Article

Less Direct, More Analytical: Eye-Movement Measures of L2 Idiom Reading

Marco S. G. Senaldi* and Debra A. Titone*

Department of Psychology, McGill University, Montréal, QC H3A 1G1, Canada
*Correspondence: marco.senaldi@mcgill.ca (M.S.G.S.); debra.titone@mcgill.ca (D.A.T.)

Abstract: Idioms (e.g., break the ice, spill the beans) are ubiquitous multiword units that are often semantically non-compositional. Psycholinguistic data suggests that L1 readers process idioms in a hybrid fashion, with early comprehension facilitated by direct retrieval, and later comprehension inhibited by factors promoting compositional parsing (e.g., semantic decomposability). In two eye-tracking experiments, we investigated the role of direct retrieval and compositional analysis when idioms are read naturally in sentences in an L2. Thus, French–English bilingual adults with French as their L1 were tested using English sentences. For idioms in canonical form, Experiment 1 showed that prospective verb-related decomposability and retrospective noun-related decomposability guided L2 readers towards bottom-up figurative meaning access over different time courses. Direct retrieval played a lesser role, and was mediated by the availability of a congruent “cognate” idiom in the readers’ L1. Next, Experiment 2 included idioms where direct retrieval was disrupted by a phrase-final language switch into French (e.g., break the glace, spill the fèves). Switched idioms were read comparably to switched literal phrases at early stages, but were penalized at later stages. These results collectively suggest that L2 idiom processing is mostly compositional, with direct retrieval playing a lesser role in figurative meaning comprehension.

Keywords: idioms; multiword expressions; L2 processing; bilingualism; eye-tracking; reading; language switching; cross-language effects

1. Introduction

When speakers acquire a new language, one key to reaching “native-like” proficiency is in mastering its idiosyncratic word-combination patterns (Boers et al. 2006; Catbenton and Segalowitz 1988; Hoey 2005; Lewis 2000; Nattinger and DeCarrico 1992; Richards and Rodgers 2014; Segalowitz 2010; Wood 2006). Even an advanced learner of English who has a solid grasp of both the lexicon and the grammar might occasionally produce combinations such as make a step, or close a lack, when speaking or writing. These phrases are understandable; however, they might sound odd to a first-language speaker, who would more likely say take a step or close a gap to convey the same ideas (Altenberg and Granger 2001; Nesselhauf 2003).

While traditional linguistic accounts focused on the generative nature of human syntax (Chomsky 1980; Fraser 1970), the increasing availability of text corpora over the last few decades has made it clear that much of our linguistic production is routinized and prefabricated (Christiansen and Arnon 2017; Erman and Warren 2000; Jackendoff 1995, 1997, 2003; Nattinger and DeCarrico 1992; Pawley and Syder 1983; Sinclair 1991; Van Lancker Sildits and Rallon 2004). Indeed, conventional word combinations (referred to as formulatic units, phraseological units or multiword expressions across different theoretical traditions) are pervasive in everyday life (Arnon and Snider 2010; Bannard and Matthews 2008; Columbus 2010, 2013; Siyanova-Chanturia et al. 2011a; Siyanova and Schmitt 2007; Tremblay and Baayen 2010; Tremblay and Tucker 2011; Tremblay et al. 2011; Wray 1999, 2013; Wulff 2008). They also tend to exhibit some degree of formal rigidity and semantic
idiosyncrasy that makes them especially challenging for L2 learners (Alexander 1987; Boers 2000; Cacciari 1993; Charteris-Black 2002; Fernando 1996; Howarth 1998; Irujo 1986, 1993; Kovecses and Szabó 1996; Littlemore 2001; Lattey 1986; Moon 1997; Siyanova-Chanturia et al. 2011b; Sonbul 2015; Yorio 1989; Zughoul 1991).

Idiomatic expressions are one prototypical and highly studied linguistic element that is firmly situated within the general class of multiword strings (e.g., *kick the bucket*, *save your skin*, *blow a fuse*). Idioms are usually defined as semantically non-compositional linguistic entities whose figurative meaning extends the meanings of their component words, though we know that they vary in many ways (Abel 2003; Cacciari 2014; Cacciari and Glucksberg 1991; Carrol 2021; Carrol and Conklin 2017; Gibbs et al. 1989; Libben and Titone 2008; Nunberg 1978; Siyanova-Chanturia et al. 2011a; Titone and Libben 2014). Idioms tend to be recognized and processed as familiar expressions by L1 speakers. However, it remains an open question whether they entail a different and more demanding processing strategy when encountered in a second language.

1.1. Idiom Representation and Processing in L1 Users

Studies of idiom comprehension conducted with L1 speakers have assessed if idioms are directly retrieved from memory (Bobrow and Bell 1973; Cacciari and Tabossi 1988; Swinney and Cutler 1979), composed incrementally during processing (Gibbs and Nayak 1989; Gibbs et al. 1989; Nunberg et al. 1994), or a combination of both (Caillies and Butcher 2007; Cutting and Bock 1997; Libben and Titone 2008; Smolka et al. 2007; Sprenger et al. 2006; Titone and Connine 1999). In line with the last hypothesis, comprehension data from cross-modal priming (Titone and Libben 2014) and eye-tracking (Titone et al. 2019) support a hybrid and multidetermined model, according to which speakers use both direct retrieval and compositional parsing over different time courses in idiom processing. Direct retrieval is mediated by item-level variables such as familiarity, which measures the subjective frequency with which speakers encounter a given idiom string in written or spoken form, regardless of whether they are confident about its meaning (Cronk and Schweigert 1992; Titone and Connine 1994). On the other hand, compositional parsing is modulated by variables like semantic decomposability, which refers to whether the figurative meaning of an idiom as a whole is semantically related to the meanings of its component words (e.g., *pop the question* vs. *kick the bucket*; Gibbs and Nayak 1989; Nunberg et al. 1994).

1.2. Direct Retrieval and Compositional Parsing in L1 Idiom Reading

One previous contribution examined multiple dimensions of idioms in one go and paved the way for the present study; thus, we detail it here. Titone et al. (2019) used eye-tracking during reading to assess how direct retrieval and compositional analysis interact in L1 idiom processing. Specifically, L1 speakers of English silently read English sentences containing idioms followed by a figurative context (Idiomatic–Idiomatic; e.g., *Dolan spilled the beans when he mentioned the surprise party to his friend*), idioms followed by a literal context (Idiomatic–Literal; e.g., *Dolan spilled the beans when he tried to pour too many into the soup pot*) and matched novel phrases followed by a literal disambiguating region (e.g., *Dolan cooked the beans before he started adding vegetables to the soup pot*) as their eye movements were recorded.

Early eye-movement measures of reading showed a processing advantage of idioms over matched literal phrases, a finding previously seen in a vast body of L1 comprehension studies (Cacciari and Tabossi 1988; Carrol 2021; Carrol and Conklin 2020; Carrol and Littlemore 2020; Conklin and Schmitt 2008; Cronk and Schweigert 1992; Giora 1997; Libben and Titone 2008; McGlone et al. 1994; Siyanova-Chanturia and Van Lancker-Sidtis 2018; Siyanova-Chanturia et al. 2011a; Tabossi et al. 2009b; Underwood et al. 2004). This early idiom superiority effect suggests that at a first stage of comprehension L1 speakers can recognize idioms prior to phrase offset by virtue of their conventional nature and retrieve them directly from memory.
Focusing on reading times of the post-idiom disambiguating region, idioms followed by a literal continuation were disadvantaged compared to idioms followed by a figurative context, especially when familiarity was high and decomposability was low. This suggests that when idioms were more prone to direct retrieval and less compositionally analyzable, L1 readers were strongly committed to a figurative reading of the phrase and were thus garden-pated when they encountered a literal continuation. However, when decomposability was higher, the processing cost of the Idiomatic–Literal condition was attenuated, as more decomposable idioms were less likely to be interpreted figuratively on the first pass. As further confirmation, at the level of idiom total reading time, higher familiarity made fully idiomatic sentences (Idiomatic–Idiomatic) faster to read than fully literal ones, in particular when decomposability was low (thus rendering the idioms more “word-like”). Overall, during later processing stages speakers appeared to engage in a word-by-word reanalysis of idioms’ internal semantics. This process was made harder by increased decomposability, which created greater ambiguity between a literal and a figurative interpretation of the phrase being processed (e.g., a highly decomposable idiom like get the message was harder to reanalyze than a low-decomposable idiom like have a lark).

To clarify which idiom component drove this decomposability interference effect, in the same study, Titone et al. (2019) separately investigated the role of verb-related and noun-related decomposability (Bortfeld 2003; Keysar and Bly 1995; Hamblin and Gibbs 1999; Libben and Titone 2008). Verb relatedness was expected to guide idiom comprehension prospectively and can be exemplified by an idiom like save your skin, where the verb save independently contributes to the figurative phrasal meaning as compared to kick in kick the bucket. By the same token, noun relatedness was expected to drive idioms’ semantic integration retrospectively. This is exemplified by an idiom like pop the question, where the noun question shares more semantic features with the figurative meaning of the phrase than, for instance, bucket in kick the bucket. In Titone et al. (2019), L1 speakers’ comprehension was especially inhibited by increased prospective decomposability. This may have occurred because before reaching the idiom noun, participants presumably activated the literal meaning of the verb, but once they recognized the idiom at phrase offset, the contextually inappropriate literal meaning of the verb was more challenging to suppress when it was still partially related to the contextually appropriate figurative meaning of the phrase.

In sum, results from Titone et al. (2019) suggested that first-language speakers generally employ early-stage direct form retrieval when comprehending idioms in context, and reanalyze them compositionally later on. This leads to an interfering effect of decomposability, in that the word-by-word reanalysis of a decomposable idiom generates ambiguity between a literal and a figurative interpretation of the phrase.

Building from this study, an underexplored question that we will try to address in the present study is how idioms are read and comprehended in a second language, and which role direct retrieval and compositional parsing play in the process. Crucially, given that second-language speakers tend to be less familiar with idiom forms in their L2, they are expected to rely less on direct retrieval and to prefer a compositional strategy when encountering idioms in a sentence.

