university of groningen / faculty of science and engineering

Idiom Decomposability: there is more to it than a single pan-herring

Bachelor's Project Thesis

Ties Lind, s2864495, t.lind.1@student.rug.nl
Supervisors: S.K. Mehlhorn, PhD & dr. J.C. van Rij-Tange

Abstract: Idiom decomposability relates to how individual words of an idiom contribute to the overall figurative meaning. For example, the idiom to *button your lips* is considered decomposable, since the meaning can be worked out from the individual words. Idioms like *to kick the bucket*, meaning "to die" are nondecomposable. Various studies have tried to develop a metric for idiom decomposability, such as a rating, by asking native speakers to judge the decomposability of idioms. Yet recent literature has shown that being familiar with an idiom influences how one might decompose it. For this reason the decomposability ratings made by native speakers may be biased. A multiple choice questionnaire was designed to test Dutch idioms on English speakers (as the idioms should be unfamiliar), to see whether there is a relationship between the decomposability rating and identifying the correct meaning. As stated in the paper by Sprenger et al. (2019), if an idiom is highly decomposable, knowledge about its individual words may help a language learner to deduce the meaning. 57 monolingual native English speakers, and 37 multilingual native English speakers were tested on whether they could identify the correct meaning of 50 Dutch idioms in a multiple choice questionnaire. If the decomposability ratings from native speakers indeed reflect how well the individual words of the idioms relate to the figurative meaning, then we should expect that idioms with higher decomposability ratings should have a higher accuracy, since the meaning should be easier to work out. The results show that in general, the decomposability rating of the idiom affects the accuracy. This result suggests that there is a relationship between the decomposability ratings and how easy the meanings for the idioms are able to be identified. As the relationship shows to be positive, we can conclude that the ratings made by the Dutch natives do reflect (at least partly) the decomposability of the Dutch idioms. However, the fact that some idioms do not follow this trend shows that there is an ongoing challenge to create a reliable metric for idiom decomposability.

1 Introduction

Idioms are multi-word expressions that often have a meaning that is different than what the individual words literally suggest. As an example, depending on the context, to *break the ice* usually means to relieve social tension as opposed to crushing frozen water. Expressions like *bury the hatchet*, *piece of cake* and *break a leg* are very common in everyday language, yet they defy the standard compositional view of language comprehension and production. They are considered an important part of *nativelike* language proficiency, and native speakers should "speak idiomatically unless there is good reason not to do so" (Searle, 1975).

In the English language alone there are estimated to be hundreds of thousands of morphologically complex lexical items, such as idioms and epithets (e.g. *Daddy Longlegs*), based on conversational talk analysis by Pawley and Syder (1983). Idioms are an interesting part of linguistic research because knowledge about how they are acquired and processed is limited compared to normal words and sentences that do not have figurative meanings. Additionally, experimenting is a challenge because of the variability in speakers' familiarity and knowledge of these phrases (Sprenger et al., 2019).

Traditionally, all idioms have been regarded as

noncompositional, where the intended figurative meaning of the phrase is not derived from the meanings of the individual words (Swinney and Cutler, 1979). For example, the meaning for *to kick the bucket* cannot be determined through analysing the individual word meanings, to result in the figurative meaning "to die". This view suggests that idioms must be analyzed as an entire unit, and their figurative meanings are directly stipulated in the mental lexicon, much like the meanings of individual words (Fraser, 1970).

However, this noncompositional view cannot explain why some idioms are understood or acquired more easily than others. Later research in psycholinguistics has shown that idioms do not form a homogeneous group, and that many idioms contain words which contribute to the overall figurative meaning of the expression. For example, for the idiom to *button your lips*, *button* refers to "close" and *lips* refers to "mouth", so the meaning "close your mouth" can be worked out. In contrast, working out the meaning for *to kick the bucket* is not so straightforward. According to this notion, Gibbs Jr et al. (1989) identified three types of idioms: decomposable idioms (e.g. *to button your lips*, *clear the air*), abnormally decomposable idioms (e.g. *spill the beans*) and nondecomposable idioms such as *to kick the bucket*. A decomposable idiom is therefore one where the individual words contribute to the idiom's figurative meaning. For a nondecomposable idiom, the meaning cannot be obtained by analysing the individual words. Abnormally decomposable idioms, according to Gibbs Jr et al. (1989), are a subgroup of decomposable idioms, but the individual words have a more metaphoric relation to the idiom's figurative meaning.

