

Chapter 1

“Spilling the bag” on idiomatic variation

Kristina Geeraert

KU Leuven

R. Harald Baayen

University of Tübingen & University of Alberta

John Newman

University of Alberta & Monash University

Recent corpus-based studies have shown that idioms can vary much more extensively than previously claimed (Moon 1998; Barlow 2000; Duffley 2013), but little research has been conducted on how we understand or regard these variants (Gibbs et al. 1989; McGlone et al. 1994). This study further examines idiomatic variation, specifically investigating the acceptability and processing of several types of variants through rating and eye-tracking experiments. Four types of idiom variants were included, in addition to the canonical form and a literal meaning of the expression (i.e. in a different context). The results show that altering an idiom does not render it completely unacceptable nor incomprehensible, but rather, certain variants are more preferred and easier to understand. These results show support for probabilistic accounts of language.

1 Introduction

Idioms have traditionally been regarded as multiword units whose meaning cannot be derived from the meaning of its parts (Bobrow & Bell 1973). The literal meaning of an idiom is the word-by-word compositional meaning of the words, whereas the idiomatic meaning is stored separately with the idiom, as if a large



word. Furthermore, if the idiom is stored whole, then idioms must also be structurally fixed. This rationale has led researchers to predominantly investigate idioms in their canonical form, and how they are understood in comparison to a literal paraphrase (Swinney & Cutler 1979; Gibbs 1980; Cacciari & Tabossi 1988; Titone & Connine 1999).

Recent corpus-based research however has shown that idioms can occur with a range of variation (Moon 1998; Barlow 2000; Langlotz 2006; Schröder 2013), such as syntactic variation (e.g. *her new-found reputation was a bubble that would burst* [from *burst one's bubble* 'shatter one's illusions about something'], *the question begged by* [*beg the question* 'invite an obvious question'], lexical variation (e.g., *set/start the ball rolling* 'set an activity in motion', *a skeleton in the closet/cupboard* 'a discreditable fact that is kept secret'), truncations (e.g., *make hay* [*while the sun shines*] 'take advantage of favourable conditions'), and even adverbial or adjectival modifications (e.g., *pulling political strings* [*pull strings* 'use one's power or influence to gain an advantage'], *rock the party boat* [*rock the boat* 'upset the status-quo']). This variation can even occur with nondecomposable idioms (Duffley 2013), or idioms thought to be semantically frozen and syntactically fixed, such as *kick the bucket* 'die' (e.g., *no buckets were kicked*, *reluctant to kick his brimming bucket of life*, and *my phone kicked the pail last week*). These studies demonstrate that idioms are not nearly as fixed as previously assumed. This variability in idioms, and MULTIWORD EXPRESSIONS (MWEs) more generally, is acknowledged as a key challenge in the automated identification of MWEs in corpora (cf. Savary et al. 2018 [this volume] and Scholivet et al. 2018 [this volume]).

Few studies have investigated idiomatic variation from an experimental perspective. Gibbs and colleagues (Gibbs et al. 1989; Gibbs & Nayak 1989) explored lexical and syntactic variation of decomposable and nondecomposable idioms using a semantic similarity rating task. They found that decomposable idioms (i.e. idioms whose constituents contribute to the meaning of the whole, as in *pop the question* 'propose marriage') were rated as more similar in meaning to a literal paraphrase than were nondecomposable idioms, or idioms whose constituents do not contribute meaning, as in *kick the bucket*. But as Duffley (2013) has shown, nondecomposable idioms can be modified in context and still retain their meaning. In addition, the semantic decomposability measure used by Gibbs and colleagues has not proven a reliable measure, with participants performing at chance in replication studies (Titone & Connine 1994b, Tabossi, Fanari & Wolf 2008). Replication studies have also shown inconsistent results – decomposable and nondecomposable idioms are not always found to be statistically different (Tabossi et al. 2008). Finally, a measure of semantic similarity between an idiom

variant and a literal paraphrase may not be the best method for determining the comprehension of idiomatic variation. Semantic similarity has been shown to be predicted by the same local contexts as observed in corpora (Miller & Charles 1991), suggesting that this measure may simply be reflecting how interchangeable the variant is with its paraphrase, or how acceptable the variant may be at conveying the meaning in the paraphrase.

Meanwhile, McGlone et al. (1994) explored the semantic productivity of idiom variation. Variants in this study produced an enhanced idiomatic meaning based on the original (e.g. *shatter the ice*, from *break the ice*, meaning ‘to break an uncomfortable or stiff social situation in one fell swoop’). In a self-paced reading study, they measured the reaction time for participants to read the final sentence of a story, which contained idioms, variants, or literal paraphrases. They found that participants were significantly faster at reading the canonical form of the idiom, but that variants were read as fast as literal paraphrases. They suggest that canonical forms of idioms are accessed whole, while variants are processed like literal language and are therefore processed slower. However, they did not control for the type of variation used in their study. They included instances of lexical variation (e.g. *shatter the ice*), quantification (e.g. *not spill a single bean* [*spill the beans* ‘reveal secret information’]), and hyperboles (e.g. *it’s raining the whole kennel* [*rain cats and dogs* ‘rain very hard’]). It is unclear whether some types of variants are easier to comprehend than others.

The current study attempts to improve upon these previous studies. We explore the acceptability and processing of several types of idiom variants through a rating task and an eye-tracking experiment, respectively. While both of these experiments have been presented independently elsewhere – the eye-tracking study was presented in Geeraert, Baayen, et al. (2017) and part of the acceptability ratings study was presented in Geeraert, Newman, et al. (2017), but presented in full below – they have been brought together here in a multi-methodological study in order to tease apart and contrast speaker judgements from online processing. Previous research has sometimes conflated these two methods, making interpretation difficult (cf. Gibbs et al. 1989; Gibbs & Nayak 1989). But here we distinctly separate them, utilizing an online measure of processing in addition to a subjective evaluative measure, to determine any converging or diverging results between the two methods, which are important for understanding idioms and idiomatic variation. These two studies utilize the same data, yet are different perspectives. Thus, this chapter provides a complete reportage of the larger study, a discussion of the variables useful for predicting each modality (with suggestions as to why), as well as a unique perspective in the idiom literature.

Two main research questions are explored. First, how do variants compare with the canonical form? This question explores whether differences between the canonical form and the variants are still present when the type of variation is controlled. Second, how do variants compare with each other? This question explores whether any differences emerge between the variant types – are certain variant types more preferred or easier to process?

This study is largely exploratory, but we do have some predictions about the results. For example, formal idiom blends are often regarded in the literature as being errors or slips of the tongue (Fay 1982; Cutting & Bock 1997). We therefore hypothesized that blends would be least preferred and more difficult to process due to this perceived error-like nature. Meanwhile, some idioms have been shown to occur in “idiom sets” (Moon 1998), such as *shake/quake/quiver in one’s boots* ‘tremble with apprehension’ or *down the drain/chute/tube/toilet* ‘completely lost or wasted’. Given this, we predict that lexical variation will not be more difficult to understand than the canonical form, and may be considered an acceptable variant strategy. A modifier inserted into the idiom should take additional time to process due to the presence of the extra word, but given their relative frequency and overall productivity in corpora (Moon 1998; Schröder 2013), may be the most preferred variant. Lastly, a partial or truncated form will likely be faster to process, due to the omission of a word, but may not be widely accepted due to their limited occurrence in corpora (Moon 1998).

The remainder of the chapter proceeds as follows: We discuss each experiment in turn, beginning with the acceptability ratings, and then the eye-tracking experiment. We discuss the design of each experiment and the results obtained. We conclude with a discussion of our findings, how the results of the two experiments converge or diverge, as well as how the results fit into the larger discussion on idioms, and specifically within a probabilistic approach to language.

2 Acceptability rating experiment

2.1 Methodology

2.1.1 Materials

Sixty idioms were extracted from the Oxford Dictionary of English Idioms (Ayto 2009) and the Collins COBUILD Idioms Dictionary (Sinclair 2011), listed in Appendix A. These idioms varied in length and syntactic structure: 20 three-word idioms consisting of a verb and a noun phrase (V-NP, e.g. *rock the boat*); 20 four-word idioms consisting of a verb and a prepositional phrase (V-PP, e.g. *jump on*

the bandwagon ‘join others doing something fashionable’); and 20 five- or six-word idioms (10 each) consisting of a verb, noun phrase, and a prepositional phrase (V-NP-PP, e.g. *hear something through the grapevine* ‘hear gossip’). Two contexts were created for each idiom: one literal and one figurative (e.g. *I used to pretend I could talk to plants, and I would hear things through the grapevine* = literal; and *I used to be a socialite, and I would hear things through the grapevine* = figurative). Both contexts had identical final clauses, with the idiom in sentence-final position. As syntactic variation is possible with idioms (Moon 1998; Schröder 2013), the contexts were not restricted to the present tense.

