ORIGINAL ARTICLE

Effects of aging and noncanonical form presentation on idiom processing: Evidence from eye tracking

Katja I. Haeuser[1],* [iD], Shari Baum[2] and Debra Titone[2]

[1]Saarland University, Saarbrücken, Germany and [2]McGill University, Montreal, Canada
*Corresponding author. E-mail: khaeuser@coli.uni-saarland.de.

Abstract

Comprehending idioms (e.g., *bite the bullet*) requires that people appreciate their figurative meanings while suppressing literal interpretations of the phrase. While much is known about idioms, an open question is how healthy aging and noncanonical form presentation affect idiom comprehension when the task is to read sentences silently for comprehension. Here, younger and older adults read sentences containing idioms or literal phrases, while we monitored their eye movements. Idioms were presented in a canonical or a noncanonical form (e.g., *bite the iron bullet*). To assess whether people integrate figurative or literal interpretations of idioms, a disambiguating region that was figuratively or literally biased followed the idiom in each sentence. During early stages of reading, older adults showed facilitation for canonical idioms, suggesting a greater sensitivity to stored idiomatic forms. During later stages of reading, older adults showed slower reading times when canonical idioms were biased toward their literal interpretation, suggesting they were more likely to interpret idioms figuratively on the first pass. In contrast, noncanonical form presentation slowed comprehension of figurative meanings comparably in younger and older participants. We conclude that idioms may be more strongly entrenched in older adults, and that noncanonical form presentation slows comprehension of figurative meanings.

Keywords: aging; eye tracking; idioms; sentence processing; reading

Language processing is an important part of human behavior that enables communication among individuals, and is crucial for nearly all social encounters. One important question is how language processing changes over the life span, particularly in healthy older adults. On the one hand, healthy aging (i.e., aging in the absence of neurological disorders such as dementia) is known to bring a myriad of changes in so-called fluid cognitive functions, including working memory capacity (Caplan & Waters, 1999, 2007; Just & Carpenter, 1992), processing speed (Salthouse, 1996), or inhibitory function (Hasher & Zacks, 1988). On the other hand, healthy aging is known to lead to greater general world and lexical–semantic

CrossMark

knowledge (Burke & Shafto, 2008; Hedden & Gabrieli, 2004), often referred to collectively as "frozen" or crystallized cognitive abilities.

Here, we report on one type of language processing that draws on both fluid and crystallized aspects of cognition: the figurative processing of idiomatic expressions (e.g., *break the ice* and *spill the beans*). Traditionally, idioms have been defined as multiword expressions whose figurative meaning is distinct from their literal interpretation (Cacciari, 2014; Cacciari & Glucksberg, 1991; Nunberg, Sag, & Wasow, 1994; Swinney & Cutler, 1979). According to noncompositional models of processing, idioms are likened to "long words," which are semantically unanalyzable and syntactically frozen (i.e., the noncompositional view of idioms; see Bobrow & Bell, 1973; Gibbs, 1980; Swinney & Cutler, 1979). However, this rather simple definition is insufficient for characterizing the full range and variety of linguistic forms that may be classified as idioms (reviewed in Libben & Titone, 2008; Titone & Connine, 1999). For example, the noncompositional view of idioms cannot easily explain why speakers show considerable agreement as to which idioms are syntactically flexible and which are not. In addition, there is evidence that idioms are syntactically modifiable without disrupting comprehension of their idiomaticity (e.g., *He didn't spill a single bean* and *Those beans, she didn't spill*; Titone & Connine, 1999). Finally, there is evidence that the component words of some idioms map onto their figurative meanings in a semantically transparent way (e.g., *break the ice* and *steal the show*; Cutting & Bock, 1997; Gibbs & Nayak, 1989; Gibbs, Nayak, & Cutting, 1989; Glucksberg, 2001; McGlone, Glucksberg, & Cacciari, 1994). Thus, the noncompositional view is insufficient for describing the full variety of idioms and how they may be differentially understood. Therefore, many researchers now embrace a hybrid view of idioms, which characterizes such expressions as multiword sequences that undergo some degree of semantic and syntactic decomposition during comprehension (Cacciari & Tabossi, 1988; Caillies & Butcher, 2007; Libben & Titone, 2008; Sprenger, Levelt, & Kempen, 2006; Tabossi, Fanari, & Wolf, 2008; Titone, Columbus, Whitford, Mercier, & Libben, 2015; Titone & Connine, 1999; Titone & Libben, 2014; Titone, Loveseth, Kasparian, & Tiv, 2019).

Crucial here is that idioms, as a class of language, are also interesting from the perspective of cognitive aging. On the one hand, idiom comprehension requires both lexical–semantic knowledge to retrieve the lexicalized configuration and its meanings from memory. On the other hand, idioms likely require some degree of executive control to inhibit unwanted activation of literal word meanings and maintain a representation of the figurative form and meaning in memory (see Bohrn, Altmann, & Jacobs, 2012; Haeuser, Titone, & Baum, 2016; Papagno, Lucchelli, Muggia, & Rizzo, 2003; Rapp, Mutschler, & Erb, 2012; Rizzo, Sandrini, & Papagno, 2007; Schettino et al., 2010).

Because of this interplay between lexical–semantic knowledge and executive control, idioms draw upon cognitive components that are normally described as hallmarks of cognitive aging: stability or even increase in lexical–semantic knowledge (*crystallized abilities*) on the one hand (Burke & Shafto, 2008), but impairments in executive functions of cognition on the other hand (*fluid abilities*; Grady & Craik, 2000; Hasher, Stoltzfus, & Zacks, 1991; Li, Lindenberger, & Sikström, 2001). Age-related decline in fluid aspects of cognition have often been described as impairments in processing speed, inhibitory control, working memory,

or context updating (Braver, Paxton, Locke, & Barch, 2009; Hasher, Zacks, & May, 1998; Lindenberger, 2014; Salthouse, 2000; Verhaegen & Cerella, 2002).

Thus, in the present study, the question we ask concerns how healthy aging impacts idiom processing when idioms are presented in a canonical form, which promotes fast form-based, figurative retrieval, compared to a noncanonical form, which disrupts the familiar idiomatic form and would likely emphasize the literal or compositional nature of the phrase. In building to this experiment, we first selectively review the literature on idiom processing in healthy aging, and on processing of idioms in their noncanonical form.

Idiom processing in aging

Previous work on idiom processing in healthy older adults has shown evidence for both age-related decline in some aspects of idiom comprehension, and stability in others. For example, older adults had difficulty deciding whether an idiom had a possible literal interpretation, suggesting a high degree of automatic activation of the figurative meaning (e.g., *drive someone around the bend* vs. *be on cloud nine*; Westbury & Titone, 2011). In addition, older adults did not show priming for literal target words that followed literally plausible idiomatic primes, as did younger adults (i.e., *tie the knot* [prime]–*rope* [literal target]; Grindrod & Raizen, 2019). Similarly, Sprenger, La Roi, and van Rij (2019) found that subjective familiarity ratings for idiomatic expressions significantly increase with age (Study 1). Of interest, in that study the idiom familiarity ratings of older adults were less modulated by absolute corpus frequency than the ratings of younger adults (Study 2), which suggests an age-related increase in idiom familiarity irrespective of whether the idiom is high- or low-frequent in absolute terms.

Taken together, these findings suggest that older adults' crystallized knowledge of figurative expressions is intact, if not improved (see Sprenger et al., 2019), but that older adults have trouble when the experimental task highlights the literal, compositional nature of the phrase (i.e., possibly arising from an impairment in a fluid aspect of cognition, such as executive control). This observation is confirmed by other studies: for example, despite the occasional word finding difficulties that affect idiom naming in old age (Conner et al., 2011), older adults outperform younger adults in idiom production (Hyun, Conner, & Obler, 2014), which also suggests that they rely on their greater crystallized knowledge of language. Consistent with this idea, older adults have shown evidence of greater idiomatic sensitivity than younger adults in some studies (Coane, Sánchez-Gutiérrez, Stillman, & Corriveau, 2014; Hung & Nippold, 2014; Westbury & Titone, 2011). For example, older adults provided lexically more elaborate and semantically rich explanations of idioms, and performed at ceiling in a phrase-to-idiom matching task (Hung & Nippold, 2014), suggesting that knowledge of idiomatic forms and meanings is largely preserved in aging. In a phrase recall task, older adults showed higher false alarm rates for canonical-form idioms; that is, they reported that they recalled seeing *kick the bucket* after seeing an idiom variant such as *kick the pail*, presumably because they were more likely to automatically activate the fixed figurative expression during initial processing of the idiom in its noncanonical form (Coane et al., 2014).

Thus, older adults seem to maintain (or even improve) their knowledge of idiomatic forms, but may have difficulty when tasks emphasize the dual, compositional nature of idioms (i.e., that idioms have both literal and figurative meanings). Therefore, as suggested earlier, one important question concerns how older adults process idioms when they are presented in a noncanonical form that disrupts the figurative configuration and highlights their literal, compositional nature (e.g., *John broke the <u>cracked</u> ice*, *bore his <u>own</u> cross*, *cleared the <u>stale</u> air*)?

