TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 9 (2017) 653–669

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12263

This article is part of the topic "More Than Words: The Role of Multiword Sequences in Language Learning and Use," Morten Christiansen and Inbal Arnon (Topic Editors). For a full listing of topic papers, see: http://onlinelibrary.wiley.com/doi/10.1111/tops.2017.9.issue-2/issuetoc

Idiom Variation: Experimental Data and a Blueprint of a Computational Model

Kristina Geeraert,[a] John Newman,[b,c] R. Harald Baayen[b,d]

[a]*Department of Linguistics, KU Leuven*
[b]*Department of Linguistics, University of Alberta*
[c]*School of Languages, Literatures, Cultures & Linguistics, Monash University*
[d]*Department of Linguistics, University of Tübingen*

Received 2 March 2016; received in revised form 7 October 2016; accepted 12 October 2016

Abstract

Corpus surveys have shown that the exact forms with which idioms are realized are subject to variation. We report a rating experiment showing that such alternative realizations have varying degrees of acceptability. Idiom variation challenges processing theories associating idioms with fixed multi-word form units (Bobrow & Bell, 1973), fixed configurations of words (Cacciari & Tabossi, 1988), or fixed superlemmas (Sprenger, Levelt, & Kempen, 2006), as they do not explain how it can be that speakers produce variant forms that listeners can still make sense of. A computational model simulating comprehension with naive discriminative learning is introduced that provides an explanation for the different degrees of acceptability of several idiom variant types. Implications for multi-word units in general are discussed.

Keywords: Idiom variation; Acceptability ratings; Computational modeling; Salience; Naive discriminative learning

1. Introduction

Idioms are traditionally described as multi-word units that have idiosyncratic meanings and that are highly restricted in their realization (Levelt, 1989; Nunberg, Sag, & Wasow,

Correspondence should be sent to Kristina Geeraert, Department of Linguistics, Blijde-Inkomststraat 21, PO Box 3308, B-3000 Leuven, Belgium. E-mail: kristina.geeraert@kuleuven.be

1994; Culicover, Jackendoff, & Audring, 2017, this volume). Early studies proposed to account for idiom processing by positing fixed multi-word units serving as access representations (Bobrow & Bell, 1973; Swinney & Cutler, 1979). However, several priming studies indicate that the structure and lexical elements of idioms are accessible just as in non-idiomatic sequences of words (Cacciari & Tabossi, 1988; Konopka & Bock, 2009; Sprenger, Levelt, & Kempen, 2006). The configuration hypothesis of Cacciari and Tabossi (1988) accounts for these findings by positing that idioms are configurations of words. These words are exactly the same as those activated in reading non-idiomatic sentences. But as an idiomatic expression is processed, word by word, evidence accumulates for its idiomatic interpretation. For speech production, Sprenger et al. (2006) proposed both conceptual nodes and superlemma nodes, with the latter linking up to the lemma nodes of idioms' individual words.

These models all presuppose that idioms are fixed expressions. A series of corpus surveys (cf. Langlotz, 2006; Moon, 1998; Schröder, 2013; Wulff, 2008), however, has brought to the fore that the forms of idioms are much less rigid than had been assumed. They are actually utilized in a variety of ways and with varying degrees of productivity. For example, Moon (1998) observed that idioms can occur with lexical variation in nouns (e.g., *a skeleton in the closet/cupboard*), verbs (e.g., *say/kiss goodbye to something*), and even particles (e.g., *rap someone on/over the knuckles*). Variations can include truncations (e.g., *don't count your chickens* [*before they're hatched*]), reversals (e.g., *can't eat/have your cake and have/eat it too*), with homophonous words (e.g., *dull as ditchwater/dishwater*), and even insertions (e.g., *we're a little late getting our Christmas act together*). Duffley (2013) found a surprising amount of variation for *kick the bucket* and *shoot the breeze* on the Internet, such as passives (e.g., *no buckets were kicked*), lexical variation (e.g., *my phone kicked the pail last week*), and even inserted concepts (e.g., *was ready to kick its digital bucket*), challenging the perspective on idioms as summarized in, for instance, Levelt (1989).