The form listed in the dictionary was regarded as the canonical form (for a different approach to establishing canonical forms of MWEs (see Savary et al. 2018 [this volume]). If more than one form was listed then the form most familiar to the first author was used, as she spoke the same variety as the participants in the study. In addition to the canonical form, these idioms were manipulated for four types of variation within the figurative context (i.e. the context was identical for all variants). First, lexical variation, where one of the lexical items within the expression was altered to a synonymous or near-synonymous word (e.g. *discover something through the grapevine*). Synonyms were selected based on their naturalness in the context to convey a similar meaning.¹ Second, partial form of the idiom, where only a portion of the idiom was presented, usually a key word or words (e.g. *use the grapevine*). In order for the sentence to still be grammatically correct, pronouns or lexically-vague words replaced the missing elements of the expression, such as *it, them, things* for nouns, or *have, be, do, use* for verbs. Third, integrated concept, where an additional concept was integrated into the idiom (e.g. *hear something through the judgemental grapevine*). These additional concepts expanded or emphasized the figurative contexts in which the idiom occurred. Finally, formal idiom blend, where two idioms were blended together (e.g. *get wind through the grapevine* – blending *hear something through the grapevine* with *get wind of something* ‘hear a rumour’). Each “experimental” idiom (i.e. the 60 idioms selected) was paired with a non-experimental idiom for use in the idiom blend condition. These paired “blending” idioms were chosen for their intuitive plausibility, but controlled for their syntax and semantics (Cutting & Bock 1997). Four types of blends were created: same syntax, similar semantics (sSYN, sSEM); same syntax, different semantics (sSYN, dSEM); different syntax, similar semantics (dSYN, sSEM); and different syntax, different semantics (dSYN, dSEM), illustrated in Table 1. Five instances of each type of blend occurred with the three syntactic types (i.e. V-NP, V-PP, or V-NP-PP), totalling 15 of each blend

¹An online thesaurus (<http://www.thesaurus.com>) was often utilized for synonymous words.

type. There is clearly a linguistic playfulness at work in the creation of the idiom blends in Table 1, just as there is in many of the non-canonical idiom forms found in naturally occurring usage (cf. Moon 1998; Duffley 2013). This playfulness, it should be noted, presents a special challenge to modelling of MWEs in the context of NLP work on multiword expressions or annotation of MWEs in corpora, as noted in Savary et al. (2018 [this volume]). Indeed, Savary et al. (2018 [this volume]) consider “wordplay proneness”, as they call it, a challenge that “goes far beyond the state of the art in semantic modelling and processing of VMWEs [verbal MWEs]”.

Table 1: Four types of blends used in the idiom blend condition.

Type of blend	Example	Source idioms	Total
sSYN, sSEM	<i>rock the applecart</i>	<i>rock the boat</i> <i>upset the applecart</i>	15
sSYN, dSEM	<i>shoot your tongue</i>	<i>shoot the breeze</i> <i>hold your tongue</i>	15
dSYN, sSEM	<i>pass the mustard</i>	<i>cut the mustard</i> <i>pass muster</i>	15
dSYN, dSEM	<i>face the wringer</i>	<i>face the music</i> <i>put through the wringer</i>	15

Half of the idioms had the beginning portion of the expression altered (verb), while the other half had alternations made to the final portion of the expression (noun). In total, there are six conditions: one in a literal context and five in a figurative context (i.e. one canonical form and four variants). The experiment utilized a Latin-square design, where every participant saw each idiom once in one of the six conditions. Six versions of the experiment were created, each one containing 10 idioms in each of the six conditions.

Conditions:

1. **Literal meaning** of the idiom in its canonical form
(*While the guys were reshingling, they suddenly went through the roof.*)
2. **Canonical form** of the idiom in a figurative context
(*Although these were new stocks, they suddenly went through the roof.*)
3. **Lexical variation** of the idiom in a figurative context
(*Although these were new stocks, they suddenly went through the ceiling.*)

4. **Partial form** of the idiom in a figurative context
(*Although these were new stocks, they suddenly went through it.*)
5. **Integrated concept** within the idiom in a figurative context
(*Although these were new stocks, they suddenly went through the investment roof.*)
6. **Idiom blend** of two idioms in a figurative context
(*Although these were new stocks, they suddenly went through the charts.*)

Since the blending idioms only occurred in one condition (i.e. idiom blend), they were used as fillers in their canonical form in the other five versions of the experiment, occurring in either a figurative or literal context. Each blending idiom was excluded as a control in the version of the experiment where it occurred in the idiom blend condition in order to avoid a bias in the materials. Therefore, in each version of the experiment, 10 of the blending idioms occurred in the idiom blend condition, while the remaining 50 appeared as fillers. Of these fillers, 20 occurred in a figurative context and 30 occurred in a literal context. This was done to increase the number of literal contexts in the experiment so that they were not so underrepresented. In sum, each participant saw 110 items: 60 experimental idioms (10 in each condition) and 50 blending idioms as fillers.

Finally, six practice sentences were created using a different six idioms. These idioms all occurred in their canonical form. Three were in a figurative context and three in a literal context. These were the same for all participants.

2.1.2 Procedure

Using E-prime 2.0 standard edition software, each sentence was presented in random order at the top of the computer screen. The text was presented in a black, bold, 24-point Courier New font, centered on a white background. Below each sentence was a VISUAL ANALOGUE SCALE (VAS), which is a continuous graphical rating scale that allows fine gradations to be measured (Funke & Reips 2012).

Participants were explicitly told that they would be reading sentences containing English expressions, but that some of the expressions had been modified in various ways. They were asked to rate the acceptability of the expression, as it occurred in the sentence, by clicking the mouse anywhere on the provided scale, which was labelled with “acceptable” on the right and “unacceptable” on the left. The mouse was repositioned to the centre of the scale on each trial. Participants were encouraged to use the whole scale before the experiment began, and were given an opportunity to take a short break halfway through the experiment.

After the participants had completed the rating task, they were asked whether they knew each idiom. As different speakers are familiar with different subsets

of idioms, this information allowed us to control, at the level of the individual, whether they knew the idiom (Cacciari et al. 2005), while maximizing the number of idioms in the study. Each idiom appeared, in its canonical form, in a black, bold, 22-point Courier New font, centered on a white background. Above the idiom was the question “Do you know this expression?” and below were two boxes, one labelled “yes” and the other labelled “no”. Using the mouse, participants clicked on the appropriate box to respond. The mouse repositioned itself to the center of the screen on each trial.

At the end of this second task, participants were presented with a few additional questions pertaining to their idiom usage (e.g. How often do you use these expressions?, Do you like using these expressions?). Participants responded to these questions using the same VAS scale as the rating task, this time labelled with a thumbs-up image on the right for a positive response and a thumbs-down image on the left for a negative one. Lastly, participants were asked to rate the acceptability of seven prescriptively “incorrect” sentences (LQs), shown below, also using this VAS scale. These sentences attempted to elicit a measure of the participant’s flexibility with language and non-standard usage.

Language questions (LQs):

1. The only option the school board has is to lay off a large *amount* of people.
2. Slot machines are thought to be more *addicting* than table games.
3. The document had to be signed by both Susan and *I*.
4. While cleaning the kitchen, Sally looked up and saw a spider on the *roof*.
5. I thought it could’ve *went* either way.
6. She *could care* less what he had to say about it.
7. You have to balance your life, *irregardless* of what anybody thinks.

2.1.3 Participants

Forty-eight undergraduate linguistics students from the University of Alberta participated in this experiment. All participants were native speakers of English. There were 37 female and 11 male participants, ranging from 17–43 years of age. All participants were reimbursed for their time with course credit.

2.2 Results

The results were analyzed with mixed-effects linear regression using the `lme4` package (Bates et al. 2015) in R (R Core Team 2012). We focus on two analyses:

the rating responses and the rating reaction times. Only the 60 experimental idioms were included in these analyses (i.e. the fillers were not included outside of the idiom blend condition).