Processing of noncanonical idioms

Most existing evidence on the production and comprehension of noncanonical form idioms comes from two sources: corpus studies of idiom production (e.g., Langlotz, 2006; Schröder, 2013; for review see Fellbaum, 2019), and acceptability ratings that focus on speakers' intuitions about noncanonical form idioms (e.g., Geeraert, Newman, & Baayen, 2017a; Gibbs & Nayak, 1989; McGlone, Glucksberg, & Cacciari, 1994; Tabossi, Wolf, & Koterle, 2009). Very few studies focus on the question of how noncanonical idioms are processed online, for example, during reading (see Geeraert, Baayen, & Newman, 2017b; Kyriacou, Conklin, & Thompson, 2019, for exceptions). Moreover, the studies that do exist have yielded somewhat conflicting findings.

In contrast, some studies show that idioms have a great deal of internal variability and can undergo some degree of semantic and syntactic modification and still be understood quickly (e.g., Geeraert et al., 2017a; Smolka & Eulitz, 2019). For example, corpus studies have demonstrated that speakers use idioms very productively in language (e.g., *by and not so large*; *let the tiger/kitten/tomcat out of the bag*; and *spill the kidney beans*; see Fellbaum, 2019; Langlotz, 2006). Similar findings have been obtained by studies that examined acceptability ratings of modified idioms (e.g., *the bucket, John kicked* or *kick the pail*; Tabossi et al., 2009). Despite the fact that some idiom variants tend to be less acceptable than others (e.g., stocks that *go through the ceiling* are judged as slightly less acceptable than stocks that *go through the investment roof*; Geeraert et al., 2017a), there is evidence that idiom variants maintaining at least some of an idiom's constituent words are judged as comparably acceptable to canonical-form idioms.

Similarly, passivized idioms (e.g., *the bucket was kicked*) are read as quickly overall as passivized control phrases (e.g., *the apple was kicked*; see Kyriacou et al., 2019), and priming for target words that are related to the figurative meaning of an idiomatic phrase (e.g., *embarrassment*) is maintained when idioms are topicalized (e.g., *The ice, John broke*, as opposed to *John broke the ice*; see Mancuso, Elia, Laudanna, & Vietri, 2019). Both results suggest that presentation of idiom variants does not globally block access to idiomatic configurations and figurative meanings. In another illustration, Smolka and Eulitz (2019) showed that idiom variants such as *She reached for the planets*, for which a single word of an idiom was replaced by a close semantic associate (in this case, *planets* for the canonical noun *stars*), were more acceptable as paraphrases of the figurative meaning compared to variants that had unrelated word replacements (e.g., *She reached for the candy*). However, both related and unrelated word paraphrases yielded lower acceptability ratings than

canonical form idioms overall, suggesting there might also be limits to idiomatic productivity (at least when there is no pragmatic context available that licenses the use of an idiom variant). Above and beyond these studies, a remaining question is whether findings from acceptability ratings and meta-linguistic judgments readily reflect processes that come to bear during online language processing (e.g., natural reading) when the only task demand is to comprehend the sentence.

With respect to online processing, a study that examined noncanonical form idioms showed that the insertion of an additional word into an idiom leads to longer processing of idiomatic configurations during online reading (even though it should be noted that in that study, not all idiom variants led to slowing). Specifically, Geeraert et al. (2017b) found that insertion of an additional concept (e.g., *stocks that go through the investment roof* or *hear something through the judgmental grapevine*) increased the number of initial fixations on the idiom region during natural reading. This slowing in online processing is interesting because the same idiom variant was rated as highly acceptable in offline ratings (Geeraert et al., 2017a). Taken together, these findings suggest that modification or variation of idiomatic forms can increase comprehenders' processing effort during online comprehension (in contrast to offline acceptability ratings), especially when the idiom variant is not licensed by prior context (Fellbaum, 2019). Thus, despite a large body of research attesting to idiom variability and productivity in offline measures, idiom modification seems to slow online comprehension, at least in younger adults.

The present study

We investigated whether idiom modification impacts natural reading, and crucially, whether it does so in a different manner for younger and older adults. To our knowledge, this is the first study to use eye tracking to investigate age differences in idiom processing. We recorded eye movements as younger and older adults read idiomatic (*break the ice*) and nonidiomatic control phrases (*store the ice*), which allowed us to investigate age-related changes during a natural task that merely asked participants to read for comprehension (i.e., read without having to make acceptability ratings or meta-linguistic judgments). Previous studies using eye tracking with younger adults (e.g., Carrol & Conklin, 2017, Experiments 1 & 2; Cieślicka, Heredia, & Olivares, 2014; Geeraert et al., 2017b; Milburn & Warren, 2019; Siyanova-Chanturia, Conklin, & Schmitt, 2011) have sometimes reported an idiom advantage, that is, faster reading for idiomatic compared to literal control phrases in some measures of eye tracking (e.g., faster processing for *spill the beans* vs. *spill the chips*; Carrol & Conklin, 2017; or faster processing for *at the end of the day* compared to *at the end of the war*; Siyanova-Chanturia et al., 2011).

Critical to the present work is a recent study by Titone et al. (2019), in which idioms and literal control phrases were presented prior to a disambiguating region that biased either a figurative or a literal interpretations of the phrase (e.g., figurative: *Penelope* hit the books *only two weeks before the dreaded final examination* vs. literal: *Penelope* hit the books *with her hand when she sat down quickly at the desk*). This design allowed the researchers to assess which meanings of an idiom (literal or figurative) readers integrated during first pass reading, in the absence

of any prior context. Nonidiomatic control sentences (e.g., *Penelope put the books on the shelf when she returned home from the library*) were also presented. Here, we use the same materials as the Titone et al. (2019) study in order to ask (a) How do older adults process idioms in these conditions; and (b) what happens when idioms are read in a noncanonical form, that is, with a modifier that disrupts the canonical configuration (e.g., *break the cracked ice* and *kick the black bucket*)?

Our rationale for inserting a modifier was that idioms presented in their noncanonical form would disrupt the familiar idiomatic configuration, thus creating greater semantic conflict between literal and figurative meanings of the phrase. We reasoned that such increased literal/figurative conflict may be particularly challenging for older adults, for whom figurative forms may be more entrenched (Grindrod & Raizen, 2019; Sprenger et al., 2019). In both conditions, we presented idioms that varied with respect to prior ratings of their subjective familiarity (taken from Libben & Titone, 2008). We expected that idiom processing would be easier for idioms rated as high versus low familiar, though the effects of familiarity might be reduced for older relative to younger adults because of their greater life-long exposure to idiomatic forms.

Our general hypotheses regarding older adults were as follows. First, given prior research showing evidence for an age-related entrenchment of idiomatic forms (Grindrod & Raizen, 2019; Sprenger et al., 2019; Westbury & Titone, 2011), we expected that older adults would show more slowing (indicating comprehension difficulties) when reading sentences that bias a canonical-form idiom toward its literal meaning (e.g., *Penelope hit the books with her hand when she sat down quickly at the desk*).

Regarding noncanonical form presentation of idioms, there were three possible outcomes pertinent to the age comparison. To the extent that greater (i.e., more crystallized) language knowledge predominates during idiom processing, we would expect that the presence of a modifier would slow idiom reading in older adults more than in younger adults. This hypothesis can also be derived from studies that argue that idiomatic configurations are more entrenched in older adults (Sprenger et al., 2019; Westbury & Titone, 2011). A second possible outcome was that idioms are so thoroughly lexicalized in older adults that they can still be accessed quickly in memory, despite the presence of a modifier. A third possible outcome was that the effects of noncanonical form presentation somehow interact with idiom familiarity, so that, for example, the modifier is less likely to have a negative effect in low-familiar idioms because their figurative configurations are less prominent and the literal meaning is more likely to be entertained during processing.

Method

Participants

Twenty-one native-English-speaking older adults participated for compensation at a rate of CAD $10/hr. The control group consisted of 25 native-English-speaking younger adults, a subset of younger participants included in Columbus et al. (2015). Because data acquisition for this study took place in Montreal, a city with a highly

Table 1. Demographic information and language background of younger and older adults (standard deviations in parentheses)

	Younger adults	Older adults
N	25	21
Age	22.80 (3.03)	65.65 (7.86)
Formal education (yrs)	15.98 (2.49)	15.50 (3.10)
L2 age of acquisition	9.26 (6.65)	>18
L2 formal instruction (yrs)	>8	<4

L2, second language.

bilingual population, we collected additional information from all participants regarding their language background. All participants had learned English as the first language from birth without exposure to an L2 before the age of 3. For all participants, English was the main language of instruction during early formal schooling (elementary and high school). All subjects rated English as the dominant language at the time of testing, and their current language exposure at the time of testing was highest for English compared to French or other known languages. All participants were matched on the number of years of formal education and had normal or corrected-to-normal vision, and no self-reported history of speech, hearing, language, and/or neurological/psychiatric disorders. All study procedures were carried out with the approval of the McGill University Research Ethics Board. Written consent was obtained from all participants. Demographic information and language background of younger and older adults is presented in Table 1.

