UNIVERSITY OF BIRMINGHAM

University of Birmingham
Research at Birmingham

Of false friends and familiar foes

Carrol, Gareth; Littlemore, Jeannette; Dowens, Margaret

Link to publication on Research at Birmingham portal

Of False Friends and Familiar Foes: Comparing native and non-native understanding of figurative

phrases

Gareth Carrol and Jeannette Littlemore, University of Birmingham

Margaret Gillon Dowens, University of Nottingham Ningbo

Corresponding author:

Dr Gareth Carrol

Department of English Language and Applied Linguistics

University of Birmingham

Edgbaston

Birmingham

B15 2TT

United Kingdom

Email: g.carrol@bham.ac.uk

Abstract

Research into figurative language identifies variables such as familiarity, transparency, decomposability and motivation, all of which play an important role in how native and non-native speakers learn, process and understand figurative phrases. However, these variables are not always defined and operationalised in the same way, and are often treated as independent. We discuss these factors as they relate to the judgements that language users make, and as they relate to the ability to correctly infer meaning in a range of familiar and unfamiliar idioms, and novel metaphors. In a rating study, we show that familiarity has a clear effect on perceptions of transparency. For less familiar idioms, judgments of decomposability after the meaning became known were strongly affected by whether or not speakers were correct in guessing the meaning. We also saw clear cross-language effects, whereby idioms that exist in the L1 for non-native speakers were seen as more familiar, more transparent and were better identified. We discuss how these factors contribute at different stages to allow speakers to make sense of both familiar and unfamiliar figurative phrases.

Introduction

Figurative language – referring broadly to any language where a speaker means something other than what is literally expressed (Gibbs & Colston, 2012) – is a pervasive element of everyday communication. Native speakers use idioms, metaphors and other tropes so frequently that they may seem fairly unremarkable, yet the challenge they present to language learners is substantial. In this paper we aim to explore some of the factors that contribute to how native (first language, L1) and non-native (second language, L2) speakers understand figurative meaning. How do native L1 and L2 speakers differ in their ability to interpret and infer non-literal meaning, and how does this relate to the broader complex of skills that make up "figurative competence" (Pollio & Pollio, 1974)?

This paper has two primary aims. The first is to provide a broad overview of the literature as it relates to the topic of figurative competence, and we purposefully incorporate a wide range of studies and topics here. For example, we briefly address the development of figurative language and semantic abilities in first language acquisition and relate this to the development of figurative competence in a second language, since there are important parallels in the kinds of skills that speakers require. We go on to discuss a range of studies that are relevant to the question of figurative competence, highlighting not only the key findings, but also attempting to reconcile the differences in terminology and conception that sometimes obfuscate the overlap between otherwise highly complementary research. As such, a subsidiary aim is to consider the different dimensions that are often described in the literature and attempt to develop a more unified way of talking about them. Our second aim is to specifically explore the question of how native (English L1) and non-native (English L2) speakers perceive and interpret figurative phrases. We concentrate here on how differences in relative familiarity and language proficiency impact other judgements, such as intuitions about the semantic properties of different phrases. We also consider cross-language overlap here, to establish how far L1 knowledge contributes to figurative understanding.

Figurative competence in L1

The ability to correctly interpret figurative meaning involves development of the same broad cognitive and linguistic abilities that underpin more general language comprehension (Levorato & Cacciari, 1992, 1995, 1999). Typically, figurative competence begins to look adult-like by around 10-11 years of age (Levorato & Cacciari, 1999; Vulchanova, Vulchanov & Stankova, 2011). This emerges from the development of the ability to consider language on a broad level, utilising developing skills of inferencing from context and semantic analysis to enable children to process a broader range of meanings (Cain, Towse & Knight, 2009). Even from a young age, children can successfully use context to help them guess the meaning of idioms (e.g. Cain & Towse, 2008; Cain et al., 2009; Gibbs, 1991). The ability to apply semantic analysis also seems to be available from an early age (e.g. Gibbs, 1987, 1991), and becomes more important as children get older (e.g. Nippold & Taylor, 1995; Levorato & Cacciari, 1999).

Developing abilities in semantic analysis and inferencing are essential if children are to move toward adult-like understanding of figurative language, and seem to be a natural part of more general linguistic and cognitive maturation. In the case of idioms, the additional factor of familiarity plays a very important role. Idioms – generally defined as non-compositional figurative phrases – are often lexically fixed, opaque and imbued with aspects of culture that make them hard to interpret on first encounter. How children begin to acquire the plethora of idioms that exist in any language remains very much an open question. The "language experience" view proposes that learning idioms and other expressions is a direct result of the amount of meaningful exposure a speaker receives (Nippold & Rudzinski, 1993). However, Reuterskiöld and Van Lancker (2012) showed that children as young as eight were significantly better at recalling idioms than novel (non-figurative) phrases after even one exposure, suggesting that there may be something inherently salient or noticeable that helps these to be learned. The idiom literature provides ample evidence that familiarity is an important aspect of how idioms are processed and understood (e.g. McGlone, Glucksberg, &

Cacciari, 1994; Schweigert, 1986, 1991; Schweigert & Moates, 1988; Swinney & Cutler, 1979), although this seems to be a property of all familiar "formulaic" language, rather than being specific to idioms (e.g. Tabossi, Fanari & Wolf, 2009). The ability to successfully interpret idiomatic language, and figurative language more generally, is therefore grounded in two key areas: a sophisticated set of semantic and inferencing skills that help to derive meaning based on analogy, analysis and context; and a broad and detailed knowledge of the conventional figurative phrases that are in common use in the language.

A key question for this paper is how these two competencies interact to aid figurative understanding. Idioms are often classified according to their semantic properties, and whether they are decomposable (can be interpreted via the component words) or non-decomposable (knowledge of the whole phrase is the only way to infer the meaning) has important implications for how they are processed and represented (e.g. Abel, 2003; Caillies & Butcher, 2007; Gibbs, 1980; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989; Nunberg, Sag & Wasow, 1994; Titone & Connine, 1994). In line with broadly usage-based models of language (e.g. Tomasello, 2003), frequency can also have an influence on representation, regardless of the semantic properties for any given phrase (see also Goldberg, 2003, for a similar view in Construction Grammar). Libben and Titone (2008) collected normative data for 219 English idioms, including characteristics such as familiarity, decomposability and meaningfulness (how confident respondents were that they actually knew the figurative meaning for any given phrase). Their results showed a significant interaction of familiarity and decomposability, such that decomposability only made a contribution to how meaningful phrases were for the lowest familiarity items. When idioms were highly familiar, the degree of decomposability made no difference to how they were understood. Libben and Titone went on to propose two routes to idiom comprehension – direct retrieval of meaning for highly familiar phrases, and semantic analysis/decomposition for low familiarity phrases.

Keysar and Bly (1995, 1999) also explored the relationship between semantic properties and familiarity for idioms. In their studies, participants were taught the meaning of archaic and unknown idioms (selected and pre-tested to ensure that they were generally unfamiliar to participants, e.g. *the goose hangs high*). Participants were taught either the correct meaning or its conceptual opposite and subsequently, they regarded the learned meaning as more transparent, regardless of whether this actually was the true meaning or not. This effect became stronger as the idiom was encountered more times. Keysar and Bly proposed an important distinction between true conceptual transparency, and transparency that results from conventional usage. In other words, once a speaker has been introduced to the meaning of an idiom, other meanings seem "less sensible" (1999, p. 1576), regardless of how plausible they might be. They further argued that, far from idiom comprehension involving the activation of underlying cognitive or conceptual structures (e.g. Nayak & Gibbs, 1990; Cacciari & Glucksberg, 1991), it is subjective knowledge and familiarity that constrain how speakers see the motivation in any given phrase. Motivation refers to whether or not there is some discernible link between the literal reading of a phrase and its figurative meaning (Boers & Webb, 2015), and this can arise from a variety of sources. Importantly, these are not fixed perceptions, hence acquiring greater familiarity with a given phrase may lead to a corresponding change in the perceptions of its semantic, conceptual or metaphorical motivation. Other studies have also found that perceptions of novel or less familiar idioms change as exposure increases (e.g. Schweigert et al., 2003; Schweigert, 2009).

The question of terminology

The studies discussed above highlight that semantic features like decomposability are not fixed properties of idioms, but are actually highly subjective and are likely to change over time. Moreover, there is considerable variation in the way that such properties are labelled, defined and operationalised. Terms such as analysability, transparency and decomposability are used fairly interchangeably, with some researchers considering these as broadly the same thing (e.g. Cain et al.,

2009). Others make a clear distinction, e.g. Cieślicka (2015) argues that transparency relates to the underlying motivation, while decomposability refers to whether the individual words make identifiable contributions to the overall meaning. For example, Cieślicka argues that *jump the gun* is transparent because there is a clear metaphorical motivation behind its meaning (starting a race before the starting pistol has been fired), but since its meaning cannot be distributed over the component words it is non-decomposable. In contrast, *pop the question* is opaque (its figurative meaning is not obvious if you do not know specifically that *the question* refers to a proposal of marriage), but the figurative meaning (once known) can be mapped onto the component words (*pop* = "say/ask" and *the question* = "marriage proposal").

The question of motivation – or *why* a phrase means what it does – is an important one. Despite their reputation as "frozen metaphors", the majority of idioms are motivated in some way. Grant and Bauer (2004), in their taxonomy of English idioms, identify very few phrases that they consider to be "core idioms": phrases where there is no way to make sense of the meaning except by guessing from context. They suggest that the majority of idioms are in some way figurative, that is, there is something that enables a hearer to reinterpret the phrase pragmatically to arrive at the intended meaning. What contributes to an idiom's motivation, however, can be very varied. Some idioms seem to reflect underlying conceptual metaphors (e.g. *blow your top* is based on the metaphor ANGER IS HEAT – Gibbs, 1993). Some rely on metonymic reference, so *pop the question* evokes to the most salient part of a marriage proposal to refer to the whole proposal. Others rely more on the link between the phrase and a specific experiential domain – as in the example of *jump the gun* above (which is only transparent if we know that races are typically started with guns). As Boers and Webb (2015) point out, where a literal reading of a phrase is congruent with the figurative scene that is evoked, and the context in which it is used, an idiom will seem more transparent. In contrast, when the imagery that is evoked by a literal reading is not immediately obvious, a phrase may seem less transparent. Ultimately, however, the degree of motivation cannot be seen until the actual meaning is known. For example, *kick the bucket* may easily evoke an image of someone

causing a mess by accidentally overturning the contents all over the floor. It may therefore seem transparent on first encounter, but this judgement would turn out to be wrong once the actual meaning ("to die") was discovered.

The contribution of individual words is also an important cue in how speakers interpret idioms. Boers and Webb (2015) provide several good examples of phrases where the misinterpretation of a key word might change the understanding of a phrase completely, e.g. *a shot in the arm* or *to follow suit*. Clearly, understanding the specific referent of component words is key to both appropriately interpreting the intended meaning, and to identifying the underlying motivation. Some studies have considered this in more detail, to establish whether the literal meanings of words are activated as an automatic part of idiom processing. Smolka, Rabanus and Rösler (2007) found that the verbs used in idioms led to activation of both literal and figurative meanings, and suggested that this is evidence that all idioms are decomposed during processing, in line with hybrid theories of idiom processing (e.g. Sprenger, Levelt and Kempen, 2006; Cutting & Bock, 1997). In contrast, Libben and Titone (2008) found that global decomposability but not the decomposability of the individual component words contributed to judgements of how meaningful the whole idiom was. Hamblin and Gibbs (1999) claim that the individual words make important contributions, even for non-decomposable idioms. For example, "die quickly" is seen as a better definition of *kick the bucket* than "die slowly", since *kick* implies a dynamic, fast motion.

Taken together, the above discussion highlights that judgements of how semantically analysable idioms are can be made based on multiple sources of information, and are therefore likely to be highly variable and idiosyncratic. Despite this, idioms are often classified in a binary way: transparent vs. opaque, or decomposable vs. non-decomposable (although Gibbs, 1980, included an additional category of abnormally decomposable idioms). Given that idioms can seem more or less motivated for any number of reasons, it may be more logical to view this as a scale rather than an either/or choice, and to differentiate between the different facets that constitute "analysability". This

question, as it relates to native and non-native speakers, will be discussed further in the next section.

Native vs. non-native figurative competence

Previous research in this area highlights that the different variables often addressed in idiom studies are not as clear-cut or independent as is sometimes assumed. It also suggests a number of important differences in how native and non-native speakers will perceive figurative language. Where familiarity ratings are gathered for non-natives, these tend to be lower and more variable than for native speakers (e.g. Abel, 2003; Carrol & Conklin, 2014, 2017; Nordmann, Cleland & Bull, 2014). In general, then, even for advanced language learners, idiomatic knowledge is much lower in language learners than native speakers. One consequence of this greatly reduced familiarity is that non-native speakers may need to rely much more heavily on processes such as semantic analysis and inferencing from context when they encounter figurative language, rather than simply being able to "retrieve" the meaning for a known idiom.