Six predictor variables are discussed below. *Condition* is a factor indicating the condition in which the idiom occurred (e.g. canonical form, lexical variation, idiom blend). *Length* specifies the number of words within the idiom’s canonical form. *PairedSemantics* is a factor specifying whether the two idioms used in the formal idiom blend have similar or different semantics (e.g. *spill the beans* & *let the cat out of the bag* ‘reveal a secret’ = similar; *shoot the breeze* ‘have a casual conversation’ & *hold your tongue* ‘remain silent’ = different). Meanwhile, *KnowIdiom* is a factor indicating the participant’s knowledge of the idiom (i.e. yes or no). And *Trial* is the scaled (i.e. standardized) order of presentation of the stimuli in the experiment. Since the stimuli was presented randomly, this order will be different for each participant.

meanTransparencyRating is the scaled average rating for the transparency (or clarity) of the idiom’s meaning as a whole. Since speakers differ in how they interpret the decomposability (i.e. compositionality) of idioms, as evidenced by the low reliability of the decomposability classification task (Titone & Connine 1994b, Tabossi, Fanari & Wolf 2008), we were interested in a measure for how clear or obvious people find the meaning of the idiom as a whole. This measure then, may provide some indication of how literal or figurative people consider an idiom. These ratings were collected in a separate experiment, specifically designed to elicit ratings of transparency. Those participants saw each idiom, along with a definition and an example sentence, and were asked to rate the transparency of the idiom (for further details, see Geeraert 2016). The average rating for each idiom was included as a separate predictor to determine whether transparency influences people’s preferences of variation.

2.2.1 Rating responses

The first model examines the rating responses. The fixed effects of this model are shown in Table 2. This model has three significant interactions with *Condition*. The first, between *Condition* and *KnowIdiom*, is shown in the top-left panel of Figure 1. As expected, participants are not sensitive to variation when an idiom is unfamiliar. But when the idiom is known, there is a clear preference for the canonical form. Two variants types, integrated concepts and lexical variation, are rated as more acceptable than the others, with a slight preference for variants which have an additional concept inserted into the idiom. The remaining variants: idiom blends, partial forms, and a literal reading of the idiom, are all rated as the least preferred variants.

Table 2: Fixed effects for the acceptability rating responses.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	88.81	6.54	13.59	
Condition=Concept	-19.68	7.34	-2.68*	187.74
Condition=Blend	-37.56	7.36	-5.10*	
Condition=Lexical	-22.69	7.36	-3.08*	
Condition=Literal	-46.46	7.36	-6.31*	
Condition=Partial	-45.25	7.37	-6.14*	
KnowIdiom=No	-30.17	3.58	-8.43*	30.24
Length	-2.02	1.49	-1.35	1.98
meanTransparencyRating	3.82	1.91	2.00	18.30
Trial	1.77	0.69	2.58	4.26
I(KnowIdiom=No Condition=Concept)	13.66	4.92	2.78*	52.10
I(KnowIdiom=No Condition=Blend)	31.76	4.74	6.71*	
I(KnowIdiom=No Condition=Lexical)	23.31	4.94	4.71*	
I(KnowIdiom=No Condition=Literal)	31.26	4.74	6.59*	
I(KnowIdiom=No Condition=Partial)	22.63	4.85	4.66*	
I(Length Condition=Concept)	1.09	1.72	0.64	2.63
I(Length Condition=Blend)	2.52	1.71	1.48	
I(Length Condition=Lexical)	0.52	1.71	0.31	
I(Length Condition=Literal)	4.63	1.71	2.71*	
I(Length Condition=Partial)	4.11	1.71	2.40	
I(meanTransparencyRating Condition=Concept)	0.72	2.25	0.32	1.32 ^a
I(meanTransparencyRating Condition=Blend)	2.01	2.22	0.90	
I(meanTransparencyRating Condition=Lexical)	3.59	2.31	1.56	
I(meanTransparencyRating Condition=Literal)	6.35	2.27	2.80*	
I(meanTransparencyRating Condition=Partial)	0.37	2.31	0.16	

* = Factors that remain significant after a Bonferroni correction

^aAn ANOVA test run during model comparison indicates that the inclusion of this interaction is significant ($p = 0.045$).

Length also occurs in a significant interaction with Condition, shown in the top-centre panel of Figure 1. Participants tend to rate idioms as less acceptable in their canonical form if they are longer. This pattern holds for most variants as well: integrated concepts, lexical variation, and formal idiom blends have slopes which are not significantly different from the canonical form and are therefore depicted in grey. Literal meanings and partial forms however are rated as more acceptable if the idiom is longer. Apparently, literal interpretations (which likely may also characterize partial forms) benefit from the presence of many words.

1 “Spilling the bag” on idiomatic variation

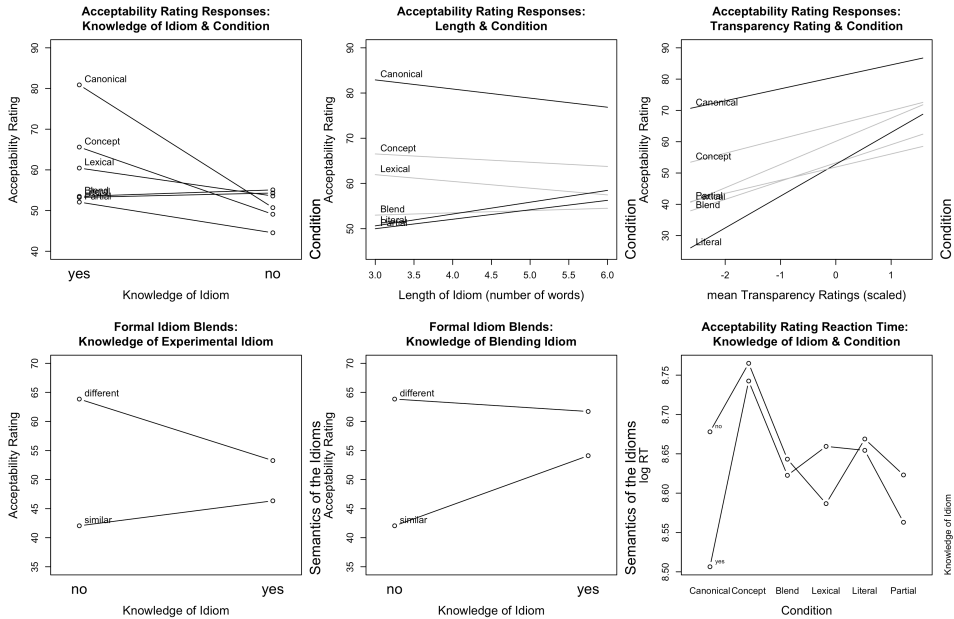


Figure 1: Interactions in the mixed-effects linear regression models for the acceptability rating responses and reaction times. Lines in grey represent factors levels which are not significantly different.

This might suggest that as expressions become longer, the non-idiomatic reading becomes stronger and begins to interfere with the idiomatic reading.

The last interaction, between meanTransparencyRating and Condition, is illustrated in the top-right panel of Figure 1. Higher ratings of acceptability are given to idioms judged to be more transparent. For the condition in which the context enforced a literal reading, the effect of transparency was stronger than for any of the other idiom variants. This result is not unexpected, given that not all idioms have a plausible literal meaning (Titone & Connine 1994b).

Trial was significant as a main effect; participants became more accepting of the stimuli (both variants and the canonical form) the further they advanced through the experiment. But participants differed in exactly how much more accepting they became, as evidenced by the by-Subject random slopes for Trial. These slopes in the random effects structure are in addition to Subject and Idiom included as random effects.² Finally, it is interesting to note that the frequency or syntax of the idiom, as well as whether modifications were made to the verb or the noun, did not affect the acceptability of idioms or variants.

²The rating response model and the RT model show the same random effects structure.

We also looked specifically at formal idiom blends, given their error-like status in the literature (Fay 1982; Cutting & Bock 1997), in order to explore whether any factors influence their acceptability. Two interactions appear significant, shown in Table 3: both between the participant’s knowledge of an idiom and the paired semantics of the two idioms involved in the blend. The bottom-left panel in Figure 1 shows the interaction with knowledge of the experimental idiom, while the bottom-centre panel shows the knowledge of the blending idiom. Both interactions indicate that blends are rated as more acceptable when the meanings of the two idioms differ, and less acceptable when they are similar. Participants significantly rate blends with similar semantics with a lower acceptability if one of the idioms is unknown. A three-way interaction between these variables (i.e. knowledge of both idioms and the semantic similarity of the idioms) is not significant, suggesting that speakers only need to be unfamiliar with one of the idioms to regard semantically similar idiom blends as less acceptable. The noticeability of the unknown idiom likely causes this increase in unacceptability, which is perhaps not as noticeable for those who are familiar with both blended idioms – presumably, they are able to access the meaning of the blend easier, as they are familiar with both idioms from which the parts belong, and therefore are not as surprised or unimpressed by the blend. Finally, `meanTransparencyRating` is also significant in this model – speakers prefer idiom blends that are more transparent and clearer in meaning.

