèves* vs. *spill the beans*/*cook the fèves* vs. *cook the beans*; Senaldi and Titone 2022). Canonical and switched idioms were embedded in contexts biasing their figurative or literal meaning. The results showed that idioms and control phrases were equally affected by the language switch in early measures (i.e., slower first pass gaze duration). However, switched idioms used figuratively (*spill the fèves* = "to reveal the secrets") were read significantly more slowly than switched idioms used literally (*spill the fèves* = "literally spill beans out of a container") in total reading time, a late reading measure. This again indicates that L2 idiom processing is mostly compositional, and that the retrieval of the figurative meaning occurs (at a cost) in later stages of processing, but only when the context warrants its activation; when the context is felicitous with a literal interpretation, idioms and literal phrases are processed similarly.

However, not all studies report differences between L1 and L2 idiom processing (Beck and Weber 2016a; Carrol et al. 2016; Conklin and Schmitt 2008), with L2 proficiency and dominance being likely driving factors, presumably because they directly relate to language experience and opportunities for lexical entrenchment (Carrol et al. 2018; Cieślicka et al. 2014). Indeed, some findings show that less proficient L2 users are more likely to resort to a literal analysis, whereas advanced L2 users are more likely to process idioms like native speakers (Cieślicka and Heredia 2017). Further evidence from self-paced reading studies suggests that idiom literalness (or literal plausibility) may interact with contextual cues in both L1 and L2 idiom processing (Beck and Weber 2019, 2020). In these studies, processing costs were observed for both L1 and L2 users when preceding context biased a figurative interpretation, but subsequent information was in line with an idiom's literal meaning, thus speaking against a literal-first bias, even for L2 users. However, when idioms were *not* primed for a figurative interpretation but were instead preceded by a literally-biasing context, figurative resolutions became costly (although unsurprisingly not for idioms that were literally implausible: e.g., *lose your cool*). While this cost was only trending for both L1 and L2 users, looking at the studies' stimuli, it appears that (for highly literal idioms), figurative continuations in the literal contexts were not necessarily incongruous. For example, the figurative meanings of the idioms *break the ice* and *spill the beans* in "*the chilly Eskimo just wanted to break the ice with his peers*" or "*the assistance chef who didn't seem to do anything well had managed to spill the beans about the surprise*", respectively, are not implausible in the context. This could explain why literal resolutions in literal contexts (at least for highly literal idioms) were not read significantly faster than figurative continuations. Literal continuations in *figurative* contexts, on the other hand, were consistently more incongruous/improbable (e.g., "*the new schoolboy wanted to break the ice on the lake sooner than later/the popular teenager who liked to gossip with his group of friends had managed to spill the beans on the stove even faster than expected*"). In addition, the idiom literality effects reported above were observed in later parts of the sentences (underlined in the previous examples), which may reflect differences in contextual congruency more than idiom meaning activation. Collectively, many of the findings reviewed thus far largely support existing L2 idiom-processing models, according to which the literal meaning is predominant, although modulations as a function of L2 proficiency, dominance, experience, and potentially idiom literalness cannot be excluded. The figurative meaning of idioms, on the other hand, seems to (fully) activate in later stages of comprehension at a cost to the processing system (e.g., longer total reading times) and after a literal-compositional analysis has been performed. Importantly, this cost seems to only emerge when the context explicitly biases a figurative meaning, suggesting that contextual bias might be an essential prerequisite for L2 idiom activation.

1.2. Milburn et al.'s (2021) Study

In contrast to the above, a recent eye-tracking visual world study by Milburn et al. (2021), reported L2 idiom activation in inappropriate contexts. The researchers recorded fixations of advanced L2 users of English with L1 Norwegian while they listened to incomplete sentences biasing an idiomatic conclusion (*Because she was exhausted, Victoria hit the* [*sack*]), a literal conclusion (*After tripping on a rock, Victoria hit the* [*dirt*]), or were unbiased (*Victoria hit the* [*sack/dirt*]). The final words (*sack/dirt*) of the sentences were omitted, and participants were asked to complete the sentences by clicking on one of four available pictures on the monitor. Target pictures consisted of a figurative target (idiomatic conclusion), a literal target (literal conclusion), and two distractor images. Figurative and literal targets were matched for cloze probability using the Corpus of Contemporary American English (COCA; Davies 2009): for example, *dirt* and *sack* were statistically equally likely to follow *hit the*. Distractor images corresponded to nouns that were either more or less predictable than the figurative and literal targets (e.g., *hit the road* [high cloze] vs. *hit the iceberg* [low cloze]). The chosen idioms were normed for familiarity and had no Norwegian equivalents, thus eliminating processing variance due to item unfamiliarity, and cross-linguistic effects, respectively (for a recent discussion on this, see Du et al. 2023).

In the figurative condition, participants looked more at the figurative targets compared to both the literal targets and (high-cloze) distractors, although the difference between literal and figurative targets was on the marginal side ($p = 0.052$). In the literal condition (*hit the* [*dirt*]), participants looked more at the literal targets across the time window, but the researchers reported a steeper parabolic curve for looks to the figurative targets (*sack*), with increased looks being observed at the beginning and end of the time window relative to both literal targets and distractors. No significant effects were noted in the unbiased condition. It was argued that idiomatic status trumped both contextual fit (as figurative targets were considered in the literal condition) and local predictability (as figurative targets were also better lures than high-cloze distractors). This effect was somewhat mitigated by higher L2 proficiency, which was attributed to an enhanced ability to suppress (the irrelevant) idiomatic meanings. Overall, the findings were taken to indicate that advanced L2 users overcame the literal bias and started overextending idiom use in contextually inappropriate situations. As participants were familiar with the idioms at hand and were highly accurate in the figurative condition, this was further interpreted as evidence of the participants having stable L2 idiom representations, but doubts regarding their contextual fit, which made it difficult for them to ignore idioms when they were explicitly present. These findings would further imply the overactivation of figurative meanings; if idioms were considered solely based on their idiomatic status, this would necessitate recognising idioms in the first instance. Idiom recognition would in turn lead to idiom activation and hence the retrieval of the associated figurative meaning. However, overextending idiom use seems counterintuitive given that, as we have seen, figurative meaning activation typically comes at a processing cost to L2 users and is reliant on contextual bias.