One question that these findings give rise to is how similar the meanings are of different idiom variants. Gibbs and colleagues (cf. Gibbs & Nayak, 1989; Gibbs, Nayak, Bolton, & Keppel, 1989) addressed this question with a semantic similarity rating task, comparing variants of "normally decomposable idioms" (idioms whose constituents directly contribute meaning to the expression, such as *lay down the law*), "abnormally decomposable idioms" (idioms whose constituents contribute meaning indirectly through a metaphorical relationship, like *meet your maker*), and "nondecomposable idioms" (idioms whose constituents do not contribute to the meaning of the expression, such as *kick the bucket*). Normally decomposable idioms with lexical and syntactic alternations were rated as more similar in meaning to literal paraphrases than nondecomposable idioms. However, Tabossi, Fanari, and Wolf (2008) conducted a replication study of Gibbs and Nayak (1989) and found that both normally decomposable idioms and nondecomposable idioms were rated as more semantically similar to their literal paraphrases when syntactically varied than were abnormally decomposable idioms. Meanwhile, Tabossi et al. (2008) and Titone and Connine (1994) conducted replication studies where they asked participants to group idioms into their decomposability categories. Both studies found that

participants were unreliable at categorizing idioms, performing at chance. Discrepancies are even observed in the literature as to which category an idiom belongs; for example, Gibbs, Nayak, Bolton, and Keppel (1989) list *button your lips* as normally decomposable, while Libben and Titone (2008) list it as abnormally decomposable.

A second question that the attested idiom variability gives rise to is how acceptable the variants of an idiom are compared to its canonical form. Corpus studies have shown that idioms can occur with a range of variation, but are all variational options equally acceptable? This study addresses this question by means of an acceptability rating experiment. We systematically varied a large number of idioms (changing words, adding words in, or replacing content words by pronouns), and observed systematic differences in acceptability depending on variant type. Although some scholars take acceptability ratings to reflect representations and/or their availability (cf. Fanselow & Frisch, 2006), we take such ratings to be simply informative about speaker preferences.

We assume that these preferences arise as a function over time of the cumulative support for the idiom's meaning. Therefore, to better understand what makes an idiom variant more, or less acceptable, we carried out a simulation study with naive discriminative learning to see if the results from the two methods converge. Naive discriminative learning makes use of wide learning networks (two-layer networks with weights estimated by the Rescorla–Wagner learning rule) to approximate implicit error-driven learning. Consistent with earlier work (Baayen, Milin, Filipovic-Durdevic, Hendrix, & Marelli, 2011; Baayen, Milin, & Ramscar, 2016; Baayen, Shaoul, Willits, & Ramscar, 2016), we adopt a minimalist "end-to-end" perspective that proceeds from sublexical form units such as letter pairs to semantic units without mediation by morpheme, word, or word n-gram form representations. We shall see that this approach opens up new ways of understanding the comprehension of idiom variants.

In what follows, we first report our experiment. We then introduce the computational model, based on the experimental stimuli, and discuss selected idioms to illustrate model predictions for idiom variant processing. We conclude with a discussion of our findings and their contribution in the larger context of the current discussion of multi-word units in lexical processing.

2. Acceptability ratings 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): 20 three-word idioms consisting of a verb and a noun phrase (*rock the boat*); 20 four-word idioms consisting of a verb and a prepositional phrase (*jump on the bandwagon*); and 20 five- or six-word idioms (10 each) consisting of a verb, noun phrase, and a prepositional phrase (*hear something through the grapevine*). Two contexts were created for each idiom: one

literal and one figurative (e.g., *I used to be a socialite, and I would hear things through the grapevine* = figurative; and *I used to pretend I could talk to plants, and I would hear things through the grapevine* = literal). Both contexts had identical final clauses, with the idiom in sentence-final position.

These idioms were manipulated for four types of variation, selected based on the literature and controllability in an experimental design. First, lexical variation, where one of the lexical items in the expression was altered to a synonymous or near-synonymous word (*discover something through the grapevine*). An online thesaurus was utilized for synonymous words (http://www.thesaurus.com/). Second, partial form of the idiom, where only a portion of the expression was presented, usually a key word (*use the grapevine*). In order for the sentence to still be grammatically correct, pronouns or lexically vague words were used to replace 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 (*hear something through the judgemental grapevine*). These additional concepts expanded or emphasized the figurative context in which the idiom occurred. Finally, a formal idiom blend, where two idioms were blended together (*get wind through the grapevine*—blending *hear something through the grapevine* with *get wind of something*). Half of the idioms had the beginning portion of the expression altered, while the other half had alternations made to the final portion of the expression. 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. Therefore, 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 (e.g., *While the guys were reshingling, they suddenly went through the roof.*)
2. **Canonical Form** of the idiom in a figurative context (e.g., *Although these were new stocks, they suddenly went through the roof.*)
3. **Lexical Variation** of the idiom in a figurative context (e.g., *Although these were new stocks, they suddenly went through the ceiling.*)
4. **Partial Form** of the idiom in a figurative context (e.g., *Although these were new stocks, they suddenly went through it.*)
5. **Integrated Concept** within the idiom in a figurative context (e.g., *Although these were new stocks, they suddenly went through the investment roof.*)
6. **Idiom Blend** of two idioms in a figurative context (e.g., *Although these were new stocks, they suddenly went through the charts.*)

The other idioms used in the idiom blend condition (the second idiom in the blend) were used as fillers in their canonical form in the other five versions of the experiment. Each 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 these "blending"

idioms occurred in a formal blend in the idiom blend condition, while the remaining 50 appeared in their canonical form as fillers—20 in a figurative context and 30 in a literal context. This increased the number of literal contexts in the experiment, reducing their underrepresentation. In sum, each participant saw 110 items: 60 experimental idioms (10 in each of the six conditions) and 50 "blending" idioms as fillers.