One highly relevant model here is Cieślicka's Literal Salience Hypothesis (2006), which proposes just such a distinction in what native and non-natives speakers are likely to do. Native speakers, at least on some level, seem to understand idioms as whole units (e.g. Libben & Titone, 2008; Rommers, Bastiaansen & Dijkstra, 2013; Sprenger et al., 2006;), and retrieve the more salient figurative meaning by default (Giora, 1997). Cieślicka proposed that L2 users are likely to encounter component words used individually and literally earlier and more frequently in the course of learning a language (compared to the same words used in idiomatic contexts), hence a literal, compositional reading will be the most salient. Over time, and as idioms are encountered more often, this may shift, although the degree of exposure required will mean that a complete shift to a more native-like system is unlikely for most language users. Evidence for this literal bias for L2 idioms comes from a number of studies (e.g. Cieślicka, 2006; Cieślicka, Heredia & Olivares, 2014; Siyanova-Chanturia, Conklin & Schmitt, 2011; Carrol & Conklin, 2017), although others have found

that this effect is not as clear cut, especially amongst groups with higher proficiency and better knowledge of English idioms (e.g. Carrol, Conklin & Gyllstad, 2016; Beck & Weber, 2016). Despite some contrasting results, we might therefore expect to see a greater degree of semantic analysis amongst non-native speakers, since this may be the only route available to allow them to work out the meaning. This in turn may have an effect on how analysable idioms are perceived to be, and Abel (2003) showed that non-native speakers in general tended to rate idioms as more decomposable than native speakers, suggesting that this was indeed the case. Nordmann, Cleland and Bull (2014) compared groups of native and non-native speakers on a range of idiom ratings (familiarity, meaningfulness, literalness, decomposability). They found significant differences between the two groups on all measures, but once familiarity was accounted for in the analysis, differences between native and non-native speakers for literalness and decomposability disappeared. They concluded (like Keysar and Bly), that increasing familiarity affects judgements such decomposability, making it "seem like there is meaning where there is none" (p. 92). They also stress the low reliability in the ratings they obtained, suggesting that since familiarity does play such an important role, its inherent variability for L2 speakers means that judgements are highly idiosyncratic. Similar results (correlations between familiarity, meaningfulness and decomposability) were found in a meta-analysis by Nordmann and Jambazova (2016), and the authors suggest that this may explain why decomposability has had such a variable role in idiom studies.

Once we account for familiarity then, would we expect to see differences between native and non-native speakers in the ability to infer and explain meaning? Logically, native speakers could be expected to still have an advantage when faced with the task of working out the underlying motivation for any given phrase. For one, they are likely to be much more aware of some of the cultural and historical associations that a phrase may evoke, which may make it seem more transparent than to speakers from other cultures. Whether this is actually the case, however, is entirely open to question. Wray (2009) suggests that since idiomatic phrases are often so conventionalised as to have lost their literal roots, what makes them analysable is actually only

relevant in the context of "post hoc linguistic game-playing" (p. 193), rather than tapping into realistic online processes. (See Wray, 2002, 2008 for more on the "needs only analysis" model of formulaic and figurative language). The conceptual metaphors that underpin different idioms may also be a source of variation between languages and cultures (Boers & Webb, 2015; Kovecses, 2005). Charteris-Black (2002) showed that the most challenging idioms for learners were those where there was linguistic but not conceptual equivalence, i.e. the same phrase has different figurative meanings in L1 and L2. Laufer (2000) found that partial overlap (same meaning but different form, e.g. English *lip service* vs Hebrew *lip tax*) and conceptual non-equivalence (no equivalent idiom in the L1) both led to avoidance of English idioms even by advanced learners.

As well as greater knowledge of the experiential and conceptual domains, native speakers might also have an advantage since their vocabulary is (in most cases) likely to be larger than that of language learners. The first pillar of the Global Elaboration Model of figurative competence outlined by Levorato and Cacciari (1995) is knowledge of the dominant and peripheral meanings of words. Since in general native speakers know more words, we might expect them to have an advantage over non-natives in working out the meanings of figurative phrases, even when these are unfamiliar. Hypothetically, this effect might manifest in an example like *draw a blank*: a non-native speaker who has not heard the phrase before might assume that "draw" has its core meaning of "produce a picture". However, a native speaker might be better able to consider alternative meanings of the word – in this case that "draw" is used to mean "pull out" – and therefore see more of a connection to the underlying meaning. (The phrase itself is thought to relate to the process of drawing lottery tickets, where some have prizes written on them and others are blank – www.phrases.org.uk.) It is again highly questionable whether this is what native speakers actually do, and it may be that such an advantage is only relevant in terms of working out the motivation *after* the meaning is known. Liontas (2015) found that lack of key vocabulary knowledge was a major barrier to language learners interpreting unfamiliar idioms in the L2; this is especially true since idioms often include low

frequency vocabulary (e.g. *bury the hatchet*), or use less common meanings of words (e.g. the *draw a blank* example discussed above).

Use of context is an area where we might expect to see less of a difference between native and non-native speakers. Liontas (2002) outlined his Idiom Diffusion Model of Second Languages and stated that the first "prediction" stage was for learners to make hypotheses as to possible meanings based on the lexical items within an idiom, and on knowledge of the possible source domains. The second "confirmation and reconstruction" stage involves the learner comparing these hypotheses with the context of use, and adjusting any inferences accordingly (i.e. drawing on semantic, pragmatic and cultural knowledge to fine-tune the hypothesis). In contrast to these predictions, Boers, Eyckmans and Stengers (2007) reported that learners were, in general, fairly poor at guessing the meaning of idioms even in rich contexts. This suggests that the linguistic and extra-linguistic knowledge required to do this still puts native speakers at a significant advantage.

Wray, Bell and Jones (2016) set out to explore the question of whether, in the absence of familiarity, native and non-native speakers show different patterns of performance or different strategies when interpreting figurative phrases. To eliminate the effect of familiarity they used old-fashioned, semantically opaque idioms that are no longer in common use in modern British English (e.g. *kick over the traces*), chosen from historical novels to ensure that they were used in realistic contexts. They presented participants with a number of idioms and asked them to guess the meaning, and to explain their reasoning as they did so. They found that in general, native speakers made greater use of context to support their guesses and also made use of analogy to try and link the phrases to known idioms that might have similar meanings. Non-natives were more likely than natives to identify specific words in the idioms that they did not know, which could indicate a greater level of analysis of the component words. It might also be indicative of a more performance-based effect, whereby English native speakers are in general less likely to admit lack of knowledge in their L1, and language learners are more used to encountering unknown words and having to either infer or ask

their meaning. As proficiency increased amongst the non-natives, the participants did begin to more closely resemble the native speakers in the strategies they used, suggesting that there is no fundamental difference other than the availability of resources (i.e. depth of vocabulary, knowledge of comparable idioms).

More broadly, however, is there reason to think that competence in other aspects of figurative language should differ according to native speaker status? In the case of metaphors, which are not "known" in the same way as idioms, should non-native competence still lag behind either competence amongst native speakers, or competence by the same speakers in their L1? Littlemore (2010) investigated the second of these questions and found significant correlations between L1 and L2 performance for a group of French-English learners in their comprehension and production of novel metaphors. Participants were asked to take part in a series of tasks investigating (i) the production of novel metaphors, (ii) metaphoric fluency (i.e. the number of alternative possible meanings that they were able to come up with for a series of novel metaphors), (iii) the ability to find meaning in metaphors, and (iv) speed of interpretation. Alongside a general pattern whereby each type of metaphorical competence in the L1 was significantly correlated with the equivalent type of metaphoric competence in the L2, there were some interesting differences. Metaphoric fluency was in general better in participants' first language, but ability to find meaning in metaphors was better in the L2. In other words, students were more likely to look for (and find) meaning in L2 metaphors. Alongside this, individual variation in the different tasks was high, and Littlemore concluded that successful second language learning must necessarily involve cognitive flexibility and a willingness to look for, play with and explore new patterns in the L2.

Cross-language influence in processing figurative language

This difference in resources has been echoed in some of the literature exploring the effects of cross-language influence on how non-native speakers understand and process figurative language, and formulaic phrases more broadly. Speakers show a clear advantage when phrases they encounter in

their L2 also exist in the same form in their L1, for example when processing collocations (e.g. Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). For idioms, this translates to an advantage in production (Irujo, 1986; Laufer, 2000), and in terms of recognition, recall and comprehension of familiar phrases (Carrol & Conklin, 2014, 2017; Carrol, Conklin & Gyllstad, 2016; Pritchett et al., 2016; Titone et al., 2015). A study by van Ginkel et al. (2016) presented data from a rating task where language learners (German L1, Dutch L2) judged idioms in their L2. Items were either L2-specific (no equivalent in German) or had a high degree of cross-language overlap (same or very similar expression exists in German and Dutch). For L2-only items, perceived transparency and L2 proficiency were significant predictors of whether participants could correctly identify idiom meanings. For phrases that did have a German (L1) equivalent, neither familiarity (in L2) nor transparency were predictors of identifying the meaning. In comparison, familiarity was a significant predictor for a set of L1 Dutch participants, but transparency played no role. The study concluded that knowledge of idioms from the L1 could be used as a reference, making these expressions easier and negating the effects of transparency. In other words, if an expression is known in the L1, all that is required is for a speaker to refer back to this knowledge to retrieve the meaning. When encountering idioms and other figurative phrases in an L2, speakers therefore use multiple cues and strategies, of which L1 knowledge is one possible contributor (e.g. Liontas, 2002).

Summary

It is clear that familiarity is rightly seen as a key driver of idiom recognition and comprehension, both for native and non-native speakers. There is some indication that semantic factors such as transparency and decomposability only become important when phrases are less well known. In other words, when a phrase is known, a speaker can simply retrieve a meaning from the lexicon, but when this meaning is not available, broader knowledge (semantic, cultural, contextual) must be utilised to infer the meaning. Since idiomatic competence generally lags well behind more general L2 vocabulary knowledge (Steinel, Hustijn & Steinel, 2007), this suggests that non-native speakers will

rely on factors like transparency more than native speakers, although when L1 knowledge can be used as a reference point, this may again reduce the need to analyse the semantic properties of the phrase. Problems may arise when there is non-equivalence, or in vocabulary terms "false friends", where the same phrase has a different meaning in L1 and L2. For phrases that are not "known" in the same way (i.e. metaphors, which are more creative and do not have a conventionalised lexical form), learners seem to show comparable performance in L1 and L2, suggesting that this might be a fairly stable individual skill (Littlemore, 2010). Proficiency and exposure are presumed to have a very important role here, since knowledge of a wider range of words and word meanings will be of great benefit when it comes to identifying potential meaning in the L2. An example of this comes from Littlemore (2004), where students were asked to interpret the phrase "skirt around the hardest topics". Guesses of the meaning reflected knowledge of the core meaning of "skirt" (i.e. one student incorrectly guessed that it might mean "to hide", and mimed a skirt covering something to demonstrate this). However, native speakers might not even consider an example like this to be metaphorical or figurative at all, and may simply see "skirt" as an alternative meaning of a known word. Ability to correctly interpret any kind of figurative language may therefore be contingent on a complex set of lexical, phrasal and conceptual knowledge, as well as the individual ability and flexibility that speakers might naturally possess.

The present study aims to address some of the issues outlined above. We aim to explore the relationship between factors such as familiarity, transparency and decomposability, to see how these contribute to a speaker's ability to infer meaning. Whilst there is some indication in the literature that these variables are not independent, this is a point that is often not directly addressed, and many studies that collect rating data from native speaker participants fail to account for it. There is also substantial variation in how researchers define factors like transparency and decomposability, and in whether these should be considered as separate variables. We aim to investigate some of these factors to help better understand their importance in figurative language.

We also aim to explore differences amongst speaker groups, focusing on effects of native/non-native speaker status and L1 influence. We aim to see how speakers vary in their judgements of the factors outlined above, and to compare how these judgments influence the ability to correctly interpret figurative meaning. We also intend to explicitly test how L1 knowledge affects speaker judgements.

Research questions

1. What is the relationship between judgements of familiarity, transparency and decomposability, and how do these influence the ability to correctly interpret the meaning of figurative phrases?

2. How do native (L1 English) and non-native speakers (L2 English) differ in their judgements?

3. What effect does L1 knowledge have on both judgements of phrases and ability to interpret figurative meanings?

Study

Materials

We selected items from three categories in order to compare different types of figurative phrase. English idioms were selected on the grounds that these should be (generally speaking) highly familiar to native speakers, but much more variable in how well-known they are amongst second language learners. The second set of items were novel metaphors, all selected to be unfamiliar (in the sense that they do not have fixed, conventionalised lexical forms), but which should all be more or less transparent based on a metaphorical reading of the phrase. The third set were idioms translated into English from other languages that should be equally unfamiliar to both the native and non-native speakers in our study, hence the contribution of familiarity should be equally limited. Within this third set of idioms, a sub-set of items from another language, Chinese, was chosen, so that we

could also test a set of participants from this language background and assess the contribution of L1 knowledge.

English idioms were generally of the form "x-det-Y", where x was either a verb (e.g. *pop the question, pull your weight, smell a rat*) or preposition (e.g. *over the moon*, *under the weather*). In some cases a preposition was used rather than a determiner (e.g. *walking on air, let off steam*). All items were selected from previous idiom studies (Carrol & Conklin, 2014, 2017), and had been normed to ensure that they are all highly familiar to English native speakers.

Novel metaphors were all of the form "A is B", and were taken from the list of metaphors devised by Katz, Paivio, Marschark and Clark (1988). The authors generated a total of 260 metaphors for which they obtained extensive rating data on dimensions such as comprehensibility, ease of interpretation and degree of metaphoricity. This data was recently reviewed in a 25 year replication by Campbell and Raney (2016) and found to be robust (ratings and the relationship between dimensions were found to still be highly reliable). We selected an initial set of 60 items from the Katz et al. (1988) list, all chosen to be toward the higher end of the "comprehensibility" scale (on the grounds that in our study, the point of the metaphor category was to provide items that were fundamentally unknown but which should be more or less transparent). We eliminated items that used low frequency vocabulary and also made some small amendments to help ensure intelligibility for non-native speakers (e.g. we changed the original metaphor *time is a physician* to *time is a doctor*, and *humour is a salve* to *humour is a medicine*). Although *A is B* style metaphors actually make up a relatively small proportion of figurative language in general (Cameron, 2003), we selected these to ensure consistency of form in this category.