Some methodological limitations cast further doubts regarding the reliability and generalisability of these results. First, most of the figurative targets were *not*, strictly speaking, incompatible in the literal condition (in their literal sense). For example, a *literal* interpretation of *hit the sack* (i.e., hit an object in the shape of a sack) would make sense in the sentence "*After tripping on a rock, Victoria hit the sack*". Equally, a literal interpretation of *kick the bucket* would make sense in the sentence "*To get into the room, Omar was about to kick the bucket*", if one imagines a physical pail blocking an entrance. In yet other instances, the *figurative* meaning of idioms was also plausible in the literal condition. For example, in the sentence "*In the hospital, Raj had to hold his horses*", the figurative meaning of *hold your horses* (i.e., "to wait") would make sense if one had to be patient instead of impulsive. Importantly, the viability of either the literal and/or the figurative meaning of idioms in the literal condition speaks against idiom overextension in inappropriate contexts. In fact, it is possible that the figurative targets were considered *because* they were "too good" of a fit. Furthermore, many literal targets were also plausible in the figurative condition, which

might explain why the difference between literal and figurative targets in the figurative condition was only marginal. For instance, in the sentence, "*James's big deadline was looming so yesterday he broke his foot*", one could imagine a scenario where someone broke their foot deliberately to gain some extra time ahead of the deadline.

Second, if idiomatic status overrides contextual fit and cloze probability, then figurative targets should have elicited significantly more looks in the unbiased condition too, and significantly more so than high-cloze distractors. This was not the case, however, further suggesting that targets were considered on the basis of their contextual fit, not of their idiomatic status. Third, figurative meanings are significantly primed when highly expected ("figurative attunement"; Beck and Weber 2016b), an effect potentially attenuated in this study as figurative targets were always present on the screen and were indeed viable completions for at least a third of the experimental trials (if we exclude the unbiased condition). Therefore, participants may have been on the lookout for idioms by way of checking if an idiomatic ending would have been preferable in the context. In addition, although the researchers carefully counterbalanced (local) cloze probability between figurative and literal continuations, differences in (global) contextual predictability, word length, or frequency were not accounted for.

Finally, we would like to raise some concerns regarding the way the time-course data were interpreted in this visual world paradigm study. Milburn et al.'s (2021) finding that figurative targets were strong lures in the literal context is solely based on the shape of the fixation curve, i.e., the slope of the curve at the beginning and end of the selected time window. However, this is problematic since indices in the visual world paradigm are not conventionalised and it remains unclear (a) why these particular indices have been selected among many other possibilities, and (b) what a "steeper curve" actually means (see Figure 1 in Milburn et al. 2021). As McMurray (2023) demonstrates with a series of simulations, properties of the fixation curve such as peaks and slopes need to be interpreted with caution as they may not reflect real underlying cognitive processes, but rather sensory or oculomotor control processes. Similarly, the correlation that Milburn et al. (2021) found between L2 proficiency and the slope of the curve also appears spurious and might not be caused by individual differences in L2 proficiency, following concerns raised by McMurray (2023). Importantly, the reliability of the fixation curves seems to deteriorate over time, so that the underlying activation is less reliably reflected later in the trial (McMurray 2023). This is of particular concern in Milburn et al.'s (2021) study which has a large analysis window of 3500 ms.

Adding issues such as ecological validity to the mix (i.e., hearing incomplete sentences and being presented with four picture options), as well as the fact that figurative targets related to the literal meaning of idioms' last words (e.g., a picture of a table for the idiom *turn the tables*), the visual world paradigm does not seem to be the ideal choice to investigate idiom processing in adult speakers. Due to the described limitations of Milburn et al. (2021), we decided not to directly replicate this study. Instead, we designed an eye-tracking reading task using the same idiomatic expressions and testing a similar population. One of the advantages of using a reading paradigm is that it has clearly established indices such as first pass reading time, total reading time, and regression likelihood, which lend themselves to confirmatory hypothesis testing.

1.3. The Present Study

In the present study, we investigated (a) if encountering a potentially idiomatic sequence leads to the (over)activation of its figurative meaning regardless of contextual bias, and (b) whether the ability to suppress irrelevant figurative meanings improves as a function of higher L2 proficiency. Idioms and matched control phrases were embedded in identical sentences designed to be felicitous with a literal interpretation (e.g., "*While harvesting potatoes, the farmer stumbled and hit the sack/dirt with his head*"). We further controlled for contextual and phrasal predictability, word length, and the frequency of the phrases' final words (*sack/dirt*). To make our results comparable with those of Milburn et al., we also recruited advanced L2 users of English with L1 Norwegian.