Six practice sentences were created using six "practice" idioms. All occurred in their canonical form, three in a figurative context and three in a literal one. 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, which is a continuous graphical rating scale that allows fine gradations to be measured (Freyd, 1923; Funke & Reips, 2012).

Participants were 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 labeled with "acceptable" on the extreme right and "unacceptable" on the extreme left. The mouse was repositioned to the center of the scale on each trial. Participants were encouraged to use the whole scale before the experiment began, and they were given an opportunity to take a short break halfway through the experiment.

After the participants had rated the acceptability of the idiom variants, they were asked whether they were familiar with the idioms. As different speakers are familiar with different subsets of idioms, the predictor KnowIdiom allowed us to control, at the level of the individual, whether they knew the idiom (see Cacciari, Corradini, & Padovani, 2005, for subject variability in idiom processing), while at the same time maximizing the number of idioms used 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 labeled "yes" and the other labeled "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.

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 to 43 years of age. All participants were reimbursed for their time with course credit.

2.2. Results

2.2.1. Variables

Five 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. `KnowIdiom` is a factor indicating the participant's knowledge of the idiom (i.e., yes or no). `And Trial` is the 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 standardized average rating for the transparency (or clarity) of the idiom's meaning as a whole. Since speakers differ in how they interpret the decomposability of idioms, as evidenced by the low reliability of the decomposability classification task (cf. Titone & Connine, 1994), 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, and it is in line with other proposals of an idiomaticity continuum (cf. Wulff, 2008). 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 (see Geeraert, 2016, for further details). The average rating for each idiom was included as a separate predictor to determine whether transparency influences people's preferences of variation. For the purposes of this study, `meanTransparencyRating` is a control variable, included to disentangle acceptability from decomposability.

2.2.2. Acceptability rating responses

The results were analyzed with mixed-effects linear regression using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2014). Only the 60 experimental idioms were included in this analysis (i.e., the fillers were not included outside of the idiom blend condition). A selection of the results will be discussed, specifically the interactions with `Condition`. The full model is summarized in the Supplementary Material file. Further details of this experiment and analysis are available in Geeraert (2016).

Although this experiment was largely exploratory, we had some expectations about the results, such as the canonical form and literal meaning being the most and least preferred, respectively. We had some predictions for the variants as well: Idiom blends would be less acceptable, due to their "error-like" status in the literature, while integrated concepts would be more acceptable, due to their frequent occurrence in corpora. And these predictions are in fact observed in Fig. 1.

The left panel of Fig. 1 shows the interaction between `Condition` and `KnowIdiom`. 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

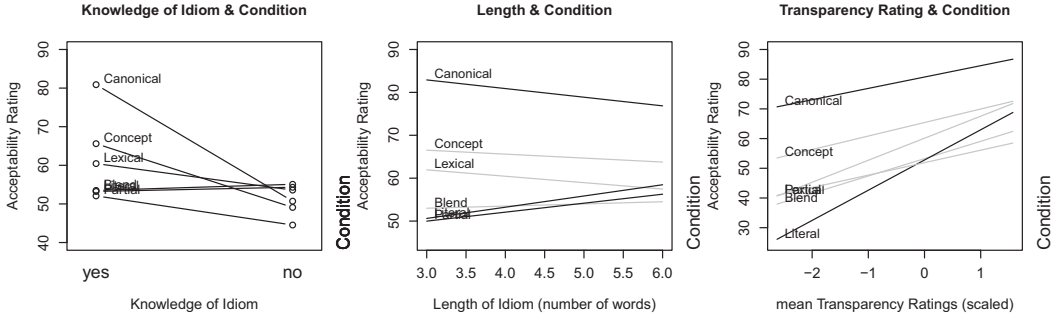

Fig. 1. Interactions in the mixed model for the acceptability ratings of idiomatic variants. Lines in gray represent slopes that do not differ significantly from the slope of the canonical form.

others, with a slight preference for variants with 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.