1.3. Previous Evidence on Idiom Reading in the L2

In contrast to L1 processing, the semantic and structural peculiarity of idioms leads to processing challenges for L2 comprehenders, especially for late L2 readers (Abel 2003; Bortfeld 1997; Charteris-Black 2002; Cie´slicka 2006; Conklin and Schmitt 2008; Ellis et al. 2008; Eskildsen 2009; Irujo 1993; Jiang and Nekrasova 2007; Laufer 2000; Li and Schmitt 2009; Matlock and Heredia 2002; Rossiter et al. 2010; Steinel et al. 2007; Tabossi et al. 2009a; Van Lancker Sidtis 2003; Weinert 1995; Wood 2006; Wray 2000). A hybrid model might account for this idiom inhibition in the L2 as a kind of reverse familiarity effect that blocks direct retrieval. In other words, speakers are presumably less exposed to idiomatic structures in a second language, and thus possess weaker whole-phrase representations of L2 idioms in their mental lexicon. Consequently, it will be harder for L2 users to access idioms directly
and holistically in memory during language processing, leaving incremental parsing as the preferred strategy to pursue. However, the relative paucity of studies that address idiom processing from a second-language perspective makes it difficult to paint a consistent picture, which is also exacerbated by differences in experimental design, methodology and individual differences across studies in L2 abilities.

Focusing on reading data, Underwood et al. (2004) tracked the eye movements of L1 and L2 speakers of English while reading formulaic (including idioms) and non-formulaic sequences embedded into short paragraphs. While L1 speakers fixated on the terminal word of formulaic sequences for a shorter amount of time with respect to non-formulaic strings, the advantage registered for L2 speakers was only in terms of fixation count and not fixation duration. Eye-tracking evidence from Siyanova-Chanturia et al. (2011a) showed that, unlike L1 speakers, L2 speakers read idioms comparably to matched novel phrases, and that figurative meanings of idioms were harder to retrieve than literal ones. In Spanish-dominant bilinguals, Cieslicka et al. (2014) found that the reading of English idioms was facilitated when they were intended literally in a neutral context and intended figuratively in a figurative context. By contrast, in a self-paced reading task Conklin and Schmitt (2008) detected a processing advantage for idioms, used both figuratively and literally, over matched literals in both L1 English speakers and proficient L2 speakers.

Evidence from other methodologies, such as cross-modal or visual-visual priming, also suggests that L2 speakers adopt a slower literal and compositional strategy when processing idioms in comparison to L1 speakers (Beck and Weber 2016a; Cieslicka 2006; Cieslicka et al. 2021; but see van Ginkel and Dijkstra 2020). As stated above, speakers' overall lower familiarity with idiomatic structures in their L2 prevents them from identifying idioms at first glance and retrieving them as a whole from the lexicon. Thus, they initially commit to a slower literal interpretation of the target phrase and revise it only later on when integrating its semantics with the sentential context.

1.4. Cross-Language Overlap as a Modulator of L2 Direct Retrieval

Although L2 speakers rely primarily on compositional processing when comprehending idioms, direct retrieval could potentially come into play at a later stage after speakers have identified the phrase as idiomatic. While familiarity mediates direct retrieval in the L1 (Carrol and Littlemore 2020; Cronk and Schweigert 1992; Libben and Titone 2008; Titone and Libben 2014; Titone et al. 2019), it may not be as accurate a measure for second-language speakers, as it is likely correlated with factors such as comprehender proficiency and exposure. A more relevant determinant of bilingual idiom processing may instead be cross-language overlap, which refers to whether an idiom possesses translational equivalents across the languages spoken by a subject. For example, a French–English bilingual speaker reading the idiom *play with fire* in their L2-English might benefit from the availability of the congruent idiom *jouer avec le feu* in their L1-French. Previous research has indeed shown facilitation for such “cognate” idioms that were dually represented across both languages spoken by participants in off-line comprehension and production (Irujo 1986, 1993; Laufer 2000; Charteris-Black 2002), on-line production (Liontas 2002) and on-line comprehension (Carrol et al. 2016). For example, Titone et al. (2015) collected speeded meaningfulness judgments from English–French bilinguals for English idioms presented with their final noun intact (e.g., *spill the beans*) or switched to French (e.g., *spill the fèves*). Overlapping idioms were less disrupted by the presence of a phrase-final language switch, which generally penalized idioms over literals. Moreover, Pritchett et al. (2016) showed L2 idioms with an L1 counterpart to be more efficiently recalled from memory. These results dovetail nicely with cognate facilitation effects that have been replicated time and again at the single-word level in bilingual readers (e.g., Gullifer et al. 2013; Libben and Titone 2009; Pivneva et al. 2014; Titone et al. 2011; Van Hell and Dijkstra 2002), a point to which we later return in the General Discussion.
1.5. The Current Study

The goal of the current eye-tracking reading study was to extend the findings of Titone et al. (2019) to idiom reading in the L2. More specifically, the research questions pursued here concerned how L2 readers process idiomatic expressions during natural sentences reading, and to what extent the predictions of a hybrid model (Libben and Titone 2008; Titone and Libben 2014) apply to L2 idiom comprehension.

Following the methods of Titone et al. (2019), and using similar materials, French–English bilingual adults identifying French as their L1 silently read English sentences containing idioms used figuratively, idioms used literally and matched literal control phrases, while their eye movements were recorded. All idioms had a verb–determiner–noun structure (e.g., spill the beans). Literal controls were composed of the same idiom noun and a matched novel verb (e.g., cook the beans). Idiomatic and literal phrases were placed in sentence-initial position and followed by a disambiguating context.

According to a hybrid model (Libben and Titone 2008; Titone and Libben 2014), speakers use both direct retrieval and compositional analysis when making sense of idioms in context. Given that L2 readers benefit from early direct retrieval to a lesser extent, we predict they will engage in a preferential compositional mode when processing idioms. Given this, prospective and retrospective effects of decomposability may also play a differential role over time, with prospective verb relatedness guiding early-stage idiom identification and retrospective noun relatedness facilitating late-stage idiom integration. Given that direct retrieval should take place only after readers have correctly identified the phrase being read as idiomatic, we also hypothesized direct retrieval to play a secondary role in the form of a late facilitating effect of cross-language overlap.

To the best of our knowledge, the interaction of semantic transparency and cross-language overlap in L2 idiom comprehension has so far only been addressed in an eye-tracking study by Cieślicka and Heredia (2017; for a related study on collocations, see Yamashita 2018). There, English- and Spanish-dominant bilinguals read idiomatic sentences in English. Overlapping opaque idioms were processed more easily by Spanish-dominant bilinguals when used in their literal sense, while overlapping transparent idioms were read significantly slower when used literally. In contrast with their study, disambiguating contexts will be here placed only after the idiomatic region, so that we can separately investigate how idiom phrases are processed on the first pass without any contextual support, and how they are re-analyzed after the biasing region has been read.

2. Experiment 1

2.1. Materials and Methods

2.1.1. Participants

A total of 37 French–English speakers (Age $M = 20.68$, $SD = 3.96$; 31 female, 6 male) who identified French as their L1 were recruited from McGill University and the Montreal community in exchange for course credit or compensation of 10 CAD/h. All participants gave informed consent, and the study was approved by the Research Ethics Board Office of McGill University. They were asked to complete a language history questionnaire modelled after the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al. 2007). Here, subjects reported information related to their relative exposure to French and English, and their self-rated abilities for French and English. Relative exposure to the L2 was measured based on the question “Please list what percentage of time you are currently and on average exposure to each language”. Participants were generally proficient in English (their L2) as indicated in Table 1. Complete data on both L1 and L2 proficiency are reported in the Appendix A. All participants had normal or corrected-to-normal vision and no self-reported history of speech or hearing disorders.
Table 1. Participant characteristics (Experiment 1). Self-ratings were collected on a 1–10 scale (/10).

| Measure                        | Mean  | SD    |
|--------------------------------|-------|-------|
| Demographics                   |       |       |
| Chronological age              | 20.68 | 3.96  |
| English self-ratings (/10)     |       |       |
| Speaking                       | 7.97  | 2.01  |
| Reading                        | 8.78  | 1.29  |
| Writing                        | 8.00  | 1.94  |
| Translating                    | 7.94  | 1.95  |
| Listening                      | 8.76  | 1.39  |
| Pronunciation                  | 7.36  | 2.42  |
| Fluency                        | 7.82  | 2.27  |
| Vocabulary                     | 7.72  | 2.11  |
| Grammar                        | 7.82  | 2.20  |
| Overall competence             | 8.01  | 1.96  |
| Language exposure (%)          |       |       |
| English                        | 46.70 | 18.19 |
| French                         | 49.95 | 19.27 |

2.1.2. Stimuli

Stimuli were identical to those used in Titone et al. (2019). A total of 60 English idioms with a verb-article-noun structure (e.g., Dolan spilled the beans) were selected from Libben and Titone’s (2008) norms. Literal control phrases were generated by replacing the idiom verb with a novel verb (e.g., Dolan cooked the beans) that was matched in character length (+/−1 character) and word frequency (+/−5 counts per million, Kučera and Francis 1967).

Each idiom was embedded in one of three sentence conditions: an idiom followed by a figuratively biased disambiguating region (Idiomatic–Idiomatic: Dolan spilled the beans when he mentioned the surprise party to his friend) or an idiom followed by a literally biased disambiguating region (Idiomatic–Literal: Dolan spilled the beans when he tried to pour too many into the soup pot). Non-idiomatic phrases were also embedded in sentences where a supporting context followed (Literal–Literal: Dolan cooked the beans before he started adding vegetables to the soup pot). All idioms were rated as highly literally plausible (M = 3.87; SD = 0.77) in Libben and Titone’s (2008) norms, so we could build both an idiomatic and a literal disambiguating context for each. This design permitted us to examine idiom recognition and the factors modulating comprehension at two critical time points, namely prior to and during the disambiguation region. Table 2 presents example sentences from each condition.