To ensure that the participants examined in this study were norm-typical, we had all participants engage in a cognitive test battery. Due to technical failures during testing, the cognitive test data were incomplete, leaving data from 23 younger and 17 older adults for analysis. Cognitive tests included the AX continuous performance task (AX-CPT) to assess context updating and maintenance (Braver et al., 2001), as well as the Anti-Saccade (Hallet, 1978; for review, see Munoz & Everling, 2004) and Stroop (arrow) tasks to assess inhibition. The critical variable for the AX-CPT task was the reaction time cost score between BX and BY trials per subject (for further information on how this score is computed, see Columbus et al., 2015). Critical variables for the Anti-Saccade and Stroop tasks were the reaction time cost scores between hard and easy trials per subject. Table 2 shows that older adults were significantly slower on all raw reaction time measures (e.g., reaction times on congruent and incongruent trials), but did not perform consistently worse when *cost* scores were taken into account (i.e., scores that set performance on congruent and incongruent trials into relation). Overall, this indicated an age-related impairment in processing speed in our group of older adults, but no impairments in inhibitory aspects of cognition (Anti-Saccade and Stroop Arrow tasks) or in the ability to maintain and update contextual information in working memory (AX-CPT task).

Table 2. Results of the cognitive test battery in younger and older adults

	Younger adults	Older adults	t (38) =	p
N	23	17		
AXCPT: RTs on BX (hard) trials (ms)	379 (121)	564 (142)	−4	.001***
AXCPT: RTs on BY (easy) trials (ms)	341 (90)	473 (124)	−4	<.001***
AX-CPT BX-BY RT Cost (ms)	38 (96)	91 (146)	−1	=.2
Stroop Arrow: RTs congruent trials (ms)	439 (76)	562 (94)	−5	<.001***
Stroop Arrow: RTs incongruent trials (ms)	463 (69)	624 (123)	−6	<.001***
Stroop Arrow RT Cost (ms)	24 (45)	63 (56)	−2	=.02*
Antisaccade RT Pro (easy) (ms)	299 (43)	336 (44)	−3	<.001***
Antisaccade RT Anti (hard) (ms)	352 (68)	408 (69)	−3	=.01*
Antisaccade RT Cost (ms)	53 (49)	71 (61)	1	=.3

NOTE: RT, reaction time. Standard deviations in parentheses. $*p < .05.$ $**p < .05.$ $***p < .001.$

Materials

Experimental materials were 54 familiar English idioms that had a verb–determiner–noun structure (e.g., *kicked the bucket*, *broke the ice*, and *lost his seat*), taken from the idiom corpus provided in Libben and Titone (2008). These idioms formed a subset of the experimental stimuli presented in Titone et al. (2019). In a first step, control phrases were created for each idiom individually by replacing the idiom's verb with another verb of approximately the same length (e.g., *tipped the bucket*, *put the books*, and *left his seat*). Verbs in literal and idiomatic phrases were matched in number of characters ($M_{idiomatic} = 5.31$, $SD_{idiomatic} = 1.59$; $M_{literal} = 5.28$, $SD_{literal} = 1.57$), $t (106) = 0.12$, $p = .9$. They were also matched in frequency, based on the Zipf scale from the SUBTLEX US data base ($M_{idiomatic} = 3.33$, $SD_{idiomatic} = 0.99$; $M_{literal} = 3.31$, $SD_{literal} = 0.88$), $t (106) = 0.09$, $p = .9$ (Brysbaert & New, 2009).

All idioms and literal control phrases were embedded in two-clause sentences, matched in length as closely as possible. The first clause of the sentence contained an agent (always a name; e.g., *Bruce . . .*) and was followed by the idiom (or matched literal control phrase), presented in the past tense (e.g., *Bruce broke the ice . . .* and *Bruce stored the ice . . .*). The second clause of the sentence was a disambiguating region that biased either a figurative reading of the phrase (*Bruce broke the ice by quickly introducing himself to everyone at the wedding*; sentence type Id-Id) or a literal reading of the phrase (*Bruce broke the ice by driving his snowmobile directly onto the thawing lake*; sentence type Id-Lit). In literal control phrases, the second clause continued the sentence in a plausible literal way (*Bruce stored the ice in his cooler so he could bring it to the holiday party*; Lit-Lit condition). Thus, each item had three experimental versions: an idiom biased toward its figurative meaning (Id-Id), an idiom biased toward its literal meaning (Id-Lit), and a literal control sentence (Lit-Lit). A full list of the experimental sentences is presented in the appendix of Titone et al. (2019); example stimuli are presented in Table 3.

Table 3. Example sentences from Id-Id, Id-Lit, and Lit-Lit conditions, including the modifier

Sentence type	Idiom	Modifier	Sentence
Id-Id	Bit the bullet	Iron	Larry bit the (iron) bullet and bought diamond earrings for his wife's birthday.
Id-Lit	Bit the bullet	Iron	Larry bit the (iron) bullet to verify the quality of casings on his ammunition.
Lit-Lit	Bit the bullet	Iron	Larry hid the (iron) bullet so the police would not find the crucial evidence.
Id-Id	Lost his seat	Empty	Otto lost his (empty) seat on the council after he was caught embezzling funds.
Id-Lit	Lost his seat	Empty	Otto lost his (empty) seat at the football game when he got up to get some beer.
Lit-Lit	Lost his seat	Empty	Otto left his (empty) seat unoccupied when he left quickly to use the bathroom.
Id-Id	Blew a fuse	Huge	Sarah blew a (huge) fuse when the material she had studied most wasn't on the test.
Id-Lit	Blew a fuse	Huge	Sarah blew a (huge) fuse when her bedside lamp had a severe electrical malfunction.
Lit-Lit	Blew a fuse	Huge	Sarah fixed a (huge) fuse that always failed when she used the plug beside the sink.

Lexical and idiom characteristics of the 54 idioms used in the present experiment are presented in Table 4.

In a final step, we chose modifiers for each item that fit the semantics of the constituent words in a meaningful way (e.g., *broke/stored the cracked ice, kicked/ tipped the black bucket, changed/hummed her sad tune,* and *bought/loved the old farm*). The goal was to select adjectives that were semantically neutral with respect to a figurative versus literal read of the sentences. On average, inserted adjectives were five characters long (range: 3–10) and relatively high frequency ($M = 3.44$; $SD = 0.84$), based on the Zipf scale from the SUBTLEX-US database (Brysbaert & New, 2009), with a minimum of 1.2 (e.g., *forced her wavering hand*) and a maximum of 4.9 (e.g., *spilled the little beans*). Note that the modifier was identical across all three experimental versions of an item.

The design resulted in six conditions per experimental item (Id-Id with modifier, Id-Id without modifier; Id-Lit with modifier, Id-Lit without modifier; and Lit-Lit with modifier, Lit-Lit without modifier). During the experiment, literal and idiomatic sentences were randomly presented over 6 experimental lists, so that each participant read only one experimental version of each item, for a total of 54 sentences per participant. That is, each list consisted of 18 Id-Id sentences (9 canonical and 9 noncanonical), 18 Id-Lit sentences (9 canonical and 9 noncanonical), and 18 Lit-Lit sentences (9 canonical and 9 noncanonical). Fillers were 80 metaphor sentences from a prior experiment (Columbus et al., 2015), yielding a total of 134 sentences per participant in one testing session.

Table 4. Lexical characteristics of the 54 verb–determiner–noun idioms (taken from Libben & Titone, 2008)

	M	SD	Min	Max
Length (*n* characters)	15	2.10	10	19
Familiarity (1–5)	3.37	0.83	1.67	4.8
Decomposability (proportion)	0.48	0.26	0.05	0.98
Literal plausibility (1–5)	3.81	0.76	1.5	4.86
Verb relatedness (1–5)	2.89	1.08	0.8	4.93
Noun relatedness (1–5)	2.45	1.02	0.54	4.50

Procedure

Participants were tested in the lab for one session that lasted approximately 2 hr. Upon signing the consent form, participants completed a language history questionnaire, then performed the sentence reading task, and subsequently completed the cognitive test battery. For the reading task, participants were informed that they would read sentences on a screen, one at a time, while their eye movements were recorded. Each trial began with a fixation cross, presented in the middle of the screen, followed by a sentence, aligned to the left side of the screen. Participants were instructed to read each sentence silently for comprehension and to press a button on a control pad to indicate when they finished reading the sentence. Yes–No comprehension questions were included on 25% of trials to ensure participants were reading for content.

Participants were tested in a quiet room on an Eye-link 1000 tower-mounted eye-tracking system (SR-Research, Ontario, Canada), using a 21-inch ViewSonic CRT monitor with a screen resolution of 1024 × 768 pixels. Viewing was binocular, but eye-tracking data were collected for the right eye. All experimental sentences were presented aligned to the left side of the screen on a single line, in yellow 10-point Monaco font on a black background. Eye movements were calibrated using a 9-point grid; recalibrations were performed when necessary. Participants' heads were stabilized using a headrest throughout the experiment. Three characters subtended approximately 1 degree of visual angle.