`Length` also occurs in a significant interaction with `Condition`, shown in the center panel of Fig. 1. Participants tend to rate idioms as less acceptable in their canonical form if they are longer. This pattern also holds for most variants: Integrated concepts, lexical variation, and formal idiom blends have slopes which are not significantly different from the canonical form and are therefore depicted in gray. Literal meaning and partial forms, however, are rated as more acceptable if the idiom is longer. Apparently, literal interpretations (which are likely to characterize also the partial forms) benefit from the presence of many words, whereas idiomatic interpretations suffer. This suggests 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 right panel of Fig. 1. Higher acceptability ratings 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 the other idiom variants. This result is not unexpected, given that not all idioms have a plausible literal meaning (cf. Titone & Connine, 1994).

The frequency of occurrence of the idiom, as well as several measures derived from word n-gram frequencies for subsequences of n words, were not predictive for the acceptability ratings (all $|t| < 1$).

In summary, this experiment explored the acceptability of idiomatic variation, using several types of variants. Modifying an idiom makes it less acceptable, but the decrease in acceptability varies substantially depending on the type of modification. Modifying the idiom with further concepts (*go through the investment roof*) and replacing words with near-synonyms (*go through the ceiling*) resulted in more acceptable idiom variants than partial forms (*go through it*) and idiom blends (*go through the charts*). In addition, idioms which are known, more obvious in meaning, and shorter were considered most acceptable. This was true for most variants as well; only partial forms and literal

meanings (i.e., the least accepted variants) deviated from this pattern, perhaps reflective of their non-idiomatic reading, or less idiomatic reading in the case of partial forms.

The questions we address in the next section are how to account for (a) the drop in acceptability when the canonical form of an idiom is altered, and (b) why changing an idiom does not necessarily render it completely unacceptable.

3. Modeling with naive discriminative learning

3.1. Model description

We introduce the framework of naive discriminative learning by means of the small data set given in Table 1. The left column lists eight sentences, three of which express the concept of dying, albeit with different words (*pass away, die*, and *kick the bucket*). The words *kick* and *bucket* are also used in their literal sense, and frequencies are chosen such that the idiom is infrequent.

The raters in our experiment received orthographic input. We therefore represent orthographic input with letter pairs. Denoting the space character by the #, we encode a sentence such as *Mary passed away* with the input features (henceforth, cues) `#M Ma ar ry y# #p pa as ss se ed d# #a aw wa ay y#`. The "interpretations" listed in the second column of Table 1 gloss the reader's understanding of the sentence. Note that *kicked the bucket*, *pass away*, and *died* are all glossed with "died."

We refer to elements, such as "died," in these glosses as "lexomes." Lexomes are conceptualized as pointers to locations in a high-dimensional lexical co-occurrence space (see, e.g., Landauer & Dumais, 1997). Whereas the lexomes themselves are static pointers, the content to which they point, the semantic vectors, are subject to continuous updating as experience unfolds. Homographs such as *bank* are associated with different lexomes, whereas forms such as *died* are associated with multiple lexomes, one for the termination of life, and one for past tense (see Baayen, Shaoul et al., 2016, for further discussion). In our model, we represent the interpretation of a sentence by its set of lexomes, an obvious simplification, but sufficient for the present purposes. Lexomes are the output units (henceforth outcomes) of our model.

Table 1
Example data set to illustrate naive discrimination learning

Sentence	Interpretation	Frequency
Mary passed away	"Mary died"	40
Bill kicked the ball	"Bill kicked the ball"	100
John kicked the ball away	"John kicked the ball away"	120
Mary died	"Mary died"	300
Mary bought some flowers	"Mary bought some flowers"	20
Ann bought a ball	"Ann bought a ball"	45
John filled the bucket	"John filled the bucket"	100
John kicked the bucket	"John died"	10

All cues are linked directly to all outcomes in a two-layer network. The weights on the connections from cues to outcomes are determined by the learning equations of Rescorla and Wagner (1972), and for this example they were estimated with the help of the equilibrium equations of Danks (2003). The total support that the model provides for a given lexome is obtained by summation of the weights on the afferent connections from all the diphones in the input (e.g., the letter pairs in *John kicked the bucket*). We refer to this total support as a lexome's activation. We assume that higher activations give rise to higher rated acceptability.

When the sentence *Ann kicked the bucket*, which is not part of the training data listed in Table 1, is presented to the network, the activation of the lexome for "to kick" is 0.007, the activation for the lexome "bucket" is −0.092, and that of "to die" is 1.154. Thus, the model predicts that this novel sentence should be understood in its idiomatic reading. However, when another novel sentence, *Ann filled the bucket*, is presented, the activations of "filled," "bucket," "kicked," and "died" are 0.870, 0.870, −0.080, and 0.171, respectively; thus, the idiomatic meaning is dispreferred. It is noteworthy that the model achieves these correct predictions in the absence of any representations of word forms or combinations of word forms. By providing the learning algorithm with access to sublexical co-occurrence patterns, it no longer is necessary to keep track of morphemes, combinations of morphemes, and combinations of words (cf. Baayen, Hendrix, & Ramscar, 2013).