Translated idioms were chosen from published norming data and previous studies by one of the authors and fell into two sub-categories. One was made up of idioms translated from Bulgarian and German, which were intended to be generally unfamiliar to all participants (see Participants section below). Bulgarian idioms were taken from Nordmann and Jambazova (2016). This paper collected

together a set of 90 Bulgarian idioms, translated word for word into English by a Bulgarian native speaker and then verified by an independent Bulgarian-English speaker. The idioms differ in their syntactic structure and length, but none exist in the same or very similar form in English. German idioms were taken from the extensive data provided in Citron et al. (2016), who collected affective and psycholinguistic norms for 619 German idioms. They also provided English translations, and the idioms vary in syntactic form and length. We selected only items that do not exist in the same or very similar form in English. Items with very low frequency vocabulary were not considered, and items were chosen to represent a range of transparency/decomposability according to the norms collected for native speakers.

The second set of translated items were all taken from Mandarin, which has a large set of *chengyu* or "fixed expressions". Candidate items were chosen from items used in Carrol and Conklin (2014, 2017) all of which had been translated and verified, and normed for their high familiarity amongst Chinese native speakers. Again, phrases varied in their length and structure, but no items existed in the same or very similar form in English. Based on the previously obtained norming data, phrases varied in how decomposable they were considered to be by Mandarin native speakers, who judged the phrases in the L1. For both sets of translated idioms, we eliminated any phrases for which there were very obvious cultural allusions (e.g. some Chinese idioms refer to Chinese names or feature mythical creatures such as dragons).

Each phrase was assigned a "correct" meaning. Assigning a clear meaning for any idiom is not necessarily straightforward, especially once we consider the nuances of meaning that an idiom might carry (for example, does *kick the bucket* mean just "die" or "die quickly"?). In this study, we were concerned with the general meaning for each phrase, rather than in differentiating very specific knowledge of minor semantic differences. For English idioms this was a short paraphrase of the meaning, as agreed by the first two authors (both native English speakers). These were verified using various online resources and published idiom dictionaries. Idioms translated from Bulgarian

and German were assigned a meaning based on those provided in Nordmann and Jambazova (2016) and Citron et al. (2016), respectively. Chinese idioms were assigned a meaning based on the entry in the Dictionary of 1000 Chinese Idioms (Lin and Leonard, 2012), from which items were originally taken. As with the idioms themselves, low frequency vocabulary was removed from the definitions to ensure that they were as simple and comprehensible as possible. For novel metaphors, there is not strictly a "correct" meaning, so we began by assigning each phrase the best interpretation of the meaning based on the intuition of the first two authors, then asked three other native speakers to also give their interpretations. Items where 80% (4 out of 5) of these raters gave the same or a very similar interpretation were retained, and this was considered to be the "correct" interpretation for the purposes of the study.

We next created incorrect answers for all items. Just like assigning the "correct" meaning to any given phrase, choosing the incorrect alternatives is not a straightforward process. The degree of similarity to the actual meaning will determine how likely participants are to choose an incorrect distractor, so we adopted a process whereby the first two authors produced what we considered to be plausible alternative meanings for each phrase, and these were compared and any disagreements resolved following discussion and refinement. No alternative was very similar to the "correct" meaning (e.g. in the case of "die" and "die quickly" for *kick the bucket*). Also, no answer was closely related to an entirely literal reading of the phrase (since all were intended to be read figuratively, this would in many cases mark the literal alternative out as a clear distractor) , and all were considered to be *possible* figurative interpretations of the intended figurative meaning. For example, for the idiom *sitting on the fence*, we produced the following four alternatives:

1. To be undecided about something (actual idiom meaning)

2. To be in a risky situation

3. To be relaxed and carefree

4. To be on the outskirts of a situation

To ensure that the metaphors in particular had a "correct" meaning (i.e. the meaning that we had assigned was preferred out of the four options), we asked a set of seven native speakers to read each metaphor and the four alternatives we had provided. Our final list contained only items where at least five out of the seven raters agreed on the most likely meaning.

We selected final lists of 22 phrases for each category (22 English idioms, 22 novel metaphors, 22 translated Bulgarian/German idioms and 22 Chinese idioms). A full list of the stimuli is provided in the appendix.

Participants

Three groups were selected to take part in the rating study. The first was a set of English native speakers (L1 English, n = 31), all undergraduate students at a UK University. The second was a set of "general" non-natives (L2 English, n = 25), excluding anyone who had either Bulgarian or German as a first language, and excluding anyone who was a native speaker of Chinese. Participants were undergraduate or postgraduate students or postdoctoral researchers working at a UK university, and came from a variety of L1 backgrounds. They had been studying English for an average of 16.7 years (SD = 4.8), and had lived in the UK for an average of 12.6 years (SD = 1.8). The third group was a set of Chinese native speakers (L1 Chinese, n = 34), all studying English at a China campus of a British university in China (with English as the medium of instruction). They had been studying English for an average of 12.4 years (SD = 2.9), and had on average spent less than one year living in an English-speaking country (mean = 0.3, SD = 0.7).

As well as being asked to provide some general background information, The L2 English and L1 Chinese speakers were asked to complete a short vocabulary test as a basic measure of their proficiency. This consisted of a modified version of the Vocabulary Size test (Nation & Beglar, 2007), and consisted of 30 items selected from the 10,000 most frequent English words. In this test,

participants are presented with a word and five possible options (four possible meanings and a "don't know" option). The L2 English group scored higher (mean = 21.9/30, SD = 3.8) than the L1 Chinese group (mean = 14.7, SD = 2.3). We also collected data on how often participants used English in their everyday lives using a questionnaire that asked them to rate usage on a variety of dimensions (reading for pleasure, reading for work/study, watching TV, etc.). We included 10 dimensions, each rated out of five, to give a possible total of 50: L2 English, mean = 40.1, SD = 6.9; L1 Chinese, mean = 36.1, SD = 7.1.

Procedure

Following a short introduction where the purpose of the task was explained and an example of a figurative phrase was given, a worked example was provided for the phrase *kick the bucket*. Participants were shown the phrase at the top of the screen and asked to indicate how familiar they were with it on a seven point Likert scale. They were told that if they had never heard it before they should choose "one", and if they knew it well and had no problem understanding the meaning, they should choose "seven". Once they had responded using the number keys, they were asked how transparent they thought the phrase was. This was explained as how easily they thought they could guess the meaning of the phrase based on the individual words. Again, participants were asked to respond using a seven point scale, where "one" = very difficult to guess the meaning and "seven" = very easy to guess the meaning. They were then presented with the phrase along with four possible meanings, and asked to select what they thought was the figurative meaning of the phrase. Following this, they were presented with the phrase and were told what it actually means. They were then asked "Now that you know this, how easy is it to see the connection between the individual words and the figurative meaning?". Again, participants were asked to respond using a seven point scale, where "one" = no connection between the words and the meaning and "seven" = a really clear connection. Hereafter, we refer to the four answers collected here as familiarity, transparency, meaning and decomposability. Participants were asked whether they were unsure

about any aspect of the study, and informed that if they wanted to take a short break at any time they could do so. They were told that the test was not timed, but that they should try not to dwell too long on any given item and they should try to answer as intuitively as possible. All questions had a time limit of 30 seconds, so if a response was not provided within this time, the study moved on to the next question. (N.B. The limit was per question, not per trial, so a participant had up to 30 seconds to respond to the question on familiarity, then up to 30 seconds on transparency, etc.).

Items were presented in random order, with the phrase always presented along the top of the screen in Courier new font, size 18pt. For questions on familiarity, transparency and decomposability the numbers from one to seven appeared horizontally along the middle of the screen, with a reminder of what the extremes represented for each question. For the question asking for the figurative meaning, the four options were presented vertically one after the other, and were arranged throughout the study so that the position of the "correct" meaning was randomly assigned. In between each item there was a short pause, and the next item was preceded with "New phrase" to make sure that participants were aware that a new trial had begun. All participants saw all 88 items. For L1 English speakers this took around 30-40 minutes. Both L2 English and L1 Chinese speaker groups took around 45-60 minutes.

Results

Items where the response timed out for any of the four questions were removed, accounting for less than 1.4% of the data. Familiarity, transparency and decomposability were all scored on a scale from one to seven. Meaning was scored as one if a participant chose the correct response and zero otherwise. A summary of the response data by group and by phrase type is presented in Table 1.

Table 1. Mean ratings (standard deviation in brackets) for familiarity, transparency and decomposability (out of seven). meaning is expressed as the percentage of phrases for which the "correct" answer was identified.

	Familiarity	Transparency	Meaning	Decomposability
	L1 English			
English idiom	6.1 (1.5)	4.3 (1.6)	.88 (0.31)	4.9 (1.6)
Novel metaphors	3.1 (1.9)	4.4 (1.5)	.90 (0.31)	6.0 (1.1)
Translated idioms (general)	1.9 (1.4)	2.6 (1.3)	.29 (0.46)	3.1 (1.8)
Translated idioms (Chinese)	1.8 (1.3)	2.8 (1.3)	.43 (0.49)	3.8 (1.8)
	L2 English			
English idiom	4.6 (2.2)	4.3 (1.7)	.73 (0.44)	4.9 (1.8)
Novel metaphors	3.5 (2.1)	5.1 (1.5)	.85 (0.36)	6.2 (1.3)
Translated idioms (general)	2.4 (1.7)	3.5 (1.5)	.37 (0.48)	3.9 (2.1)
Translated idioms (Chinese)	2.3 (1.5)	3.6 (1.5)	.39 (0.49)	4.3 (2.0)
	L1 Chinese			
English idiom	3.6 (2.1)	4.1 (1.8)	.47 (0.50)	4.4 (2.0)
Novel metaphors	3.8 (2.2)	4.6 (1.9)	.71 (0.45)	5.4 (1.8)
Translated idioms (general)	2.9 (2.0)	3.6 (1.7)	.35 (0.48)	3.7 (2.0)
Translated idioms (Chinese)	4.6 (2.3)	4.8 (2.0)	.63 (0.48)	5.1 (1.9)

Data were analysed using linear mixed effect models, constructed using the lme4 package (version 1.1-13; Bates, Maechler, Bolker & Walker, 2015) in R (version 3.4.0, R Core Team, 2017). Ratings for familiarity, transparency and decomposability were centred. For meaning, a mixed logistic regression model was used. For each of the following analyses we fitted an omnibus mixed effects model with fixed effects of phrase type (Type) and participant group (Group). We included random intercepts for subject and item and, following Barr et al. (2013), by-subject random slopes for the effect of Type and by-item random slopes for the effect of Group, where these did not lead to any convergence issues. Where differences are reported these are based on the differences of least squares means extracted from the models using the difflsmeans function in the lmerTest package (version 2.0-33; Kuznetsova, Brockhoff and Christensen, 2016).

Familiarity

Figure 1 shows the patterns of familiarity for the three speaker groups for each phrase type.

Figure 1. Judgements of familiarity (centred) per phrase type for L1 English (left panel), L2 English (middle panel) and L1 Chinese (right panel) groups. Id = English idioms, Met = metaphors, Id-T = general translated idioms, Id-C = Chinese idioms. Error bars represent 95% confidence intervals.
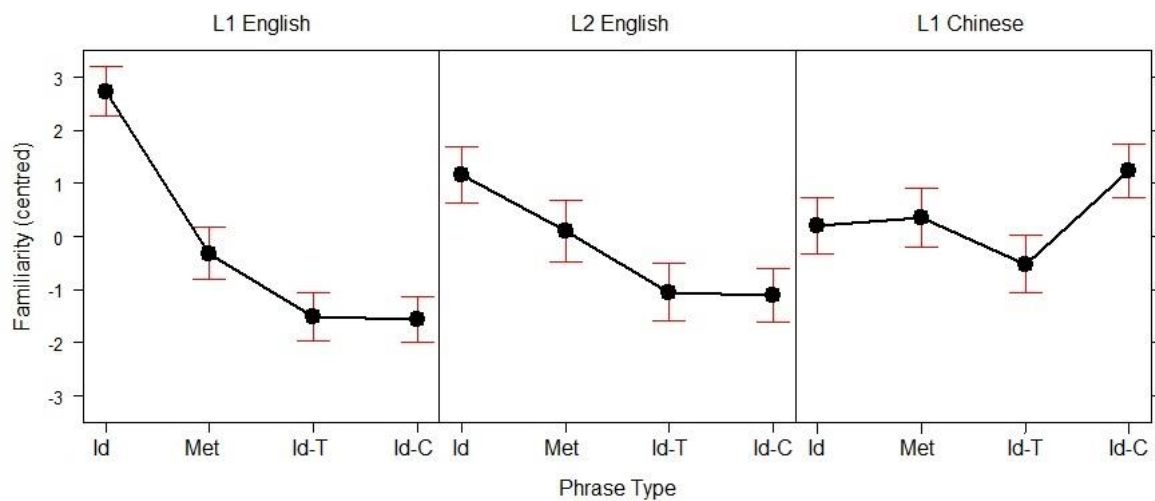
Figure 1 shows that, unsurprisingly, the L1 English group were more familiar with English idioms than the L2 English group(t = 5.23, p < .001), who were in turn more familiar than the L1 Chinese (t = 2.84, p = .005). Metaphors were perceived as similarly familiar for all three groups, with no difference between L1 English and L2 English (t = -1.29, p = .199), no difference between L2 English and L1 Chinese (t = -0.71, p = .481) and a marginal difference between L1 English and L1 Chinese (t = -1.88, p = .061). General translated idioms were no different for L1 English and L2 English groups (t = -1.54, p = .127) or L2 English and L1 Chinese (t = 1.56, p = .122), but were perceived as more familiar to L1 Chinese than L1 English (t = 2.91, p = .004). There was a clear effect of L1 knowledge for the translated Chinese idioms, where there was no difference in familiarity for L1 English and L2 English

(t = -1.69, p = .093), but L1 Chinese showed significantly higher ratings compared to both groups: L1 English, t = 8.86, p < .001; and L2 English, t = 7.59, p < .001.