Table 2. Example sentences from Idiomatic–Idiomatic (Id-Id), Idiomatic–Literal (Id-Lit), and Literal–Literal (Lit-Lit) conditions (Experiment 1).

| Condition | Sentence                                                                 |
|-----------|--------------------------------------------------------------------------|
| Id-Id     | Dolan spilled the beans when he mentioned the surprise party to his friend.|
| Id-Lit    | Dolan spilled the beans when he tried to pour too many into the soup pot.|
| Lit-Lit   | Dolan cooked the beans before he started adding vegetables to the soup pot.|

Ratings of verb relatedness, noun relatedness and literal plausibility for each idiom were also obtained from Libben and Titone’s (2008) norms. In the idiom norms, verb and noun relatedness ratings reflected whether the verb or the noun independently contributed to the overall idiom meaning, while literal plausibility reflected an idiom’s potential to have a literal interpretation. Both verb/noun relatedness and literal plausibility were assessed on a 1–5 Likert scale. Unlike in Titone et al. (2019), familiarity ratings were collected from the same French–English speakers who took part in the present study, to ensure that the idioms in our dataset were known to the participants. At the end of the experiment, speakers were presented with the 60 idioms tested and were instructed to rate their familiarity with each
idiom on a three-point Likert-scale (1 = Unfamiliar with idiom; 2 = Know idiom, but not confident in defining it; 3 = Familiar with idiom, confident in defining it). Participants were also asked whether each idiom, in their opinion, possessed a translational equivalent in French and, if so, what that French counterpart would be (e.g., the English idiom *break the ice* overlaps with the French *briser la glace*). An average cross-language overlap score was then computed for each idiom, indicating the proportion of speakers who rated that idiom as overlapping with a corresponding French one. Table 3 reports the average normative characteristics of our idioms and stimuli.

**Table 3.** Item and stimuli characteristics (Experiment 1).

| Variable                  | Condition                        | Measurement                          | Min   | Max   | Mean  | SD    |
|---------------------------|----------------------------------|--------------------------------------|-------|-------|-------|-------|
| Idiom norms               | Familiarity (/3)                  |                                      | 1.00  | 3.00  | 1.99  | 0.81  |
|                           | Verb Relatedness (/5)             |                                      | 0.80  | 4.93  | 2.86  | 1.07  |
|                           | Noun Relatedness (/5)             |                                      | 0.54  | 4.50  | 2.48  | 1.01  |
|                           | Literal Plausibility (/5)         |                                      | 1.50  | 5.00  | 3.87  | 0.77  |
|                           | Cross-language Overlap (/1)       |                                      | 0.00  | 0.92  | 0.28  | 0.24  |
| Noun                      | Length (characters)               |                                      | 3     | 8     | 4.67  | 1.24  |
| Idiom phrase              | Idiom                             | Length (characters)                  | 11    | 20    | 15.78 | 2.16  |
|                           | Literal                           | Length (characters)                  | 11    | 20    | 15.68 | 2.13  |
| Disambiguating region     | Idiomatic–Idiomatic               | Length (characters)                  | 42    | 63    | 54.25 | 4.32  |
|                           | Idiomatic–Literal                 | Length (characters)                  | 43    | 62    | 53.80 | 4.31  |
|                           | Literal–Literal                   | Length (characters)                  | 43    | 61    | 53.42 | 4.15  |

2.1.3. Apparatus

We used an Eye-Link 1000 tower-mounted system (SR-Research, Ottawa, ON, Canada) to record eye movements (1000 Hz sampling rate). Viewing was binocular but eye movements were recorded from the right eye only. Stimuli were presented on a 21” ViewSonic CRT monitor with a screen resolution of 1024 × 768 pixels, using UMass EyeTrack 7.10 software (https://blogs.umass.edu/eyelab/software/, accessed on 23 March 2022). Text was presented on a single line in yellow 10-point Monaco font on a black background. Participants’ eyes were positioned 71 cm from the monitor, and thus 1 degree of visual angle comprised approximately 3 characters of text.

2.1.4. Procedure

Each participant completed a consent form and online language background questionnaire before the eye-tracking session began. Participants were then instructed to read sentences naturally while their eye movements were recorded. Before each trial, participants were instructed to direct their gaze towards a small fixation circle at the center of the screen (calibration point). The experimenter initiated the trial only once the participant’s gaze was fixed on the circle. A fixation square appeared in a location equivalent to that of the sentence-initial word; gazing at the fixation square automatically triggered sentence presentation.

Participants viewed one of three sets of sentences in a fully counterbalanced and randomized fashion. Within each set, each idiom appeared in only one of the three conditions (Idiomatic–Idiomatic, Idiomatic–Literal, Literal–Literal). No participant saw the same idiom more than once. In addition to the 60 experimental sentences, each participant read 19 filler sentences, 10 practice sentences and 22 comprehension questions. The filler and practice sentences were literal sentences between 11 and 17 words in length. Yes/no comprehension questions followed roughly 20% of trials. Participants responded “yes” or “no” to the question by pressing specific buttons on the controller pad.

After the reading task, participants completed a survey where they rated the 60 idioms tested for familiarity and cross-language overlap, as described above.
2.2. Results

Overall accuracy on the comprehension questions was 86%, indicating that participants were attentive throughout the experiment. Prior to analysis, 2.5% of the total fixations were removed due to track loss (e.g., eye blinks) or fixation durations shorter than 80 milliseconds.

We then fitted linear mixed-effects (LME) models to the eye movement data in R (R Core Team 2021) with the `lme4` package (Bates et al. 2015), leaving out zero values (which occurred if readers had skipped a region). Specifically, we analyzed first pass gaze duration of the whole idiom (Idiom First Pass Gaze Duration) to inspect early processing stages, before readers reached the disambiguating region, and total reading time of the whole idiom (Idiom Total Reading Time) to investigate later stages of processing, involving sentence-level semantic integration and reanalysis. As concerns zero values for each of the two measures, the idiom region was skipped on first pass (Idiom First Pass Gaze Duration) on 11 trials out of 2217 total trials, and completely skipped (Idiom Total Reading Time) on only 1 trial.

In a first set of models, condition (categorical), cross-language overlap (continuous), verb relatedness (continuous), and their interaction were fixed effects. In a second set of models, verb relatedness was replaced with noun relatedness (continuous), to separately assess the effect of prospective (verb-related) and retrospective (noun-related) decomposability. Control variables included trial order (continuous) and literal plausibility (continuous). While the former accounted for training or fatigue effects emerging throughout the experimental session, the latter controlled for the effect of an idiom’s potential literal interpretation in driving readers towards or away from a compositional strategy.

When modeling idiom first pass gaze duration, the Idiomatic–Idiomatic and Idiomatic–Literal conditions were merged into a unique idiomatic condition because they were identical prior to the disambiguating region. Condition was thus treated as a two-level factor and sum-coded (literal: −0.5 vs. idiomatic: +0.5). After readers reached the disambiguating region, condition became a three-level categorical variable (Idiomatic–Idiomatic, Idiomatic–Literal, Literal–Literal). Two treatment-coded models were thus fitted for idiom total reading time. In the first model, the Literal–Literal condition acted as baseline and was contrasted against the Idiomatic–Idiomatic condition and Idiomatic–Literal conditions individually. In the second model, the Idiomatic–Idiomatic baseline was contrasted against the Idiomatic–Literal condition (not tested in the first model) and the Literal–Literal condition (already tested in the first model). All continuous predictors were scaled to reduce collinearity. Cross-language overlap was correlated with both verb relatedness ($\rho = 0.37$) and noun relatedness ($\rho = 0.31$), while all remaining correlations among all predictors were lower than the absolute value of 0.20. Subjects and items were set as random intercepts across the models. When we followed the procedure to incorporate a maximally specified random structure (by-subjects and by-items slopes and intercepts; Barr et al. 2013), either the models failed to converge or, when the models did converge and were compared against the intercept-only model via log-likelihood model comparisons, they were not warranted. Moreover, operating without random slopes prevented them from capturing potentially interesting variation due to the individual effects of item-related variables such as verb/noun relatedness and cross-language overlap.

For each model, the estimated coefficient ($\beta$), standard error (SE), t, and p values are reported. We evaluated significance using Satterthwaite approximations implemented in the `lmerTest` package (Kuznetsova et al. 2017).

Table 4 reports means and standard deviations for raw Idiom First Pass Gaze Duration and raw Idiom Total Reading Time in each experimental condition.
Table 4. Subject-aggregated means and standard deviations (in milliseconds) for raw Idiom First Pass Gaze Duration and raw Idiom Total Reading Time (Experiment 1).

| Measure                        | Condition          | Mean   | SD    |
|--------------------------------|--------------------|--------|-------|
| Idiom First Pass Gaze Duration | Idiomatic          | 576.93 | 143.72|
|                                | Literal            | 562.47 | 150.10|
| Idiom Total Reading Time       | Idiomatic–Idiomatic| 893.74 | 259.86|
|                                | Idiomatic–Literal  | 876.30 | 302.11|
|                                | Literal–Literal    | 835.22 | 247.27|

2.2.1. Pre-Disambiguation Effects (Idiom First Pass Gaze Duration)

Inspecting early measures of reading allowed us to understand how idioms were processed by L2 readers at first glance, before reaching a disambiguating context that clarified their figurative or literal meaning.

A significant three-way interaction emerged between verb relatedness, condition, and cross-language overlap ($\beta = -31.46, SE = 15.77, t = -2.00, p = 0.046$), whereas no effects of noun relatedness reached significance ($p > 0.1$). As shown in Figure 1, when verb-mediated prospective decomposability was high, gaze duration for the idiom condition was longer if idioms did not possess an L1 (French) counterpart. By contrast, when verb relatedness was low, increased L1–L2 overlap sped up reading in both the idiomatic and literal condition, with a slight disadvantage for the idiom condition when overlap was maximal.

Thus, increased prospective decomposability facilitated the recognition of the idiomatic form on the first pass relative to the literal condition for those items that were not dually represented across both languages known by the readers. Conversely, when L2 users could not benefit from verb-related decomposability, overlap with an equivalent idiom in the readers’ L1 did not impact the idiomatic condition differently than the literal one unless it was maximal. At this point, idioms incurred a specific slowdown, suggesting idiom form priming with respect to the literal condition. These early results overall suggest a somewhat independent role of compositional parsing and direct retrieval in the first stage of L2 idiom comprehension.