3.2. Modeling idiomatic versus literal

This kind of approach scales up to large real data sets at the lexical level (cf. Baayen et al. 2011; Baayen, Shaoul et al., 2016). But exploring discriminative learning on real idiom data is, unfortunately, not straightforward since standard corpora do not provide proper mark-up that would allow the automatic identification of idioms. Therefore, we constructed a mini-corpus with the aim of providing a preliminary exploration of the potential of discrimination learning for idiom processing. Our mini-corpus contains the 120 idioms included in our rating study and 500 pseudo-sentences, for a total of 620 "sentences." To create the pseudo-sentences, we made a list of all words that occur in the idioms, sorted them by increasing length, assigned them Zipfian frequencies using a log-normal-poisson model with mean 2 and standard deviation 1, and then placed the word tokens into sequences of 5 or 6. The order of the mini-corpus sentences was randomized. Each idiom was paired with a single idiomatic lexome, and each non-idiomatic sentence received as lexomes identifiers for its constituent words. The total number of different lexomes was 354. Letter trigrams were used as cues, resulting in 974 distinct types.

Next, we trained a discrimination network on this corpus, using sentence-by-sentence application of the learning rule of Rescorla and Wagner (1972). We set the learning rate to 0.001 for non-idiomatic sentences and to 0.008 for idioms. This higher learning rate for idioms is motivated by the idea that the idiomatic meaning is highly surprising and highly salient when first learned. The Rescorla–Wagner equations have a parameter that

can be changed depending on the salience of an outcome, which we used to create this higher learning rate for idioms.

We follow the model presented in Baayen, Shaoul et al. (2016) for auditory comprehension by utilizing simplified temporal dynamics of reading. Essentially, we use the network as a memory which is accessed from a buffer that can contain, in the present simulation, five consecutive letter trigrams. Reading a sentence in this simplified set-up—simplified because we abstract away from the eye movements that characterize real reading—then reduces to moving the sentence trigram by trigram through this window. For example, for the idiom *spill the beans*, the contents of the buffer window, for the first six timesteps, are:

```
#sp
#sp spi
#sp spi pil
#sp spi pil ill
#sp spi pil ill ll#
spi pil ill ll# l#t
pil ill ll# l#t #th
```

The trigrams in the buffer activate the lexomes by passing activation through the network. Typically, the appropriate lexomes are among those with the highest activations. Here, we zoom in on the lexomes in the sentences and display their activation as a function of time. Fig. 2 presents the resulting activation functions for idioms and idiom fragments. The horizontal black lines represent a threshold above which a lexome counts as recognized. The thick solid red line represents the activation of the idiom's lexome. The remaining lines present the activations of the lexomes for the individual words in the idiom. The activation of individual lexomes generally rises and then falls, depending on when and how many of their trigrams are present in the windowing buffer. Thus, the combination of a network memory and a temporally restricted window moving across the sentence results in the appropriate lexomes presenting themselves and then fading away, without having to segment the input into form units of different sizes and granularity.

What is remarkable is that in this approach the idiom's lexome is available from the start and has an activation that remains high over time (left panel). This behavior is a consequence of the high learning rate for idioms, which causes letter trigrams present in an idiom's orthographic form to acquire strong connection strengths to that idiom's lexome. Since at any point in time there are multiple letter trigrams providing strong support for the idiom's lexome, it remains strongly activated as long as relevant letter trigrams are present in the input.

The relative stability of idiom activation over time can be observed for a wide range of learning rates. The between-word trigrams in the idiom, such as `r#s` in *hear something through the grapevine*, also contribute to this activation. Scrambling of the word order makes the between-word boundary trigrams unavailable, and their absence both reduces the activation of the idiom lexome and renders its activation function more volatile (see Fig. S1). Nevertheless, the model still provides good support for the idiomatic

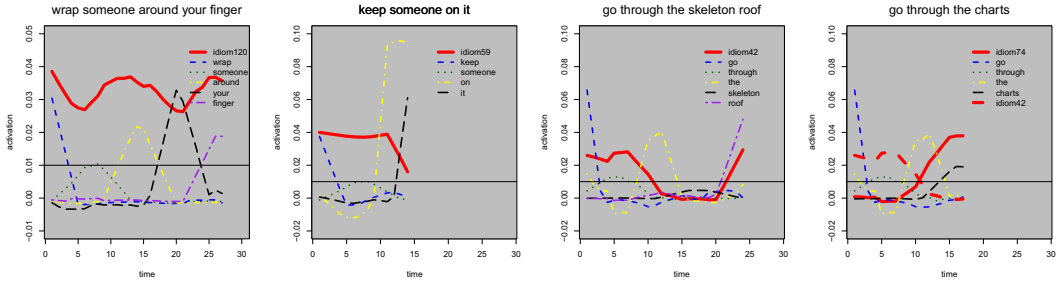

Fig. 2. Activation of idiom and word lexomes over time for idioms and idiom variants.

reading. It seems to us, albeit impressionistically, that this prediction may be correct, as the idiomatic reading remains strongly available to us even under scrambling (cf. *nose pay the through*).