If Milburn et al.'s hypothesis of an overactivation of idiomatic phrases is correct, we would expect processing delays for idioms, as readers concurrently entertain two phrasal interpretations and engage in disambiguation processes: e.g., *hit the sack* = "go to bed" (nontarget meaning) vs. "literally hit a sack" (target meaning). These delays should be evident in the form of slower reading times and increased rereading probability for idioms vs. control phrases (*hit the dirt*), as the latter do not have alternative interpretations. Specifically, we would expect slower processing in late measures, in line with previous research reporting slower total reading times for idioms used figuratively (Carrol and Conklin 2017; Cieślicka and Heredia 2017; Senaldi and Titone 2022; Siyanova-Chanturia et al. 2011). On the other hand, if no processing differences are observed, this would suggest that advanced L2 users still have literal-compositional analysis as their preferential processing route, and do not activate idiomatic meanings unless the context explicitly calls for a nonliteral interpretation. Furthermore, a negative interaction with L2 proficiency (i.e., shorter (re)reading times for idioms as proficiency increases) would indicate that greater L2 proficiency aids the suppression of irrelevant (figurative) meanings. Importantly, unlike previous studies, in this study, we *only* used idioms in their literal sense to avoid any artificial priming from overexposure to figurative meanings.

2. Materials and Methods

2.1. Participants

Fifty native speakers of Norwegian with English as an L2 (mean age = 29.5, *SD* = 3.99; females = 38) took part in the main eye-tracking experiment conducted at the Centre of Multilingualism in Society across the Lifespan, University of Oslo. Participants received compensation for their participation and signed a consent form prior to the start of the experimental session. The study was reviewed and approved by the the Norwegian Agency for Shared Services in Education and Research (SIKT; reference number 933733). All materials, data, and data analysis code are provided in Supplementary Materials.

As a proxy of our participants' proficiency in English, we opted for LexTALE (Lemhöfer and Broersma 2012), which was also employed by Milburn et al. (2021). LexTALE is an online vocabulary test in the form of a lexical decision task. In the English version, participants are required to decide if a word they see on the screen is a real English word or not, and click YES or NO respectively. Familiarity with the meaning of the words is not required. Scores are given at the end of the task in percentages. Participants scored very highly (*M* = 84.75, *SD* = 13.15), with only one participant scoring below 50% (range: 47.50–100%). Our participants were therefore of similar L2 proficiency as the participants in Milburn et al., who scored on average 82.98% (range: 62.5–98.75%).

Self-reported data collected via a shortened version of the LeapQ questionnaire (Marian et al. 2007) in Norwegian showed that our participants started learning English at an average of 6.18 years old (similar to Milburn et al.'s participants, where the average was 6.06 years). In terms of usage on an average week, our participants reported using English 17% of the time with friends, 5% with family members, and 37% at work/school. They further reported having spent on average 0.44 years in an English-speaking country. In line with their LexTALE scores, participants reported having above average proficiency in speaking (*M* = 4.6/6), listening (*M* = 5.16/6), reading (*M* = 4.88/6), and writing (*M* = 4.56/6) in English.

2.2. Materials

We used the same 20 idiomatic expressions (e.g., *turn the tables*, *hit the sack*) as in Milburn et al. (2021) and paired them with 20 literal control phrases by replacing the final words of the idioms with other nouns that matched in terms of word length (*turn the pages*, *hit the door*; *t*(19) = −0.50, *p* = 0.62), word frequency (*t*(19) = −2.02, *p* = 0.06), and transitional probability (*t*(19) = 0.57, *p* = 0.57). The latter refers to the statistical likelihood of the final noun to appear after the verb phrase (*turn the*) and was estimated using COCA's Collocations function (Davies 2009). We additionally measured the contextual cloze probability of the final nouns via a cloze task administered to a different set of 12 participants with the same

demographics as those of the main study. Participants were presented with the stimuli sentences leading up to (but excluding) the final word of the idiom/control phrase and were asked to provide the next word or two words that came to mind. Pairwise *t*-tests showed no difference in the cloze probability between the final words of the idioms and those of the control phrases (*t*(19) = 0.01, *p* = 0.99), suggesting comparable contextual predictability.

The idioms were literally plausible and had already been normed for familiarity for this particular group of L2 users by Milburn et al. (2021). However, to ensure that the participants who took part in our eye-tracking study knew the idioms, they were also asked to complete an idiom multiple-choice test at the end of the experimental session. Idioms were presented out of context and were accompanied by four choices: one correct definition (target), two incorrect definitions (distractors), and the option to select "I don't know". For example, for the idiom *fall off the map* the choices were "to become unpopular after having been popular" (target), "to become lost in unfamiliar surroundings/to become unable to do things one was able to do in the past" (distractors), and "I don't know". Average accuracy on the idiom test was high across participants (*M* = 0.71, *SD* = 0.19), indicating familiarity with the items. Looking at the familiarity scores by-item, it appeared that for two idioms, accuracy was below chance (*hold the key*, 0.34; *wear the pants*, 0.48), but this was likely due to distractors being too good or too close to the target meaning (i.e., "to be entrusted with a secret" and "to be the breadwinner in a household", respectively). We did not norm idioms for meaning transparency given that subjective familiarity highly correlates with this construct, making it hard to disentangle separate effects (Carrol et al. 2018). Idioms and control phrases were embedded in short sentences designed to favour *literal* interpretations, thus eliminating confounds from figurative attunement (Beck and Weber 2016b)[1] and hence the expectation of encountering nonliteral meanings (see Table 1 for stimuli examples).

Table 1. Example stimuli sentences across conditions.