Results

The eye-movement record yields a number of measures that are associated with variations in the processing time-course of a target word or region. These are commonly separated into groups of early and late measures (Radach & Kennedy, 2004; Rayner, 1998, 2009; see also Carrol & Conklin, 2014, for a methodological overview of eye tracking in idiom research). Early measures of reading are supposed to indicate very early, presemantic effects of lexical access in memory (i.e., accessing the form of a lexical entry, irrespective of its meaning). Late measures are thought to tap into later occurring semantic effects, such as comprehension of a word and its integration into the discourse context.

We focused on two early measures and two late measures that are common in eye tracking and idiom research (see Carrol & Conklin, 2014; Cieslicka, 2014; Milburn & Warren, 2019; Rayner, Warren, Juhasz, & Liversedge, 2004; Siyanova-Chanturia et al., 2011; Titone et al., 2019; Warren, McConnel, & Rayner, 2008). Early measures included first-pass gaze duration on phrase-final nouns (i.e., the sum of all fixation durations made on the noun before exiting either to the left or to the right), and go-past times of phrase-final nouns (i.e., the amount of time that the reader looks at a target word as well as any time spent rereading earlier parts of the sentence before moving ahead).[1] Late measures included total reading time (henceforth, TRT) on the idiom region (i.e., the sum of all fixation and refixation durations made within the idiom region, e.g., *kicked/tipped the bucket*), and the proportion of regressive eye movements from the disambiguating region back to earlier regions of the sentence (a binomial variable).

For our statistical analyses, we computed separate linear mixed-effects models for each eye-tracking measure, using the lme4 library in *R* (R Development Core Team, 2018). This method eliminates the need for separate analyses of random effects variables (i.e., separate analyses if variance for subjects and items). In addition, linear mixed-effects models do not require that continuous variables (such as idiom familiarity or idiom decomposability) be artificially categorized (i.e., there is no requirement for artificial median splits of continuous factors, except perhaps to simplify presentation of the data).

We included three predictors for each model. These were *sentence type* (idiomatic vs. literal), *age group* (younger vs. older adults) and *idiom familiarity* (a scaled continuous variable), including all two- and three-way interactions among these variables. We present findings from canonical form idioms (*broke the ice*) separately from findings from noncanonical form idioms (*broke the cracked ice*). The reason for reporting results from the modified sentences in a separate section is that those sentences are fundamentally different in a way that has direct bearing on how the target word or phrase is read. Because all noncanonical sentences had inserted adjectives that canonical sentences did not have, all noncanonical sentences were systematically longer, subject to different parafoveal preview experiences, and the target region was different in terms of lexical characteristics.

In all models, *sentence type* was a categorical factor consisting of two levels (literal and idiomatic) for early reading measures (gaze duration [GD] noun and go-past noun), given that readers encountered idioms or control phrases *prior* to accessing the disambiguating region. In contrast, for TRT idiom and regressions out of the disambiguating region, there were three levels of the factor sentence type (Lit-Lit, Id-Lit, and Id-Id), given that these measures included fixations that readers made after accessing the disambiguating region. To compare all three factors with one another, there were necessarily two versions of each model for the late reading measures.

In the first version, Lit-Lit was the baseline, enabling two fixed effect comparisons: (a) Id-Id against the baseline Lit-Lit, and (b) Id-Lit against the baseline Lit-Lit. In the second version, Id-Id was the baseline, also enabling two fixed effect comparisons: (a) Id-Lit against the baseline Id-Id, which was our new contrast of interest, and (b) Lit-Lit against the baseline Id-Id, which was perfectly redundant

with the first model version and, by-definition, always had the same qualitative outcome. Thus, in our reporting below, we exclusively focus on the Id-Lit versus Id-Id contrast from the second model version.

Of note, while researchers have at their disposal many methods of regression coding for factors involving three or more levels (e.g., treatment coding with refit baseline, or alternatively Helmert or backward difference coding; Schad, Vasishth, Hohenstein, & Kliegl, 2020), only the approach used here (to our knowledge) makes it possible to statistically evaluate all three contrasts, which was crucial for our experimental goals (i.e., Helmert coding compares the mean of two levels of a factor against another; backward difference coding compares the first two levels of a factor, followed by the next two levels of that factor; thus, neither coding scheme was appropriate for our intended experimental manipulation and hypotheses). The fixed effect for age group in all models was also treatment coded (with younger adults as the reference category). Note that in the event that any treatment-coded model failed to converge, we substituted deviation coding, followed by post hoc treatment-coded submodels to decompose any significant interactions.

Of importance, because the majority of models reported below had treatment-coded categorical predictors (where levels of a factor are compared to a fixed reference category) model coefficients show simple, rather than analysis of variance-style main effects. Thus, *b*'s indicate the estimated difference in reading times between the baseline category and the comparison category. For example, when $b = 80$ for the factor age group, this means that there was an 80 ms difference in reading times of older adults compared to younger adults.

In addition to the fixed effects mentioned above, all models contained global idiom decomposability as a control variable (this value was scaled and taken from Libben & Titone, 2008). We also controlled for the length (in number of characters) of the respective region in each model (Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner & Duffy, 1986). Both control variables were scaled (i.e., converted to standard scores with $M = 0$ and $SD = 1$). All reported models contained random intercepts for subjects and items, and the findings reported below did not change when these models were refit to include the maximal random effects structure warranted by the design (see Barr, Levy, Scheepers, & Tily, 2013, for method used), or when continuous DVs were log-transformed (Gelman & Hill, 2007). *P* values in all model reports were calculated using the Satterthwaite approximation, as implemented in the *lmerTest* package in *R*. Confidence intervals for significant model parameters indicate 2.5% and 97.5% limits and were estimated using the Wald method in *R*'s core function *confint*.

Behavioral results: Comprehension questions

Accuracy of the comprehension questions was high in both younger ($M = 0.96$, range: 0.8–1) and older adults ($M = 0.94$, range: 0.76–1), with no significant differences between the groups, $t(44) = -1.13$, $p = .26$. Thus, we may conclude that participants were attentive during the experiment, and successfully read the sentences for comprehension.
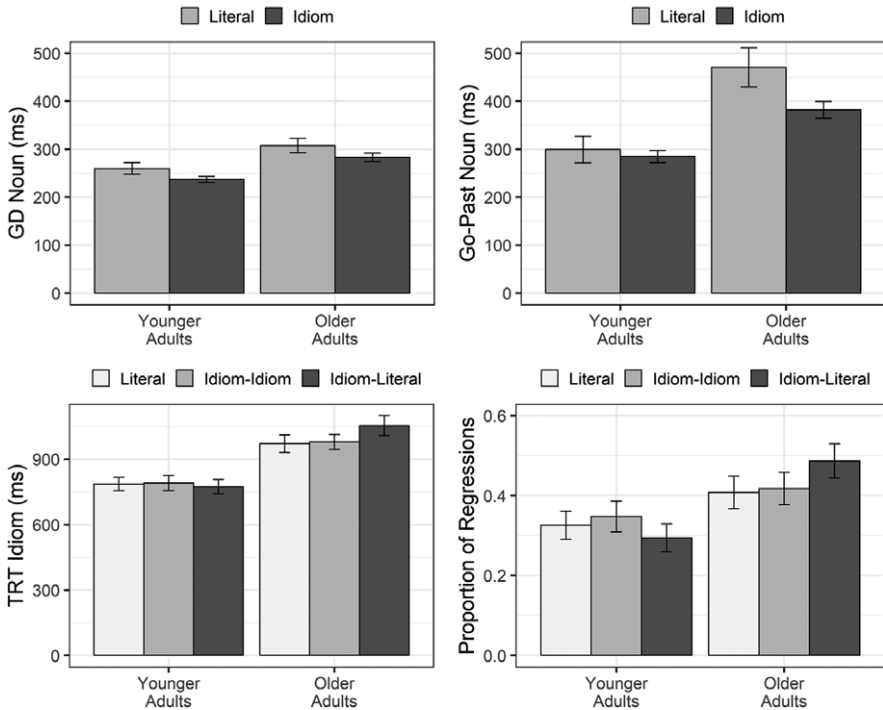
Figure 1. Canonical idioms (e.g., *broke the ice* and *kicked the bucket*) and literal control phrases (e.g., *stored the ice* and *tipped the bucket*): gaze durations and go-past times on the phrase-final noun (both in ms), as well as TRT idiom (ms) and proportion of regressions out of the disambiguating region. Error bars indicate standard errors of the mean, adjusted for the within-subjects factor sentence type (Morey, 2008).

Reading times for canonical form idioms (broke the ice)

GD for phrase-final nouns

There was a simple effect of age group, $b = 43.63$, 95% confidence interval (CI) [4.76, 82.50], $SE = 19.83$, $t = 2.20$, $p < .05$ (see Figure 1, upper left panel), indicating that older adults fixated phrase-final nouns approximately 44 ms longer than younger adults, regardless of whether sentences were idiomatic or literal. There was also a simple effect of sentence type, $b = -21.47$, 95% CI [−41.18, −1.75], $SE = 10.06$, $t = -2.13$, $p < .05$, suggesting that fixations on idiom nouns (e.g., *Mary kicked the* bucket …) were consistently shorter than fixations on nouns of matched literal phrases (e.g., *Mary tipped the* bucket …), regardless of age group or idiom familiarity. No other effects were significant.