Figure 1. Partial effects plot of first pass gaze duration of the whole idiom region as a function of cross-language overlap (x-axis) and verb relatedness (median split across the left and right panels) for the idiomatic (Id) and literal (Lit) condition. Error shadings reflect ±1 standard error of the mean.
2.2.2. Post-Disambiguation Effects (Idiom Total Reading Time)

Total reading time of the idiom, as a whole, indexes late processing stages that take place after readers have entered the disambiguating region and have been biased towards a figurative or literal interpretation of the target phrase. At this level, a slowdown incurred by the fully idiomatic condition (Idiomatic–Idiomatic) would suggest that second-language readers initially committed to a literal and compositional interpretation of the target idiom and were subsequently garden-pathed by the presence of a figurative context. Once again, it is paramount for our research question to understand how item-level variables modulating direct retrieval and compositional analysis interact at this later stage of comprehension.

Verb-relatedness models returned a significant two-way interaction between condition and cross-language overlap for the Idiomatic–Idiomatic vs. Literal–Literal contrast ($\beta = 67.14$, $SE = 22.18$, $t = 3.03$, $p = 0.002$) and the Idiomatic–Idiomatic vs. Idiomatic–Literal contrast ($\beta = 52.78$, $SE = 22.21$, $t = 2.38$, $p = 0.018$). In noun-relatedness models, the same interaction between condition and cross-language overlap emerged only for the Idiomatic–Idiomatic vs. Literal–Literal contrast ($\beta = -57.32$, $SE = 23.29$, $t = -2.46$, $p = 0.014$). As shown in Figure 2, which refers to verb-relatedness models, readers were slower at reading the fully idiomatic condition (Idiomatic–Idiomatic) with respect to the Idiomatic–Literal and fully literal (Literal–Literal) condition if they could not rely on the direct retrieval of a corresponding French idiom. Conversely, when L2 idioms overlapped with the L1, processing time was fastest for full idiomatic sentences relative to the other conditions.

![Figure 2. Partial effects plot of total reading time of the whole idiom region as a function of cross-language overlap (x-axis) for the Idiomatic–Idiomatic (Id-Id), Idiomatic–Literal (Id-Lit) and Literal–Literal (Lit-Lit) conditions. Error shadings reflect ±1 standard error of the mean.](image)

A two-way interaction between noun relatedness and condition also emerged for the Idiomatic–Idiomatic vs. Literal–Literal contrast ($\beta = -46.65$, $SE = 22.30$, $t = 2.092$, $p = 0.037$) and for the Idiomatic–Idiomatic vs. Idiomatic–Literal contrast ($\beta = 71.23$, $SE = 22.30$, $t = 3.194$, $p = 0.001$). Accordingly, as shown in Figure 3, when idioms were followed by a figurative context (Idiomatic–Idiomatic), they experienced a processing cost with respect to the Literal–Literal and Idiomatic–Idiomatic conditions if retrospective (i.e., noun) decomposability was low. The more the noun contributed independently to the overall meaning of the idiom, the more the idiomatic condition was facilitated with respect to the other two conditions.
In sum, L2 readers processed idiomatic and literal phrases similarly in early measures, unless the semantics of idiom verbs was highly related to idioms’ figurative meaning. In this case, the idiomatic condition became slower to process, likely because comprehenders were left to deliberate between a literal and a verb-driven figurative interpretation of the phrase. More specifically, this early figurative priming affected L2-unique idioms, for which direct retrieval of a corresponding L1 form was not possible and compositional analysis was the only viable option. At this preliminary stage of comprehension, the role of direct retrieval was less, with slower reading times, and thus increased figurative priming, only for those idioms where low verb-related decomposability was low but an equivalent form existed in the readers’ L1 (French).

After L2 readers reached the disambiguating region, the fully idiomatic condition became generally more effortful to reanalyze and integrate at the sentence level, compared to the literal control sentence, presumably because readers had to revise an initial literal interpretation. Nonetheless, increased retrospective noun-based decomposability and the existence of a congruent idiom form in the L1 kicked in to ease this processing load, even to the point of making idiomatic sentences maximally fast to read.

Collectively, reading data from this first experiment suggest that L2 idiom processing is predominantly compositional, with prospective and retrospective decomposability weighing in at different time points to guide L2 idiom recognition and sentence-level semantic integration. Direct retrieval seemingly played a secondary role, and emerged in the form of cross-language overlap facilitation. More specifically, the retrieval of a corresponding idiomatic form from the speakers’ L1 increased idiomatic priming in early reading measures only if prospective decomposability was low, while it sped up idiom processing across the board at a late stage.

To confirm whether idiom direct retrieval is really of secondary importance during L2 reading, we conducted a follow-up eye-tracking experiment, presented below. This follow-up experiment investigated how L2 readers processed idiomatic phrases where the possibility of direct retrieval was disrupted through a language-switching manipulation. Specifically, French–English bilingual adults who were dominant in French read English
sentences in the same conditions as Experiment 1. Idiomatic and matched literal phrases could appear either in their intact form (e.g., *Dolan spilled the beans*) or with their nouns switched to French (e.g., *Dolan spilled the fèves*). The rationale of this manipulation is that momentary language switches alter the original idiomatic form, and thus its direct retrieval, while minimally affecting its meaning and syntax.

Titone et al. (2015) previously collected speeded meaningfulness judgments on the same manipulation from English–French bilinguals who were dominant in either English or French. Switched idioms were found to be judged more slowly than switched literal controls. Also, in a previous eye-tracking study (Senaldi et al. n.d.), we tested the same set of materials as the present work on English–French bilinguals who spoke English as their L1, and found the idiomatic condition to be more impacted than the literal condition by the presence of a language switch in early reading measures. The processing cost of reading switched idioms was nonetheless mitigated by increased cross-language overlap, which allowed readers to retrieve the canonical form of the underlying idiom from their L1. Those results confirmed that direct retrieval is the default processing when idioms are encountered in the L1, and that it is also the preferential repair strategy adopted by the language processing system when recovering the underlying form of modified idioms.

Thus, if L2 speakers are indeed less reliant on direct retrieval, we can expect that phrase-final exogenous switches, which will force readers into a word-by-word compositional mode, will not hamper the idiomatic condition to a greater extent than the literal condition.

3. Experiment 2

3.1. Materials and Methods

3.1.1. Participants

A total of 38 French-dominant French–English speakers (Age $M = 24.24$, $SD = 4.30$; 25 female; 13 male) from McGill University and the Montreal community took part in the study for course credit or compensation of 10 CAD/h. All participants gave informed consent, and the study was approved by the Research Ethics Board Office of McGill University. All participants had normal or corrected-to-normal vision and no self-reported history of speech or hearing disorders. As in Experiment 1, participants filled a language history questionnaire based on LEAP-Q (Marian et al. 2007). Participants’ characteristics for Experiment 2 are reported in Table 5. Complete data are reported in Appendix A. Data for 5 participants were excluded from the following analysis due to excessive track loss, resulting in a final sample of 33 subjects.

Table 5. Participant characteristics (Experiment 2). Self-ratings were collected on a 1–7 scale (/7).

| Measure                      | Mean | SD  |
|------------------------------|------|-----|
| Demographics                 |      |     |
| Chronological age            | 24.24| 4.30|
| English self-ratings (/7)    |      |     |
| Speaking                     | 5.24 | 1.26|
| Reading                      | 5.66 | 1.05|
| Writing                      | 5.05 | 1.29|
| Translating                  | 4.79 | 1.41|
| Listening                    | 5.03 | 0.87|
| Pronunciation                | 4.68 | 1.16|
| Fluency                      | 5.13 | 1.28|
| Vocabulary                   | 4.79 | 1.19|
| Grammar                      | 5.11 | 1.41|
| Overall competence           | 5.18 | 1.16|
| Language exposure (%)        |      |     |
| English                      | 39.84| 23.71|
| French                       | 57.89| 23.92|
3.1.2. Stimuli

The stimuli for Experiment 2 were previously used in Senaldi et al. (n.d.) with English-dominant bilinguals. Items consisted of 42 English idioms with a verb–article–noun structure (e.g., *Dolan spilled the beans*). These idioms were a subset of the 60 idioms tested in Experiment 1. The literal control condition for each idiom was kept identical (e.g., *Dolan cooked the beans*).

For each of the three contextual conditions already tested in Experiment 1 (Idiomatic–Idiomatic, Idiomatic–Literal and Literal–Literal), a language-switched condition was created by translating the idiom noun into French (e.g., *she spilled the fèves; she cooked the fèves*). French nouns were matched in both character length and Lexique word frequency (New et al. 2001) with the original English nouns. Each of the 42 items was thus embedded into six sentential conditions: an idiom followed by a figurative disambiguating region, with its noun being either in English (Idiomatic–Idiomatic—English) or in French (Idiomatic–Idiomatic—French); an idiom expression followed by a literal context, with either an English (Idiomatic–Literal—English) or a French noun (Idiomatic–Literal—French); and a non-idiomatic control phrase followed by a literal context, with either an English (Literal–Literal—English) or a French noun (Literal–Literal—French). Table 6 presents example sentences from each condition.

Table 6. Example sentences of Idiomatic–Idiomatic (Id-Id), Idiomatic–Literal (Id-Lit), and Literal–Literal (Lit-Lit) conditions, both intact (En) and switched (Fr), from Experiment 2.

| Condition     | Sentence                                                                 |
|---------------|---------------------------------------------------------------------------|
| Id-Id-En      | Dolan *spilled the beans* when he mentioned the surprise party to his friend. |
| Id-Lit-En     | Dolan *spilled the beans* when he tried to pour too many into the soup pot.   |
| Lit-Lit-En    | Dolan *cooked the beans* before he started adding vegetables to the soup pot. |
| Id-Id-Fr      | Dolan *spilled the fèves* when he mentioned the surprise party to his friend. |
| Id-Lit-Fr     | Dolan *spilled the fèves* when he tried to pour too many into the soup pot.   |
| Lit-Lit-Fr    | Dolan *cooked the fèves* before he started adding vegetables to the soup pot. |

As in Experiment 1, ratings of verb relatedness, noun relatedness and literal plausibility for each idiom were obtained from Libben and Titone’s (2008) norms. At the end of the experiment, we collected ratings of familiarity (on a 1–7 Likert scale) and cross-language overlap. Table 7 reports the average normative characteristics of our idioms and stimuli.