Condition	Phrase	Stimulus
ID-LIT	Fall off the map	Susan used colourful pins to mark out the places she wanted to visit on the atlas, but they annoyingly **fell off the map** a second later.
LIT-LIT	Fall off the wall	Susan used colourful pins to mark out the places she wanted to visit on the atlas, but they annoyingly **fell off the wall** a second later.
ID-LIT	Raise the roof	When remodelling the house, the builders had to **raise the roof** quite a bit.
LIT-LIT	Raise the costs	When remodelling the house, the builders had to **raise the costs** quite a bit.

Note. ID-LIT refers to idiomatic phrases used literally; LIT-LIT refers to literal control phrases.

2.3. Procedure

The stimuli sentences were counterbalanced across two lists using a Latin square design. Each list contained 20 experimental phrases (10 idioms used literally, 10 literal control phrases), and each participant only saw a phrase in one of the two conditions. There were also 20 filler sentences. Idioms and matched control phrases never occurred over a line break.

To record eye-movements while reading we used an EyeLink 1000+ desktop-mount eye-tracker (SR Research, Ontario, Canada), with a sampling rate of 1000 Hz. Participants were seated in front of a computer monitor. To minimise head movements, we used a chin- and forehead-rest. The eye-tracker was calibrated using a 9-point grid, and re-calibration was performed whenever necessary. The stimuli sentences were triple-line spaced, in black font (Courier New, size 16) on a white background, and they were displayed in the middle of the screen, one at a time. Drift correction was carried out before the presentation of each sentence. Participants were asked to read the sentences silently as fast as possible without compromising comprehension, and to press ENTER when they finished reading each sentence to proceed to the next trial. Seventy per cent of the filler sentences were followed by

YES/NO questions to ensure attention to the task. To answer these questions, participants had to press one of the corresponding keys on the keyboard. The trial presentation was randomised by participant.

The eye-tracking reading task always initiated the experimental session and was followed by the idiom familiarity test, the adapted LeapQ questionnaire, and the LexTALE vocabulary task. We pre-registered our study design prior to data collection on OSF (https://osf.io/y4hz9, accessed on 15 January 2024).

3. Results

Accuracy on the YES/NO comprehension questions was high (92%), demonstrating that participants understood the task and were attentive. We focused the analyses on the whole idiom/control phrase region (*hit the sack/dirt*), as well as on the final words (*sack/dirt*) separately (Carrol and Conklin 2014). Fixation durations shorter than 80 ms were merged with the largest nearest fixations (for a distance up to 0.5 degrees of visual angle) or were removed. A further 1.6% of the data was lost due to track loss, and 1.6% of the cases in the whole phrase dataset and 0.3% of the cases in the final word dataset were discarded following the removal of outliers.

For both the whole phrase and final word regions, we examined *first pass reading time* (the duration of all fixations during first pass), *total reading time* (the duration of all fixations, including those of refixations), and *regression likelihood* (the likelihood of re-fixating the region of interest from subsequent parts of the sentence). For the final word region only, we additionally examined *first fixation duration*.[2] For the whole phrase region only, we also examined *spillover* (the duration of the first fixation made to the right after leaving the region of interest), as unresolved processing can sometimes carry over into the following region.

The data were analysed using mixed-effects models and the *lme4* package, version 1.1–32 (Bates et al. 2014), in R version 4.2.2 (R Core Team 2023). The main fixed effect of interest was Condition, a two-level factor (ID-LIT: idioms used literally vs. LIT-LIT: literal control phrases used literally) with ID-LIT set as the baseline. We included LexTALE scores as a predictor tapping into L2 proficiency and specified an interaction with Condition. Idiom familiarity test scores correlated highly with LexTALE scores ($r = 0.60$), and since idioms were, on the whole, familiar to our participants, we only included the LexTALE scores as a predictor to avoid collinearity in the models. LexTALE has been normed extensively as a task and is often used as a reliable proxy of L2 proficiency. Finally, the Trial index (i.e., trial sequence order) was included to control for any fatigue/learning effects as experience with the reading task accumulated. To normalise distributions, reading time measures were log-transformed, whereas LexTALE scores and the Trial index were centred. Regression likelihood into the phrase region, a binary measure, was estimated using logistic regression (Jaeger 2008). The remaining models were estimated using linear regression.

Maximal models initially included by-participant and by-item random intercepts and slopes for Condition and Trial index. For items only, we also added intercepts of the LexTALE scores and an interaction between Condition and LexTALE score. In the final models, the random effect structure was simplified as necessary to eliminate convergence/singular fit issues. Importantly, the effects of the final models were the same as those of the maximal models. The final models were further compared to models lacking an effect of Condition (null models) using likelihood ratio tests with the *anova()* function. All model outputs are provided in Appendix A, and the means (calculated with *emmeans* package (Lenth 2018)) are provided in Table 2.

There was no effect of Condition in any of the models of the phrase region. Full models with Condition did not result in a better fit when compared to null models lacking an effect for Condition (all *p*-values ≤ 0.26). There was an effect of LexTALE scores across the board, with higher scores leading to faster first pass ($\beta = -0.01$, $SE = 0.002$, $t = -4.08$; $p < 0.001$), total reading times ($\beta = -0.07$, $SE = 0.003$, $t = -2.30$, $p = 0.02$), and smaller spillover effects ($\beta = -0.004$, $SE = 0.001$, $t = -2.68$, $p = 0.008$). Interestingly, higher LexTALE scores led to an *increased* likelihood for regression into the whole phrase region ($\beta = 0.04$, $SE = 0.01$, $t = 3.34$,

$p < 0.001$), which is likely a processing trade-off given the faster reading times. There was no significant interaction between LexTALE and Condition in any measure examined. No effect of Condition or of any other factor was noted in the first fixation duration of the final words. There was no significant effect of Trial index in either the phrase or the final word analyses.