In summary, the model predicts that the idiomatic reading is supported from the start, that it receives continuous support, and that the literal interpretation has to be constructed step by step as the relevant lexomes become available over time. This behavior of the model is achieved simply by including initial surprise at an idiom's meaning by means of an increased learning rate at the first (and in this simulation, the only) exposure to the idiom.

3.3. Modeling idiomatic variation

Having trained the model on idioms encountered only in their canonical form, we now inspect how it handles idiom variants.

The effect of shortening an idiom and maintaining only part of its original form is illustrated in the second panel of Fig. 2 (see also Fig. S2). Replacing *their toes* by *it* in the idiom *keep someone on their toes* results in an abrupt decline in the activation of the idiom's lexome, ending up lower than the activations of *on* and *it*. This may underlie the low acceptability ratings of partial forms, which appear not to provide support for an idiomatic interpretation.

Next, consider an idiom being extended with an additional concept (i.e., *go through the investment roof*). As *investment* is not a word in the simulation's vocabulary, we replaced it with *skeleton* (from *a skeleton in the closet*). The result is shown in the third panel of Fig. 2 (cf. also Fig. S3). When the intruding word comes in, the activation of the idiom lexome drops below the recognition threshold. When the final word of the idiom, *roof*, moves into the buffer, the activation of the idiom lexome re-emerges. This temporary cancelation of the idiom lexome by the intruding word may explain why this manipulation of the idiom results in intermediate acceptability. Lexome recovery makes this variation strategy more appreciated than other options, but not nearly as preferred as the canonical form. This suggests that full, uninterrupted support over time is crucial for maximal acceptability.

The rightmost panel of Fig. 2 illustrates the activation function for the idiom blend *go through the charts*, which fuses *go through the roof* with *off the charts*. The lexomes of the two idioms show a cross-over pattern in the expected temporal order (see also Fig. S4). The relatively low ratings of blends in our experiment suggest it does not help to have at least one idiomatic lexome active: The lexome of an idiom should receive continuous support, instead of coming and going support.

For the lexical substitution condition (*go through the ceiling*), the model behaves in a similar way as for partial forms and therefore predicts lower ratings than actually observed. The reason is clear: The model as implemented here has no knowledge of *roof* and *ceiling* being semantically related. The model could be extended with an algorithm evaluating the semantic similarity of *ceiling* and *roof* by means of the cosine of the semantic vectors indexed by these two lexomes. The greater this similarity, the more acceptable the idiom variant would be found. Such an extension, however, is beyond the scope of this study.

Words taken from different idioms, for example, the idioms *cut the mustard*, *beat around the bush*, *go through the roof*, and *burn a hole in your pocket*, brought together in a novel sentence *mustard around the pocket*, show a pattern of activations in which idiom lexomes have the same ephemeral activation as the lexomes supported by individual words (cf. Fig. S5). The model predicts that the idioms *cut the mustard* and *burn a hole in your pocket* are never far away, even when reading the words *mustard* or *pocket* in isolation, as would be expected (cf. Sprenger et al., 2006).

Recall that when the context biases a literal interpretation, acceptability ratings for the canonical form of the idiom plummet (see Fig. 1). The literal meaning has to be given preference, even though the idiomatic meaning is strongly supported by the trigram cues. This incongruency is, we think, the cause of the large decrease in acceptability. Modeling the resolution of this conflict would take us beyond what can be accomplished with the present bottom-up network, and it is best understood in terms of further top-down processes known to be involved in decision-making and conflict resolution (see Ramscar & Gitcho, 2007).

We do not know whether the results based on the present tiny data set will replicate when the model is trained on realistic amounts of data. Given that for morphological processing, we do know that wide learning networks of the kind used here do scale up properly (Baayen et al. 2011; Baayen, Milin et al., 2016), we are optimistic that some headway can be made, once corpora with proper mark-up for idioms become available.

4. Discussion

Traditionally, idioms have been described as having a rigid prescribed form (see, e.g., Levelt, 1989, for a summary of the relevent literature). Corpus studies have shown that idioms are much more variable than previously thought. We addressed, by means of a rating experiment, the acceptability of both the canonical forms of idioms (literal and idiomatic) and several idiom variant types that are attested in corpora. Acceptability

varies substantially with each variant type. And among the lowest ratings were those for the literal condition, the condition in which the canonical form is placed in a context that enforces a literal interpretation.