Transparency

We analysed perceived transparency first in a model including only effects of Type and Group, then in a model including familiarity ratings. Inclusion of familiarity as a covariate made a significant improvement to the model ($\chi2$ (1) = 2699, p < .001). In this model, familiarity was a significant predictor of transparency (t = 58.25, p < .001), suggesting a very close relationship between the two variables, i.e. more familiar phrases are in general judged to be more transparent than less familiar phrases. Inclusion of familiarity as part of a three-way interaction with Type and Group made a further improvement ($\chi2$ (11) = 166.3, p < .001). Figure 2 shows perceptions of transparency by Type and Group, firstly with familiarity not included, and secondly when familiarity is included as an interaction term in the model.

Figure 2. Judgements of transparency (centred) per phrase type for L1 English (left panel), L2 English (middle panel) and L1 Chinese (right panel). The solid line shows the ratings for transparency on its own, while the dashed line shows the ratings once familiarity is included in the analysis. Id = English idioms, Met = metaphors, Id-T = general translated idioms, Id-C = Chinese idioms. Error bars represent 95% confidence intervals.
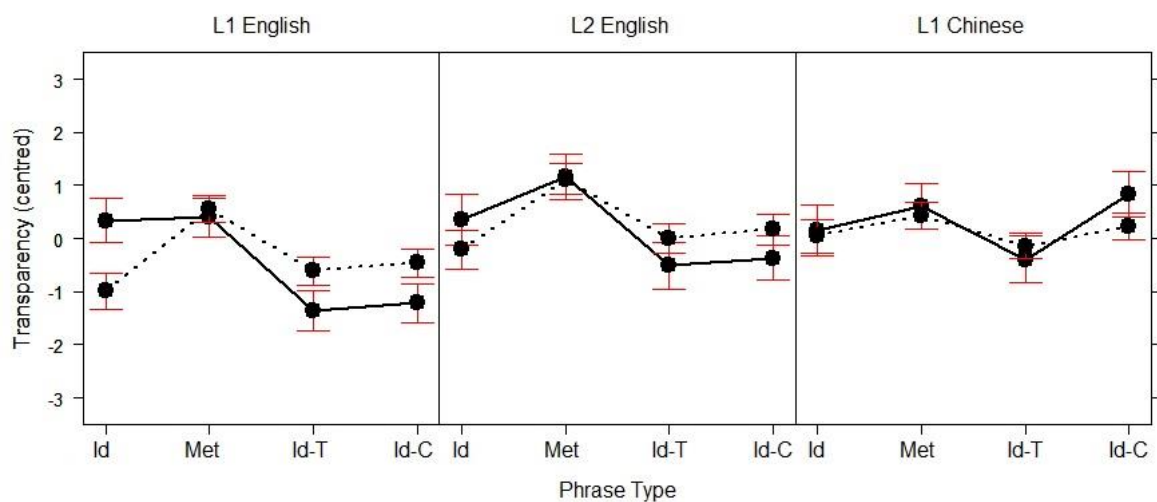

Figure 2 suggests that familiarity does make significant contributions to perceived transparency for all three speaker groups. As for the data as a whole, for all phrase types familiarity was a significant predictor of transparency, suggesting a clear positive relationship between the two variables. To explore these effects further and to avoid overfitting the model, we fitted separate models for each phrase type.

There was no difference between groups in a model without familiarity included, but once familiarity is included as an interaction with Group, there were differences between L1 English and L2 English (t = 2.39, p = .02) and L1 English and L1 Chinese groups (t = 4.24, p <.001), but not between the two non-native speaker groups (t = 1.65, p = .10). For all three groups, familiarity made a significant contribution to perceptions of transparency, but this effect is most clearly observed for L1 English

speakers (left panel, Figure 2). Once this was accounted for, the two non-native groups saw the English idioms as more transparent than the L1 English group.

Metaphors were seen as more transparent by L2 English speakers, both in the model without familiarity included and when this was included as an interaction term. With familiarity included, perceived transparency was higher for L2 English than both L1 English (t = 3.64, p < .001) and L1 Chinese (t = 4.47, p < .001). There was no difference between L1 English and L1 Chinese (t = 0.73, p = 0.500).

For general translated idioms, the addition of familiarity seemed to raise perceptions of transparency for all three groups, opposite to the effect seen for L1 English judging English idioms. When familiarity was included as an interaction term, L1 English speakers judged items to be less transparent than both L2 English (t = -3.74, p < .001) and L1 Chinese (t = -2.73, p = .008); there was no difference between the two non-native groups (t = 1.28, p = .203).

For Chinese idioms, the pattern for L1 Chinese speakers judging Chinese idioms (right panel Figure 2) mirrors the pattern for L1 English speakers judging English idioms (left panel Figure 2), whereby familiarity makes a clear contribution to perceptions of transparency. Once familiarity was accounted for there were differences in perceived transparency between L1 English and L2 English (t = 4.19, p < .001) and L1 English and L1 Chinese (t = 3.62, p < .001), but not between the two non-native groups (t = 0.74, p = .500).

Meaning

A series of mixed logistic regression models were fitted to explore the likelihood of each speaker group identifying the correct meaning for the different phrase types (see Table 1 for summary of correct responses here). An omnibus model was fitted with fixed effects of Group and Type. We added in familiarity and transparency to see whether these made significant contributions and compared the resulting models. Both familiarity ($\chi 2$ (1) = 171.4, p < .001) and transparency ($\chi 2$ (1) =

133.6, $p < .001$) made significant improvements on their own. Inclusion of both variables also made a significant improvement compared to either individually (compared to familiarity only: $\chi2$ (1) = 23.5, $p < .001$; compared to transparency only: $\chi2$ (1) = 61.2, $p < .001$). Both variables showed a positive overall relationship with meaning: for familiarity, $t = 7.76$, $p < .001$; for transparency, $t = 4.86$, $p < .001$). In order to explore this further, we fitted separate models for each phrase type. For each model we first added in familiarity to see whether this improved the model, then added transparency to see if this made any further improvement. The variables were added in this order on the grounds that if a phrase is known then its meaning can simply be retrieved directly, hence only for unknown phrases should relative transparency have an effect. Figure 3 shows the contribution of familiarity and transparency per speaker group for each phrase type.

For idioms, the best fitting model included an interaction between group and familiarity (comparison with model including fixed effect of Group only: $\chi2$ (3) = 88.8, $p < .001$.). Including transparency made no further improvement, either as a fixed effect ($\chi2$ (1) = 0.85, $p = .357$) or as a separate interaction with group ($\chi2$ (3) = 5.53, $p = .137$). The model confirms that familiarity was the main driver of identifying the correct meaning for L1 English speakers ($z = 6.89$, $p < .001$). There were interactions between familiarity and Group for L2 English ($z = -2.40$, $p = .016$) and L1 Chinese ($z = -4.45$, $p < .001$), hence familiarity was less important for identifying the correct meaning for these two groups (left panel, Figure 3).

For metaphors, familiarity made an improvement to a model with Group only ($\chi2$ (1) = 19.8, $p < .001$) and addition of transparency further improved the fit ($\chi2$ (1) = 31.1, $p < .001$). Inclusion of either as an interaction term made no further improvements. Analysis of this model confirmed that transparency was a significant predictor of identifying the "correct" meaning ($z = 5.61$, $p < .001$), but once transparency was included, familiarity was not a significant predictor ($z = 0.25$, $p = .805$). Model comparison showed that the best fitting model contained only transparency as a fixed effect, with inclusion of familiarity making no improvement ($\chi2$ (1) = 0.06, p = .808). For all three groups, the

relative transparency determined whether the "correct" answer was identified (z = 7.26, p < .001). In this model there was no difference between L1 English and L2 English (z = -0.54, p = .588) but L1 Chinese scored significantly lower than both L1 English (z = -5.94, p < .001) and L2 English (z = -3.71, p < .001).

The same pattern was observed for general translated idioms, where familiarity made a significant improvement ($\chi2$ (1) = 8.71, p = .003) and the addition of transparency made a further marginal improvement $\chi2$ (1) = 2.98, p = .085). This model suggested that when both variables were included, transparency contributed to identifying the correct answer (z = 1.74, p = .082) but familiarity did not (z = 1.31, p = .190). Model comparison again showed that the best fitting model contained only transparency as a fixed effect, and including familiarity made no further improvement ($\chi2$ (1) = 1.70, p = .192). In this final model, transparency made a significant contribution to identifying the correct answer for all three groups (z = 3.18, p = .001). There were marginal between-group differences for L1 English and L2 English (z = .89, p = .059) and L1 English and L1 Chinese (z = 1.70, p = .089).

For Chinese idioms, familiarity made an improvement to the basic model as a fixed effect ($\chi2$ (1) = 27.9, p < .001) and inclusion of transparency further improved the model ($\chi2$ (1) = 19.8, p < .001). Both variables were significant predictors of identifying the correct answer: familiarity, z = 2.82, p = .005; transparency, z = 2.34, p = .019. Although including interactions did not make any improvement to the overall model, there was some indication that L1 Chinese did show an effect of familiarity but not transparency, whereas the pattern was reversed for the other groups (bottom row, Figure 3).

Figure 3 summarises the effects of familiarity and transparency for each phrase type, based on a model including both variables as interactions with Group. In this, it is clear that when phrases are known, familiarity is the main driver of whether or not the meaning is correctly identified. When phrases are unknown or less familiar, transparency is much more important.

Figure 3. Effects of familiarity (left) and transparency (right) for each phrase type: English idioms (top row), metaphors (second row), translated idioms (third row) and Chinese idioms (bottom row). Grey shading indicates 95% Confidence Intervals. Mean correct is expressed on the logit scale.
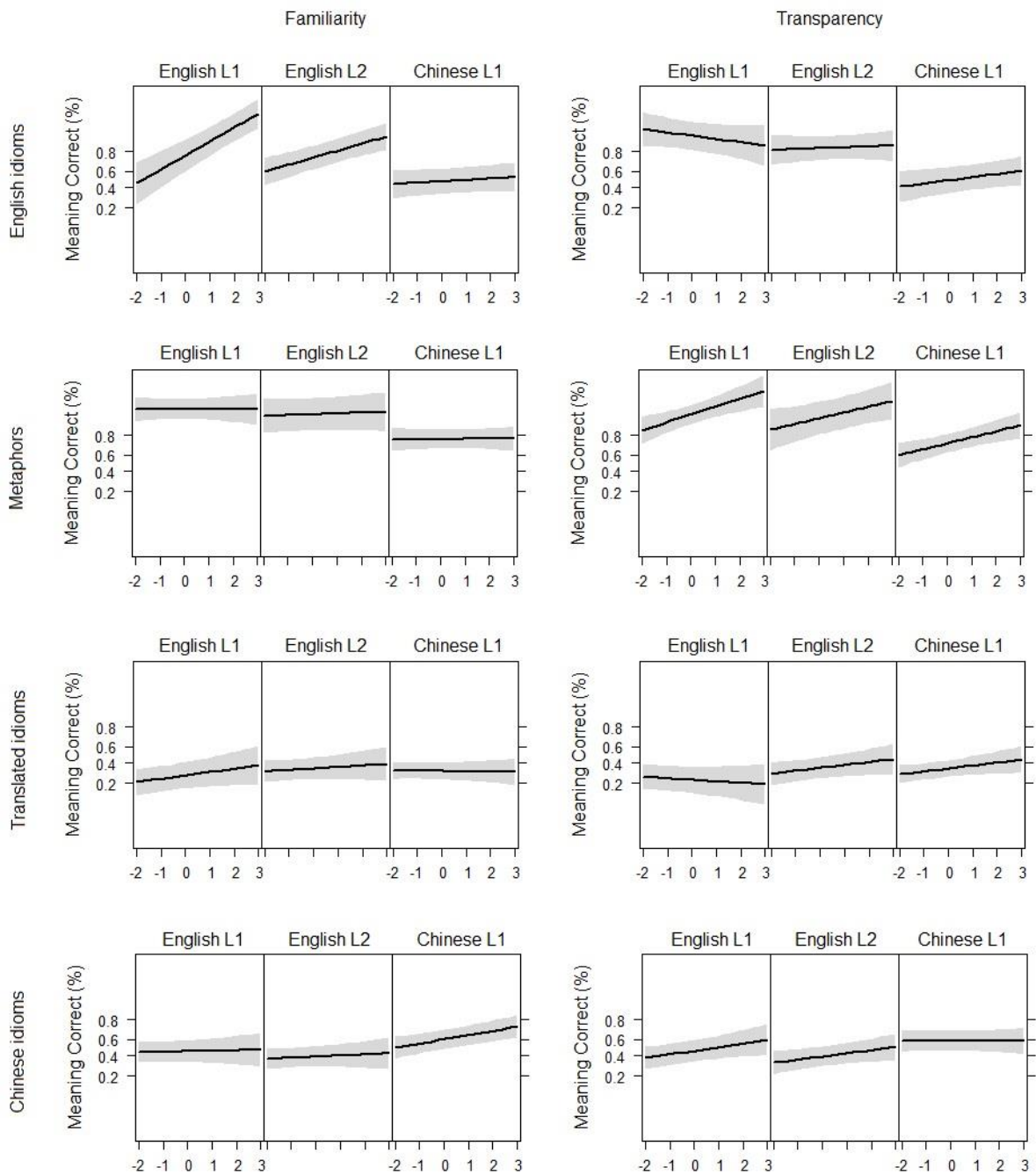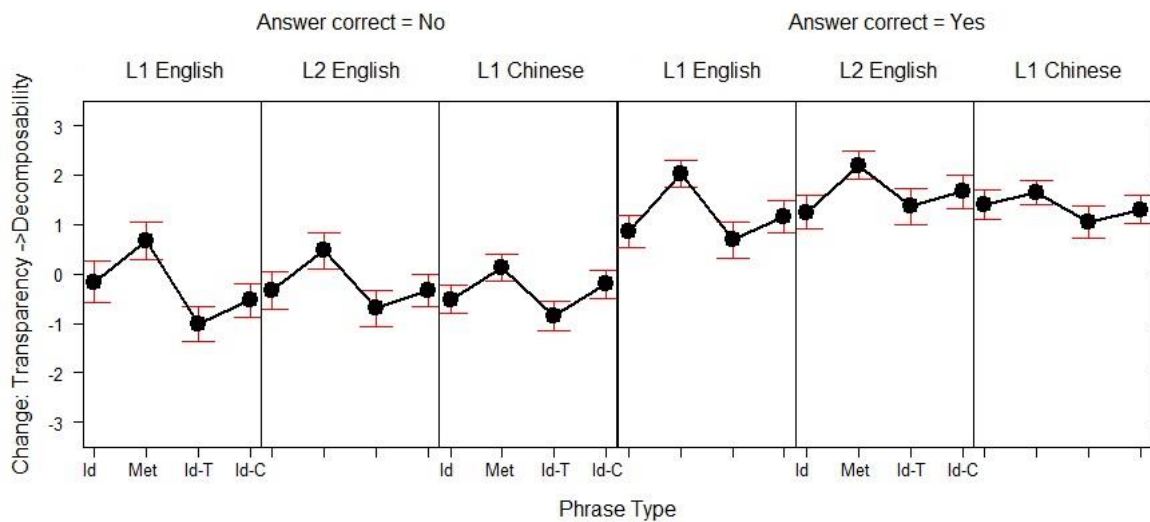
Decomposability