Table 2. Mean reading times and regression probability for idioms used literally (ID-LIT) vs. literal control phrases (LIT-LIT).

	Condition			
	ID-LIT		**LIT-LIT**	
	Mean	*SE*	**Mean**	*SE*
Final word region				
First fixation duration	223.0	4.7	221.0	4.6
Idiom region				
First pass reading time	455.0	23.9	466.0	23.0
Total reading time	638.0	31.5	627.0	29.1
Spillover	232.0	6.2	235.0	6.3
Regression probability	0.2	0.0	0.2	0.0

Note. All reading times are reported in milliseconds while regression probability as probability between 0 and 1.

4. Discussion

This study aimed to test the replicability of Milburn et al.'s (2021) finding that advanced L2 users overextend idiom use in inappropriate contexts, although more advanced users might be better able to suppress it. To test this, we designed an eye-tracking study recording participants' eye movements while reading sentences containing idioms intended literally or literal control phrases. If idiom overextension held, great(er) activation of the phrases' (nontarget) figurative meanings would mean inflated reading times for idioms relative to the control phrases, as meaning disambiguation would add to the processing time. However, our results showed undifferentiated processing across the board, with idioms and control phrases being read at the same rate in both early and late eye-tracking measures. The absence of an idiom-specific processing cost speaks against idiom overextension in unbiasing contexts. Furthermore, while greater L2 proficiency facilitated processing in general, it did not interact with idiom status. The results are, therefore, at odds with Milburn et al.'s, but in line with previous literature advocating a predominantly literal-compositional analysis of L2 idioms (e.g., Carrol and Conklin 2017; Cieślicka and Heredia 2017; Senaldi and Titone 2022; Siyanova-Chanturia et al. 2011), although it should be noted that in the present study *only* the literal meaning of the idioms was explored as there was no figurative condition in the study design.

There could be several, not mutually exclusive, reasons for these divergent findings. First, what Milburn et al. reported might not be true effects, but by-products of the study design and analysis discussed in the Introduction or the result of a Type I error. Second, the presence of idiomatic meanings (which is the norm in idiom studies in general) could have activated figurative meanings, even when the context suggested a literal interpretation. Despite multiple differences between our experiments (visual world vs. reading study), idioms were explicitly visible to participants in both paradigms, through the figurative target pictures in Milburn et al. and through the sentence stimuli in our reading task. However, one important difference is that in our study, idioms were never explicitly used in the figurative sense, as sentences were designed to always be felicitous with a literal interpretation. With findings reporting priming for figurative meanings as a function of increased exposure to idiomatic phrases (Beck and Weber 2016b), we aimed to reduce the likelihood of figurative meanings becoming activated simply because participants expected idioms to occur. This may have resulted in our idioms passing mostly "under the radar", with participants not (measurably) activating idiomatic meanings.

This has potentially important implications for research in idiom processing, not only in the L2 but also in the L1. Compared to experimental settings, idioms are less frequent in

real language use, and although studies employ both filler and control items by default, their idiosyncrasy might render them particularly marked and noticeable even after a few encounters (Siyanova-Chanturia and Martinez 2015). This markedness might be further pronounced when idioms are highly opaque and/or when they have an improbable literal meaning (e.g., [to achieve something] by the skin of one's teeth) because of the obvious lack of compositionality or the lack of real-world applicability, respectively (Mancuso et al. 2020; Titone and Connine 1999). When idioms are repeatedly used in their figurative sense, as is typical in idiom studies, this may lead to the additional priming of both idiom forms and their figurative meanings, potentially leading to speeded processing. Indeed, in the Introduction we mentioned how idiomatic phrases (kick the bucket) often enjoy faster processing compared to nonidiomatic control phrases (kick the ball), a phenomenon referred to as the idiom-processing advantage. To begin with, idioms are more familiar, frequent, and predictable than the average control phrase due to their conventionality. Adding priming to the equation would only further boost their ease of processing. Effectively, however, we do not currently know how (L1 or L2) idioms would be processed in an experimental design where idiomatic meanings are not highly expected and especially when idioms are also literally plausible. It might be that more naturalistic, infrequent encounters with idioms in their figurative sense actually delay comprehension compared to nonidiomatic phrases. This is not unlikely given that idioms evoke stronger and more widespread activation in the brain (Boulenger et al. 2008, 2012; Lauro et al. 2007; Zempleni et al. 2007), highlighting the additional processing effort involved in their processing.

