



Familiarity breeds dissent: Reliability analyses for British-English idioms on measures of familiarity, meaning, literality, and decomposability



Emily Nordmann^{a,*}, Alexandra A. Cleland^a, Rebecca Bull^b

^a School of Psychology, William Guild Building, University of Aberdeen, Aberdeen AB24 2UB, UK

^b National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, 637616, Singapore

ARTICLE INFO

Article history:

Received 27 August 2013

Received in revised form 24 February 2014

Accepted 25 March 2014

Available online 15 April 2014

PsychInfo classification:

2720

2260

Keywords:

Idioms

Familiarity

Reliability

ABSTRACT

To date, there have been several attempts made to build a database of normative data for English idiomatic expressions (e.g., Libben & Titone, 2008; Titone & Connine, 1994), however, there has been some discussion in the literature as to the validity and reliability of the data obtained, particularly for decomposability ratings. Our work aimed to address these issues by looking at ratings from native and non-native speakers and to extend the deeper investigation and analysis of decomposability to other aspects of idiomatic expressions, namely familiarity, meaning and literality. Poor reliability was observed on all types of ratings, suggesting that rather than decomposability being a special case, individual variability plays a large role in how participants rate idiomatic phrases in general. Ratings from native and non-native speakers were positively correlated and an analysis of covariance found that once familiarity with an idiom was accounted for, most of the differences between native and non-native ratings were not significant. Overall, the results suggest that individual experience with idioms plays an important role in how they are perceived and this should be taken into account when selecting stimuli for experimental studies. Furthermore, the results are suggestive of the inability of speakers to inhibit the figurative meanings for idioms that they are highly familiar with.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Idioms are expressions that are more than the sum of their parts. The figurative meaning of an idiom can be difficult to ascertain from a compositional analysis of the literal meaning of the component words (e.g., *kick the bucket* literally means *to strike a container with one's foot* and figuratively means *to die*). Much of the research to date has focused upon how, and if, both of these meanings are activated during comprehension and production, and how the lexical representation of idioms can be incorporated into existing models of language. Item generation for psycholinguistic research can often be an arduous process given the large number of characteristics that even single words can be matched upon. Idiomatic expressions provide the additional concepts of decomposability and literality and consequently designing well-matched experimental stimuli can become very difficult. The purpose of the current work was not to add to the body of literature concerned with the lexical representation of idioms, but rather to investigate the methods used to obtain normative data that is then used to select experimental stimuli.

Whilst there are large databases of information about single words (e.g., CELEX (Baayen, Piepenbrock, & Van Rijn, 1993) and SUBTLEX (Brysbaert & New, 2009)), normative data for idioms exists in the

form of a handful of published papers (e.g., Bonin, Méot, & Bugaiska, 2013; Caillies, 2009; Libben & Titone, 2008; Tabossi, Arduino, & Fanari, 2011; Titone & Connine, 1994) that cover several hundred expressions in different languages. Whilst the paucity of data is unfortunate, what is of greater concern is the lack of reliability reported for these data. In particular, concerns have been raised over the relevance of the construct of semantic decomposability given that studies show few consistent effects of decomposability (Tabossi, Wolf, & Koterle, 2009). Whilst the bulk of the discussion concerning the reliability of ratings has focused upon the aspect of decomposability, it is also worth considering the reliability of ratings obtained for other features of idiomatic expressions to determine whether or not decomposability represents a special case. The initial focus of the current work was therefore to investigate whether the concepts of familiarity, meaning and literality are subject to the same level of inter-rater variability.

1.1. Familiarity

Familiarity is considered to be a measure of the frequency with which a speaker or listener encounters a word or idiomatic expression (Gernsbacher, 1984). Titone and Connine (1994) describe familiarity as the “subjective frequency” (p. 255) of an idiom which can be viewed as distinct from ‘objective’ frequencies obtained from word counts in, for example, newspapers. Aside from the current lack of an objective frequency corpus for idioms, the inherent variability of individual

* Corresponding author. Tel.: +44 1224 273646; fax: +44 1224 273426.

E-mail address: emily.nordmann@abdn.ac.uk (E. Nordmann).

experience with such expressions means that a subjective rating of familiarity is more likely to provide useful information. Familiarity also incorporates how well the meaning of a phrase is known or understood, and all norming studies to date have found strong positive correlations between familiarity and meaning judgements (e.g., Libben & Titone, 2008; Tabossi et al., 2011; Titone & Connine, 1994).

1.2. Decomposability

Decomposability is generally defined as whether the literal meanings of the component words of an idiomatic expression have a connection to the overall figurative meaning. A normally decomposable idiom is defined as one where the component words of a phrase are connected to the overall figurative meaning in a literal way, e.g., *play the market*. Abnormally-decomposable idioms are those items whereby the relationship between the component words and the figurative meaning is more metaphorical, e.g., *save your skin*. Non-decomposable idioms are those phrases that are seen as most stereotypically idiomatic, whereby the literal meaning of the component words bear no relation to the overall figurative meaning, e.g., *kick the bucket*. From a theoretical perspective, being able to manipulate the decomposability of stimuli for behavioural studies is important as it allows conclusions to be drawn about the lexical representation of idiomatic expressions. If decomposability could be demonstrated to affect production latencies then this would be strong evidence for a special type of idiom representation where such information could be encoded. Conversely, if decomposability appears to have no effect on language processing then this would support research that claims that idioms utilise the same levels of lexical representation and processing mechanisms that literal sentences do (e.g., Konopka & Bock, 2009). So far the experimental evidence of an effect of decomposability has been mixed (e.g., Cutting & Bock, 1997; Konopka & Bock, 2009; cf. Nordmann, Cleland, & Bull, 2013; Sprenger, Levelt, & Kempen, 2006) and this lack of consensus may be a result of the data used to group stimuli as decomposable or non-decomposable.