Table 7. Item and stimuli characteristics (Experiment 2).

| Variable | Condition | Measurement | Min  | Max  | Mean | SD  |
|----------|-----------|-------------|------|------|------|-----|
| Idiom norms | Family | Familiarity (/7) | 1.59 | 6.55 | 4.29 | 1.38 |
| Verb Relatedness (/5) | 0.80 | 4.93 | 2.73 | 1.07 |
| Noun Relatedness (/5) | 0.54 | 4.47 | 2.47 | 1.02 |
| Literal Plausibility (/5) | 2.07 | 5.00 | 3.96 | 0.72 |
| Cross-language overlap (/1) | 0.23 | 0.95 | 0.60 | 0.21 |
| Noun | English | Length (characters) | 3   | 8   | 4.76 | 1.23 |
| French | 3   | 9   | 5.43 | 1.38 |
| Idiom phrase | Idiomatic—English | Length (characters) | 8   | 17  | 12.79 | 2.19 |
| Idiomatic—French | 7   | 19  | 13.40 | 2.41 |
| Literal—English | 8   | 17  | 12.71 | 2.11 |
| Literal—French | 7   | 19  | 13.38 | 2.34 |
| Disambiguating region | Idiomatic—Idiomatic—English | Length (characters) | 36  | 51  | 44.88 | 3.30 |
| Idiomatic—Literal—English | 36  | 51  | 43.85 | 3.66 |
| Literal—Literal—English | 34  | 50  | 43.45 | 3.15 |
| Idiomatic—Idiomatic—French | 36  | 51  | 44.90 | 3.30 |
| Idiomatic—Literal—French | 36  | 51  | 43.88 | 3.67 |
| Literal—Literal—French | 34  | 50  | 43.48 | 3.16 |
3.1.3. Apparatus

The apparatus was the same as in Experiment 1.

3.1.4. Procedure

The procedure was the same as in Experiment 1, with the only difference being in the structure of the experimental lists. Each subject was randomly assigned to one of six fully counterbalanced lists. Each idiom appeared in only one of the six conditions in each list, and no participant saw the same idiom more than once. Each subject read 42 experimental sentences (7 per condition), 128 filler sentences taken from Libben and Titone (2009) containing interlingual homographs and cognates, 10 practice sentences, and 23 comprehension questions, for a total of 203 sentence stimuli. Half of the filler sentences contained switches into French and half did not. Practice sentences conveyed literal meanings and were similar in form to the Literal–Literal sentences and the homograph/cognate fillers. Comprehension questions appeared on roughly 10% of all the trials. Each of the six lists was arranged into two different randomized orders, resulting in a total of 12 different stimuli sets.

3.2. Results

Participants were overall highly accurate in responding to comprehension questions ($M = 89.4\%$), indicating that they were attentive throughout the experiment.

As in Experiment 1, eye movement data were analyzed with LME in R (R Core Team 2021) via the lme4 package (Bates et al. 2015). Idiom first pass gaze duration and idiom total reading time were, once again, the reading measures of interest. The idiom region was skipped on first pass (Idiom First Pass Gaze Duration) on 11 trials out of 1124 total trials, while it was never completely skipped at the level of total reading time.

For each reading measure, we first built core models to test for the effect of the two main independent variables, i.e., phrase idiomaticity and the presence of a temporary language switch. In these models, condition (categorical), switch (categorical), and their interaction were fixed effects. Control variables included trial order (continuous) and literal plausibility (continuous). The treatment of the variable condition in pre-disambiguation (Idiomatic vs. Literal) and post-disambiguation (Idiomatic–Idiomatic vs. Idiomatic–Literal vs. Literal–Literal) measures followed the same procedure as in Experiment 1. When building core models, we started with a maximally specified random structure (by-subjects and by-items slopes and intercepts), and consecutively removed random slopes in case of convergence issues (Barr et al. 2013). This allowed us to assess whether the effect of condition and switching held over and above individual variation among subjects and items.

In the next step of our analysis, follow-up models were built to zoom in on the effect of item-specific variables modulating direct retrieval (cross-language overlap) and word-by-word semantic parsing (verb relatedness and noun relatedness). In a first group of models, verb relatedness (continuous), average cross-language overlap (continuous), condition, switch and their four-way interaction were set as fixed effects. In a separate set of models, noun relatedness (continuous) replaced verb relatedness.

All continuous variables were scaled to reduce collinearity. While verb and noun relatedness were not significantly correlated ($p > 0.05$), cross-language overlap was significantly correlated both with verb relatedness ($\rho = 0.50$) and noun relatedness ($\rho = 0.32$). Random slopes were removed from follow-up models, so that they could not absorb potential variation due to the effects of item-specific variables.

For each model we reported the same statistics as in the previous experiment. Table 8 reports means and standard deviations for raw Idiom First Pass Gaze Duration and raw Idiom Total Reading Time in each experimental condition.
Table 8. Subject-aggregated means and standard deviations (in milliseconds) for raw Idiom First Pass Gaze Duration and raw Idiom Total Reading Time (Experiment 2) across sentence conditions containing nonswitched English (En) nouns and nouns switched to French (Fr).

| Measure                      | Condition        | Nonswitch (En) | Switch (Fr) |
|------------------------------|------------------|----------------|-------------|
|                              | Mean             | SD             | Mean        | SD          |
| Idiom First Pass Gaze Duration | Idiomatic        | 620.09         | 159.17      | 705.83      | 196.47      |
|                              | Literal          | 603.27         | 184.63      | 675.56      | 248.71      |
| Idiom Total Reading Time     | Idiomatic–Idiomatic | 1053.86       | 575.03      | 1249.48     | 539.22      |
|                              | Idiomatic–Literal | 1085.79       | 459.89      | 1083        | 380.49      |
|                              | Literal–Literal  | 1006.35        | 479.89      | 1048.47     | 469.13      |

3.2.1. Pre-Disambiguation Effects (Idiom First Pass Gaze Duration)

A significant two-way interaction between noun relatedness and switch emerged in noun-relatedness models ($\beta = 53.23$, $SE = 24.12$, $t = 2.21$, $p = 0.028$), whereby the early processing of nonswitched phrases was overall facilitated by increased retrospective decomposability. This result is anyway not revealing to our research question in that it does not show any differential impact of language switching on idioms and literals. As for the rest, a main inhibiting effect of language switching was found both in verb relatedness models ($\beta = 54.11$, $SE = 26.99$, $t = 2.01$, $p = 0.045$) and in core models ($\beta = 80.16$, $SE = 27.33$, $t = 2.93$, $p = 0.005$).

Consistent with our predictions, first-pass gaze durations were comparable for switched L2 idioms and switched L2 literals, with only a general hindering effect of language shift. Since temporarily changing nouns’ language impeded direct retrieval, the absence of an idiom disadvantage confirms that a direct retrieval route is not pursued by default in the first stage of L2 idiom comprehension.

3.2.2. Post-Disambiguation Effects (Idiom Total Reading Time)

A significant two-way interaction between condition and switch emerged for the Idiomatic–Idiomatic vs. Idiomatic–Literal contrast in both the noun relatedness models ($\beta = -162.25$, $SE = 81.32$, $t = -2.00$, $p = 0.046$) and in the core models ($\beta = -158.10$, $SE = 76.72$, $t = -2.06$, $p = 0.040$). Accordingly, shifting language demands impacted the fully idiomatic condition more than when idioms were followed by a literal context. A main inhibiting effect of switch was also found in the noun-relatedness ($\beta = 138.07$, $SE = 57.30$, $t = 2.41$, $p = 0.016$), verb-relatedness ($\beta = 129.01$, $SE = 63.44$, $t = 2.03$, $p = 0.042$) and core models ($\beta = 152.73$, $SE = 56.86$, $t = 2.69$, $p = 0.008$).

Collectively, the results from Experiment 2 confirm that L2 readers initially parsed idioms word for word, committing to a literal analysis. After reading the disambiguating context, they identified the target phrase as idiomatic and revised their initial interpretation, incurring greater processing costs. Given that the presence of a temporary language shift prevents direct retrieval and pushes comprehenders towards a compositional mode, a specific slowdown of the switched idiomatic condition would imply that the language processing system attempted a direct-retrieval strategy but was hindered by the presence of a formal manipulation. The fact that L2 speakers did not process switched idioms more effortfully than switched literals in early measures confirms that both conditions were parsed compositionally at first. A specific disadvantage for the switched idiomatic condition (and thus an attempt at direct retrieval) emerged only at the level of total reading time, where also evidence from Experiment 1 showed a direct-retrieval effect mediated by cross-language overlap.

4. Discussion

The two eye-tracking experiments presented here investigated how L2 users read and processed idioms in a natural sentential context. In particular, we assessed to what extent the predictions of a hybrid model (Libben and Titone 2008; Titone and Libben 2014)
extend to idiom processing in the L2, by analyzing the interplay of direct retrieval and compositional parsing during comprehension.

Experiment 1 drew upon the materials and methods of Titone et al. (2019) to test a sample of French–English bilingual adults, dominant in French, while they read English sentences containing idioms intended figuratively, idioms intended literally, and matched literal phrases followed by a disambiguating context. When testing L1 readers on the same materials, Titone et al. (2019) detected a processing advantage for idioms over literal control sentences in early stages of comprehension, suggesting that L1 speakers readily identified idioms as familiar sequences and retrieved them directly from memory. In later comprehension stages, idioms appeared to be reanalyzed compositionally, with increased verb-related decomposability making idioms’ semantic integration more laborious. In this study, L2 readers did not process idioms differently from literal control sentences at an early stage, except when prospective verb relatedness was higher, and a congruent idiom “cognate” could not be retrieved from French or, vice versa, when verb relatedness was lower, and readers could rely on an equivalent French idiom. In the former case, high verb relatedness made interpretation of the phrase ambiguous between a figurative or a literal reading, and the idiomatic condition was read more slowly. In the latter case, when L2 users could not resort to verb-mediated decomposability to access the figurative meaning and identify the phrase as idiomatic, increased overlap with a congruent French idiom brought about a slight disadvantage of the idiomatic condition, signaling higher figurative priming. However, after readers encountered the disambiguating context, idiomatic sentences were facilitated in comparison to literal sentences by increased retrospective noun-related decomposability and by the availability of a translationally equivalent idiom in French.