Table 3: Fixed effects for the acceptability ratings of idiom blends.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	63.62	6.06	10.50	
<code>meanTransparencyRating</code>	5.21	2.30	2.26	3.00
<code>KnowExpIdiom=Yes</code>	-10.58	4.64	-2.28*	1.02
<code>KnowBlendingIdiom=Yes</code>	-2.13	4.77	-0.45	0.59
<code>Semantics=Similar</code>	-21.80	7.43	-2.93*	2.00
<code>I(Semantics=Similar KnowExpIdiom=Yes)</code>	14.87	6.47	2.30*	3.23
<code>I(Semantics=Similar KnowBlendingIdiom=Yes)</code>	14.19	6.50	2.18*	2.74

* = Factors that remain significant after a Bonferroni correction

2.2.2 Rating reaction times

We also analyzed the REACTION TIMES (RTs) for how quickly the participants responded to the acceptability rating task. Faster reaction times are associated with easier judgements of acceptability. The fixed effects for this model are shown in Table 4. Only one interaction, between KnowIdiom and Condition, is significant in this model, illustrated in the bottom-right panel in Figure 1. The RTs associated with each condition are similar for both those who know the idiom and those who do not. Significantly longer RTs are observed with integrated concepts, while significantly shorter RTs are observed with partial forms. These results may simply reflect the fact that the integrated concept condition has an additional word inserted into the idiom, whereas the partial form condition has a word omitted from the expression. This RT difference likely corresponds to length of the expression. The most notable observation is that participants are significantly faster rating the canonical form of the expression if the idiom is known. If the idiom is unknown, the RT to rate the canonical form does not differ significantly from the other variants. These results illustrate that the canonical form has an advantage if it is familiar, but that variants of the same length as the canonical are rated as quickly as if one does not know the expression.

Table 4: Fixed effects for the acceptability rating reaction times.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	8.51	0.04	226.18	
Condition=Concept	0.24	0.02	9.65*	93.85
Condition=Blend	0.12	0.02	4.70*	
Condition=Lexical	0.15	0.02	6.22*	
Condition=Literal	0.15	0.02	5.92*	
Condition=Partial	0.06	0.02	2.27	
KnowIdiom=No	0.17	0.04	4.40*	1.69
Trial	-0.08	0.01	-8.16	39.96
I(KnowIdiom=No Condition=Concept)	-0.15	0.05	-2.82*	13.10
I(KnowIdiom=No Condition=Blend)	-0.15	0.05	-2.91*	
I(KnowIdiom=No Condition=Lexical)	-0.24	0.05	-4.67*	
I(KnowIdiom=No Condition=Literal)	-0.16	0.05	-3.08*	
I(KnowIdiom=No Condition=Partial)	-0.11	0.05	-2.18	

* = Factors that remain significant after a Bonferroni correction

In sum, this experiment explored the acceptability of idiomatic variation, using several types of variants. The canonical form is the most preferred and participants are quicker at rating this form, but only when the expression is known. Modifying an idiom makes it less acceptable, but the decrease in acceptability varies according to the type of alternation – integrating an additional element (*go through the investment roof*) or replacing a word with a near-synonym (*go through the ceiling*) were considered more acceptable variants. We now turn our attention to how these variants are understood.

3 Eye-tracking experiment

3.1 Methodology

3.1.1 Materials

This experiment utilized the same materials as the previous experiment.

3.1.2 Procedure

This experiment used the Eye-Link 1000, desk-top mounted video-based eye-tracking device, manufactured by SR Research. The eye-tracker sampled the pupil location and size at a rate of 1000Hz, and was calibrated using a 9-point calibration grid. Calibration occurred at the beginning of the experiment, after the practice, and again after every 22 sentences, for a total of five blocks. The computer screen resolution was set to 1920 x 1080 pixels.

The stimuli were presented in two parts. Participants first saw the “context clause” (e.g., *Although these were new stocks,*), followed by the “idiom clause” (e.g. *they suddenly went through the roof.*) on a separate screen. Each trial began with a fixation cross presented for 1,000 msec on the left side of a light-grey screen. Next, they saw the context clause, also on a light-grey background, in a bold, black, Courier New 30-point font. Every clause was displayed in full and fit on one line. To exit this screen, participants had to trigger an invisible boundary in the bottom right corner. A blank, light-grey screen was presented for 1,000 msec before the fixation cross preceding the idiom clause appeared. The sequence of screens for the idiom clause was identical to the context clause.

Ten percent of the stimuli were followed by a true/false comprehension question, which pertained to the immediately preceding sentence, and were presented randomly throughout the experiment. Participants pushed one of two buttons on a game controller to answer these questions, which were clearly labelled on the

question screen. The experiment began with a practice session, which consisted of six practice sentences and three questions. These were the same for all participants, although their order varied.

All participants had normal or corrected-to-normal vision. The right eye of each participant was tracked. Participants sat approximately 85cm from the computer screen, with the camera placed on the desk about 35cm in front of the computer screen. The participants sat in a sound-proof booth, while the experimenter sat outside the booth, running the experiment. The lights were kept on. The experiment was self-paced and took about 45 minutes to complete. Each participant was given an opportunity for a short break half-way through the experiment.

After the participants had completed the eye-tracking portion, they were asked to complete three additional tasks: (1) to indicate their knowledge of each expression; (2) to answer questions pertaining to their idiom usage; and (3) to rate the acceptability of the seven prescriptively “incorrect” sentences (LQs). These tasks were identical to the ones in the acceptability rating experiment.

3.1.3 Participants

Sixty linguistics undergraduate students from the University of Alberta participated in this experiment. All were native speakers of English, and all were different participants than those who participated in the previous study. There were 43 female and 17 male participants, ranging from 17–29 years of age. All participants were reimbursed for their time with course credit.

3.2 Results

The results were analyzed using mixed-effects linear regression. We focus on the total fixation duration (i.e. the total amount of time spent fixating on the AREA OF INTEREST, or AOI) within two AOIs: the idiom as a whole (i.e. the summed fixations on all words within the idiom) and the altered word within the idiom (i.e. the synonymous word in lexical variation, the additional word in the integrated concept, the semantically vague “replacement” word in partial forms, and the word from another idiom in the idiom blend). As above, the analyses only include the 60 experimental idioms.

Ten predictor variables appeared significant in the models. Condition, Know Idiom, Length, meanTransparencyRating, and Trial are the same variables used in the previous experiment. Gender is a factor specifying whether the participant is male or female. PortionAltered is a factor specifying which part of the idiom (i.e. beginning/verb or ending/noun) was manipulated in the variant. And

meanVariationRating is a scaled mean measure of acceptability for a particular idiom with a each type of variation – these averaged ratings were extracted from the previous experiment and included here to determine if participants’ preferences influence their ease of comprehension.

Two measures reflecting the semantic contribution of the constituents were utilized in analyzing these results. meanTransparencyRating (described above) and LSA.Score.Paraphrase, which is a measure of similarity using LATENT SEMANTIC ANALYSIS (LSA), between the words in the idiom and its paraphrase (e.g., *spill the beans* ‘reveal a secret’). This score was obtained from a pairwise comparison of two texts (i.e. an idiom and its paraphrase), which compares the local contexts in order to obtain a value of similarity (Landauer et al. 1998).³ This measure allows us to control for the idiom’s compositionality. If the exact words in the idiom have little to do with the expression’s meaning, then the LSA score will be small (e.g., *cut the mustard* ‘be acceptable’ = 0.07). But if the words used share meaning or contribute to the idiom’s meaning, then the LSA score will be larger (e.g., *stop something in its tracks* ‘stop something’ = 0.87).