Theories that account for idiom processing by means of fixed form representations (Bobrow & Bell, 1973) or by means of fixed lexical configurations (Cacciari & Tabossi, 1988) face a problem reminiscent of the question of how to account for acoustically reduced variants in auditory comprehension, such as [hlɛrɛs] for "hilarious" (Johnson, 2004). Numerous problems arise when the lexicon is enriched with additional variant forms, which led Johnson (p. 23) to conclude that "models of auditory word recognition, that aim to account for anything beyond laboratory speech, must abandon traditional 'dictionary' assumptions about the auditory mental lexicon." We face a similar quandary here. One might consider adding additional form representations or additional lexical configurations to the mental dictionary, but this would predict idiom variants to be fully acceptable, which as we have shown is not the case. But without representations or constellations for idiom variants, one would need additional mechanisms generating precise predictions about which partial matches to idiom representations or constellations are close enough to support an idiomatic interpretation. Given the creativity of idiom variants, as exemplified by *my phone kicked the pail last week* (from Duffley, 2013; an idiom variant claimed to be generally not possible by Culicover et al., 2017, this volume), it is unclear that such a mechanism actually can be made to work.

The simulation study presented here illustrates a very different way of thinking about the problem. (Albeit, this simulation has currently only been run on the experimental stimuli.) As the corpora literature shows, words and syntactic structures in an idiom are alive and kicking. The only idiom-specific representation that we allow in our model is the idiom's lexome, a pointer to its semantic vector that it may share with other forms (e.g., "die" for *pass away* and *kick the bucket*). A minimalist end-to-end approach with wide learning provides a proof of concept that idioms' lexomes may receive strong support from sublexical orthographic units such as letter triplets under the assumption that their surprising and salient interpretation gives rise to deeper learning, modeled with a higher learning rate. As a consequence, the letter triplets of an idiom come to provide continuing support for its lexome, whereas for literal word sequences, the lexomes of individual words come and go. The more this continuous support for an idiom's lexome is interrupted in an idiom variant, the less acceptable that variant becomes. The dynamics of this model are close to what we think is targeted by the configuration hypothesis (Cacciari & Tabossi, 1988), but our model sidesteps the problem of having to engineer fixed lexical configurations into a mental dictionary that are somehow fuzzy enough to allow for idiom variants.

An important methodological issue is worth elaborating here. Theories that posit units for idioms, and use these units to explain how the idiosyncratic meaning of an idiom is retrieved, have to explain how such units themselves are accessed. This question is typically not addressed. We are so familiar with being able to look up words in a dictionary, or to search for patterns in files, that we take for granted that accessing units is trivial. However, models such as the interactive activation model (McClelland & Rumelhart,

1981) were developed precisely because human look-up has all kinds of properties that are foreign to look-up with the algorithms implemented on our computers. Our wide learning approach offers an alternative algorithm to that of the interactive activation model but shares its goal of approximating human look-up. The crucial difference is that we go straight for what needs to be looked up, or better, discriminated, namely, the message encoded in the signal, operationalized here as the idiom's lexome. As we have shown, this is possible without having form units serving as gatekeepers to meaning. Research on morphological processing (Baayen et al., 2011) and acoustic reductions (Baayen, 2010) suggests that with sufficiently fine-grained input representations, additional layers of mediating form representations against which the input would have to be matched become redundant. Crucially, effects of frequency of occurrence, taken as evidence in traditional approaches for form units at various levels, are correctly predicted by wide learning networks, even though these networks do not know about such units. In other words, precisely by addressing the question of how the message encoded in the speech signal is to be discriminated from other messages that might have been encoded, many of the effects that in standard approaches are explained by positing layers of "hidden" representations simply come for free.

Our way of thinking about idioms is also relevant to the vexed problem of how to understand frequency effects for sequences of words that do not have blatantly idiosyncratic meanings (see Arnon & Christiansen, 2017, this volume, for an overview). In our approach, positing form units for word n-grams puts the cart before the horse, and does not help explain frequency effects for word n-grams, as these very frequency effects arise in our theory as an epiphenomen of discrimination.

Baayen et al. (2011, 2013) sought to explain frequency effects for word n-grams in terms of the support provided by letter n-grams for the individual lexomes in these word n-grams, but their explanation is in all likelihood insufficient to cover the full range of evidence. As a consequence, our theory forces us to incorporate lexomes for non-idiomatic word n-grams (cf. Lensink, Schiller, Verhagen, & Baayen, 2016), which are required on independent grounds for, for example, formulaic expressions such as *good morning* and *tickets please*.