These results suggest that L2 idiom comprehension is primarily compositional, with different idiom components individually contributing to their semantics over different time courses. As suggested by extant evidence on L1 idiom processing (Cacciari and Tabossi 1988; Carrol 2021; Libben and Titone 2008; Siyanova-Chanturia et al. 2011a; Titone et al. 2019), L1 users’ greater exposure to idiomatic structures likely leads to the formation of automatized whole-phrase representations in memory that can be directly accessed during processing. In this regard, L2 speakers have comparably less experience with L2 idioms to form entrenched phrase-level representations that allow them to identify idioms at first glance in context. Consequently, when encountering an idiom in their L2, readers adopt (or continue) by default a literal and compositional strategy. Thus, at this point, every piece of semantic information that can cumulatively lead to the phrasal figurative meaning will be leveraged to successfully comprehend the idiom. Before the disambiguating region is reached, increased verb-mediated decomposability gives readers a clue that the intended phrasal meaning might be different from the preferred literal one, therefore slowing down the idiomatic condition. After the disambiguating region is read, idiom processing can be facilitated retrospectively if the meaning of the noun is closely related to the idiom’s phrasal meaning.

A facilitatory role of semantic decomposability ties in with previous work demonstrating that L2 users are more sensitive to the internal semantics of idioms, probably due to their analytical bias. In off-line rating tasks, L2 learners were found to judge idioms as more decomposable and transparent than L1 speakers (Abel 2003; Hubers et al. 2020), while Steinel et al. (2007) observed that both imageability and transparency facilitated L2 idioms in a paired-associate learning task. Similarly, Skoufaki (2008) found that advanced English learners defined the meanings of unknown idioms more consistently when they were highly transparent. By contrast, other studies found intuitions on decomposability and transparency to be dependent on L2 learners’ familiarity with the same meanings, rather than on specific properties of idioms’ constituents (Keysar and Bly 1995; Malt and Eiter 2004). Finally, neither van Ginkel and Dijkstra’s (2020) priming data nor Beck and Weber’s (2021) novel-phrase recall data revealed any transparency effects.
Given a default compositional approach of L2 readers when making sense of idioms in context, direct retrieval exerted only a secondary role in early comprehension when compositional parsing was impeded, and took on a more general facilitating effect in later stages, after the disambiguating context made the idiomatic interpretation of the phrase explicit. More specifically, readers were facilitated in integrating idioms’ semantics at the sentence level by the possibility to retrieve a corresponding idiom form from their L1 (e.g., *briser la glace* for *break the ice*). This “cognate” effect for idioms coheres with past research showing facilitating effects of cross-language overlap on the processing of L2 idioms (Carrol et al. 2016; Charteris-Black 2002; Irujo 1986, 1993; Laufer 2000; Liontas 2002; Pritchett et al. 2016; Titone et al. 2015) and single-word cognates (Gullifer et al. 2013; LIBben and Titone 2009; Pivneva et al. 2014; Titone et al. 2011; Van Hell and Dijkstra 2002). Nonetheless, as noted by Titone et al. (2015), the concept of cross-language overlap (i.e., “cognate” status) involves an abstract and structural mapping that goes beyond orthography or phonology when it comes to idioms. In this regard, Beck and Weber (2016a) teased apart lexical level idioms, which have word-for-word translational counterparts, and post-lexical level idioms, which point to the same idiomatic concepts but do not share the same lexemes across languages, finding no processing difference between the two.

Interestingly, while some scholars have interpreted this general advantage of congruent “cognate” idioms as an indication that the bilingual lexicon might be integrated also beyond the single-word level (Zeng et al. 2020), a few studies have found that even L1-unique idioms enjoy some form of processing advantage when encountered as word-for-word translations in the comprehender’s L2 (Carrol and Conklin 2014, 2017; Carrol et al. 2016; Carrol et al. 2018). Woven together, the evidence accumulated thus far suggests that comprehenders can generally leverage their L1 knowledge when required to identify multiword representations in the L2 (see also Conklin and Carrol 2018).

In Experiment 2, a language-switching manipulation was added to the same conditions of the previous experiment. Idioms and literals thus appeared either in their canonical form or with their nouns translated into French (e.g., *spill the fèves, cook the fèves*). Participants were once again French–English bilingual adults identifying French as their L1. The purpose of this follow-up experiment was to confirm the secondary role of direct retrieval in L2 idiom reading by testing a formal manipulation that intentionally undermined it. Previous research similarly addressed the impact of canonical form disruption on idiom processing, focusing mostly on L1 speakers (Geeraert et al. 2018; Gibbs et al. 1989; Haueuer et al. 2021; Kyriacou et al. 2020, 2021; McGlone et al. 1994; Smolka and Eulitz 2020). In this study, inserting a momentary language shift on idiom-final nouns would push readers into a compositional parsing mode, therefore hindering direct retrieval. Therefore, a specific disadvantage incurred by switched idioms over switched literals would confirm that a failed attempt at direct retrieval had taken place in the idiom condition. Vice versa, if idioms and literals were equally impacted by language switching, this would suggest that a compositional strategy had been adopted in both conditions.

When Senaldi et al. (n.d.) tested the same manipulation on L1 readers, language switching inhibited idioms more than literals in early processing stages. Parallel findings were obtained by Titone et al. (2015) with speeded meaningfulness judgments. While a general switching cost is consistent with previous experimental literature (Altarriba et al. 1996; Bultena et al. 2015; Gullifer and Titone 2019; Guzzardo Tamargo et al. 2016; Litcofsky and Van Hell 2017), this specific idiom disadvantage confirmed a disruption in direct retrieval during early stage L1 idiom recognition. In the present study, L2 idioms and literals were not differentially impacted by the presence of a language switch in early measures; however, switched idioms in a figurative context (Idiomatic–Idiomatic) incurred greater processing costs with respect to idioms with a literal continuation (Idiomatic–Literal) in total reading time. Hence, evidence from Experiment 2 confirms that direct form retrieval characterizes only late-stage L2 idiom processing, after comprehenders read the entire idiom phrase and entered the disambiguating region.
Of note, in the idiom literature, another variable that is frequently examined as a modulator of direct retrieval is familiarity (Cronk and Schweigert 1992; Libben and Titone 2008; Titone and Libben 2014). In the present work, familiarity ratings had been collected from participants to confirm that they were familiar overall with the items used in the study. Average familiarity scores for Experiment 1 (1.99 out of 3) and Experiment 2 (4.29 out of 7) suggest that this was the case. However, as noted in Section 1.3, we believe that familiarity might be less reliable a predictor when it comes to L2 processing, in that it might be influenced by a host of related factors like subjects’ exposure to the L2 and overall proficiency. For this reason, we have rather focused on the role of cross-language overlap, which could be conceived of as L2 readers’ familiarity with equivalent idiom forms in their L1. As a confirmation of this, cross-language overlap and familiarity appeared to be overall correlated both in Experiment 1 ($\rho = 0.79$) and Experiment 2 ($\rho = 0.65$). To ensure that this did not impact our interpretation of the results, we re-ran all the models in Experiments 1 and 2, adding scaled average self-reported familiarity as a control variable. Likelihood ratio tests indicated that adding familiarity as a control variable did not significantly change models’ fit for Experiment 1. For Experiment 2, inserting familiarity in the models significantly changed fit except in models predicting idiom first pass gaze duration with noun relatedness. Most importantly, in all the models, for both experiments, predictors’ significance did not change after controlling for this measure.

Also of note, one of the control variables included in both experiments was literal plausibility (or ambiguity), in order to check if the possibility to interpret a given idiom literally (e.g., *scratch your head* vs. *give your word*) could represent a potential confound. Overall, we selected idioms with a high degree of literal plausibility to make the Idiomatic–Literal condition possible and natural for all items. Therefore, since our chosen idioms do not cover the full spectrum of literal plausibility, exploring its interaction with the other main experimental variables (condition, cross-language overlap and verb/noun relatedness) would not be an ideal indicator of the impact of this variable.

Another related variable to examine in future research on L2 idiom reading is meaning dominance, which builds on the notion of literal plausibility to measure how often an ambiguous idiom string is employed in the figurative versus literal sense in language use (Milburn and Warren 2019). While meaning dominance can give a more nuanced picture of the effects of idioms’ semantic ambiguity on L2 reading, there might be a chance that it is less informative a predictor than with L1 speakers. While a second-language speaker could easily label an idiom like *scratch your head* as potentially literal, knowing in which proportion it occurs in each of its two senses would require extensive exposure to its use in the L2, which might not be the case for all L2 learners. Moreover, making a meaning dominance judgment could be challenging in the case of highly decomposable idioms, where the literal and the figurative meaning can be quite similar and hard to disentangle.

Taken together, the heavier reliance of L2 users on compositional parsing that emerged in both experiments is also consistent with usage-based and constructionist views (Bybee 2006; Goldberg 2006; Tomasello 2003; Wulff 2008), which conceive of the lexicon as a network of linguistic structures of varying complexity, from single words to idioms and abstract syntactic patterns. Since this framework predicts language experience to be a major determinant of constructions’ entrenchment in the mind, ambient exposure of L2 users to idiomatic and multiword language would not suffice to form entrenched and automatized representations that can be retrieved directly in processing. On a related note, the L2 acquisition framework put forth by Wray (2000, 2002) postulates a dynamic balance between holistic memory-based and analytic composition-based processes, depending on how late the second language is acquired. In accordance with this model, late bilinguals lean towards the analytical end of the spectrum, which could explain the literal/compositional bias exhibited during idiom processing.

The emphasis of the present work was specifically on the cognitive underpinnings of bilingual idiom reading; however, converging findings of a slower literal and compositional processing of L2 idioms also come from other experimental paradigms. Van Lancker Sidtis
(2003) found that L2 speakers performed worse at distinguishing idiomatic and literal meanings of aurally presented sentences; however, in a cross-modal priming by Cieślicka (2006), L2 speakers exhibited increased priming for literal rather than idiomatic targets. Building on the graded salience hypothesis (Giora 1997, 1999, 2003), which predicted a processing advantage for the phrasal meaning (literal or figurative) that is more salient (i.e., frequent, familiar and/or conventional) in the speakers’ lexicon, Cieślicka (2006, 2010) put forth a literal salience resonant model for L2 idiom processing. Accordingly, literal meanings of L2 idioms enjoy a more salient status, even after L2 idioms are incorporated into the learners’ lexicon, since learners are likely to use and come across literal meanings more often. A hybrid, multidetermined model would say that this more salient status arises mechanistically from the fact that holistic idiomatic representations are less entrenched in the minds of L2 speakers (e.g., Titone et al. 2015). Other cross-modal and visual priming studies found facilitation for both literally and figuratively related targets over unrelated targets in both L1 and L2 speakers (Beck and Weber 2016a; van Ginkel and Dijkstra 2020), although L2 speakers in Beck and Weber (2016a) exhibited a figurative disadvantage compared to literal phrases. Figurative attunement, i.e., incremental facilitation for figurative targets, seems nonetheless achievable in L2 speakers by increasing the proportion of idiomatic sentences in the experimental list (Beck and Weber 2016b). Increased late-stage attunement to L2 idioms’ figurative reading was also observed by Cieślicka et al. (2021) when L2 speakers processed an anaphoric referent placed downstream from an idiomatic phrase, which is supposed to suppress a contextually inappropriate literal reading (e.g., I always miss the boat when it comes to jokes, and that makes it nearly impossible for me to attend comedy shows).