Go-past times for phrase-final nouns

There was a significant interaction between sentence type and age group, $b = -81.33$, 95% CI [−144.42, −18.24], $SE = 32.19$, $t = -2.53$, $p < .05$. Figure 1 (upper right panel) suggests that this interaction was driven primarily by older
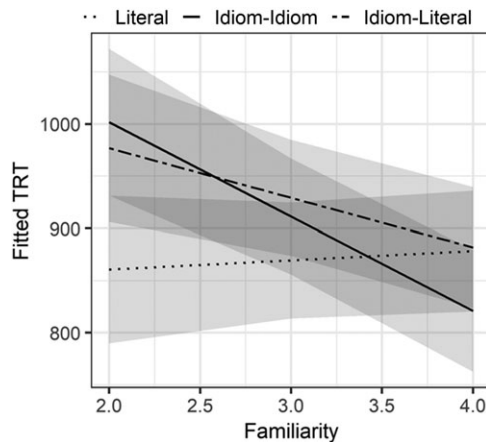
Figure 2. Partial effects plots for TRT idiom in canonical items (*kick the bucket*), illustrating the relationship between sentence type and familiarity. Gray bands represent 95% confidence intervals.

adults, who were relatively faster when reading nouns in idioms versus literal phrases (*kicked the bucket* vs. *tipped the bucket*), regardless of idiom familiarity. In contrast, younger adults' reading times showed no difference between idioms and the literal condition. This interpretation of the data was confirmed by post hoc follow-up models splitting the data by age group. There was a significant simple effect of sentence type only in the model for older adults, $b = -90.46$, 95% CI [−143.84, −37.07], $SE = 27.24$, $t = -3.32$, $p < .001$, but not in the model for younger adults, $b = -6.78$, $SE = 19.04$, $t = -0.36$, $p > .1$.

Total reading time of the idiom
For total reading time, the first version of the model (baseline: Lit-Lit) showed a significant interaction between sentence type Id-Id and idiom familiarity, $b = -97.00$, 95% CI [−173.62, −20.37], $SE = 39.10$, $t = -2.48$, $p < .05$. The partial effects plot for this interaction (see Figure 2) shows that, irrespective of age group, high-familiar idioms presented with a figuratively biasing context region were read faster than their matched literal items (Lit-Lit items), whereas low-familiar idioms were read more slowly. The second version (baseline: Id-Id) showed no additional effects of interest.

Proportion of regressions out of the disambiguating region
Because treatment-coded models failed to converge for this dependent variable, we deviation coded all categorical factors for this generalized linear model (i.e., each level of a condition was compared to the grand mean of the model rather than another reference level of condition).

The first model version showed two main findings of interest. First, there was a significant interaction between sentence type Id-Lit (vs. the grand mean) and age group, $b = 0.83$, 95% CI [0.09, 1.57], $SE = 0.38$, $z = 2.19$, $p < .05$. Second, there was a significant interaction between sentence type Id-Id (vs. the grand mean)
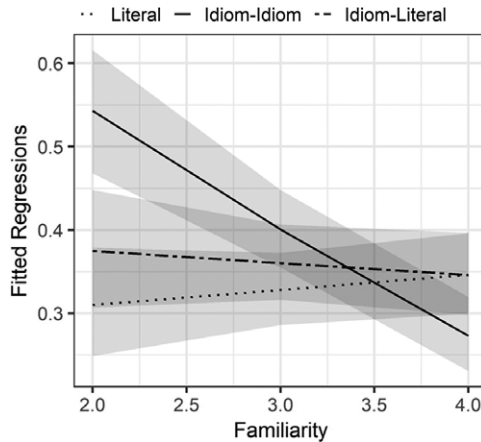
Figure 3. Partial effects plot for proportion of regressions out of the disambiguating region in canonical-form idioms (*kick the bucket*), illustrating the relationship between sentence type and familiarity. Gray bands represent 95% confidence intervals.

and familiarity, $b = -0.63$, 95% CI [−1.01, −0.25], $SE = 0.19$, $z = -3.27$, $p < .01$. To better understand the source of these interactions, we split items by sentence type (i.e., Lit-Lit, Id-Id, and Id-Lit) and performed a series of post hoc treatment coded models (which now converged). The model for the literal condition (Lit-Lit) showed no simple effects or interactions. However, the model for the idiomatic condition, Id-Id, yielded a main effect of familiarity, $b = -0.55$, 95% CI [−0.93, −0.18], $SE = 0.19$, $z = -2.90$, $p < .01$, indicating that, regardless of age group, participants were less likely to regress out of figuratively biasing context regions to the extent that the preceding idiom increased in familiarity, irrespective of age group (see Figure 3). Finally, the model for the Id-Lit condition showed a main effect of age group, $b = 1.06$, 95% CI [0.49, 1.63], $SE = 0.29$, $z = 3.66$, $p < .001$, indicating that older adults showed a greater proportion of regressions when the disambiguating region biased an idiom toward its literal interpretation (see Figure 1, bottom right panel). The second model version showed no additional effects or interactions.

Summary of findings for canonical idioms
The data for canonical form idioms yielded three key findings. First, there was early facilitation for idioms (GD noun), in particular for older adults (go-past noun), suggesting that older adults may have accessed idiomatic configurations in memory more quickly as a group. Second, there was late contextual facilitation as familiarity of idioms increased (TRT idiom; proportion of regressions out of the disambiguating region), across both age groups. Third, older but not younger adults had difficulty reading subsequent contexts that biased idioms' literal interpretations (proportion of regressions), suggesting that they exhibited special difficulty suppressing an idiom's figurative interpretation when a following context called for them to do so.
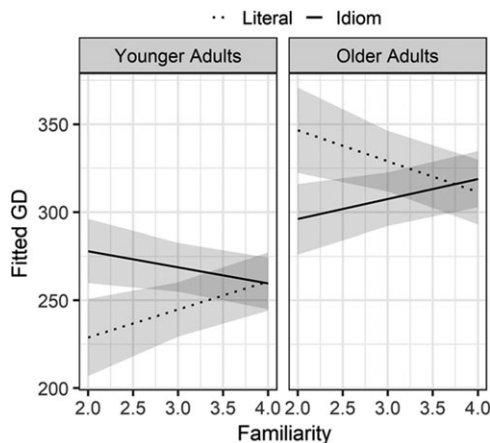
Figure 4. Partial effects plot for gaze durations on the noun in noncanonical items (*kick the black bucket*), illustrating the relationship between sentence type and familiarity. Gray bands represent 95% confidence intervals.

Reading times of noncanonical form idioms (broke the cracked ice)

GD for phrase-final nouns
The model for GD nouns showed a significant three-way interaction between sentence type, age group, and idiom familiarity, $b = 44.30$, 95% CI [11.31, 77.30], $SE = 16.84$, $t = 2.63$, $p < .01$ (see Figure 4). To investigate the source of this interaction, we computed post hoc follow-up models that split subjects by age group.[2] The model for older adults showed no significant simple effects or interactions, but the model for younger adults showed a significant interaction between sentence type and idiom familiarity, $b = -20.69$, 95% CI [−39.89, −1.49], $SE = 9.80$, $t = -2.11$, $p < .05$. The partial effects plot of this interaction (Figure 4) shows that younger adults read nouns in noncanonical idioms more quickly as familiarity increased. No other effects were significant.

Go-past times on phrase-final nouns
There was a significant simple effect of age group, $b = 141.10$, 95% CI [73.22, 208.97], $SE = 34.63$, $t = 4.07$, $p < .001$ (see Figure 5), upper right panel), indicating that older adults had longer reading times overall. No other effects were significant.

Total reading time on the idiom
The first version of the model (baseline: Lit-Lit) showed a significant simple effect of age group, indicating longer reading times for older adults overall, $b = 412.29$, 95% CI [111.11, 713.46], $SE = 153.66$, $t = 2.68$, $p < .01$ (see Figure 5, bottom left panel). There was also a significant simple effect of sentence type Id-Id, $b = 148.28$, 95% CI [39.16, 257.41], $SE = 55.68$, $t = 2.66$, $p < .01$ see Figure 5, bottom left panel), indicating that noncanonical form idioms biased toward their figurative
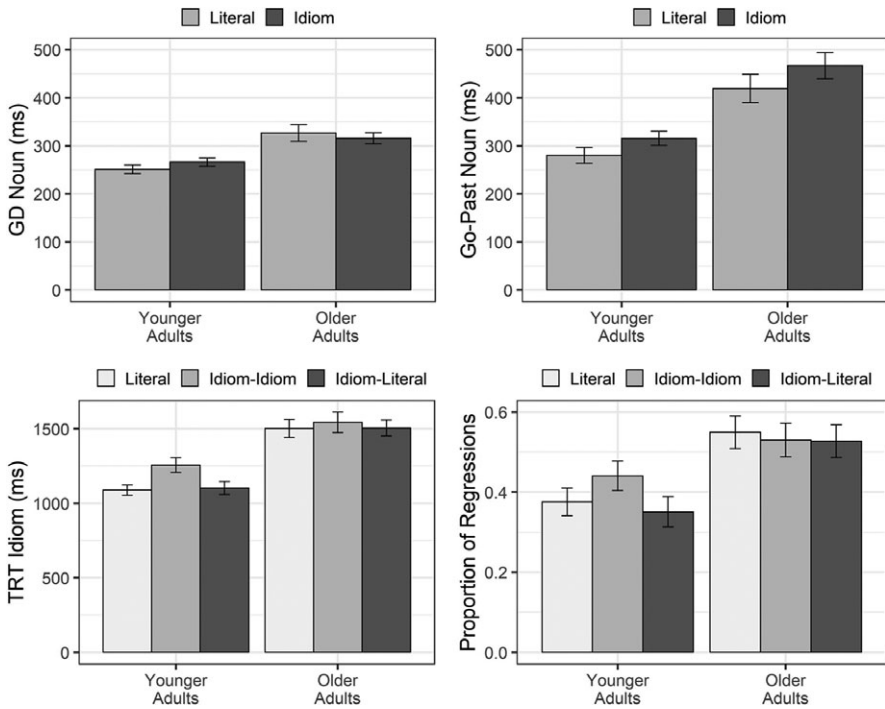
Figure 5. Noncanonical idioms (e.g., *broke the cracked ice* and *kicked the black bucket*) and literal control phrases (e.g., *stored the cracked ice* and *tipped the black bucket*): gaze durations and go-past times on the phrase-final noun (both in ms), as well as TRT idiom (ms) and proportion of regressions out of the disambiguating region. Error bars indicate standard errors of the mean, adjusted for the within-subjects factor sentence type (Morey, 2008).