Furthermore, as far as the idiom advantage is concerned, it is yet to be determined whether it truly reflects facilitated (L1) idiom comprehension, or merely faster phrasal recognition (due to increased priming from multiple sources), without this necessarily leading to easier integration (for a more detailed discussion, see Kyriacou et al. 2021). It might be, for instance, that when idioms are embedded within longer discourse, this phrase-level advantage is offset by longer processing times outside the idiom region (i.e., in regions preceding or following the idiom). This is often difficult to measure in practice given that experimental stimuli would need to be felicitous with both the meaning of idioms and that of control phrases, while simultaneously maintaining the surrounding context constant to eliminate noise from lexico-syntactic variations. However, there is some indirect evidence in support of this argument from research focusing on idiom predictability, a phrase-internal construct with strong priming effects: the more predictable an idiom, the faster it is recognised, and hence, the faster it is processed (e.g., Cacciari et al. 2007; Cacciari and Tabossi 1988; Fanari et al. 2010; Tabossi et al. 2009). It has been found that predictable idioms are read faster than unpredictable ones in the idiom region (somewhat mirroring the difference between idioms and control phrases), but crucially, predictable idioms, intended literally as well as figuratively, lead to longer reading times in post-idiom disambiguating regions, compared to unpredictable idioms intended literally (Cacciari and Corradini 2015). This suggests that faster idiom processing does not necessarily facilitate their integration with subsequent context. As this experiment focused on figurative vs. literal meaning activation and integration, it did not include a nonidiomatic control condition for comparison. In addition, the post-idiom regions were not the same across conditions, thus processing differences owing to linguistic variability are not unlikely. Nevertheless, to explore whether idiom overactivation in our study might have caused integration problems and hence interference outside the idiom region, we conducted some post-hoc analyses in the pre- and post-idiom regions. Importantly, in our study, these regions were identical for both idioms and literal control phrases. We looked at total reading time, as this measure also includes time spent in the region as a result of (regressive) refixations. Model specifications and predictors were the same as in the main analysis. Again, we found no effect of Condition in either the pre- ($\beta = 0.007$, $SE = 0.02$, $t = 0.36$, $p = 0.71$) or post-idiom regions ($\beta = 0.02$, $SE = 0.03$, $t = 0.65$, $p = 0.52$). In line with our main findings, therefore, we found no evidence of idiom (over)activation in other regions.

One could argue that the lack of an idiom effect, and, in particular, the lack of an idiom-processing cost could be due to high L2 proficiency, a factor believed to promote more

nativelike processing of formulaic sequences (Beck and Weber 2016a; Carrol et al. 2016; Cieślicka and Heredia 2017; Conklin and Schmitt 2008). Our participants were indeed highly proficient in English (as were the participants in Milburn et al.'s study). However, assuming the L1 idiom advantage holds, then more nativelike processing should have manifested in facilitated processing for L2 idioms, meaning that idioms should have been read faster than controls despite being used in their literal sense. This was not the case, lending further support to models of L2 idiom processing, according to which, literal meaning takes precedence (Abel 2003; Cieślicka 2006; Liontas 2002). Of course, some activation of figurative meanings cannot be entirely precluded, but while it is possible that our participants recognised and activated idioms to a certain extent, the results do not suggest idiom overactivation or overextension in inappropriate contexts. Therefore, idiom status alone does not automatically lead to overextension of L2 figurative meanings.

5. Conclusions

The present study has demonstrated that advanced L2 users do not (over)activate figurative meanings when a literal interpretation suffices. Unlike Milburn et al.'s (2021) findings, therefore, the present data do not support the idea that advanced L2 users overcome the literal bias and start overextending idiom use in inappropriate contexts. In addition, L2 proficiency did not differentially affect the processing of idioms in a way that would suggest enhanced suppression skills for nontarget meanings. These results consolidate the previously attested primacy of literal meanings in L2 idiom processing. Future studies should investigate the potential implications of more varied L2 proficiency levels, as well as consider influences from idiom-specific properties, and especially literalness and its interaction with contextual cues.

To our knowledge, our study is the first comparing the processing of potentially idiomatic strings to matched control phrases without having a 'figurative condition' in the study design: i.e., a condition where the figurative meaning of (literally plausible) idioms is explicitly biased. While this has not been the main focus of the present study, future studies could investigate the hypothesis that overexposure to figurative meanings in experimental tasks directly impacts processing of L1 and L2 idioms, and whether speeded idiom recognition is truly accompanied by facilitated integration in context. Such data could provide us with new insights into how generalisable the (internally largely consistent) findings on idiom processing are to nonexperimental contexts, where figurative meanings are considerably less frequent and potentially more disruptive.

Appendix A

	Final Word Region								Idiom Region												
	First Fixation Duration				First Pass Reading Time				Total Reading Time				Spillover				Regression (in) Likelihood			
Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*
(Intercept)	5.4	0.02	262.25	**<0.001**	6.12	0.05	118.78	**<0.001**	6.46	0.05	133.32	**<0.001**	5.44	0.03	206.74	**<0.001**	−1.15	0.22	−5.17	**<0.001**
Condition [LIT-LIT]	−0.01	0.02	−0.33	0.739	0.02	0.03	0.69	0.488	−0.02	0.04	−0.41	0.681	0.02	0.02	0.77	0.443	−0.3	0.29	−1.03	0.302
LexTALE score	0	0	−1.15	0.25	−0.01	0	−4.08	**<0.001**	−0.01	0	−2.3	**0.021**	0	0	−2.68	**0.007**	0.05	0.01	3.34	**0.001**
Trial index	0	0	−1.23	0.219	0	0	−0.56	0.575	0	0	1.71	0.088	0	0	−1.59	0.112	0	0.01	0.34	0.734
condition [LIT-LIT] × LexTALE score	0	0	0.09	0.926	0	0	−0.44	0.661	0	0	−1.06	0.288	0	0	1.32	0.188	0.01	0.02	0.31	0.753
condition [LIT-LIT] × Trial index	0	0	0.7	0.482	0	0	1.6	0.109	0	0	−0.35	0.728	0	0	1.31	0.191	0	0.01	−0.2	0.845
Random Effects																				
σ^2	0.08				0.19				0.16				0.08				3.29			
τ_{00}	$0.01_{participant}$				$0.05_{participant}$				$0.06_{participant}$				$0.01_{participant}$				$0.78_{participant}$			
	0.00_{item}				0.02_{item}				0.02_{item}				0.00_{item}				0.35_{item}			
τ_{11}					$0.01_{participant.conditionLIT\text{-}LIT}$				$0.01_{participant.conditionLIT\text{-}LIT}$								$0.21_{participant.conditionLIT\text{-}LIT}$			
					$0.01_{item.conditionLIT\text{-}LIT}$				$0.02_{item.conditionLIT\text{-}LIT}$								$0.87_{item.conditionLIT\text{-}LIT}$			
ρ_{01}					$0.16_{participant}$				$-0.05_{participant}$								$-0.07_{participant}$			
					-0.72_{item}				-0.81_{item}								-0.78_{item}			
ICC	0.1				0.29				0.33				0.16				0.27			
N	$50_{participant}$				$50_{participant}$				$50_{participant}$				$50_{participant}$				$50_{participant}$			
	20_{item}				20_{item}				20_{item}				20_{item}				20_{item}			
Observations	793				976				976				621				976			

Note: Values in bold denote significant effect with *p < 0.005*.