Now consider an n-gram such as *the president of the United States*. This n-gram reduces readers' uncertainty about presidents to the past and present presidents of a particular country. At the time of writing, the lexome most likely to receive strongest support from these words is the same lexome that is activated by the orthographic input *Barack Obama*. This lexome has an onomasiological motivation, and there is no principled difference here compared to the lexomes required for compounds such as *cream cheese recipe* and *hogwash*.

Incomplete word n-grams such as *the president of the* and *president of the United* are, from this perspective, highly similar to idiom variants, to shortened words such as *condo* and *ridic* ("ridiculous"), and to acoustic reductions such as [hlɛrɛs]. Adding units for all these variants is not necessary. All that we need to do in order to explain the frequency effect of a word n-gram is to integrate over the activations of all lexomes that this word n-gram is consistent with. By giving up the axiom that a frequency effect would provide

a decisively indicative litmus test for the existence of some representation, we can unburden the mental lexicon of hundreds of millions of form units.

Acknowledgments

This study was supported in part by an Izaak Walton Killam memorial scholarship awarded to the first author and an Alexander von Humboldt research chair awarded to the third author (grant no. 3015004901).

References

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. In M. Christiansen & I. Arnon (Eds.), *More than words: The role of multiword sequences in language learning and use*, *Topics in Cognitive Science*, *9*(3), 621–636.

Ayto, J. (Ed.) (2009). *From the horse's mouth: Oxford dictionary of English idioms*. Oxford, England: Oxford University Press.

Baayen, R. H. (2010). Assessing the processing consequences of segment reduction in Dutch with naive discriminative learning. *Lingue & Linguaggio*, *9*, 95–112.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, *56* (3), 329–347.

Baayen, R. H., Milin, P., Filipovic-Durdevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438–481.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition, and Neuroscience*, *31*(1), 106–128.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Cacciari, C., Corradini, P., & Padovani, R. (2005). Speed of processing effects on spoken idiom comprehension. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* July 21–23, Stresa, Italy (pp. 372–377). Mahwah, NJ: Lawrence Erlbaum. Available at: http://csjarchive.cogsci.rpi.edu/proceedings/2005/docs/p372.pdf.

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

Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. In M. Christiansen, & I. Arnon (Eds.), *More than words: The role of multiword sequences in language learning and use*, *Topics in Cognitive Science*, *9*(3), 552–568.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, *47*(2), 109–121.

Duffley, P. J. (2013). How creativity strains conventionality in the use of idiomatic expressions. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), *Language and the creative mind* (pp. 49–61). Stanford, CA: CSLI Publications.

Fanselow, G., & Frisch, S. (2006). Effects of processing difficulty on judgments of acceptability. In F. Keller, G. Fanselow, C. Fery, R. Vogel, & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives* (pp. 291–316). Oxford, England: Oxford University Press.

Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology*, *14*, 83–102.

Funke, F., & Reips, U.-D. (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, K. (2016). Climbing on the bandwagon of idiomatic variation: A multi-methodological approach. PhD thesis, University of Alberta.

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

Gibbs, R. W., Nayak, N. P., Bolton, J. L., & Keppel, M. E. (1989). Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition*, *17*(1), 58–68.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*, 68–101.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

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

Lensink, S. E., Schiller, N. O., Verhagen, A., & Baayen, R. H. (2016). Parts and wholes – on the cognitive reality of multi-word units. Paper presented at the 2nd Usage-Based Linguistics Conference, Tel Aviv, June 15.

Levelt, W. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103–1121.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of the basic findings. *Psychological Review*, *88*, 375–407.

Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford, England: Oxford University Press.

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491–538.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Science*, *11*(7), 274–279.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York: Appleton-Century-Crofts.

Schröder, D. (2013). *The syntactic flexibility of idioms: A corpus-based approach*. Munich: AVM.

Sinclair, J. (Ed.). (2011). *Collins COBUILD idioms dictionary*. Glasgow, UK: HarperCollins.

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

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, *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*(2), 313–327.

Titone, D. A., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbolic Activity*, *9*(4), 247–270.

Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. London/New York: Continuum.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Fixed effects for the acceptability rating responses.

Table S2. Random effects for the acceptability rating responses.

Fig. S1. Idiom and word activations for words in canonical order (left) and scrambled order (right).

Fig. S2. Idiom and word activations for the idiom *keep someone on their toes* and its shortened, partial form *keep someone on it*.

Fig. S3. Idiom and word activations for the idiom *go through the roof* and its integrated concept variant *go through the skeleton roof*.

Fig. S4. Idiom and word activations for the idiom *go through the roof* and its blend with *off the charts*.

Fig. S5. Idiom and word activations for words from the idioms *go through the roof* (idiom42), *cut the mustard* (idiom15), *beat around the bush* (idiom2), and *burn a hole in your pocket* (idiom8).

Fig. S6. Activation of idiom and word lexomes over time for idioms of different lengths.