5. Conclusions

In Experiment 1, we showed that idiom comprehension in an L2 is mostly driven by compositional processes, with prospective verb-related decomposability and retrospective noun-related decomposability jointly guiding readers towards a bottom-up access to the figurative meaning. Unlike L1 idiom processing, direct retrieval plays a later role in comprehension, and it is mediated by the availability of a translationally congruent “cognate” idiom in the comprehenders’ L1. Additional support for these findings emerged in Experiment 2 using language-switched idioms. Here, idioms whose direct retrieval was disrupted by a language shift appeared to be penalized only in later processing stages, while in earlier stages they were processed akin to switched literal phrases.

Of note, the present study examined L2 readers as one monolithic group because of the complexity already present in the design. Therefore, further research is necessary to shed light on how individual differences in L2 proficiency affect readers’ propensity for a holistic versus incremental comprehension route. While more proficient L2 readers are expected to behave more similarly to L1 readers, preliminary findings from Milburn et al. (2021) suggest that, after overcoming an initial literal bias, even more advanced L2 learners might go through a temporary phase where they overapply a figurative direct-retrieval strategy also to literal contexts. This intriguing result deserves additional empirical follow-up work. The differential impact of individual bilingual experiences on idiom processing emerged also in previous work by López and Vaid (2018). Here, bilingual speakers with more extensive brokering (i.e., informal translation) experience were facilitated in making semantic relatedness judgments on idioms across language boundaries. These findings could have interesting implications for the role of cross-language overlap in bilingual speakers with different individual experiences.

Finally, idioms in this study were placed in sentence-initial position with no prior semantic context; however, putting them after a disambiguating context might facilitate L2 readers in interpreting them figuratively right from the start. Whether this would translate to enhanced decomposability effects, more direct access to idiom forms, or both, is fodder for future investigations.

Taken together, this study, in combination with prior work using similar materials and procedures for L1 readers, suggests that the cognitive mechanisms of idiom processing
involve different concentrations of idiom form retrieval and compositional processes over time, calibrated by readers’ current state of L1 and L2 knowledge, and how idiom representations overlap across an L1 and L2. Indeed, this mechanistic account is the central core of a hybrid multidetermined model (Libben and Titone 2008; Titone and Libben 2014; Titone et al. 2019). Notably, an advantage of this mechanistic account is that it is not restricted to the comprehension of literary idioms of the type studied here, but is also more general, and can be applied to other forms of figurative and non-figurative language (e.g., metaphor as in Columbus et al. 2015; and also multiword expressions that are not typically thought of as “idioms” but function cognitively within the same processing space as detailed by Wray 2002, 2013), and may underlie dominant assumptions about how language generally is understood. In this way, a hybrid, multidetermined account, supported empirically here and in past work, may be seen as an extension of general language processing operations. Similarly, the comprehension of idiomatic or other multiword expressions may be seen as a natural and fundamental, core part of “language” proper, as cogently advanced by usage-based models of language (Bybee 2006; Goldberg 2006; Tomasello 2003; Wulff 2008).

Author Contributions: Conceptualization, D.A.T.; data curation, D.A.T.; formal analysis, M.S.G.S. and D.A.T.; funding acquisition, D.A.T.; methodology, D.A.T.; resources, D.A.T.; writing—original draft, M.S.G.S. and D.A.T.; writing—review and editing, M.S.G.S. and D.A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NSERC Discovery Grant to Debra Titone.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Board Office of McGill University (protocol code 30-0806, 5 September 2006).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Anonymized data, R scripts for models and stimuli can be downloaded at the OSF repository (https://osf.io/gv8zf/).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Full participant characteristics (Experiment 1). Self-ratings were collected on a 1–10 scale (/10).

| Measure                  | Mean | SD  |
|--------------------------|------|-----|
| Demographics             |      |     |
| Chronological age        | 20.68| 3.96|
| English self-ratings (/10)|      |     |
| Speaking                 | 7.97 | 2.01|
| Reading                  | 8.78 | 1.29|
| Writing                  | 8.00 | 1.94|
| Translating              | 7.94 | 1.95|
| Listening                | 8.76 | 1.39|
| Pronunciation            | 7.36 | 2.42|
| Fluency                  | 7.82 | 2.27|
| Vocabulary               | 7.72 | 2.11|
| Grammar                  | 7.82 | 2.20|
| Overall competence       | 8.01 | 1.96|
| French self-ratings (/10)|      |     |
| Speaking                 | 9.97 | 0.16|
| Reading                  | 9.82 | 0.55|
| Writing                  | 9.27 | 1.42|
| Translating              | 9.20 | 1.07|
| Listening                | 9.77 | 0.53|
| Pronunciation            | 9.70 | 0.69|
Table A1. Cont.

| Measure            | Mean | SD  |
|--------------------|------|-----|
| Fluency            | 9.82 | 0.50|
| Vocabulary         | 9.43 | 0.96|
| Grammar            | 8.82 | 1.85|
| Overall competence | 9.70 | 0.69|
| Language exposure (%) |     |     |
| English            | 46.70 | 18.19|
| French             | 49.95 | 19.27|

Table A2. Full participant characteristics (Experiment 2). Self-ratings were collected on a 1–7 scale (/7).

| Measure            | Mean | SD  |
|--------------------|------|-----|
| Demographics       |      |     |
| Chronological age  | 24.24 | 4.30|
| English self-ratings (/7) | | |
| Speaking           | 5.24 | 1.26|
| Reading            | 5.66 | 1.05|
| Writing            | 5.05 | 1.29|
| Translating        | 4.79 | 1.41|
| Listening          | 5.03 | 0.87|
| Pronunciation      | 4.68 | 1.16|
| Fluency            | 5.13 | 1.28|
| Vocabulary         | 4.79 | 1.19|
| Grammar            | 5.11 | 1.41|
| Overall competence | 5.18 | 1.16|
| French self-ratings (/7) | | |
| Speaking           | 7    | 0   |
| Reading            | 7    | 0   |
| Writing            | 6.82 | 0.51|
| Translating        | 6.53 | 0.86|
| Listening          | 7    | 0   |
| Pronunciation      | 6.97 | 0.16|
| Fluency            | 6.95 | 0.32|
| Vocabulary         | 6.82 | 0.46|
| Grammar            | 6.68 | 0.62|
| Overall competence | 6.95 | 0.32|
| Language exposure (%) | | |
| English            | 39.84 | 23.71|
| French             | 57.89 | 23.92|

Notes

1 A few interactions and main effects emerged that did not involve switching and were thus not relevant to our research question. Specifically, a significant two-way interaction was detected between verb relatedness and condition for the Idiomatic–Idiomatic vs. Literal–Literal contrast \( (\beta = 129.75, SE = 44.93, t = 2.89, p = 0.004) \), and between cross-language overlap and noun relatedness \( (\beta = 69.36, SE = 32.83, t = 2.11, p = 0.037) \). A main effect of condition for the Idiomatic–Idiomatic vs. Literal–Literal contrast, with a general slowdown of the fully idiomatic condition compared to the fully literal one, was detected in verb-relatedness models \( (\beta = -95.52, SE = 45.05, t = -2.12, p = 0.034) \), noun-relatedness models \( (\beta = -147.65, SE = 40.36, t = -3.66, p < 0.001) \) and core models \( (\beta = -128.84, SE = 38.15, t = -3.38, p = 0.001) \).

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]

Alexander, Richard J. 1987. Problems in understanding and teaching idiomaticity in English. Anglistik und Englischunterricht 32: 105–22.

Altarriba, Jeanette, Judith F. Kroll, Alexandra Sholl, and Keith Rayner. 1996. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. Memory & Cognition 24: 477–92. [CrossRef]
Altenberg, Bengt, and Sylviane Granger. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing. Applied Linguistics 22: 173–95. [CrossRef]

Arnon, Inbal, and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. Journal of Memory and Language 62: 67–82. [CrossRef]

Barnard, Colin, and Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children’s repetition of four-word combinations. Psychological Science 19: 241–48. [CrossRef]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68: 255–78. [CrossRef]

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67: 1–48. [CrossRef]

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 38th Annual Conference of the Cognitive Science Society, Philadelphia, PA, USA, August 10–13.

Beck, Sara D., and Andrea Weber. 2021. Phrasal Learning Is a Horse Apiece: No Recognition Memory Advantages for Idioms in L1 and L2 Adult Learners. Frontiers in Psychology 12: 1100. [CrossRef]

Bobrow, Samuel A., and Susan M. Bell. 1973. On catching on to idiomatic expressions. Memory & Cognition 1: 343–46. [CrossRef]

Boers, Frank. 2000. Enhancing metaphorical awareness in specialised reading. Journal of Memory and Language 42: 436–58. [CrossRef]

Boers, Frank, John Eyckmans, Jenny Kappel, and Helène Stengers. 2006. Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. Language Teaching Research 10: 245–61. [CrossRef]

Bortfeld, Heather. 1997. A Cross-Linguistic Analysis of Idiom Comprehension by Native and Nonnative Speakers. Ph.D. thesis, State University of New York, Stony Brook, NY, USA.

Bortfeld, Heather. 2003. Comprehending idioms cross-linguistically. Experimental Psychology 50: 217. [CrossRef] [PubMed]

Bultena, Sybrine, Ton Dijkstra, and Janet G. Van Hell. 2015. Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. Bilingualism: Language and Cognition 18: 453–69. [CrossRef]

Bybee, Joan. 2006. From usage to grammar: The mind’s response to repetition. Language 82: 711–33. [CrossRef]

Cacciari, Cristina. 1993. The place of idioms in a literal and metaphorical world. In Idioms: Processing, Structure, and Interpretation. Edited by Cristina Cacciari and Patrizia Tabossi. Hillsdale: Erlbaum, pp. 27–55.