There was a clear correlation between initial transparency scores and subsequent decomposability scores (r =.54, p < .001). Comparison of the two scores showed that there was overall a significant difference between the two (mean transparency = 3.9/7, mean decomposability = 4.6/7, t(7804) = -30.92, p < .001), hence ratings for decomposability (after the meaning was known) were in general higher than for transparency (before the meaning was known). To explore this further we constructed a model with the size of the change as the dependent variable and included fixed effects of Type and Group. We included the initial transparency rating, which significantly improved the model as a fixed effect (χ2 (1) = 2257.5, p < .001) and further improved it as part of a three-way interaction with Type and Group (χ2 (11) = 55.5, p < .001). We next added familiarity, which improved the model as a fixed effect (χ2 (1) = 6.24, p = .013), but made no further improvement as part of an interaction with Type and Group. There was a positive overall relationship between familiarity and the size of the change from decomposability to transparency: t = 2.51, p = .012. We finally considered whether adding meaning would make an additional improvement, on the grounds that whether participants got the answer right or wrong might be important in how they re-evaluated their original rating for transparency. Model comparison showed that inclusion of meaning made an improvement (χ2 (1) = 2182.9, p < .001), and when it was included familiarity was no longer a significant predictor. The best fitting model included interactions between Type, Group and transparency, and Type, Group and meaning. This model suggests that participants were more likely to show an increase from transparency to decomposability if they successfully identified the correct meaning, compared to when they were incorrect. Figure 4 shows patterns according to Phrase Type and Speaker group, and according to whether or not participants correctly identified the meaning of each phrase.

Figure 4. Change from transparency -> decomposability scores for each phrase type, for items where

the meaning was correctly identified (right three panels) and was not (left three panels). Id = English

idioms, Met = metaphors, Id-T = general translated idioms, Id-C = Chinese idioms. Error bars

represent 95% confidence intervals.

The findings presented in Figure 4 suggest that correctly identifying the meaning of a phrase led to a

greater increase in the score from transparency to decomposability than if the meaning was not

identified. The final model confirmed that this difference was greater for the two non-native groups

(effect for L1 English, t = 6.26, p < .001; interaction with L2 English, t = 2.76, p = .006; interaction with

L1 Chinese, t = 4.71, p < .001). To better interpret the effects, we fitted separate models for each

phrase type. Each model included the interaction of Group, transparency and meaning.

For English idioms, all three groups showed a non-significant negative change from transparency to

decomposability when they did not identify the correct meaning, but there were no between-Group

differences (L1 English, t = -0.72, p = .470; L2 English, t = -0.90, p = .370 Group: L1 Chinese, t = -1.68,

p = .095). There was a significant effect of meaning for L1 English (t = 6.35, p < .001), and significant

interactions between group and meaning for both L2 English (t = 3.17, p = .002) and L1 Chinese (t =

4.79, p <.001). For all three groups, when the meaning was not correctly identified (therefore

presumably not known in the first place), subsequently learning the meaning made little difference to how decomposable the idioms were perceived to be (relative to the original rating for transparency). In contrast, correctly identifying the meaning (either because this was known or because it was correctly inferred) led to a positive change, which was more pronounced for both non-native groups compared to the L1 English speakers.

For metaphors, there was a significant change from transparency to decomposability for L1 English (t = 4.73, p < .001). L2 English showed a similar change compared to L1 English (t = -1.36, p = .174) but L1 Chinese were significantly less likely to show an increase for incorrect responses (t = -2.99, p = .003). There was again a significant effect of correct answers for L1 English (t = 9.85, p < .001), and an interaction between Group and meaning for L2 English (t = 2.53, p = .012) but not L1 Chinese (t = 1.12, p = .261). All three groups therefore showed a larger change when the correct answer was identified, although this was stronger for L2 English speakers.

For general translated items, incorrectly identifying the meaning had a significantly negative effect for L1 English (t = -4.96, p < .001) and there were no between-group differences (L2 English, t = 1.43, p = .155; L1 Chinese, t = 0.74, p = .461), hence this decrease was comparable across all three groups. In contrast, correctly identifying the answer led to an increase in perceived transparency for L1 English (t = 10.48, p <.001), and there were no interactions of Group and meaning for L2 English (t = 1.61, p = .108) or L1 Chinese (t = 1.04, p = .301).

For Chinese idioms, L1 English speakers showed a negative effect of incorrectly identifying the answer (t = -3.17, p = .002), and there was no effect of Group for L2 English (t = 1.28, p = .204). There was a marginal Group effect for L1 Chinese (t = 1.7, p = .084), suggesting that the decrease was negligible for this speaker group. All three groups showed an effect of meaning, whereby correctly identifying the answer led to a higher decomposability rating (relative to the initial transparency rating) for L1 English (t = 12.40, p < .001), and no interactions for L2 English (t = 1.08, p = .279) or L1 Chinese (t = -1.50, p = .134).

Overall, the results suggest that there is a clear difference according to whether or not participants correctly identified the meaning of the phrase. When initial familiarity or transparency led to a response being correct, this led to higher decomposability ratings. When the meaning was incorrectly identified, metaphors still showed a small increase in subsequent decomposability ratings, but for all idiom types, an incorrect answer led to decomposability ratings that were lower than the initial rating for transparency.

Other factors

We finally explored the effects of relative proficiency level and other learner variables, since this varied between the two non-native speaker groups. We included vocabulary score as a predictor in a series of models to assess its impact, both between groups and as a factor within the two groups (i.e. do higher proficiency learners show different patterns in general?). Adding vocabulary score into a model for familiarity made no improvement ($\chi2$ (1) = 2257.5, $p < .001$). This is perhaps surprising, since we might assume that higher proficiency would lead to greater familiarity with idioms, but this does not seem to be the case. Similarly, number of years studying English did not make a significant improvement to the model ($\chi2$ (1) = 0.18, $p = .676$), but usage total (self-rating of how often English is used in participants' everyday lives) did make a significant improvement ($\chi2$ (1) = 6.89, $p = .009$). For transparency ratings, none of vocabulary score ($\chi2$ (1) = 0.76, $p = .384$), years studying English ($\chi2$ (1) = 0.21, $p = .645$) or usage ($\chi2$ (1) = 2.64, $p = .105$) made a significant improvement to the model.

Inclusion of additional variables in logistic regression models led to convergence errors when random slopes were included, so for identifying correct meaning we compared models with random intercepts only. Vocabulary score ($\chi2$ (10) = 15.4, $p < .001$) and usage total ($\chi2$ (1) = 11.5, $p < .001$) made significant improvements to a model with fixed effects of Type and Group, but years studying English did not ($\chi2$ (1) = 0.01 $p = .965$). For vocabulary score, inclusion of an interaction with type and group made a further improvement ($\chi2$ (7) = 31.7, $p < .001$). This meant that higher vocabulary led to more correct answers by L2 English speakers for English idioms and metaphors, but not for

either of the translated sets of items. For L1 Chinese, higher vocabulary had no effect for any of the phrase types. Inclusion of usage total as an interaction term was also an improvement ($\chi2$ (7) = 21.0, p = .004). As with vocabulary score, higher usage led to more correct answers for L2 English for English idioms and metaphors, but had no effect on translated idioms. For L1 Chinese, usage had no effect for any phrase type.

Finally, we compared models for the change from transparency to decomposability ratings. Vocabulary score made no improvement to these models ($\chi2$ (1) = 0.02, p = .969) but years studying English ($\chi2$ (1) = 4.63, p = .031) and usage total ($\chi2$ (1) = 6.63, p = .010) both did. Neither made any further improvement as an interaction with Group and Type. This suggests that once the meaning had been provided for all phrases, participants with more experience of using English were more inclined to see meaning (relative to the original rating for transparency) than those with less experience.

General Discussion

We set out to explore some of the factors involved in how native and non-native speakers understand figurative language. Our results show a clear relationship between familiarity and transparency, whereby perceived transparency is affected by how well known a phrase is. This issue has been raised in the idiom literature before (Keysar & Bly, 1995, 1999; Libben & Titone, 2008; Nordmann et al., 2014) but it is an aspect that has been overlooked in many studies. For L1 English speakers, highly familiar idioms seemed more transparent than they really are, and unknown (translated) idioms seemed less transparent. Once familiarity was accounted for (Figure 2, left panel), the three sets of idioms in this study were seen as equally transparent by L1 English speakers. Logically, this is exactly what we would expect, since there is no reason to expect English idioms (in general) to be inherently more or less transparent than idioms taken from any other language. The two non-native groups showed similar, if less pronounced patterns. That is, when idioms were highly familiar, they were also judged as more transparent, and unknown phrases also seemed less

transparent. The L1 Chinese group showed minimal effects for English idioms, but crucially did show a clear effect for Chinese idioms (familiarity led to higher transparency ratings). We will discuss differences between native and non-native groups, and the effects of L1 influence in more detail later in this section.

For metaphors, perceived familiarity made no contribution to judgements of transparency for any group. Whether these are "familiar" or not is an interesting question. They are not fixed expressions, therefore the form is not familiar, but in many cases the underlying ideas are common, e.g. *humour is a medicine,* which may be recognised as similar to idioms such as *laughter is the best medicine*. Underlying conceptual metaphors are important in how some idioms are processed (Gibbs, 1993), but based on the results here, this does not extend to phrases that simply approximate a known figurative idea. This underlines the importance of the specific, conventionalised form of idioms, rather than the message they represent (see e.g. Gibbs et al. 1997 and McGlone, Glucksberg & Cacciari, 1994, both of whom found an advantage for idioms over literal paraphrases that share the same underlying conceptual metaphor).

For the general translated idioms, lack of familiarity seemed to have the opposite effect than it had on English idioms. That is, phrases that were entirely unknown were perceived as less transparent, and once (lack of) familiarity was accounted for, perceived transparency was higher for all three groups. The same pattern was seen for L1 English and L2 English speakers judging Chinese idioms, but notably not for L1 Chinese (who were familiar with this set of items). Taken together, the effect of familiarity is clear: when phrases are well known, the perception of how transparent they are is inflated, since the link between the meaning and the form is well established and difficult to ignore. When phrases are entirely unknown, they also seem less transparent. This may indicate that speakers either struggle to generate possible meanings, or else are less confident in the meanings that they might be able to think of (i.e. they may be able to think of several, equally plausible meanings, but not decide which is most likely). Once idioms are known, a link is established between

the phrase and one specific meaning, hence other meanings (even if they are ostensibly equally valid) begin to seem less logical (e.g. Keysar & Bly, 1995, 1999).