As idioms are MWEs, multiple frequency measures were obtained: the frequency of the idiom, frequencies of the individual words, and all possible combinations of adjacent words (e.g. word1 and word2; word2 and word3; word1 and word2 and word3). To avoid collinearity, a PRINCIPAL COMPONENTS ANALYSIS (PCA) was conducted on these frequency measures. Only the first principal component (henceforth PC1. logFrequency) is significant. Finally, a second PCA was conducted on the rating responses for the seven LQs above. Only PC2 (henceforth PC2.LQ) was significant. This latent variable may reflect the participant’s flexibility with language usage.

3.2.1 Idiom as AOI

The first model examines the summed fixation durations on the idiom as a whole. The fixed effects for this model are shown in Table 5. The first interaction, between Condition and KnowIdiom, is shown in the left panel of Figure 2. The canonical form, and the majority of variants, show the same general pattern: shorter fixation durations on known idioms. These variants (except integrated concepts) are therefore shown in grey, as they do not significantly differ from the canonical form. Partial forms however show a different pattern. Fixation durations are relatively similar regardless of whether the participant is familiar with the ex-

³The LSA scores were obtained from the English Lexicon Project (Balota et al. 2007), available at <http://lsa.colorado.edu>.

1 “Spilling the bag” on idiomatic variation

pression or not; thus a facilitation effect for knowing the idiom is not observed as it is with the other variants. This particular variant is fixated upon less than the canonical form, likely due to it being shorter in length (i.e. fewer number of words). This is in line with longer fixations observed on integrated concepts – an additional word is integrated into the idiom, making it longer in length and requiring additional fixations.

Table 5: Fixed effects for the idiom as AOI.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	6.71	0.09	75.97	
Condition=Concept	0.49	0.10	5.04*	130.12
Condition=Blend	0.08	0.10	0.75	
Condition=Lexical	0.01	0.10	0.05	
Condition=Literal	-0.19	0.10	-1.94	
Condition=Partial	-0.75	0.16	-4.80*	
KnowIdiom=Yes	-0.18	0.04	-4.32*	34.84
Length	0.11	0.02	6.76	40.19
PortionIdiomAltered=Ending	-0.06	0.02	-2.52*	3.50
PC2.LQ	-0.07	0.03	-2.42	3.60
LSA.Score.Paraphrase	0.24	0.07	3.49	8.21
meanVariationRating	-0.06	0.01	-7.23	43.80
Gender=Male	-0.17	0.08	-2.17	2.53
TrialScaled	-0.04	0.01	-3.78	10.80
I(KnowIdiom=Yes Condition=Concept)	0.06	0.05	1.16	1.26
I(KnowIdiom=Yes Condition=Blend)	0.08	0.06	1.42	
I(KnowIdiom=Yes Condition=Lexical)	0.08	0.06	1.52	
I(KnowIdiom=Yes Condition=Literal)	0.03	0.06	0.55	
I(KnowIdiom=Yes Condition=Partial)	0.17	0.06	2.75*	
I(Length Condition=Concept)	-0.05	0.02	-2.62	14.11
I(Length Condition=Blend)	-0.01	0.02	-0.36	
I(Length Condition=Lexical)	0.00	0.02	0.20	
I(Length Condition=Literal)	0.02	0.02	1.04	
I(Length Condition=Partial)	0.08	0.03	2.48	

* = Factors that remain significant after a Bonferroni correction

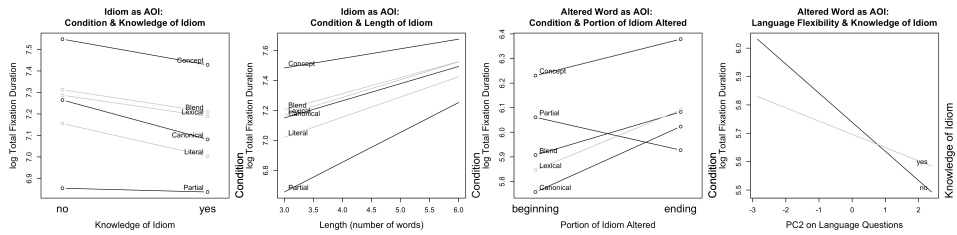


Figure 2: Interactions in the mixed-effects linear regression models for the summed total fixation duration on the whole idiom and the altered word as an AOI. Lines in grey represent factor levels which are not significantly different or slopes which are not significant.

The second interaction, shown in the second panel of Figure 2, is between Condition and Length. Longer idioms show longer summed fixation durations, as expected, due to the increased number of words in the idiom. Lexical variation, formal idiom blends, and the literal meaning of the idiom are not significantly different from the canonical form (shown in grey). The other two variants show a pattern that is significantly different from the canonical form. Integrated concepts show a slight inhibitory effect of length, where an additional concept is more difficult to integrate into shorter idioms (i.e. extra time is required). Whereas partial forms of shorter idioms have even fewer words to fixate upon and therefore show considerably shorter fixation durations. Thus, durations on integrated concepts and partial forms are more comparable to the canonical form when the idiom is longer.⁴

Interestingly, the literal meaning of the idiom shows shorter fixation durations than the canonical form, albeit not quite significantly shorter ($t = -1.94$). The literality of the expression (Titone & Connine 1994a) may be contributing to this result. Nevertheless, a general pattern is evident based on these two above interactions with Condition: variants of the same length as the canonical form are not processed significantly different from this canonical form.

Six main effects are observed in this model. Longer fixation durations are observed on the whole idiom if the beginning (the verb) was altered (i.e. Portion Altered). This is not dependent on the type of variation; all variants are easier to process if the change comes later in the idiom. This is a different result than

⁴PC1.logFrequency was also significant in the idiom as AOI model. However, this variable is strongly correlated with Length ($r = -0.9$). This correlation is unsurprising given that PC1.logFrequency was created using adjacent co-occurrence frequencies. Model comparison shows that Length is the more significant predictor in this model, producing a considerably lower AIC value, and therefore was retained at the expense of PC1.logFrequency.

that of Gibbs and colleagues (Gibbs et al. 1989; Gibbs & Nayak 1989) who found no difference with similarity ratings in whether the noun or verb was altered.

MeanVariationRating is also significant. Variants which received higher acceptability ratings are fixated on less long, suggesting preferred variants are easier to understand and interpret (or perhaps variants easier to interpret are preferred). Longer fixation durations appear on idioms which have higher LSA scores for the idiom’s paraphrase (i.e. LSA.Score.Paraphrase). This finding seems initially surprising, as previous analyses on the comprehension of idioms suggest that idioms are easier to understand when the individual components contribute meaning to the whole (Gibbs et al. 1989). However, the LSA scores indicate how similar the local contexts are between the idiom and its paraphrase (i.e. how interchangeable is the idiom and its paraphrase). When the LSA score is high (i.e. the paraphrase is easily interchangeable), looking time increases as the contexts are not distinctive for the idiom. But if the LSA score is low, then the idiom and its paraphrase are less interchangeable, making the context more distinctive and the idiom more predictable. Interestingly, meanTransparencyRating is not significant. The degree to which the idiom is considered obvious in meaning does not seem to influence the comprehension of idioms or variants.

A main effect is also observed for PC2.LQ, a latent variable representing the participants’ “flexibility” with language (i.e. the more they consider nonstandard or erroneous forms acceptable). Shorter fixations are observed on idioms, both the canonical form and variants, if speakers are more flexible with language. It is interesting to note that this finding is not restricted to only the variants. Gender also shows a significant main effect – males tend to fixate less long on the idiom than females, although we are not quite sure why. Finally, a main effect of Trial is also significant; participants fixate less long on the idiom the further into the experiment they get. But the degree to which each participant is affected by the order of presentation varies, as evidenced by significant by-Subject random slopes for Trial.⁵ By-Item random slopes for Condition with correlation parameters are also significant in this model. These slopes indicate that participants’ fixation durations vary depending on which idiom occurred in which condition – participants found certain idioms easier or more difficult to understand depending on the condition in which they occurred.

3.2.2 Altered word as AOI

We next investigate the fixation duration on the altered word (i.e. the word in the idiom that was manipulated). The fixed effects for this model are shown in Table 6.

⁵Both idiom as AOI and altered word as AOI models have the same random effects structure.