interpretation were read more slowly than literal control sentences (Lit-Lit), for both age groups. No other effects were significant.

The second model version (baseline: Id-Id) showed a significant simple effect of sentence type Id-Lit, $b = -137.80$, 95% CI [$-247.77, -27.80$], $SE = 56.11$, $t = -2.56$, $p < .05$, indicating that noncanonical idioms presented in sentences that had a literally biasing disambiguating region (Id-Lit) were read more quickly than noncanonical idioms presented in sentences that had a figuratively biasing disambiguating region (Id-Id; see Figure 5, bottom left panel). In sum, the TRT data demonstrated that both age groups had integration difficulties when reading non-canonical idioms, when their subsequent context was biased toward the idioms' figurative meaning.

Proportion of regressive eye movements out of the disambiguating region
As with canonical idioms, treatment coded models failed to converge for this dependent variable, thus we substituted deviation coding (which compared each level to the grand mean). There was a significant simple effect of age group, indicating a greater proportion of regressive eye movements for older versus younger adults

overall, $b = 0.82$, 95% CI [0.15, 1.49], $SE = 0.34$, $z = 2.40$ (see Figure 5; right bottom panel). No other effects were significant.

Summary of findings for noncanonical idioms
Presenting idioms in a noncanonical form (i.e., with an adjective before the phrase-final noun, e.g., *kicked the black bucket* and *broke the cold ice*) led to processing difficulty for both age groups. Noncanonical idioms that had disambiguating regions biased toward their figurative meaning were read more slowly than literal control sentences (TRT idiom), and more slowly than the same noncanonical idioms that had disambiguating regions biased toward their literal interpretation (TRT idiom). Based on these data, we conclude that both age groups primarily accessed literal meanings when reading noncanonical idioms on the first pass, such that they experienced comprehension difficulties when later encountering a disambiguating region biased toward the figurative meanings.

Discussion

We investigated how younger and older adults naturally read idioms presented in their canonical form (e.g., *break the ice*) or their noncanonical form (e.g., *break the cracked ice*) using eye movement measures of reading. Specifically, younger and older adults read idioms at the beginnings of sentences (e.g., *Bruce broke the [cracked] ice …*) that were followed by disambiguating regions which biased the idioms' figurative or literal interpretations (e.g., figurative: *… by quickly introducing himself to everyone at the wedding*; literal: *… by driving his snowmobile directly onto the thawing lake*). Nonidiomatic control sentences were also presented that were matched to the idioms in length by altering the verb of the idiom and leaving all else the same (e.g., *Bruce stored the cracked ice in his cooler so he could bring it to the holiday party*). These verbs were specifically selected to avoid semantic ambiguity in the literal condition (see also Columbus et al., 2015; Titone et al., 2019). There were several key findings.

We expected all participants to have difficulty reading idioms in their noncanonical form, but expected this effect to be enhanced for older adults who may represent idioms in a more figuratively entrenched manner given prior work suggesting that would be the case (e.g., Geeraert et al., 2017b). These predictions regarding canonical-form idioms were borne out to some degree, in that there was evidence for increased entrenchment of idiomatic forms in early and late measures of reading for older adults. In contrast, our predictions regarding noncanonical idioms and aging were not borne out: the modifier had similar effects in younger and older adults' sentence reading in that it slowed comprehension globally, but not disproportionally in older adults.

Age-related differences in idiom comprehension globally
Similar to prior work, we found that age-related processing difficulties with idiomatic sequences emerged primarily in sentences that later emphasized the dual nature of an idiom's meaning (e.g., when an idiom is presented with a literally

biasing context region). In contrast, when idioms were presented in sentences that later emphasized their figurative meanings, we found evidence for greater entrenchment of figurative forms in older adults. Specifically, during early access, older adults read nouns in idioms more quickly than nouns in literal control phrases, a finding that was not evident in younger adults. During late-stage integration, older adults showed similar effects of familiarity as younger adults, in that idioms presented in sentences that had later figurative biases were read faster than literal control sentences as familiarity of the idiom increased. Of note, age group did not further modulate these effects. We believe this pattern is consistent with Sprenger et al. (2019), who found that familiarity ratings of idioms generally increase with age, irrespective of whether the idiom is high or low frequent as estimated by corpus measures of frequency. Thus, this aspect of our findings suggests that semantic representations of figurative meanings, and how they are accessed in memory, are intact, or even improve with age, maybe because older adults have a life-long experience with language at their disposal that has entrenched idiomatic expressions (see Cacciari, Corrardini, & Ferlazzo, 2018, for evidence suggesting that higher levels of verbal knowledge facilitate first-pass idiom recognition).

However, older adults did exhibit reading difficulties when the dual nature of idioms was emphasized. This was apparent primarily in later measures of reading where older adults exhibited difficulty integrating canonical idioms read in sentences where a later disambiguating region was biased toward the literal reading of the phrase (i.e., canonical Id-Lit items). This result implies that older adults primarily accessed figurative meanings when reading the idiom on the first pass, and became confused when they reached a literally biased disambiguating region whose interpretation could not be aligned with the figurative meaning. Collectively, this suggests that older adults may have intact figurative meaning representations of idioms, but that they have difficulty managing the consequences of simultaneously activating literal and figurative meanings of idioms when forces conspire to make that happen. Of note, this aspect of the data aligns with studies demonstrating age differences primarily in tasks that capitalize on quick, rapid processes during online language comprehension (e.g., Grindrod & Raizen, 2019), and in tasks that emphasize the dual nature of idiomatic expressions (Westbury & Titone, 2011).

Noncanonical versus canonical form idioms

We were also interested in the online processing of noncanonical versus canonical form idioms given evidence of idiom variability and productivity in corpus studies (Fellbaum, 2019; Langlotz, 2006), and in offline ratings from native speakers (Geeraert et al., 2017a; Gibbs & Nayak, 1989; McGlone et al., 1994; Tabossi et al., 2009). Here, the findings indicated that inserting an adjectival modifier into an idiom (e.g., *kick the black bucket* and *break the cracked ice*) induced processing difficulties during reading for both younger and older adults. This slowing in reading times (compared to entirely literal Lit-Lit sentences) was approximately 160 ms in total reading time of the idiom, when data from younger and older adults were collapsed. When noncanonical idioms were presented with a figuratively biasing disambiguating region, readers spent more time fixating idioms, and were more likely to regress back to the idiom, compared to when an adjective was inserted into

a nonidiomatic control phrase (e.g., *tip the black bucket*). Conversely, noncanonical idioms presented with literally biasing disambiguating regions were read more quickly than the same idiom presented with figuratively biasing disambiguating regions.

Collectively, this data pattern suggests that readers primarily accessed the literal meanings of idioms when encoding noncanonical expressions on the first pass, such that when they encountered a literally biased disambiguating region, reading was fast, but when they encountered a figuratively biased disambiguating region, reading was impeded. Of note as well, this effect was present for both age groups, suggesting that the modifier disrupted the canonical configuration of the idiom and induced slower reading times in both younger and older adults alike (see Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013, for converging young-adult results on adjectival insertion in lexical bundles; e.g., *in the hands of* vs. *in the capable hands of*).

Thus, despite the suggestion of corpus studies that idioms often undergo a great deal of variation in language production (Fellbaum, 2019; Langlotz, 2006), and the observation that idiom variants can still be understood figuratively in offline plausibility or meaningfulness judgments (e.g., Smolka & Eulitz, 2019; Tabossi et al., 2009), the current findings from online reading suggest that noncanonical form presentation can impede online comprehension of the figurative meaning, at least when there is no prior discourse context available that could bias or pragmatically license use of an idiom variant. Hence, our data suggest that there are limits to idiomatic variability, at least when it comes to online processing of such expressions. This leads us to ponder potential factors that could explain why our findings differ from that of past work.