Familiarity and transparency both made contributions to how successful participants were at identifying meaning, and this pattern was similar for all three groups. For English idioms, familiarity but not transparency was the main driver. For L1 English this is as expected – idioms are in general so well known that their meaning is simply retrieved, regardless of how transparent they are. This may represent a ceiling effect, but it is a similar finding to Libben and Titone (2008), where decomposability played no role for highly familiar idioms. It may also reflect how idioms are understood during "normal" language processing more generally: when idioms are very well known they can be accessed directly, regardless of their semantic properties (e.g. Abel, 2003). For the two non-native groups, familiarity is still a key driver of whether the correct meaning was identified. A logical assumption would be that since non-native speakers are in general less familiar with idioms, both familiarity and transparency would be important (known phrases could be retrieved, unknown phrases must be analysed), however this was not borne out by the analysis. Figure 3 (top row) does suggest that familiarity is less important for L2 English and L1 Chinese speakers, but there is no corresponding increase in the effect of transparency. Lack of context may have been important here, and if idioms were presented in supporting contexts then this may also contribute to both perceived transparency and, in turn, to greater non-native success in identifying the meaning. Overall, in the absence of any context, knowledge of L2 idioms is the main predictor of success; since non-natives simply know fewer idioms, they are less successful at identifying the figurative meanings.

For metaphors, only relative transparency was important for "successful" meaning prediction. When a specific phrase is not known (even if it expresses a familiar idea), speakers need to actively work out the meaning rather than simply retrieving a lexicalised entry; the transparency of the phrase will determine how much effort will be required to do this (and therefore how likely it is to be "successful"). All three groups showed effects of transparency, indicating that, broadly, native and

non-native speakers were able to follow the same lines of reasoning to successfully identify the "correct" meanings. Especially in the absence of any supporting context, the only option is to employ the kind of semantic analysis skills that are a part of normal L1 competence (Levorato & Cacciari, 1995, 1999; Nippold & Taylor, 1995). When relative transparency was accounted for, L1 English and L2 English speakers showed no differences in identifying the meanings of metaphors, suggesting that this is not a skill that is particularly advantaged in L1. Translated idioms followed a similar pattern. Since they were unknown to all participants, it is unsurprising that familiarity played no role in how transparent they were perceived to be, or how easily the meaning could be identified. In addition, because the translated idioms were in general much less transparent than the metaphors, participants were much less successful at identifying the correct answer. All three groups did perform at above chance level for the translated idioms (all groups, p < .001), suggesting that they were able to utilise their semantic analysis skills to infer the most likely meaning for at least some of the phrases.

Ability to identify the meaning of Chinese idioms was driven by both familiarity and transparency, although visual inspection of Figure 3 suggests that this pattern was largely driven by the L1 Chinese group showing effects of familiarity, and the other two groups showing effects of transparency. In line with the other phrase types, when a phrase is known (which was more likely to be the case for the L1 Chinese group), this negates the need to undergo semantic analysis in order to work out the meaning. In contrast, when a phrase is fundamentally unknown (for the L1 English and L2 English groups), speakers are more successful for phrases that are more transparent.

As we define them here, we saw a clear difference between judgements of transparency (before the meaning was known) and decomposability (after the meaning was known). This highlights an important consideration in how we define these variables, which is the stage at which the judgement is being made. In the literature, ratings of decomposability are very often collected after the meaning is known: a phrase is given along with its meaning, and participants are asked to

explicitly judge the connection between the two (e.g. Gibbs & Nayak, 1989; Titone & Connine, 1994; Libben & Titone, 2008). As Boers and Webb (2015) point out, this means that motivation – whether this is based on identifying an underlying conceptual metaphor, making a connection with a particular domain, or mapping figurative meanings onto component words – can only be judged once the meaning is known. It is therefore a judgement of whether or not an idiom makes sense, once the explicit meaning is provided, and this will vary according to the knowledge and connections that each speaker is able to activate. However, this is far from the situation faced by language learners the first time they encounter an idiom in their L2. They must decide, based simply on the information available - the phrase itself, knowledge of the individual words, cultural information that may be relevant, any supporting context, etc. – whether they can infer a logical meaning. The potential pitfalls here are manifold: misunderstanding or not knowing one or more component words (Liontas, 2015), difference in conceptual basis (Charteris-Black, 2002), lack of appreciation of the cultural allusions (Boers & Webb, 2015), mismatch with an L1 idiom (Laufer, 2000; Irujo, 1986), etc. The only objective measure of this judgement is whether or not it allows the learner to guess the correct meaning: a truly transparent phrase is one where the meaning can be successfully inferred in advance, rather than simply one where a connection can be seen once the meaning has been revealed.

It was notable in our results that the biggest factor affecting the change from transparency to decomposability was whether or not a participant identified the meaning of the phrase correctly. This was in general true for all groups and all phrase types. When participants were correct, the increase was substantially more than when they were incorrect. One way to interpret this is as a confirmation effect, whereby participants use the information available them to make their best guess as to the meaning; if this guess is shown to be correct, it validates the reasoning that led to it, and increases the perceived link between the phrase and its meaning. Again, it is clear that knowledge of and increasing familiarity with the actual meaning affects how it is perceived, relative to other attributes of the phrase (Keysar and Bly, 1995, 1999; Schweigert et al., 2003; Schweigert,

2009). In contrast, incorrectly guessing the meaning led to either a negligible change (for English idioms), a less positive change (for metaphors) or a negative change (for translated idioms). This pattern suggests a similar thought process in how participants are judging the phrases and validating their own analysis: they make their best guess as to the meaning based on the connections they can see, and if this turns out to be incorrect, their confidence in that analysis is reduced. In particular for translated idioms where no knowledge of the meaning is available in advance, the distinction between getting the answer wrong (which produced a negative change) and getting the answer right (which produced a positive change) was marked for all three groups.

These results come together to suggest that semantic properties like transparency and decomposability should be clearly differentiated in the literature, according to the stage at which the judgement is being made. It may be that we can define separate stages of subjective transparency (an inference as to the likely meaning the first time a phrase is encountered), objective transparency (did this inference actually lead to the correct meaning?), and analysability (how clear is the connection between the phrase and the meaning, once this is known?). Analysability may in turn be made up of the separate but related components of decomposability and motivation. Decomposability is the extent to which the individual words make an identifiable contribution to the meaning, while motivation describes the connection that can be seen between the literal reading of the phrase and its figurative meaning, regardless of what information this is based on. In the case of phrases with an underlying conceptual metaphor, such as *blow your top*, this may contribute to both motivation *and* decomposability. Decomposability might further be seen as a global feature (where each word makes a contribution), or a more localised feature (where only some words contribute, or the contribution is relatively subtle). Figure 5 shows how these separate components may interact.

Figure 5. Factors contributing to transparency (before meaning is known) and analysability (after meaning is known); we can further divide this into the overall motivation (link between the phrase and the meaning) and decomposability (contribution of the individual words).
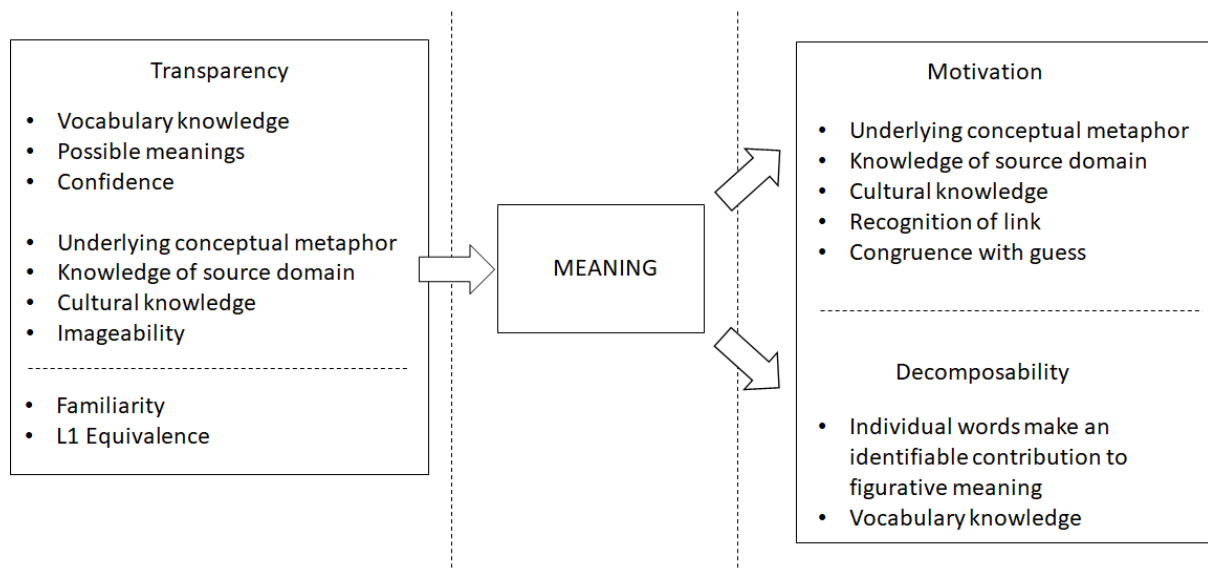
An important broader question is what exactly contributes to these judgements at each stage, i.e. what is it that makes a phrase seem transparent at all? We have discussed some aspects of this in the introduction, but the three main routes seem to be conceptual motivation (connection with an underlying metaphor), experiential motivation (knowledge of the cultural or historical relevance), and contribution of the component words. The first of these might be seen as most transparent, as in idioms like *blow your top* or *over the moon*, which reflect conceptual metaphors like ANGER IS HEAT and HAPPY IS UP, respectively. The second set (e.g. *on the ropes*, *bury the hatchet*) can, in general, be seen as iconic or metonymic acts, hence may be entirely non-transparent if the act itself, or the domain it comes from, is not known. Only the last of these – the contribution of the component words - seems related to decomposability, although it is important to remember that

these may not be mutually exclusive. For example, a phrase like *bury that hatchet* is motivated by its reference to the iconic act of making peace by Native Americans, but the word *bury* contributes something to the sense of "covering up" (see also the example of *kick the bucket* Hamblin and Gibbs, 1999, discussed previously). Again, the importance of familiarity must be restated here. Native speakers need not have any knowledge of the etymology of idioms in order to use and understand them, and in many cases phrases may be acquired and be part of the lexicon for many years without a speaker ever even considering where the meaning comes from (Wray, 2009).

One aspect of this that we have not considered is imageability – the ability to form a mental picture of the phrase – and whether this makes a contribution to perceived transparency.[1] Imageability is known to be beneficial for learning idioms in the L2 (Steinel et al., 2007), but Boers and Webb (2015) caution that this may also lead L2 learners to over-apply literal meanings in an attempt to infer meaning in unknown idioms. We therefore investigated whether imageability made a contribution in our study by consulting the concreteness norms collected by Brysbaert, Warriner and Kuperman (2014). Concreteness and imageability are highly correlated (Paivio et al., 1968), although they do vary in many cases (Richardson, 1976). Ratings are also generally for single words, rather than whole phrases, so we apply this post-hoc analysis cautiously. Nevertheless, we aggregated concreteness ratings for each of the phrases in our study by obtaining the rating for each word from Brysbaert et al. (2014), then calculating an average per phrase. Addition of these ratings made no improvement to the model for transparency, either as a fixed effect or interaction term. This suggests that on its own, concreteness does not predict transparency, but it may be important as part of the complex of factors that are taken into account when language users make such judgements.

Overall, it is clear that as we define them here, transparency and decomposability do represent different but related aspects of analysability. Multiple sources of information contribute to these judgements, and the stage at which they are being made is important. Crucial, though, is the fact

---

[1] We are grateful to our colleague at the University of Birmingham, Paula Perez-Sobrino, for suggesting this.

that judgements cannot be seen as independent of familiarity, which makes a clear contribution to perceptions for native and non-native speakers.

We should highlight that a number of factors not considered here might also be very important in our understanding of how figurative meaning is processed and understood. One factor in particular that undoubtedly plays a vital role is context. Previous work has shown that a biasing context supports the appropriate interpretation of an ambiguous (literally plausible) idiom for both native speakers (e.g. Fanari, Cacciari & Tabossi, 2010) and non-native speakers (e.g. Cieślicka, 2011). We have also discussed in the introduction various ways in which context should help in guessing or working out the meaning for newly encountered idioms for both L1 children and L2 learners (e.g. Boers & Webb, 2015; Cain & Towse, 2008; Cain et al., 2009; Gibbs, 1991; Liontas, 2002; Wray et al., 2016). In this study we purposely chose to present phrases with no supporting context. This allowed us to more easily test the contributions of familiarity and semantic properties without *muddying the waters*, but there is inevitably something artificial about any study that considers idioms out of the rich contexts that they generally occur in. This includes local, sentential contexts, but also a host of discourse level and even extra-linguistic cues that speakers can use to successfully infer meaning for any given phrase. Other factors may also be important, and the interplay of different variables is acknowledged in the Constraint-based model proposed by Libben and Titone (2008; also Titone & Connine, 1999), whereby all available information (familiarity, semantic decomposability, literalness, context) is used to reach the correct interpretation of a phrase. One underexplored dimension here is the degree to which an idiom may evoke particular emotions in a hearer, and how this might affect perceptions, interpretation, etc.[2] Related to this is the emerging field of embodied simulation, whereby linguistic interpretation is actively supported by contributions from the sensory and motor systems. Evidence is mixed as to how this applies to figurative language. For example, Desai et al. (2013) found that action verbs provoke activation of sensory-motor areas in literal sentences and metaphorical sentences, but not idiomatic ones. They suggested that degree of abstraction and

---

[2] We are grateful to an anonymous reviewer for suggesting this.

degree of conventionalisation might both modulate this effect, but it is certainly worth considering as part of how figurative competence operates. Considering and operationalising the range of potential variables here is no small undertaking, but it is essential if we are to present a unified view of the resources that language users employ when working out the meanings of idioms.