Cacciari, Cristina. 2014. Processing multiword idiomatic strings: Many words in one? The Mental Lexicon 9: 267–93. [CrossRef]

Cacciari, Cristina, and Sam Glucksberg. 1991. Understanding idiomatic expressions: The contribution of word meanings. In Understanding Word and Sentence. Edited by Gregory Simpson. Amsterdam and New York: North-Holland, pp. 217–40. [CrossRef]

Cacciari, Cristina, and Patrizia Tabossi. 1988. The comprehension of idioms. Journal of Memory and Language 27: 688–83. [CrossRef]

Caillies, Stephanie, and Kirsten Butcher. 2007. Processing of idiomatic expressions: Evidence for a new hybrid view. Metaphor and Symbol 22: 79–108. [CrossRef]

Carrol, Gareth. 2021. Psycholinguistic approaches to figuration. In Figurative Language–Intersubjectivity and Usage. Edited by Augusto Soares da Silva. Amsterdam and Philadelphia: John Benjamins, pp. 307–37.

Carrol, Gareth, and Kathy Conklin. 2014. Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. Bilingualism: Language and Cognition 17: 784–97. [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] [PubMed]

Carrol, Gareth, and Jeannette Littlemore. 2020. Resolving figurative expressions during reading: The role of familiarity, transparency, and context. Discourse Processes 57: 609–26. [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]

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]

Charteris-Black, Jonathan. 2002. Second language figurative proficiency: A comparative study of Malay and English. Applied Linguistics 23: 104–33. [CrossRef]

Chomsky, Noam. 1980. Rules and representations. Behavioral and Brain Sciences 3: 1–15. [CrossRef]

Christiansen, Morten H., and Inbal Arnon. 2017. More than words: The role of multiword sequences in language learning and use. Topics in Cognitive Science 9: 542–51. [CrossRef]

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

Ciesielska, Anna. 2010. Formulaic language in L2. In Cognitive Processing in Second Language Acquisition: Inside the Learner’s Mind. Edited by Martin Pütz and Laura Sicola. Amsterdam and Philadelphia: John Benjamins, pp. 149–68.

Ciesielska, Anna B., 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.
New, Boris, Christophe Pallier, Ludovic Ferrand, and Rafael Matos. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE™—A lexical database for contemporary french: LEXIQUE™. L’année Psychologique 101: 447–62. [CrossRef]
Nunberg, Geoffrey D. 1978. The Pragmatics of Reference. Bloomingston: Indiana University Linguistics Club.
Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. Language 70: 491–538. [CrossRef]
Pawley, Andrew, and Frances Hodgetts Syder. 1983. Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. Journal of Pragmatics 7: 551–79. [CrossRef]
Pivneva, Irina, Julie Mercier, and Debra Titone. 2014. Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition 40: 787. [CrossRef]
Pritchett, Lena K., Jyotsna Vaid, and Sumeyra Tosun. 2016. Of black sheep and white cows: Extending the bilingual dual coding theory to memory for idioms. Cogent Psychology 3: 1135512. [CrossRef]
R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 23 March 2022).
Richards, Jack C., and Theodore S. Rodgers. 2014. Approaches and Methods in Language Teaching. Cambridge: Cambridge University Press.
Rossiter, Marian J., Tracey M. Derwing, Linda G. Manimtim, and Ron I. Thomson. 2010. Oral fluency: The neglected component in the communicative language classroom. Canadian Modern Language Review 66: 583–606. [CrossRef]
Segalowitz, Norman. 2010. Cognitive Bases of Second Language Fluency. New York: Routledge. [CrossRef]
Senaldo, Marco S. G., Junyan Wei, Jason W. Gullifer, and Debra A. Titone. n.d. Scratching your tête over language-switched idioms: Evidence from eye-movement measures of reading. In revision.
Sinclair, John. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.
Siyanova-Chanturia, Anna, and Norbert Schmitt. 2007. Native and non-native use of multi-word vs. one-word verbs. International Review of Applied Linguistics in Language Teaching 45: 119–39. [CrossRef]
Siyanova-Chanturia, Anna, and Diana Van Lancker-Sidtis. 2018. What online processing tells us about formulaic language. In Understanding Formulaic Language. Edited by Anna Siyanova-Chanturia and Ana Pellicer-Sánchez. New York: Routledge, pp. 38–61.
Siyanova-Chanturia, Anna, Kathy Conklin, and Norbert Schmitt. 2011a. 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]
Siyanova-Chanturia, Anna, Kathy Conklin, and Walter J. B. Van Heuven. 2011b. 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: 776–84. [CrossRef] [PubMed]
Skoufaki, Sophia. 2008. Investigating the source of idiom transparency intuitions. Metaphor and Symbol 24: 20–41. [CrossRef]
Smolka, Eva, and Carsten Eulitz. 2020. Can you reach for the planets or grasp at the stars? Modified noun, verb, or preposition constituents in idiom processing. In The Role of Constituents in Multidword Expressions. Edited by Sabine Schulte im Walde and Eva Smolka. Berlin: Language Science Press, pp. 179–204.
Smolka, Eva, Stefan Rabanus, and Frank Rösler. 2007. Processing verbs in German idioms: Evidence against the configuration hypothesis. Metaphor and Symbol 22: 213–31. [CrossRef]
Sonbul, Suhad. 2015. Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. Bilingualism: Language and Cognition 18: 419–37. [CrossRef]
Sprenger, Simone A., Willem J.M. Levelt, and Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. Journal of Memory and Language 54: 161–84. [CrossRef]
Steinel, Margarita P., Jan H. Hulstijn, and Wolfgang Steinel. 2007. Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. Studies in Second Language Acquisition 29: 449–84. [CrossRef]
Swinney, David A., and Anne Cutler. 1979. The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior 18: 523–34. [CrossRef]
Tabossi, Patrizia, Kinou Wolf, and Sara Koterel. 2009a. Idiom syntax: Idiosyncratic or principled? Journal of Memory and Language 61: 77–96. [CrossRef]
Tabossi, Patrizia, Rachele Fanari, and Kinou Wolf. 2009b. Why are idioms recognized fast? Memory & Cognition 37: 529–40. [CrossRef]
Titone, Debra A., and Cynthia M. Connine. 1994. Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. Metaphor and Symbol 9: 247–70. [CrossRef]
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 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]
Titone, Debra A., Maya Libben, Julie Mercier, Veronica Whitford, and Irina Pivneva. 2011. Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. Journal of Experimental Psychology: Learning, Memory, and Cognition 37: 1412. [CrossRef]
Titone, Debra A., Georgie Columbus, Veronica Whitford, Julie Mercier, and Maya Libben. 2015. Contrasting bilingual and monolingual idiom processing. In Bilingual Figurative Language Processing. Edited by Roberto R. Heredia and Anna B. Cieslicka. Cambridge: Cambridge University Press, pp. 171–207.

Titone, Debra A., Kyle Lovseth, Kristina Kasparian, and Mehrgol Tiv. 2019. Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimenterale 73: 216–30. [CrossRef]

Tomasello, Michael. 2003. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge: Harvard University Press.

Tremblay, Antoine, and R. Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In Perspectives on Formulaic Language: Acquisition and Communication. Edited by David Wood. London and New York: Continuum, pp. 151–73.

Tremblay, Antoine, and Benjamin V. Tucker. 2011. The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. The Mental Lexicon 6: 302–24. [CrossRef]

Tremblay, Antoine, Bruce Derwing, Gary Libben, and Chris Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. Language Learning 61: 569–613. [CrossRef]

Underwood, Geoffrety, Norbert Schmitt, and Adam Galpin. 2004. The eyes have it: An eye-movement study into the processing of formulaic sequences. In Formulaic Sequences: Acquisition, Processing, and Use. Edited by Norbert Schmitt. Amsterdam and Philadelphia: John Benjamins, pp. 153–72.

van Ginkel, Wendy, and Ton Dijkstra. 2020. The tug of war between an idiom’s figurative and literal meanings: Evidence from native and bilingual speakers. Bilingualism: Language and Cognition 23: 131–47. [CrossRef]

van Hell, Janet G., and Ton Dijkstra. 2002. Foreign language knowledge can influence native language performance in exclusively native contexts. Psychonomic Bulletin & Review 9: 780–89. [CrossRef]

Van Lancker Sidtis, Diana. 2003. Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. Applied Psycholinguistics 24: 45–57. [CrossRef]

Van Lancker Sidtis, Diana, and Gail Rallon. 2004. Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. Language & Communication 24: 207–40. [CrossRef]

Weinert, Regina. 1995. The role of formulaic language in second language acquisition: A review. Applied Linguistics 16: 180–205. [CrossRef]

Wood, David. 2006. Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. Canadian Modern Language Review 63: 13–33. [CrossRef]

Wray, Alison. 1999. Formulaic language in learners and native speakers. Language Teaching 32: 213–31. [CrossRef]

Wray, Alison. 2000. Formulaic sequences in second language teaching: Principle and practice. Applied Linguistics 21: 463–89. [CrossRef]

Wray, Alison. 2002. Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.

Wray, Alison. 2013. Formulaic language. Language Teaching 46: 316–34. [CrossRef]

Wulff, Stefanie. 2008. Rethinking Idiomaticity: A Usage-Based Approach. New York: Continuum.

Yamashita, Junko. 2018. Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. Journals of Second Language Studies 1: 60–78. [CrossRef]

Yorio, Carlos A. 1989. Idiomaticity as an indicator of second language proficiency. In Bilingualism across the Lifespan. Edited by Kenneth Hyltenstam and Loraine K. Obler. Cambridge: Cambridge University Press, pp. 55–72.

Zeng, Tianjiao, Holly P. Branigan, and Martin J. Pickering. 2020. Do bilinguals represent between-language relationships beyond the word level in their lexicon? Journal of Neurolinguistics 55: 100892. [CrossRef]

Zughoul, Muhammad Raji. 1991. Lexical choice: Towards writing problematic word lists. International Review of Applied Linguistics 29: 45–60. [CrossRef]