Notes

[1]   Of note, figurative attunement in Beck and Weber (2016b) was defined as the experimental condition in which proportionally more idiomatic sequences were encountered as opposed to literal phrases. In a cross-modal priming task, the researchers found that L2 users became faster at responding to target words associated with an idiom's figurative meaning if they were exposed to a greater number of idioms in the stimuli. In our study the number of idioms did not fluctuate across conditions. However, we expected that if idioms were never explicitly used in their figurative sense, this should minimise activation, and hence, the priming of nonliteral meanings.

[2]   Originally, we also planned to analyse skipping likelihood for the final words (i.e., how likely it is for the final word to be skipped during first pass reading) as previous studies have shown increased likelihood for skipping the final words of (L1) idioms as opposed to control phrases (Carrol and Conklin 2017, 2020). In the present study final words in idioms and in control phrases were skipped at a similar rate (N_{ID-LIT} = 132, $N_{LIT-LIT}$ = 134), likely because we controlled for (both contextual and transitional) probability, something not always accounted for in previous studies (Kyriacou et al. 2021). We, therefore, did not investigate skipping likelihood further.

References

Abel, Beate. 2003. English idioms in the first language and second language lexicon: A dual representation approach. *Second Language Research* 19: 329–58. [CrossRef]

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv* arXiv:1406.5823.

Beck, Sara D., and Andrea Weber. 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. [CrossRef]

Beck, Sara D., and Andrea Weber. 2016b. L2 Idiom Processing: Figurative Attunement in Highly Idiomatic Contexts. Paper presented at the 38th Annual Conference of the Cognitive Science Society, Philadelphia, PA, USA, August 10–13; pp. 1817–22.

Beck, Sara D., and Andrea Weber. 2019. Context matters, figuratively, for L2 readers: Evidence from self-paced reading. Paper presented at the 11th International Conference on the Mental Lexicon, Edmonton, AB, Canada, September 25–28; Edmonton: University of Alberta Libraries.

Beck, Sara D., and Andrea Weber. 2020. Context and Literality in Idiom Processing: Evidence from Self-Paced Reading. *Journal of Psycholinguistic Research* 49: 837–63. [CrossRef] [PubMed]

Boulenger, Véronique, Olaf Hauk, and Friedemann Pulvermüller. 2008. Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex* 19: 1905–14. [CrossRef]

Boulenger, Véronique, Yury Shtyrov, and Friedemann Pulvermüller. 2012. When do you grasp the idea? MEG evidence for instantaneous idiom understanding. *Neuroimage* 59: 3502–13. [CrossRef]

Cacciari, Cristina, and Paola Corradini. 2015. Literal analysis and idiom retrieval in ambiguous idioms processing: A reading-time study. *Journal of Cognitive Psychology* 27: 797–811. [CrossRef]

Cacciari, Cristina, and Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27: 668–83. [CrossRef]

Cacciari, Cristina, Roberto Padovani, and Paola Corradini. 2007. Exploring the relationship between individuals' speed of processing and their comprehension of spoken idioms. *European Journal of Cognitive Psychology* 19: 417–45. [CrossRef]

Carrol, Gareth, and Kathy Conklin. 2014. Eye-tracking multi-word units: Some methodological questions. *Journal of Eye Movement Research* 7: 1–11. [CrossRef]

Carrol, Gareth, and Kathy Conklin. 2017. Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs'. *Bilingualism: Language and Cognition* 20: 299–317. [CrossRef]

Carrol, Gareth, and Kathy Conklin. 2020. Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Language and Speech* 63: 95–122. [CrossRef]

Carrol, Gareth, Jeannette Littlemore, and Margaret Gillon Dowens. 2018. Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua* 204: 21–44. [CrossRef]

Carrol, Gareth, Kathy Conklin, and Henrik Gyllstad. 2016. Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition* 38: 403–43. [CrossRef]

Cieślicka, Anna. 2006. Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research* 22: 115–44. [CrossRef]

Cieślicka, Anna, and Roberto R. Heredia. 2017. How to save your skin when processing L2 idioms: An eye movement analysis of idiom transparency and cross-language similarity among bilinguals. *Iranian Journal of Language Teaching Research* 5: 81–107.

Cieślicka, Anna, Roberto R. Heredia, and Marc Olivares. 2014. It's all in the eyes: How language dominance, salience, and context affect eye movements during idiomatic language processing. In *Essential Topics in Applied Linguistics and Multilingualism*. Berlin/Heidelberg: Springer, pp. 21–41.

Conklin, Kathy. 2019. Processing single-word and multiword items. In *The Routledge Handbook of Vocabulary Studies*. London: Routledge, pp. 174–88.