Non-native differences and cross-language influence

Unsurprisingly, the native and non-native speaker groups differed in how familiar they were with English idioms. It was notable, however, that for the non-natives, usage but not proficiency was important in determining familiarity, supporting the idea that it is only through exposure to the language that a broad knowledge of formulaic language can be developed (Schmitt, 2014). Once familiarity was accounted for, the L2 English group (and to some extent the L1 Chinese) seemed, in general, to see the phrases as more transparent than L1 English speakers. This is in line with Littlemore's (2010) findings that non-natives may be more inclined to see meaning in metaphor in their L2.

As with the more general discussion in the previous section, what contributes to this increased transparency remains open to further investigation. When we ran our post-hoc analysis of the influence of imageability, we considered whether this might be more important for non-native speakers, but this turned out not to be the case. We should certainly entertain the possibility that the characteristics of the two groups are important: the general non-native group in particular was made up of high-achieving L2 learners, who may be much more used to encountering unknown words and phrases and actively trying to work out the meaning – Wray et al. (2016) highlighted a similar performance-based effect for non-native speakers in their study. Alternatively, this might be taken as further evidence that native speakers do approach figurative language in a more holistic way – they simply know the meaning of phrases, hence the degree to which they need to analyse them is greatly reduced. Whilst there is evidence that the literal meanings of component words are activated when native speakers encounter idioms (e.g. Holsinger & Kaiser, 2013; Smolka, Rabanus &

Rösler, 2007; Titone & Connine, 1994), familiarity with the phrase may mean that the overall figurative meaning supersedes the contribution of individual words (although this may vary as a function of decomposability).

In contrast, learners are more inclined to undertake a literal word-by-word analysis by default (e.g. Cieślicka, 2006). This, combined with a lower level of familiarity with the phrase level meaning, may mean that non-natives have no choice but to analyse the component words in an effort to guess the meaning. In turn, this may mean that learners see possible connections that native speakers might never consider (as in the example of "skirt", discussed in the introduction), or which native speakers have simply never had to consider before. Less knowledge may therefore lead non-native speakers to see phrases (and L2 language in general) as more transparent – they may assume that they have seen all of the meaning, even when this is not the case, although this is presumed to vary according to factors such as proficiency, metalinguistic ability, etc. In addition, native speakers are generally less likely to engage in such behaviour (actively looking for meaning and working out unknown words and phrases) in their much more familiar L1. Overall, despite some indication that non-natives perceived phrases across the board as more transparent than native speakers, the patterns of results according to phrase type were comparable. This supports the idea that the skills required to successfully analyse and interpret figurative language are not fundamentally different for native and non-native speakers; differences that do exist are likely to be in terms of vocabulary knowledge, cultural knowledge and general familiarity with idioms, which contribute to perceptions of transparency in the same way as for native speakers.

On the question of how L1 knowledge is used, the pattern of results for L1 Chinese speakers was directly comparable to the pattern shown by L1 English speakers for English idioms. That is, phrases that are idioms in the L1 were seen as more familiar, more transparent and more decomposable, and the meaning was identified correctly more often for these phrases, relative to the performance of the L1 English and L2 English groups on these phrases. In line with Irujo (1986), Laufer (2000) and

Charteris-Black (2002), amongst others, L1 knowledge is clearly used to support the comprehension and production of figurative language in the L2 when congruent phrases exist. An increasing body of work is also supporting the idea that L1 knowledge is activated automatically during the online processing of formulaic and figurative language in the L2 (Carrol & Conklin, 2014, 2017; Carrol, Conklin & Gyllstad, 2016; Pritchett et al., 2016; Titone et al., 2015; Wolter & Gyllstad, 2011, 2013). Kecskes (2015) suggested that the idiom principle remains the most salient mechanism in language production in L1 and L2, but that a lack of resources (limited knowledge of figurative and formulaic expressions) leads to a limited use of formulaic language in L2. Our results speak to both of these arguments: L1 Chinese speakers showed limited knowledge of English idioms, both in terms of their subjective familiarity (mean = 3.6/7, SD = 2.1) and their ability to identify the meaning (mean = 47%, SD = 50); in contrast, their performance on idioms they knew from the L1 showed greater familiarity (mean = 4.6/7, SD = 2.3) and a higher success at identifying the correct meaning (mean = 63%, SD = 48).

Idioms translated from Chinese were also perceived as more transparent than either the English idioms or the general translated idioms, which may again be evidence for the non-independence of semantic judgements. In this case, greater familiarity with the idioms, even though they were presented in an unfamiliar form, led to higher perceived transparency. One additional consideration here is the nature of Chinese idioms, which are often based on old folk stories or historical events, and learning an idiom often involves also learning the story behind it. The connection between the phrase itself and the meaning may therefore seem more obvious if this story is known (much like knowing the etymology of a phrase like *bury the hatchet* may make it seem more transparent). It may therefore be the case that it was not simply the effects of familiarity with the phrases that made them seem more transparent, but that the additional information available to L1 Chinese speakers contributed to a richer context here. One argument against this, which suggests that fundamentally the same processes are at work for the L1 English and L1 Chinese speakers judging their own L1 idioms, is that once familiarity is accounted for, both sets of speakers consider all three

idiom types to be equally transparent (Figure 2). This suggests that the conventional transparency identified by Keysar and Bly (1995, 1999) – i.e. a sense of perceived transparency that arises from familiarity with the idiom – is what is being identified in both cases, rather than true conceptual transparency.

Finally, given that a fixed, conventional form is such a vital aspect of how idioms are recognised, the L1 effects we see here and elsewhere in the literature are perhaps more surprising than they seem. Holsinger (2013) compared processing for idiomatic phrases (*kick the bucket*) vs. phrases where the noun was replaced with a semantic associate (*kick the pail*). When these were presented in context-free sentences, the semantic associate showed no evidence that the idiomatic meaning was considered; however, when the phrases were placed in a biasing context, there was evidence that participants considered the idiomatic meaning of both phrases. In our study, Chinese L2 participants saw versions of known (in the L1) idioms in isolation (i.e. in the equivalent of Holsinger's context-free condition), but since these were translated, these were by definition not the conventionalised forms that normally produce faster processing for well-known phrases (e.g. Tabossi et al., 2009). Other studies have shown that even minor variations can remove the formulaic advantage for idioms (McGlone, Glucksberg & Cacciari, 1994) and other formulaic phrases like binomials (Siyanova-Chanturia, 2010), and there is some evidence from EEG studies that recognition of idioms generates patterns of brain activity that are generally associated with template matching (e.g. Molinaro & Carreiras, 2010; Vespignani et al., 2010) or identification of known sequences (e.g. Zhang, Yang, Gi & Ji, 2013). Despite this evidence that even minor violations lead to disruption during online processing and recognition of idioms, the translated Chinese phrases were recognised and understood as idioms, albeit in an offline task. This suggests that translation does not block the recognition of well-known word combinations, despite them appearing in an entirely unfamiliar form, and that learners are fully able and willing to draw on L1 knowledge of aspects like semantic analysability when making judgements, even in an L2.

Conclusions

In this study we set out to explore some specific questions about the nature of figurative competence in native and non-native speakers. We conclude by addressing our research questions directly:

1. Familiarity has a direct influence on perceptions of transparency, and semantic judgements therefore cannot be treated as independent. Although this has been brought up before in the literature, it is far from universally acknowledged in idiom research, and has important methodological implications. In addition, variables such as transparency, decomposability and motivation should be much more rigorously defined and operationalised. These judgements show clear differences according to the stage at which they are being made, and may be better considered as interactions between a specific speaker and a given phrase than inherent properties of idioms themselves.

2. Once familiarity is accounted for, native and non-native speakers show similar patterns in how they interpret and identify the meaning of figurative expressions. When expressions are entirely unknown, their relative transparency determines how successfully language users will determine their meaning. Where differences do exist, these are likely due to the greater range of vocabulary and cultural knowledge that native speakers generally have. However, the more analytical approach taken by many language learners means that they may see phrases as inherently more transparent than native speakers.

3. Cross-language influence has a clear effect on both judgments and the ability to identify meaning for L2 speakers. When they encounter an expression in L2 that has the same words and meaning as an idiom in the L1, speakers perceive them to be more familiar (regardless of whether they have encountered the specific expression in the L2 before), and are more likely to successfully identify the meaning. Crucially, they also perceive them as more transparent, just as native speakers perceive more meaning in idioms that they are highly

familiar with. This adds to the growing body of work that implicates L1 knowledge in L2 processing at a multiword level, in both online and offline language tasks.

As well as continuing to develop our understanding of how native speakers and language learners cope with and acquire idioms and figurative meaning more generally, future work could usefully encompass some of the many other aspects raised here (e.g. embodied simulation, cultural knowledge, emotional engagement). Perhaps only by doing so will we be able to fully understand this pervasive and intriguing aspect of language use.

References

Abel, B. (2003). English idioms in the first language and second language lexicon: a dual representation approach. *Second Language Research, 19*(4), 329-358. doi: 10.1191/0267658303sr226oa

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., & Dai, B. (2014). *lme4: Linear mixed-effects models using Eigen and S4*.

Beck, S. D., & Weber, A. (2016). Bilingual and Monolingual Idiom Processing Is Cut from the Same Cloth: The Role of the L1 in Literal and Figurative Meaning Activation. *Frontiers in Psychology, 7*, 1350. doi: 10.3389/fpsyg.2016.01350

Boers, F., & Webb, S. (2015). Gauging the Semantic Transparency of Idioms: Do Natives and Learners See Eye to Eye? In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 368-392). New York: Cambridge University Press.

Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: more than mere mnemonics? *Language Teaching Research, 11*(1), 43-62. doi: 10.1177/1362168806072460

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

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

Caillies, S., & Butcher, K. (2007). Processing of Idiomatic Expressions: Evidence for a New Hybrid View. *Metaphor and Symbol, 22*(1), 79-108.

Cain, K., & Towse, A. S. (2008). To Get Hold of the Wrong End of the Stick: Reasons for Poor Idiom Understanding in Children With Reading Comprehension Difficulties. *Journal of Speech, Language, and Hearing Research, 51*, 1538-1549.

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: an investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology, 102*(3), 280-298. doi: 10.1016/j.jecp.2008.08.001

Cameron, L. (2003). *Metaphor in Educational Discourse*. London: Continuum.

Campbell, S. J., & Raney, G. E. (2016). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behaviour Research Methods, 48*(1), 330-340. doi: 10.3758/s13428-015-0575-2

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784-797. doi: 10.1017/s1366728913000795

Carrol, G., & Conklin, K. (2017). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea "has legs". *Bilingualism: Language and Cognition, 20*(2), 299-317. doi: 10.1017/S1366728915000103

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in Translation. *Studies in Second Language Acquisition, 38*(03), 403-443. doi: 10.1017/s0272263115000492

Charteris-Black, J. (2002). Second Language Figurative Proficiency: A Comparative Study of Malay and English. *Applied Linguistics, 23*(1), 104-133.

Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22*(2), 115-144. doi: 10.1191/0267658306sr263oa

Cieślicka, A. B. (2011). Suppression of literal meanings in L2 idiom processing: Does context help? *Studies in Second Language Learning and Teaching*, *1*(1), 13-36.

Cieślicka, A. B. (2015). Idiom Acquisition and Processing by Second/Foreign Language Learners. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 208-244). New York: Cambridge University Press.

Cieślicka, A. B., Heredia, R. R., & Olivares, M. (2014). It's All in the Eyes: How Language Dominance, Salience, and Context Affect Eye Movements During Idiomatic Language Processing. In M. Pawlak & L. Aronin (Eds.), *Essential Topics in Applied Linguistics and Multilingualism* (pp. 21-41): Springer International Publishing.

Citron, F. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods, 48*(1), 91-111. doi: 10.3758/s13428-015-0581-4

Cutting, J., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition, 25*(1), 57-71.

Desai, R., Conant, L., Binder, J., Park, H., & Seidenberg, M. (2013). A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *Neuroimage, 83*, 862-869.

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology, 22*(30), 321-334.

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

Gibbs, R. W. (1987). Linguistic factors in children' s understanding of idioms. *Journal of Child Language, 14*, 569-586.

Gibbs, R. W. (1991). Semantic Analyzability in children's understanding of idioms. *Journal of Speech and Hearing Research, 34*, 613-620.

Gibbs, J. R. W. (1993). Why idioms are not dead metaphors. In C. Cacciari & P. Tabossi (Eds.), *Idioms. Processing, Structure, and Interpretation* (pp. 57-78). New Jersey: Erlbaum Associates.

Gibbs, J. R. W., & Colston, H. (2012). *Interpreting Figurative Meaning*. New York: Cambridge University Press.

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., & Cutting, C. (1989). How to Kick the Bucket and Not Decompose: Analyzability and Idiom Processing. *Journal of Memory and Language, 28*, 576-593.

Gibbs, J. R. W., Bogdanovich, J., Sykes, J., & Barr, D. (1997). Metaphor in Idiom Comprehension. *Journal of Memory and Language, 37*, 141-154.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8*(3), 183-206.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Science, 7*(5), 219-224. doi: 10.1016/s1364-6613(03)00080-9

Grant, L., & Bauer, L. (2004). Criteria for Re-defining Idioms: Are we Barking up the Wrong Tree? *Applied Linguistics, 25*(1), 38-61.