Table 6: Fixed effects for the altered word as AOI.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	5.70	0.06	98.48	
Condition=Concept	0.47	0.06	8.28*	58.40
Condition=Blend	0.15	0.06	2.67*	
Condition=Lexical	0.09	0.06	1.54	
Condition=Partial	0.30	0.07	4.61*	
PortionIdiomAltered=Ending	0.27	0.06	4.49*	17.88
KnowIdiom=Yes	-0.04	0.03	-1.29	0.21
PC2.LQ	-0.10	0.03	-3.12	2.59
PC1.logFrequency	0.03	0.01	4.70	15.43
meanVariationRating	-0.07	0.02	-4.27	14.09
TrialScaled	-0.04	0.01	-2.79	5.28
I(PortionIdiomAltered=Ending Condition=Concept)	-0.12	0.08	-1.46	9.81
I(PortionIdiomAltered=Ending Condition=Blend)	-0.09	0.08	-1.17	
I(PortionIdiomAltered=Ending Condition=Lexical)	-0.02	0.08	-0.26	
I(PortionIdiomAltered=Ending Condition=Partial)	-0.40	0.09	-4.42*	
I(PC2.LQ KnowIdiom=Yes)	0.06	0.02	2.27*	3.09

* = Factors that remain significant after a Bonferroni correction

As there is no altered word in the literal condition, this section focuses on the four idiom variants: lexical variation, partial forms, idiom blends, and integrated concepts, and how they compare to the canonical form.

The interaction between Condition and PortionAltered is seen in the third panel of Figure 2. The overall pattern is that longer fixation durations occur at the end of the idiom, which is also true for the canonical form. Since the idiom occurs at the end of a sentence, these longer fixations may reflect a sentence wrap-up effect (Rayner et al. 2000; Hirotoni et al. 2006). Nevertheless, the altered word for most variants shows significantly longer fixations than the canonical form. This is not true of lexical variation, which is the only variant that does not have significantly longer fixations than the canonical form ($t = 1.54$). Thus, a lexically altered variant is just as easy to process as the canonical form. Partial forms however, appear considerably different from the canonical form. Longer fixations are observed on the altered word when the beginning has been altered, as in *use the grapevine*. But when the ending is altered (e.g., *spilled it*), fixations on the altered word are not significantly different from the canonical form ($t = -1.44$). Altering the verb does not always result in significantly longer fixations (cf. the non-significantly different lexical variant when the beginning is altered),

however altering the verb to a semantically vague verb (i.e. *be, do, used* – in order to make the sentence grammatical) does significantly inhibit processing.

The second interaction, shown in the last panel of Figure 2, is between knowledge of the idiom (i.e. *KnowIdiom*) and the participant’s flexibility with language (i.e. *PC2.LQ*). Flexibility with language only appears facilitative for those who do not know the idiom, illustrated by the non-significant slope (in grey) for those who know the expression. Other strategies are apparently relied upon to interpret the idiom when knowledge of it is not available.

Additional main effects are also observed. Fixation durations are longer on the altered word when the co-occurrence frequencies of the idiom are higher. Thus, altering part of a more frequent sequence causes greater processing costs. In addition, participants have shorter fixation durations when the variant is rated as more acceptable (i.e. *meanVariationRating*). The more the variation strategy is preferred with a particular idiom, the easier it is to interpret. Finally, the further the participants get into the experiment (i.e. *Trial*), the shorter their fixation durations on the altered word.

We also specifically looked at idioms blends, to determine whether the syntax or the semantics of the two merged idioms affects the processing of this variant. Interestingly, neither of these variables were predictive of fixation duration – we can understand idiom blends regardless of the syntax or the semantics of the two idioms used in the blend.

Some of these alternations may have been surprising to the participants, resulting in effects that continued beyond the altered word. We therefore ran a model to explore any spillover effects from the altered word, shown in Table 7. As the idiom occurred in sentence-final position, spillover effects from an altered noun (i.e. the end of the idiom) are not able to be determined; thus, this model only focuses on spillover effects from an altered verb. We examined the fixation duration on the first content word after the verb when the verb was manipulated (i.e. the alternation occurred at the beginning of the idiom).

Spillover effects are observed for all variant types (i.e. *Condition*), but the longest durations are for integrated concepts and partial forms. Incorporating an additional word into an idiom results in a processing cost likely due to the surprisal of this extra word. Integrating this additional information into the idiom and context requires extra time. The largest spillover effect is with partial forms. It appears that the semantically vague words used in these sentences (to make them grammatical) make these partial forms more difficult to comprehend and cause considerable spillover effects. It remains to be determined whether partial forms from more naturalistic language produce this same effect.

Table 7: Fixed effects for the first content word after the verb.

	Estimate	Std. Error	t-value	Δ AIC
Intercept	5.95	0.08	73.41	
Condition=Concept	0.27	0.07	3.76*	11.6
Condition=Blend	0.17	0.06	2.75*	
Condition=Lexical	0.14	0.05	2.92*	
Condition=Partial	0.30	0.06	4.62*	
PC1.logFrequency	0.04	0.01	3.54	6.38
KnowIdiom=Yes	-0.11	0.05	-2.32*	3.20

* = Factors that remain significant after a Bonferroni correction

The last two effects are PC1.Frequency and KnowIdiom. The higher the co-occurrence frequencies of the idiom, the longer the fixation duration on the first content word after the alternation. Modifying a frequent multiword sequence inhibits processing. However, these spillover effects are reduced if the idiom is familiar (i.e. KnowIdiom).

4 Discussion

This study employed a multi-methodological approach to investigate the acceptability and processing of idiomatic variation. One advantage of using multiple methods is that they can reveal greater insights, by contrasting converging and diverging results between the different methods. Converging results can provide greater confidence that a particular result or predictor variable is robust; whereas diverging results can uncover differences due to a specific modality or shed light on findings concerning the larger picture that would otherwise be overlooked or thought contradictory (Arppe & Järvikivi 2007). The findings between the acceptability rating and eye-tracking experiments presented together in this chapter do in fact show converging and diverging results worthy of discussion.

Interestingly, the findings between the experiments primarily show diverging results with regard to our two research questions. For example, our first research question asks how the variants compare with the canonical form, and we see from the acceptability ratings that the canonical form is rated as more acceptable than variants or a literal reading, with speakers clearly preferring this form. However, the processing differences are not nearly as straightforward. Some vari-

ants are processed differently than the canonical form. The variant showing the greatest difference from the canonical form is the partial form of the idiom (e.g., *use the grapevine*). This idiom variant is fixated on less than the canonical form, as expected, largely due to the omission of a word (or words) from the expression. Yet despite this shorter fixation on the whole idiom, participants fixated significantly longer on the “replacement” verbs (i.e. the semantically vague verbs used to connect the idiom to the sentence) and significant spillover effects were observed on the first content word after these verbs. A similar inhibitory effect was not observed if the ending was modified (e.g., *spilled it*). These results are likely due to the design of the experiment. The tightly controlled stimuli used in this study made these partial forms unnatural and difficult to interpret. A study investigating partial forms in naturally occurring language may shed more light on the degree of difficulty for processing this variant.

Idioms with additional concepts integrated into the expression are also processed differently from the canonical form. These variants require additional processing time, as anticipated, but this longer reading time is largely attributable to the extra word in the expression. The longer duration on the whole idiom is very similar to the altered word AOI, suggesting that this variant experiences very little processing costs over and above having to read an extra word.

However, modification of an idiom’s form does not always result in a processing disadvantage. Some variants – lexical variation, formal idiom blends, and a literal reading of the idiom – are not processed significantly slower than the canonical. Differences between these variants and the canonical form are observed, such as longer fixations on the altered word (at least for idiom blends) or some spillover effects if the verb was altered, but these differences do not result in longer reading times for the idiom as a whole. These findings are partly in line with our predictions. Only idiom blends were predicted to be processed slower than the canonical form, due to the potential surprisal at or unrecognizability with this so-called error. But as observed, they do not present difficulties in comprehension. Thus, intentional or not, altering a word within an idiom to a synonymous or non-synonymous word does not result in a processing cost.

Our second research question asks how variants compare with each other. Once again, diverging results between the two methods are evident. Lexical variation and idiom blends are processed quite similarly, showing comparable fixation durations, to each other and to the canonical form. The length of the original idiom is maintained in these variants, possibly explaining these comparable durations. However, they do not share similar acceptabilities. Lexically modified variants are considered much more acceptable than idiom blends. In fact, idiom

blends are even less preferred when the two idioms used to make the blend share similar semantics, possibly explaining why blends are often viewed as errors (Fay 1982; Cutting & Bock 1997). Meanwhile, integrated concepts, which add extra information into the idiom, show longer reading times than the other variants, yet are the most preferred. This higher acceptability was expected, given their relatively frequent occurrence in corpora (Moon 1998; Schröder 2013), and leaves us wondering whether semantically productive lexical variants (cf. McGlone et al. 1994) would show higher levels of acceptability (on par with integrated concepts) compared with the synonymous lexical variants utilized here (following Gibbs et al. 1989). Finally, partial forms and a literal reading of the idiom are not acceptable variation strategies, even though they have comparable (or shorter) reading times to the canonical form.