Connection with past research

As with any study, several additional considerations must be addressed when integrating present and past data. One important consideration when comparing the present results to past work is task demands, which differ substantially across online and offline studies. Most prior experimental studies assessing idiom variation and productivity have used acceptability judgments (Gibbs et al., 1989; Tabossi et al., 2009) or similarity ratings between idioms and paraphrases (McGlone et al., 1994; Smolka & Eulitz, 2019), which tap into native speakers' overt intuitions about the acceptability of certain phrases. In such ratings, people are asked to establish a post hoc mapping between an idiom and its variant. These conditions might not be accurate reflections of online language processing, where linguistic input unfolds rapidly, and such that readers may have little time to search for hidden or obscure semantic relationships. Consistent with this view, Geeraert et al. (2017a) collected both offline acceptability ratings and online reading measures for modified idioms such as *hear something through the judgmental grapevine*. Despite the fact that the native speakers tested rated this specific idiom variant as highly acceptable offline, their reading data indicated slowing during online processing (Geeraert et al., 2017b).

These findings echo advances in linguistic theory positing that meta-linguistic grammaticality or acceptability judgments have limited quantitative reliability, and that the nature of these judgments renders them inadequate proxies for online

language processing (Gibson & Fedorenko, 2013; Linzen & Oseki, 2018; Schütze, 2016). Thus, to advance psycholinguistic knowledge and theory about idioms (as well as all other linguistic phenomena), we believe that the field needs studies that smartly combine results from meta-linguistic judgments and online processing data.

Another important consideration when comparing experimental studies of idiom productivity is pragmatic context. We know that prior discourse can alter idiom comprehension and justify use of variants (Fellbaum, 2019; Langlotz, 2006). A classic example is the 1988 *New York Times* editorial entitled "On being wrong: Convicted minimalist spills bean" where US-writer Frederick Barthelme defends his minimalist writings (and the ones of others) against allegations of insufficient depth of character and lack of big ideas. In this particular example, the use of the variant "spills bean" is perfectly motivated by the prior context of the word "minimalist." For conditions when prior context sufficiently licenses the use of an idiom variant, there may be few limits to idiom productivity (Fellbaum, 2019). For example, prior context allows for modification of noncompositional idioms such as *fall off the wagon* or *kick the bucket*, for which there is no obvious relationship between literal and figurative meanings (e.g., *falling off the exercise wagon*; *I fell off the daily blog wagon*; *before I fell off that wagon and started smoking again*; or *I am young but have experienced more bucket kicking within my immediate family and circle of family friends than I can shake a fist at*; see Fellbaum, 2019, for more examples).

One puzzling aspect of our data is the question of why older adults failed to show even greater slowing than younger adults when reading noncanonical idioms. We initially hypothesized this would be the case, arising from greater life-long experience with language that should have crystallized words and idiomatic configurations in memory in older compared to younger adults (see Sprenger et al., 2019, for experimental evidence that supports this hypothesis). However, even though our data confirm the notion of greater entrenchment of canonical idiom configurations in older adults (see the results from go-past noun and regressions out of the disambiguating region), the modifier had a similar effect in both age groups, in that it slowed reading comprehension globally.

One potential explanation is that the idioms presented in this study, as a group, were so thoroughly lexicalized, even in younger speakers of English, that the modifier could simply not make a dent in that activation. This interpretation is somewhat bolstered by the gaze duration data on phrase-final nouns in noncanonical phrases, which became faster as idiom familiarity increased. This could suggest that during reading of noncanonical high familiar idioms, people could still access the idiomatic configuration in memory (despite the presence of a modifier) relatively quickly, whereas only in less frequently encountered low-familiar idioms, the modifier resulted in slowing. A different pattern of results might have occurred with metaphors or proverbs, given that these are more compositional in nature and more likely to be built by means of their literal constituents during comprehension, as opposed to holistically retrieved from memory (Columbus et al., 2015).

Another important consideration given our data involves the nature of the non-idiomatic control condition (e.g., *tipped/kicked the bucket*; *stored/broke the ice*; and *hid/bit the bullet*), which consisted of phrases for which an idiom-initial verb had been replaced by a verb of similar length and frequency, which was semantically

distinct from the idiomatic verb. We referred to this condition as a "literal control" in keeping with prior literature on idiom reading, which frequently used single-word substitutions in order to yield a nonidiomatic control condition (e.g., *at the end of the day* vs. *at the end of the war* [literal control in Siyanova-Chanturia et al., 2011]; *spill the beans* vs *spill the chips* [literal control in Carrol & Conklin, 2017]; also see Carrol & Conklin, 2020; Cieslicka et al., 2014; Titone et al., 2019). In idiom research, literal baselines that preserve most of the idioms' component words are more or less the only possible way to create a set of well-controlled materials that enables comparisons between phrase types (see Carrol & Conklin, 2020, for a systematic investigation of noun vs. verb substitutions in idiom reading).

Nevertheless, the exact choice of a literal baseline is extremely important because it directly impacts what may be interpreted with respect to the idiom condition (see Titone & Connine, 1999, for a discussion of how literal control conditions can impact the interpretation of idiom priming studies). With respect to studies of idiom variation, this is highlighted in a recent paper by Kyriacou et al. (2019), which found that passivization of high-familiar idioms led to faster TRT reading times in an idiomatic condition (*the bucket was kicked*) compared to a literal control condition (*the apple was kicked*), when previously biasing the figurative meaning of the idiom (*Old John seemed to respond well to the new treatment at first, but eventually* . . .). Perhaps more on point given our specific materials, Smolka and Eulitz (2019) found that verb-modified idioms such as *grasp for the stars* were rated as well formed in capturing the figurative meaning of *reach the stars* in offline acceptability judgments, thus making it possible that our verb substitutions in creating the literal control conditions did not preclude idiom-driven figurative activation. Taken together, and applied to the present study, the fear of this alternative interpretation would be that the Lit-Lit sentences used here (e.g., *Larry hid the bullet so the police would not find the crucial evidence*) might have comparably activated the figurative meaning of *bit the bullet* as sentences containing the idiom itself, e.g., those beginning with, *Larry bit the bullet* . . .).[3]

While this account is certainly worthy of consideration, we do not believe it substantially impacts our interpretation of the data for several reasons. First, the details of the present study are fundamentally different from these studies in a manner that may preclude direct comparison. Kyriacou et al. (2019) presented participants with sentences for which a strong prior context semantically biased an upcoming idiom's figurative interpretation, and the idiom was sentence final. This would have maximized the likelihood of people figuratively interpreting both canonical and noncanonical idioms (see prior discussion on the role of pragmatic context). By contrast, in the present study, all idioms and literal control phrases were presented at the relative beginnings of sentences in a manner that had no prior biasing context, thus minimizing the chance of initial figurative activation, even for the idioms themselves. Moreover, Smolka and Eulitz (2019) conducted a rating task where participants were instructed to overtly semantically evaluate the degree of idiom well-formedness, using idioms that may have been inherently semantically decomposable (e.g., *reach for the stars*), and for which the lexical substitution was highly synonymous with the idiomatic verb (e.g., *reach* vs. *grasp*; see the prior section on the role of task demands). In our study, the overall set of items was highly

heterogeneous with respect to semantic decomposability (we controlled for this variable in our statistical models). More important, we selected the verbs of literal phrases to be highly semantically distinct from the verbs of idiomatic phrases (e.g., *bit the bullet* vs. *hid the bullet* and *kick the bucket* vs. *carry the bucket*) in order to create a successful literal control condition.

Second, and perhaps most important, to the extent that our literal control condition generated figurative activation, we would never have observed any reading time differences between the Lit-Lit condition and any of the idiom conditions, which we did. For example, noncanonical Id-Id sentences were read significantly more slowly than noncanonical Lit-Lit sentences (see Carrol & Conklin, 2020, for converging results in young-adult eye tracking). Moreover, the impact of increasing idiom familiarity across measures was more apparent for idiomatic sentences than for literal sentences. Thus, as observed here and elsewhere (see also Columbus et al., 2015; Titone et al., 2019), to the extent that readers failed to show differences between the literal control and idiom conditions, the most likely interpretation is that they did so because they failed to generate figurative activation, rather than the potential interpretation that they generated figurative activation for both idiomatic and literal sentences.

A final consideration with the findings reported here is that participants could have become aware of the presence of figurative language material over the course of the experiment, and this could have impacted how they read the sentences. However, to the extent that this occurred, we would have expected a systematic shift in how people responded to the different conditions over trials, which did not occur when we computed additional models including *trial number* as a fixed effect. Under these conditions, neither the size nor the direction of effects changed.

To conclude, our goal was to investigate online processing of canonical and non-canonical idioms in younger and older adults using eye-movement measures of natural reading. We found that older adults can access idiomatic forms in memory more readily than younger adults, which suggests that idiomatic configurations are more entrenched as we age, which in turn aids the online processing of idiomatic forms. With respect to idiom variation, we found that presentations of noncanonical idioms slows access to the figurative configuration in memory and also slows late-stage comprehension of the phrase in discourse, presumably because the modifier emphasizes the literal, compositional nature of the idiom.