Hamblin, J., & Gibbs, J. R. W. (1999). Why You Can't Kick the Bucket as You Slowly Die: Verbs in Idiom Comprehension. *Journal of Psycholinguistic Research, 28*(1), 25-39.
Holsinger, E. (2013). Representing Idioms: Syntactic and Contextual Effects on Idiom Processing. *Language and Speech, 56*(3), 373-394. doi: 10.1177/0023830913484899

Holsinger, E., & Kaiser, E. (2013). Processing (Non)Compositional Expressions: Mistakes and Recovery. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*(3), 866-878. doi: 10.1037/a0030410

Irujo, S. (1986). A piece of cake: learning and teaching idioms. *ELT Journal, 40*(3), 236-242.

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 Literary and 260 Nonliterary Metaphors on 10 Psychological Dimensions. *Metaphor and Symbolic Activity, 3*(4), 191-214. doi: 10.1207/s15327868ms0304_1

Keysar, B., & Bly, B. (1995). Intuitions of the Transparency of Idioms: Can One Keep a Secret by Spilling the Beans? *Journal of Memory and Language, 34*, 89-109.

Keysar, B., & Bly, B. (1999). Swimming against the current: Do idioms reflect conceptual structure? *Journal of Pragmatics, 31*, 1559-1578.

Kövecses, Z. (2005). *Metaphor in culture: Universality and Variation*. Cambridge: Cambridge University Press.

Kecskes, I. (2015). Is the Idiom Principle Blocked in Bilingual L2 Production? In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 28-52). New York: Cambridge University Press.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2014). *lmerTest: Tests in Linear Mixed Effects Models*.

Laufer, B. (2000). Avoidance of Idioms in Second Language: The Effect of L1-L2 Degree of Similarity. *Studia Linguistica, 54*(2), 186-196.

Levorato, M. C., & Cacciari, C. (1992). Children's comprehension and production of idioms. *Journal of Child Language, 19*, 415-433.

Levorato, M. C., & Cacciari, C. (1995). The Effects of Different Tasks on the Comprehension and Production of Idioms in Children. *Journal of Experimental Child Psychology, 60*, 261-283.

Levorato, M. C., & Cacciari, C. (1999). Idiom Comprehension in Children: Are the Effects of Semantic Analysability and Context Separable? *European Journal of Cognitive Psychology, 11*(1), 51-66. doi: 10.1080/713752299

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

Lin, M., & Leonard, S. (2012). *Dictionary of 1000 Chinese Idioms (Revised Edition)*. New York: Hippocrene Books.

Liontas, J. (2002). Context and idiom understanding in second languages. In S. Foster-Cohen, T. Ruthenberg, & M.-L. Poschen (Eds.), *EUROSLA Yearbook: Annual conference of the European second language association* (Vol. 2, pp. 155-185). Amsterdam/Philadelphia: John Benjamin Publishing Company.

Liontas, J. (2015). Straight from the Horse's Mouth: Idiomaticity Revisited. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp.301-340). New York: Cambridge University Press.

Littlemore, J. (2004). Interpreting metaphors in the language classroom. *Les Cahiers de l'APLIUT, 23*(2), 57-70.

Littlemore, J. (2010). Metaphoric competence in the first and second language: similarities and differences. In M. Putz & L. Sicola (Eds.), *Cognitive Processing in Second Language Acquisition* (pp. 293-316). Amsterdam: John Benjamins.

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

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology, 83*(3), 176-190. doi: 10.1016/j.biopsycho.2009.12.006

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Nayak, N. P., & Gibbs, J. R. W. (1990). Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General, 119*, 315-330.

Nippold, M. A., & Rudzinski, M. (1993). Familiarity and Transparency in Idiom Explanation: A Developmental Study of Children and Adolescents. *Journal of Speech and Hearing Research, 36*, 728-737.

Nippold, M. A., & Taylor, C. L. (1995). Idiom understanding in youth: Further examination of familiarity and transparency. *Journal of Speech and Hearing Research, 38*, 426-433.

Nordmann, E., & Jambazova, A. (2016). Normative data for idiomatic expressions. *Behaviour Research Methods*. doi: 10.3758/s13428-016-0705-5

Nordmann, E., Cleland, A. A., & Bull, R. (2014). Familiarity breeds dissent: Reliability analyses for British-English idioms on measures of familiarity, meaning, literality, and decomposability. *Acta Psychologica, 149*, 87-95. doi: 10.1016/j.actpsy.2014.03.009

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

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, part 2), 1-25.

Pollio, M., & Pollio, H. (1974). The development of figurative language in children. *Journal of Psycholinguistic Research, 3*, 185-201.

Pritchett, L. K., Vaid, J., Tosun, S., & Walla, P. (2016). Of black sheep and white crows: Extending the bilingual dual coding theory to memory for idioms. *Cogent Psychology, 3*(1). doi: 10.1080/23311908.2015.1135512

Reuterskiöld, C., & Van Lancker Sidtis, D. (2012). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy, 29*(2), 219-231. doi: 10.1177/0265659012456859

Richardson, J. (1976). Imageability and concreteness. *Bulletin of the Psychonomic Society, 7*(5), 429-431.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience, 25*(5), 762-776. doi: 10.1162/jocn_a_00337

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64*(4), 913-951. doi: 10.1111/lang.12077

Schweigert, W. (1986). The comprehension of familiar and less familiar idioms. *Journal of Psycholinguistic Research, 15*, 33-45.

Schweigert, W. (1991). The Muddy Waters of Idiom Comprehension. *Journal of Psycholinguistic Research, 20*(4), 305-314.

Schweigert, W. A. (2009). The effects of multiple presentations on the ratings and memorability of novel figurative phrases. *J Psycholinguist Res, 38*(1), 25-42. doi: 10.1007/s10936-008-9080-4

Schweigert, W., & Moates, D. (1988). Familiar Idiom Comprehension. *Journal of Psycholinguistic Research, 17*(4), 281-296.

Schweigert, W., Cintron, J., Sullivan, K., Ilic, E., Ellis, S., Dobrowits, C., & Roberts, C. (2003). Novel Figurative Phrases and Idioms: Phrase Characteristics over Multiple Presentations. *Journal of Psycholinguistic Research, 32*(4), 455-475.

Siyanova, A. (2010). *On-line Processing of Multi-word Sequences in a First and Second Language: Evidence from Eye-tracking and ERP.* Unpublished PhD thesis, University of Nottingham.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251-272. doi: 10.1177/0267658310382068

Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing Verbs in German Idioms: Evidence Against the Configuration Hypothesis. *Metaphor and Symbol, 22*(3), 213-231.

Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*(2), 161-184. doi: 10.1016/j.jml.2005.11.001

Steinel, M., Hulstijn, J., & Steinel, W. (2007). Second Language Idiom Learning in a Paired-associate Paradigm. *Studies in Second Language Acquisition, 29*, 449-484.

Swinney, D., & 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. (2009). Why are idioms recognized fast? *Memory and Cognition, 37*(4), 529-540. doi: 10.3758/MC.37.4.529

Titone, D., & Connine, C. (1994). Comprehension of Idiomatic Expressions: Effects of Predictability and Literality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(5), 1126-1138.

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

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 171-207). New York: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

van Ginkel, W., Hubers, F., Cucchiarini, C., Dijkstra, T., & Strik, H. (2016). *Norming Studies for Idiom Processing: Native and Non-Native Benchmarks*. Paper presented at FLaRN 2016, Vilnius, Lithuania.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive Mechanisms in Idiom Comprehension. *Journal of Cognitive Neuroscience, 22*(8), 1682-1700.

Vulchanova, M., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: A developmental study. *Vigo International Journal of Applied Linguistics, 8*, 207-234.

Wolter, B., & Gyllstad, H. (2011). Collocational Links in the L2 Mental Lexicon and the Influence of L1 Intralexical Knowledge. *Applied Linguistics, 32*(4), 430-449.

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition, 35*(3), 451-482.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. (2009). Formulaic Sequences and Language Disorder. In M. Ball, M. Perkins, N. Müller, & S. Howard (Eds.), *The Handbook of Clinical Linguistics* (pp. 184-197). Oxford: Blackwell.

Wray, A., Bell, H., & Jones, K. (2016). How native and non-native speakers of English interpret unfamiliar formulaic sequences. *European Journal of English Studies, 20*(1), 47-63. doi: 10.1080/13825577.2015.1136163

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly, 44*(4), 647-668.

Zhang, H., Yang, Y., Gu, J., & Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *Journal of Neurolinguistics, 26*(1), 89-112. doi: 10.1016/j.jneuroling.2012.05.002

Appendix: stimuli used in the rating study

Phrase	Type	Meaning
A blackmailer is a leech	Metaphor	Blackmailers suck the life out of you
Alcohol is a crutch	Metaphor	Alcohol becomes something people rely on
A good lover is a teddy bear	Metaphor	A partner is comforting
An accountant is a juggler	Metaphor	Accountants are good at balancing many things
A museum is a history book	Metaphor	Museums tell us about the past
Humor is a medicine	Metaphor	Humour can be healing
A star is a signpost	Metaphor	Stars can show people the way
Danger is a spice	Metaphor	Danger can be exciting and enticing
Evolution is a lottery	Metaphor	Evolution is random
Time is a doctor	Metaphor	Time heals wounds and problems
Discipline is a fertilizer	Metaphor	Being disciplined helps you to develop
History is a mirror	Metaphor	History can reflect how things are now
Anger is a storm	Metaphor	Anger can be wild and unpredictable
A degree is a doorway	Metaphor	Qualifications lead to better opportunities
Education is a lantern	Metaphor	Education helps us to see things better
A mind is a sponge	Metaphor	The mind can absorb lots of information
Hard work is a ladder	Metaphor	Hard work helps you rise to the top
Money is a lubricant	Metaphor	Money helps to get things done
A smile is a knife	Metaphor	A smile can be deceptive and hurtful
A friend is a ray of sunshine	Metaphor	Friends can brighten up your day
Books are treasure chests	Metaphor	Books contain precious information
A rumour is a plague	Metaphor	Rumours spread and can be destructive
He's a big stick	Bulgarian	He's a person with high social status
To give the word to someone	Bulgarian	To invite someone to speak
He/She's missed his/her first 7 years	Bulgarian	He/she lacks upbringing
A duck drank my mind	Bulgarian	I feel silly
It came out salty	Bulgarian	I'm paying higher than the normal price
He/ she is naked water	Bulgarian	He/she is incompetent and unqualified
I live five for four	Bulgarian	I live recklessly
I stepped on the lion's tail	Bulgarian	I had an argument with someone dangerous
I'm gathering my hammers	Bulgarian	I'm getting ready to leave

To call a deer a horse	Chinese	To deliberately misrepresent things
Cover your ears to steal a bell	Chinese	To not be honest with yourself
Draw a snake and add feet	Chinese	Ruin something by fiddling too much
Without shirt or shoes	Chinese	Sloppy and untidy
White clouds change into grey dogs	Chinese	Life changes in unpredictable ways
To connect two and three	Chinese	One thing happening after another
Wine and meat friends	Chinese	Friends who are only there when it suits them
Chase the wind and grasp at shadows	Chinese	To make groundless accusations
Beat the grass to scare the snake	Chinese	Act rashly and alert an enemy
Bring sticks to put out a fire	Chinese	Make a situation worse
Eyes bright like torches	Chinese	Focused and alert
One gun and a horse	Chinese	All by yourself
Different mouths but one sound	Chinese	Many people saying the same thing
Seven hands and eight feet	Chinese	Too many people making things difficult
Kill the chicken to scare the monkey	Chinese	Put on a show of strength
To add oil and vinegar	Chinese	To exaggerate a story
Chicken feathers and garlic skins	Chinese	Unimportant things
Trick the sky to cross the sea	Chinese	To cheat someone
Three long and two short	Chinese	Disasters that were not expected.
Neither three nor four	Chinese	Someone you can't really trust
They won't share the same sky	Chinese	They really hate each other
A horse doesn't stop its hooves	Chinese	To continue non-stop
To tighten your belt	English	To stop spending so much money
To pull someone's leg	English	To play a joke on someone
To twist someone's arm	English	To persuade someone to do something
To be under the weather	English	To be feeling unwell
To drop the ball	English	To let other down by making a mistake
To be over the moon	English	To be really happy
Spill the beans	English	To reveal a secret
To let off steam	English	To release stress
To smell a rat	English	To be suspicious
That rings a bell	English	That reminds me of something
To hit the roof	English	To get very angry
To break the ice	English	To make the first move in a social situation

To pop the question	English	To ask someone to marry you
To change your tune	English	To start acting or thinking differently
Find your feet	English	Get used to something new
Steal the show	English	Be more impressive than everyone else
To draw a blank	English	To not be able to remember something
Walking on air	English	Feeling very happy
To jump the gun	English	To start doing something too quickly
To play with fire	English	To do something risky
Eat your words	English	Have to admit you were wrong
To bite your tongue	English	To stop yourself from speaking
That's cold coffee	German	To be already well-known news
To leave the church in the village	German	To not exaggerate about something
To pour someone clear wine	German	To reveal the truth
To hand over the spoon	German	To die
To bring someone around the corner	German	To kill someone
He can spoon the soup out	German	He can solve a problem himself
To bite into the sour apple	German	To do something you know will be unpleasant
Buy something for an apple and an egg	German	To buy something for a cheap price
To have hair on your teeth	German	To be easily annoyed
Step into the grease bowl	German	To embarrass oneself unintentionally
To give someone a basket	German	To reject someone who shows an interest in you
To earn a pig's money	German	To earn a lot of money
To talk into the blue	German	To talk without really knowing what you want to say