The findings from these two research questions present two main observations. First, variants which add an extra element or are truncated in some way show longer or shorter reading times, respectively, while modifications that maintain the same length as the canonical form show comparable reading times to the canonical form. Second, variants that preserve more of the canonical form (e.g. integrated concepts, lexical variation) are considered more acceptable, although preference remains with the canonical form (which likely facilitates the learning of idioms and leads to faster recognition).

One cautionary note must be made. These aggregated results show patterns and preferences, but they do not imply that all idioms can be altered using all variation strategies. Much variability, particularly when it comes to comprehension, is also observed. Including the mean acceptability for each idiom in each condition as a control variable in the comprehension models resulted in preferred variants showing shorter reading times. In other words, the way in which an idiom is modified can affect how easy it is to understand. Variability is also observed in the random effects structure of the comprehension models, which have by-Item random slopes with correlation parameters for *Condition*, indicating that specific idioms can be easier or more difficult to process depending with which condition they occurred. Thus, while variation is possible and general patterns can be observed, there are also idiom-specific preferences that factor into how an idiom is altered, understood, and appreciated.

Converging and diverging results are also observed with the predictor variables in the analyses. Two variables converge between the two methods. Length is shown to be an important predictor and yet is rarely included in the idiom literature (cf. Fanari et al. 2010). Longer idioms require additional processing time, as expected, and there can be some facilitation or inhibitory processing effects de-

pending on the type of variation encountered. A literal reading gains additional approval when the idiom is longer (and perhaps also more transparent), while shorter idioms are even more preferred with the idiomatic reading. Perhaps the extra words in longer idioms clearly identify the metaphorical links associated with the idiom, making a literal reading also more interpretable.

The participant’s knowledge of the expression is another important predictor that converges between the two studies. Participants fixated less on the idiom (i.e. shorter reading times) and were faster to rate the expression when they knew the idiom. They also considered the idiom and its variants as more acceptable when it was familiar. Yet surprisingly, research on idioms tends to include an average measure of familiarity (cf. Titone & Connine 1994b), as a control for frequency or as a measure of subjective familiarity. This study demonstrates that a speaker-specific measure of familiarity is important for idiom research, as it incorporates speaker-specific experiences into the model.

Not all participant-related variables show converging results. The language questions (LQs) that were collected to provide a latent measure of the participant’s flexibility with language only appeared significant in predicting the comprehension of idioms, and not their acceptability. Participants who are more flexible (i.e. more accepting of non-standard or erroneous forms) have an easier time processing idioms and variants. This of course makes sense; these speakers are not distracted by the specific form used, but focus solely on the message being conveyed.

Frequency was also not predictive of acceptability. Even highly frequent idioms can be regarded as acceptable when altered. But frequency is predictive of comprehension. This variable only appeared in the altered word model (due to the high correlation with length in the idiom model), and revealed that alternations made to frequent idioms result in a processing cost. When a sequence of words that typically occur together has been modified, additional time is required to interpret the new sequence, as the advantage it once received due to its predictability is no longer available. The opposite pattern is observed for the semantics of formal idiom blends – a significant predictor of acceptability, but not comprehension. Speakers find blends unacceptable when the merged idioms share similar semantics, but appear to have no difficulty interpreting them.

A divergence is also seen with the variable `PortionAltered` (i.e. where in the idiom the alternation occurred: beginning/verb or ending/noun). This variable is not predictive of acceptability – participants’ judgements were not affected by where in the idiom the alternation occurred (i.e. modifications to nouns and verbs are equally acceptable). However, this variable is predictive of comprehension –

alternations made earlier in the idiom (the verb) result in greater processing costs. Gibbs et al. (1989) found no difference between modifications made to nouns or verbs in their similarity rating task, providing further confirmation that a subjective rating of similarity is not measuring comprehension. In addition, these results may also provide support for a time-dependent nature of idiom processing (Titone & Libben 2014). As one advances through the idiom, the predictability of the idiom becomes greater and the idiomatic meaning accumulates resulting in greater priming effects for later words. It seems reasonable then that changes made later in the expression will be less costly – the meaning is more predictable even if changes have been made.

Finally, a divergence is also evident between which compositionality measure was determined to be predictive for each modality. A measure of transparency is predictive of the acceptability rating responses; speakers prefer idioms that are transparent and clear in meaning. But an objective measure of contextual similarity is predictive of comprehension. Idioms are faster to process in unique or distinctive contexts (i.e. lower LSA scores), because they are more predictable. Thus, evaluative judgements are influenced by the clarity of the expression, whereas comprehension is affected by the local contexts in which the idiom occurs.

These (largely diverging) results, in regards to the predictor variables, nicely capture patterns between the two methods. Clarity of the expression and motivation for the alternation are important for the acceptability of idioms and variants, whereas the placement of the alternation and the local context (i.e. distinctiveness, as well as disruptions to this context) are important for comprehension.

This study has shown that not all variants are processed significantly different than the canonical form and that the predictability of idioms is important, especially during processing. Yet these findings conflict with traditional views of idioms, which claim that idioms cannot be modified without losing their idiomatic meaning, or that idioms are stored and accessed whole along with their idiomatic meaning, since they do not equal the sum of their parts. These traditional approaches proposed a dual-route model to account for the processing of idioms – literal language would be understood incrementally through ordinary linguistic processing and idioms would be accessed directly along with their meaning (cf. Swinney & Cutler 1979; Cacciari & Tabossi 1988). For instance, they could be activated by accessing the “idiom key” (Cacciari & Tabossi 1988), which is the idiomatic configuration indicating that sufficient input has been received. But how does one receive sufficient input if the form has been altered? One proposal is to store each variant, but this inefficient method of handling variation would result in a large burden being placed on the mental lexicon (Baayen et al.

2013). McGlone et al. (1994) proposed that idioms are accessed whole for faster processing, but that they could be understood through ordinary linguistic processing, while variants are understood like literal language (which is why they are processed slower), using various strategies in order to understand them. But this study showed that not all variants are processed slower – variants of the same length as the canonical are processed just as quickly.

This study also shows the importance of predictability in understanding idioms – when the local context is distinctive, idioms are faster to process; when alternations are made later in the expression, variants are easier to process; when frequent sequences are altered, variants are slower to process; and even the more flexible a speaker is with language, the easier idioms are to process. These results are in line with other aspects of predictability or probability seen elsewhere in language. Idioms that have a higher cloze probability have an idiomatic meaning that is available earlier (Cacciari & Tabossi 1988; Titone & Connine 1994a). The combination of words can lead to certain predictions or expectations (Elman 2011): subject-verb combinations lead to predictions about the upcoming object (e.g. *lumberjack cuts* primes *wood*, whereas *surgeon cuts* primes *bone*), and the type of semantic theme can be predicted based on voice (e.g. *she arrested* primes *crook*, but *she was arrested* primes *cop*). Speakers have even been shown to make accurate probabilistic predictions about the type of syntactic choice made by others; for example, in the dative alternation: *because he brought the pony to my children* vs. *because he brought my children the pony* (Bresnan 2007).

These predictions would not be possible if language was understood in a truly compositional way. Some scholars are therefore challenging the traditional view of the mental lexicon, as a list of dictionary entries, and instead are proposing probabilistic approaches to language, where words do not possess meaning but are simply cues to meaning, modulated by context and experience (Elman 2004; 2011; Ramscar & Baayen 2013). These approaches highlight the vast amount of information speakers have available to them, besides simply the meaning of the word – speakers are able to draw upon past experience, cultural norms, event and world knowledge and even the feelings of the speaker to interpret the meaning being communicated.