1.3. Literality

The defining feature of an idiomatic expression is the figurative meaning the phrase carries, regardless of its relationship to the literal meaning of the component words. However, independent of decomposability it is possible to judge idioms on whether or not they have the potential for a literal interpretation. For example, *kick the bucket* has a well-formed literal meaning, that is, *to strike a container with one's foot*. This literal interpretation is available despite the fact that as an idiomatic expression with the figurative meaning *to die*, *kick the bucket* is highly semantically non-decomposable. In contrast, it is difficult to imagine how one would literally *talk a mile a minute* even though the idiomatic meaning of *to speak quickly* is decomposable.

In terms of norming studies, Libben and Titone (2008) and Tabossi et al. (2011) both found that participants' literality ratings correlated negatively with familiarity ratings (higher literality rating correlated with lower familiarity), although Titone and Connine (1994) found no such relationship. The link between literality and decomposability is also unclear. Titone and Connine found literality to be negatively correlated with abnormal decomposability, that is, those expressions whose component words were only metaphorically related to their figurative meaning were less likely to have a well-formed literal meaning. In contrast, Libben and Titone found that literality was negatively correlated with all measures of decomposability (global decomposability, normal, and abnormal decomposability) whilst Tabossi et al. (2011) found no relationship between literality and any of the decomposability measures.

1.4. Reliability analyses

A number of studies (Gibbs & Nayak, 1989; Libben & Titone, 2008; Tabossi, Fanari, & Wolf, 2008; Titone & Connine, 1994) have asked participants to make categorical judgements regarding decomposability and, if it was indicated that an item was thought to be decomposable, they were then asked to make a further categorical decision between normal and abnormal decomposability. Those studies that have used categorical methods of idiom classification have generally found low levels of agreement. Titone and Connine (1994) regarded an idiom as having been reliably categorised if agreement between participants reached 67% according to an approximation of the binomial distribution. With this criterion they found that only 40% of the idioms used were reliably classified as being non-decomposable, normally decomposable, or abnormally decomposable. These results stand in stark contrast to Gibbs and Nayak (1989) who showed agreement levels above 75% for 37 out of 40 idiomatic expressions. However, as noted in Tabossi et al. (2008), the idioms selected for use in Gibbs and Nayak were not chosen at random but selected by the authors as likely to be non-decomposable, normally decomposable, or abnormally decomposable. Therefore, it is unclear whether the high agreement levels observed by Gibbs and Nayak reflect participants' true ability to reliably categorise idioms, or if the pre-selection of those items that were most likely to strongly fit into one of the categories influenced the results.

A primary concern of our work regards the use of categorical ratings for decomposability. Gibbs and Nayak (1989) suggest that decomposability exists on a continuum rather than there being a precise definition that can be used to sort idioms into dichotomous categories. The question raised is whether the low agreement ratings observed are a product of the categorical rating system, when a continuous scale, for example a Likert scale, would be more appropriate to capture the nuances of decomposability.

Tabossi et al. (2008) and Tabossi et al. (2011) utilised a 7-point Likert scale in order to obtain decomposability ratings. Those idioms that were rated as having a mean score of four or less were categorised as non-decomposable. The results from both studies indicated that the use of a Likert scale to rate decomposability does not appear to increase agreement between participants. Tabossi et al. (2008) reported that for a subset of 16 idioms agreement was high (83.54%). However, for the remaining 64 expressions agreement was on average 56.14%, a proportion not reliably different from chance. Similarly, Tabossi et al. (2011) found that only 28% of their idiomatic expressions could be reliably categorised using a 67% agreement level. As a further measure of agreement, Tabossi et al. (2008) also provided the results of an inter-rater reliability analysis using Kendall's coefficient of concordance. The authors raised concerns that the dichotomisation of idioms into decomposable and non-decomposable may have led to a lower estimate of reliability, however, the results were in line with the previous analyses and found poor reliability of .13.

One of the arguments for the use of Likert scales to assess decomposability is that such scales may be able to capture the nuances of idiom decomposability better than asking participants to make a dichotomous categorisation. However, in the literature to date, Likert scores are still then categorised according to a cut-off point (e.g., in Tabossi et al. (2008) a mean rating of four or less is considered non-decomposable). By subsequently categorising the Likert scores the subtlety of the scale is lost, particularly for those items that are not considered at the extreme ends of the scale. For example, using a cut-off of a mean rating of four or less to determine decomposability category, if an idiom was rated by 50% of participants as decomposable (more than 4) and by 50% as non-decomposable (less than 4), this item would not be considered to have been reliably categorised. However, this analysis does not allow for the possibility that 100% of scores may lie, for example, between 3.5 and 4.5 on the Likert scale. By using percent agreement, those idioms that are not strongly decomposable or non-decomposable become classified as unreliable. Hayes and Krippendorff (2007) suggest

Download English Version:

<https://daneshyari.com/en/article/7277682>

Download Persian Version:

<https://daneshyari.com/article/7277682>

[Daneshyari.com](https://daneshyari.com)