Our findings thus add to a growing body of research suggesting that older adults are able to leverage their greater crystallized knowledge during (online) language processing (e.g., Pichora-Fuller, Schneider, & Daneman, 1995; Wingfield, Aberdeen, & Stine, 1991; Wingfield & Lash, 2016). Noncanonical form presentation of idiomatic phrases affects idiom reading in younger and older adults alike by slowing quick access to figurative meanings. Future studies could investigate whether age differences in online processing of noncanonical idioms are more likely to occur for other types of idiom variants that were not investigated in this study (e.g., idioms with noun substitution, verb substitutions, or passivized idioms; see Carrol & Conklin, 2020; Geeraert et al., 2017b; Kyriacou et al., 2019; Smolka & Eulitz, 2019).

Notes

1. We are not reporting skipping rates of phrase-final nouns (i.e., kicked the *bucket*) because no differences emerged with respect to this variable. Nor were there effects of canonical form presentation, age group, or familiarity.

2. Upon reviewer request, we ran follow-up models splitting the GD noun data by sentence type (i.e., we ran one model for literal items, and another for idiomatic items). Neither model showed any significant effects.

3. We thank an anonymous reviewer for pointing this out.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**, 255–278.

Bobrow, S., & Bell, S. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, **1**, 343–346.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, **50**, 2669–2683.

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., . . . Jagust, W. J. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, **130**, 746–763.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, **106**, 7351–7356.

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

Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 373–443). New York: Psychology Press.

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *Mental Lexicon*, **9**, 267–293.

Cacciari, C., Corrardini, P., & Ferlazzo, F. (2018). Cognitive and personality components underlying spoken idiom comprehension in context. An exploratory study. *Frontiers in Psychology*, **9**, 659.

Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 217–240). Amsterdam: North-Holland.

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

Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol*, **22**, 79–108.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, **22**, 77–94.

Caplan, D., & Waters, G. (2007). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, **13**, 403–413.

Carrol, G., & Conklin, K. (2014). Eye-tracking multi-word units: Some methodological questions. *Journal of Eye Movement Research*, **7**, 1–11.

Carrol, G., & Conklin, K. (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.

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

Cieslicka, A. B., Heredia, R. R., & Olivares, M. (2014). It's all in the eyes: How language dominance, salience, and context affect eye movements during idiomatic language processing. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism* (pp. 21–42). Cham, Switzerland: Springer International.

Coane, J. H., Sánchez-Gutiérrez, C., Stillman, C. M., & Corriveau, J. A. (2014). False memory for idiomatic expressions in younger and older adults: Evidence for indirect activation of figurative meanings. *Frontiers in Psychology*, **5**, 764.

Columbus, G., Sheikh, N., Côté-Lecaldare, M., Haeuser, K., Baum, S., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading. *Frontiers in Human Neuroscience*, **8**, 1057.

Conner, P., Hyun, J., O'Connor Wells, B., Anema, I., Goral, M., Monereau-Merry, M., ... Obler, L. (2011). Age-related differences in idiom production in adultdhood. *Clinical Linguistics and Phonetics*, **25**, 899–912.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, **25**, 57–71.

Fellbaum, C. (2019). How flexible are idioms? A corpus-based study. *Linguistics*, **57**, 735–767.

Geeraert, K., Newman, J., & Baayen, R. H. (2017a). Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science*, **9**, 653–669.

Geeraert, K., Baayen, R. H., & Newman, J. (2017b). Understanding idiomatic variation. In *Proceedings of the 13th Workshop on Multiword Expressions* (pp. 80–90). Valencia, Spain: Association for Computational Linguistics.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel hierarchical models* (Vol. **1**). New York: Cambridge University Press.

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

Gibbs, R. W., Jr., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, **21**, 100–138.

Gibbs, R. W., Jr., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory & Language*, **28**, 576–593.

Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, **28**, 88–124.

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press.

Grady, C. L., & Craik, F. I. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology*, **10**, 224–231.

Grindrod, C. M., & Raizen, A. L. (2019). Age-related changes in processing speed modulate context use during idiomatic ambiguity resolution. *Aging, Neuropsychology, and Cognition*, **26**, 842–864.

Haeuser, K. I., Titone, D. A., & Baum, S. R. (2016). The role of the ventro-lateral prefrontal cortex in idiom comprehension: An rTMS study. *Neuropsychologia*, **91**, 360–370.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, **18**, 1279–1296.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 163–169.

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, **22**, 193–225.

Hasher, L., Zacks, R. T., & May, C. P. (1998). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance. Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.

Hedden, T., & Gabrieli, J. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience*, **5**, 87–96.

Hung, P., & Nippold, M. (2014). Idiom understanding in adulthood: Examining age-related differences. *Clinical Linguistics & Phonetics*, **28**, 208–221.

Hyun, J., Conner, P. S., & Obler, L. K. (2014). Idiom properties influencing idiom production in younger and older adults. *Mental Lexicon*, **9**, 294–315.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122–149.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, **16**, 262–228.

Kyriacou, M., Conklin, K., & Thompson, D. (2019). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, **63**, 404–435.

Langlotz, A. (2006). *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: Benjamins.

Li, S. C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, **5**, 479–486.

Libben, M., & Titone, D. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, **36**, 1103–1121.

Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, **346**, 572–578.

Linzen, T., & Oseki, Y. (2018). The reliability of acceptability judgments across languages. *Glossa*, **3**, 100.

Mancuso, A., Elia, A., Laudanna, A., & Vietri, S. (2019). The role of syntactic variability and literal interpretation plausibility in idiom comprehension. *Journal of Psycholinguistic Research*. Advance online publication.

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

Milburn, E., & Warren, T. (2019). Idioms show effects of meaning relatedness and dominance similar to those seen for ambiguous words. *Psychonomic Bulletin & Review*, **26**, 591–598.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, **28**, 762–788.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, **4**, 61–64.

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, **5**, 218–228.

Nunberg, G., Sag, I., & Wasow, T. (1994). Idioms. *Language*, **70**, 491–538.

Papagno, C., Lucchelli, F., Muggia, S., & Rizzo, S. (2003). Idiom comprehension in Alzheimer's disease: The role of the central executive. *Brain*, **126**, 2419–2430.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, **97**, 593–608.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current deficits and an agenda for future research. *European Journal of Cognitive Psychology*, **16**, 3–26.

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *Neuroimage*, **63**, 600–610.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372–422.

Rayner, K. (2009). The thirty-fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, **62**, 1457–1506.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, **14**, 191–201.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **30**, 1290–1301.

R Development Core Team. (2018). R: A language and environment for statistical computing. (Version 3.5.1; Feather Spray) [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Rizzo, S., Sandrini, M., & Papagno, C. (2007). The dorsolateral prefrontal cortex in idiom interpretation: An rTMS study. *Brain Research Bulletin*, **71**, 523–528.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, **103**, 403–428.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, **54**, 35–54.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, **110**, 104038.

Schettino, A., Lauro, L. R., Crippa, F., Anselmetti, S., Cavallaro, R., & Papagno, C. (2010). The comprehension of idiomatic expressions in schizophrenic patients. *Neuropsychologia*, **48**, 1032–1040.

Schröder, D. (2013). *The syntactic flexibility of idioms: A corpus-based approach*. München, Germany: Akademische Verlagsgemeinschaft München.

Schütze, C. T. (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology (Classics in Linguistics 2)*. Berlin: Language Science Press.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, **27**, 251–272.

Smolka, E., & Eulitz, C. (2019). Can you reach for the planets or grasp at the stars? Modified noun, verb, or preposition constituents in idiom processing. Psyarxiv preprints. doi:10.31234/osf.io/3sve4

Sprenger, S., La Roi, A., & van Rij, J. (2019). The development of idiom knowledge across the lifespan. *Frontiers in Communication*, **4**, 29.

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

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

Tabossi, P., Fanari, R., & Wolf, K. (2008). Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 313–327.

Tabossi, P., Wolf, K., & Koterle, S. (2009). Idiom syntax: Idiosyncratic or principled? *Journal of Memory and Language*, **61**, 77–96.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 171–207). Cambridge: Cambridge University Press.

Titone, D., & Connine, C. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, **31**, 1655–1674.

Titone, D., & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *Mental Lexicon*, **9**, 473–496.

Titone, D., Lovseth, K., Kasparian, K., & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? Evidence from eye movement measures of reading. *Canadian Journal of Experimental Psychology*, **73**, 216–230.

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience & Biobehavioral Reviews*, **26**, 849–857.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 1001–1010.

Westbury, C., & Titone, D. (2011). Idiom literality judgments in younger and older adults: Age-related effects in resolving semantic interference. *Psychology and Aging*, **26**, 467–474.

Wingfield, A., Aberdeen, J. S., & Stine, E. A. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology*, **46**, 127–129.

Wingfield, A., & Lash, A. (2016). Audition and the comprehension of spoken language in adult aging: Stability in the face of change. In K. W. Schaie and S. Willis (Eds.), *Handbook of the psychology of aging* (pp. 165–185). London: Elsevier.