In one such framework, Implicit Grammar (Baayen & Ramscar 2015), learning a language is about learning which cues (i.e. sounds, morphemes, words, contexts) are informative, or discriminative, for a particular outcome (i.e. meaning). Thus, learning occurs when cues successfully predict outcomes, but also when predictions fail to result in those outcomes. Under this view, idioms and their variants would be processed similarly to literal language: being a sequence of words

which are cues to the intended meaning (cf. Geeraert, Newman, et al. 2017). This is likely why speakers prefer the canonical form – using perfectly good cues for accessing the intended idiomatic meaning. But altering these cues is still possible. Integrated concepts still use the canonical form, but with an additional word inserted into the expression. This extra information takes additional time to integrate, but does not alter the already established cues, making this the most preferred variant. Whereas lexical variation and idiom blends alter one of the cues in the idiom, causing relearning in order to discriminate the new cue with the intended meaning, and making these variants less appreciated. This approach may also explain why idioms can become shorter or truncated over time: certain cues are better at discriminating the intended meaning, while others become irrelevant and are eventually dropped. But before this natural development happens, omitting (potentially useful) cues is considerably less appreciated.

Acknowledgements

This study was generously supported by an Izaak Walton Killam memorial scholarship, awarded to the first author. In addition, we would like to thank Lauren Rudat for her suggestions on improving the stimuli, and to the editors and anonymous reviewers for their suggestions on improving the text.

Abbreviations

AOI	area of interest	PCA	principal components analysis
LQ	language question	RT	reaction time
LSA	latent semantic analysis	VAS	visual analogue scale
MWE	multiword expression		

Appendix A:

Canonical form for the 60 idioms used in the two studies

beat around the bush	bend the rules
burn a hole in your pocket	burn your bridges
bury the hatchet	chomp at the bit
cut the mustard	dragged through the mud
drink someone under the table	drive someone up the wall
drown your sorrows	face the music
fall by the wayside	flip your lid
fly off the handle	foot the bill
get the show on the road	get under someone's skin
give up the ghost	give up the ship
go against the grain	go behind someone's back
go through the roof	go with the flow
grind to a halt	hang up your boots
have a card up your sleeve	have many irons in the fire
have someone over a barrel	have your finger on the pulse
hear something through the grapevine	jump on the bandwagon
keep someone on their toes	keep your eye on the ball
keep your nose to the grindstone	lie through your teeth
line your pockets	lose your marbles
nip something in the bud	paint yourself into a corner
pick your brain	pull someone's leg
pull the strings	pull up your socks
put your foot in your mouth	rock the boat
run the gauntlet	shake in your boots
shoot the breeze	shoot the messenger
skate on thin ice	spill the beans
spin your wheels	stick to your guns
swallow your pride	sweep something under the rug
tear a strip off someone	wash your hands of something
wear your heart on your sleeve	wrap someone around your finger

References

- Arppe, Antti & Juhani Järvikivi. 2007. Every method counts – combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3. 131–160.
- Ayto, John (ed.). 2009. *From the horse's mouth: Oxford dictionary of English idioms*. Oxford: Oxford University Press.
- Baayen, R. Harald, Peter Hendrix & Michael Ramscar. 2013. Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech* 56. 329–347.
- Baayen, R. Harald & Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 100–120. Berlin: Mouton de Gruyter.
- Balota, David A., Melvin J. Yap, Michael J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson & Rebecca Treiman. 2007. English lexicon project. *Behavior Research Methods* 39(3). 445–459.
- Barlow, Michael. 2000. Usage, blends and grammar. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 315–345. Stanford, CA: CSLI Publications.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI:10.18637/jss.v067.i01
- Bobrow, Samuel A. & Susan M. Bell. 1973. On catching on to idiomatic expressions. *Memory & Cognition* 1. 343–346.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Berlin/New York: Mouton de Gruyter.
- Cacciari, Cristina, Paola Corradini & Roberto Padovani. 2005. Speed of processing effects on spoken idioms comprehension. In Bruno G. Bara, Lawrence Barsalou & Monica Bucciarelli (eds.), *Proceedings of the XXVII annual meeting of the cognitive science society*, 21–23. New Jersey: Lawrence Erlbaum.
- Cacciari, Cristina & Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27. 668–683.
- Cutting, J. Cooper & Kathryn Bock. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition* 25(1). 57–71.

- Duffley, Patrick J. 2013. How creativity strains conventionality in the use of idiomatic expressions. In Mike Borkent, Barbara Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 49–61. Stanford, CA: CSLI Publications.
- Elman, Jeffrey L. 2004. An alternative view of the mental lexicon. *Trends in Cognitive Sciences* 8(7). 301–306.
- Elman, Jeffrey L. 2011. Lexical knowledge without a lexicon. *The Mental Lexicon* 6. 1–33.
- Fanari, Rachele, Cristina Cacciari & Patrizia Tabossi. 2010. The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology* 22(3). 321–334.
- Fay, David. 1982. Substitutions and splices: A study of sentence blends. In Anne Cutler (ed.), *Slips of the tongue and language production*, 163–195. Amsterdam: Mouton de Gruyter.
- Funke, Frederik & Ulf-Dietrich Reips. 2012. Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods* 24(3). 310–327.
- Geeraert, Kristina. 2016. *Climbing on the bandwagon of idiomatic variation: A multi-methodological approach*. University of Alberta Doctoral Dissertation.
- Geeraert, Kristina, R. Harald Baayen & John Newman. 2017. Understanding idiomatic variation. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE2017)*, 80–90. Association for Computational Linguistics. April 4, 2017.
- Geeraert, Kristina, John Newman & R. Harald Baayen. 2017. Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science* 9(3). 1–17. DOI:10.1111/tops.12263
- Gibbs, Raymond W. 1980. Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition* 8(2). 149–156.
- Gibbs, Raymond W. & Nandini P. Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology* 21. 100–138.
- Gibbs, Raymond W., Nandini P. Nayak, John L. Bolton & Melissa E. Keppel. 1989. Speakers’ assumptions about the lexical flexibility of idioms. *Memory & Cognition* 17(1). 58–68.
- Hirofani, Masako, Lyn Frazier & Keith Rayner. 2006. Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language* 54. 425–443.
- Landauer, Thomas K., Peter W. Foltz & Darrell Laham. 1998. Introduction to latent semantic analysis. *Discourse Processes* 25. 259–184.

- Langlotz, Andreas. 2006. *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: John Benjamins.
- McGlone, Matthew S., Sam Glucksberg & Cristina Cacciari. 1994. Semantic productivity and idiom comprehension. *Discourse Processes* 17. 167–190.
- Miller, George A. & Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1). 1–28.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English*. Oxford: Oxford University Press.
- R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>. ISBN 3-900051-07-0.
- Ramsar, Michael & R. Harald Baayen. 2013. Production, comprehension, and synthesis: A communicative perspective on language. *Frontiers in Psychology* 4. 233. DOI:10.3389/fpsyg.2013.00233
- Rayner, Keith, Gretchen Kambe & Susan A. Duffy. 2000. The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology* 53A(4). 1061–1080.
- Savary, Agata, Marie Candito, Verginica Barbu Mititelu, Eduard Bejček, Fabienne Cap, Slavomír Čéplö, Silvio Ricardo Cordeiro, Gülşen Eryiğit, Voula Giouli, Maarten van Gompel, Yaakov HaCohen-Kerner, Jolanta Kovalevskaitė, Simon Krek, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Lonneke van der Plas, Behrang QasemiZadeh, Carlos Ramisch, Federico Sangati, Ivelina Stoyanova & Veronika Vincze. 2018. PARSEME multilingual corpus of verbal multiword expressions. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, 87–147. Berlin: Language Science Press. DOI:10.5281/zenodo.1469555
- Scholivet, Manon, Carlos Ramisch & Silvio Cordeiro. 2018. Sequence models and lexical resources for MWE identification in French. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, 263–297. Berlin: Language Science Press. DOI:10.5281/zenodo.1469567
- Schröder, Daniela. 2013. *The syntactic flexibility of idioms: A corpus-based approach*. Munich: AVM.
- Sinclair, John (ed.). 2011. *Collins COBUILD idioms dictionary*. Harper Collins.
- Swinney, David A. & Anne Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour* 18. 523–534.

- Tabossi, Patrizia, Rachele Fanari & Kinou Wolf. 2008. Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(2). 313–327.
- Titone, Debra A. & Cynthia M. Connine. 1994a. Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20(5). 1126–1138.
- Titone, Debra A. & Cynthia M. Connine. 1994b. Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbolic Activity* 9(4). 247–270.
- Titone, Debra A. & Cynthia M. Connine. 1999. On the compositional and non-compositional nature of idiomatic expressions. *Journal of Pragmatics* 31. 1655–1674.
- Titone, Debra A. & Maya R. Libben. 2014. Time-dependent effects of decomposability, familiarity and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon* 9. 473–496.