Conklin, Kathy, and Norbert Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29: 72–89. [CrossRef]

Davies, Mark. 2009. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14: 159–90. [CrossRef]

Du, Lingli, Irina Elgort, and Anna Siyanova-Chanturia. 2023. Cross-language influences in the processing of L2 multi-word expressions. *Cross-Language Influences in Bilingual Processing and Second Language Acquisition* 16: 187.

Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–88. [CrossRef]

Erman, Britt, and Beatrice Cecilia Warren. 2000. The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse* 20: 29–62. [CrossRef]

Fanari, Rachele, Cristina Cacciari, and Patrizia Tabossi. 2010. The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology* 22: 321–34. [CrossRef]

Foster, Pauline. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. London: Routledge, pp. 75–97.

Hubers, Ferdy, Catia Cucchiarini, and Helmer Strik. 2020. Second language learner intuitions of idiom properties: What do they tell us about L2 idiom knowledge and acquisition? *Lingua* 246: 102940. [CrossRef]

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–46. [CrossRef]

Kyriacou, Marianna, Kathy Conklin, and Dominic Thompson. 2021. When the Idiom Advantage Comes Up Short: Eye-Tracking Canonical and Modified Idioms. *Frontiers in Psychology* 12: 675046. [CrossRef] [PubMed]

Lauro, Leonor J. Romero, Marco Tettamanti, Stefano F. Cappa, and Costanza Papagno. 2007. Idiom comprehension: A prefrontal task? *Cerebral Cortex* 18: 162–70. [CrossRef]

Lemhöfer, Kristin, and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44: 325–43. [CrossRef] [PubMed]

Lenth, R. 2018. *Emmeans: Estimated Marginal Means, Aka Least-Square Means*. R Package Version 1.2. Seoul: Scinapse.

Libben, Maya R., and Debra A. Titone. 2008. The multidetermined nature of idiom processing. *Memory & Cognition* 36: 1103–21.

Liontas, John. 2002. Context and idiom understanding in second languages. *EUROSLA Yearbook* 2: 155–85. [CrossRef]

Mancuso, Azzurra, Annibale Elia, Alessandro Laudanna, and Simonetta Vietri. 2020. The Role of Syntactic Variability and Literal Interpretation Plausibility in Idiom Comprehension. *Journal of Psycholinguistic Research* 49: 99–124. [CrossRef]

Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50: 940–67. [CrossRef]

McMurray, Bob. 2023. I'm not sure that curve means what you think it means: Toward a [more] realistic understanding of the role of eye-movement generation in the Visual World Paradigm. *Psychonomic Bulletin & Review* 30: 102–46.

Milburn, Evelyn, Mila Vulchanova, and Valentin Vulchanov. 2021. Collocational frequency and context effects on idiom processing in advanced L2 speakers. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 75: 169. [CrossRef] [PubMed]

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Oxford University Press.

Pulido, Manuel F. 2021. Native language inhibition predicts more successful second language learning: Evidence of two ERP pathways during learning. *Neuropsychologia* 152: 107732. [CrossRef] [PubMed]

Pulido, Manuel F. 2022. Why are multiword units hard to acquire for late L2 learners? Insights from cognitive science on adult learning, processing, and retrieval. *Linguistics Vanguard* 8: 237–47. [CrossRef]

Pulido, Manuel F., and Paola E. Dussias. 2020. Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism: Language and Cognition* 23: 652–67. [CrossRef]

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team. Available online: http://www.r-project.org/ (accessed on 10 December 2023).

Rommers, Joost, Ton Dijkstra, and Marcel Bastiaansen. 2013. Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience* 25: 762–76. [CrossRef]

Schmitt, Norbert. 2012. Formulaic language and collocation. *The Encyclopedia of Applied Linguistics*. [CrossRef]

Senaldi, Marco S. G., and Debra A. Titone. 2022. Less Direct, More Analytical: Eye-Movement Measures of L2 Idiom Reading. *Languages* 7: 91. [CrossRef]

Senaldi, Marco S. G., Junyan Wei, Jason W. Gullifer, and Debra Titone. 2022. Scratching your tête over language-switched idioms: Evidence from eye-movement measures of reading. *Memory & Cognition* 50: 1230–56.

Siyanova-Chanturia, Anna, and Ron Martinez. 2015. The idiom principle revisited. *Applied Linguistics* 36: 549–69. [CrossRef]

Siyanova-Chanturia, Anna, Kathy Conklin, and Norbert Schmitt. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27: 251–72. [CrossRef]

Tabossi, Patrizia, Rachele Fanari, and Kinou Wolf. 2009. Why are idioms recognized fast? *Memory & Cognition* 37: 529–40.

Titone, Debra A., and Cynthia M. Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31: 1655–74. [CrossRef]

Titone, Debra A., and Maya R. Libben. 2014. Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon* 9: 473–96. [CrossRef]

Underwood, Geoffrey, Norbert Schmitt, and Adam Galpin. 2004. The eyes have it. In *Formulaic Sequences: Acquisition, Processing, and Use*. Edited by Norbert Schmitt. Amsterdam: John Benjamins Publishing Company, pp. 153–72.

Vespignani, Francesco, Paolo Canal, Nicola Molinaro, Sergio Fonda, and Cristina Cacciari. 2010. Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience* 22: 1682–700. [CrossRef] [PubMed]

Wray, Alison, and Michael R. Perkins. 2000. The functions of formulaic language: An integrated model. *Language & Communication* 20: 1–28.

Zempleni, Monika-Zita, Marco Haverkort, Remco Renken, and Laurie A. Stowe. 2007. Evidence for bilateral involvement in idiom comprehension: An fMRI study. *Neuroimage* 34: 1280–91. [CrossRef]