This more contemporary model, referred to as the Idiom Decomposition Hypothesis (IDH), suggests that idioms are represented and processed differently depending on their degree of decomposability. According to Gibbs Jr et al. (1989), decomposable idioms are processed as inferences: each individual component of the phrase is retrieved from the mental lexicon and combined with the other components of the phrase according to their syntactic relations, as opposed to remembering the figurative meaning of the phrase as a whole, as shown with nondecomposable idioms. In other words, the concept of idiom decomposability is important because it challenges the traditional view of how idioms are represented in the mental lexicon. Instead of all idioms being regarded as long words where the figurative meaning must be learned and remembered, idiom decomposability may affect the ease with which a specific item is acquired and understood, and could explain why, for example, Cain et al. (2005) found that children with poor reading comprehension skills were significantly worse at working out the meaning of nondecomposable idioms, compared to decomposable idioms. Thus, if we assume the IDH, we would expect that speakers who are unfamiliar with idioms should work out the meaning of decomposable idioms more easily than nondecomposable idioms, since each separate word can be analyzed semantically. In order to test this idea, idioms should be categorized based on their degree of decomposability.

In an endeavour to categorize idioms based on their decomposability, numerous studies (e.g. Gibbs Jr et al., 1989; Gibbs et al., 1989; Hamblin and Gibbs, 1999; Titone and Connine, 1994) have asked participants to make categorical judgements regarding the decomposabiltiy of idioms, yet this method has shown low levels of agreement between studies, even showing contrasting results. For example, in the studies by Gibbs and colleagues participants were asked to categorize idioms as either normally decomposable, abnormally decomposable, or nondecomposable. A normally decomposable idiom is one where the individual components of a phrase are connected to the overall figurative meaning, e.g. *put on some weight*, *lay down the law*; the words *weight* and *law* refer to their literal definitions. Abnormally decomposable idioms are ones where the component words are related to the figurative meaning in a more metaphoric way; e.g. *break the ice* and *spill the beans*, where *ice* refers to tension, and *beans* refers to secrets or gossip. Nondecomposable idioms, as described previously, are phrases where the component parts have no relation to the figurative meaning, e.g. *cook his goose*, *chew the fat*. One of the several instances of the disagreement between studies is shown with the idiom *break the ice*, which is judged as normally decomposable in Gibbs Jr and Nayak (1989) and Gibbs et al. (1989), while Gibbs Jr et al. (1989)(b) regard it as abnormally decomposable, yet Hamblin and Gibbs (1999) view it as nondecomposable. The study by Gibbs Jr et al. (1989) has further been critiqued by Abel (2003) since the participants had

to judge the decomposability of a list idioms, only choosing which of the three categories they belong to. Yet these idioms were preselected by the authors based on these three categories, implying a balanced distribution of English idioms between these categories.

Consequently, attempting to improve idiom categorization, the use of 5 or 7-point Likert scales were used in following studies such as by Tabossi et al. (2008) and Sprenger et al. (2019). However, the use of these scales to rate decomposability does not increase agreement between participants (Nordmann et al., 2014). Moreover, although it is argued that such scales may be able to capture the nuances of idiom decomposability better than categorisation, the inconsistencies of how decomposability is measured between studies is still problematic.

An example of an experiment that has used a Likert scale to quantify the decomposability of idioms is in a study by Sprenger et al. (2019). In this experiment, monolingual Dutch natives were asked to rate the decomposability of Dutch idioms on a scale of 1-5 where 1 meant "no relation between the individual words and the figurative meaning", and 5 meant "strong relation between the individual words and the figurative meaning". The paper subsequently concludes that the probability with which an idiom is acquired is affected by the degree to which it is decomposable. However, the ratings made by the Dutch natives could have been biased due to, for example, how familiar they are with the idioms.

Although the degree of decomposability has been common way of dividing and categorizing idioms, its psychological relevance has received very limited empirical support (Tabossi et al., 2009), and the reliability of this measure is questioned because of the many other factors that influence idiom processing, such as familiarity and imageability, as evidenced by numerous experimental papers which will be discussed below. Familiarity is defined as the frequency with which a speaker encounters an idiomatic expression, often referred to as subjective frequency (Titone and Connine, 1994). Imageability is defined as the extent to which an idiom can be associated with a specific image (Paivio et al., 1966).

Data on the properties of decomposability, imageability, and familiarity have usually been gathered through studies involving subjective judgements, however, inconsistent definitions and varying operationalization of these properties could influence the reliability of the judgements, making it difficult to compare the results of different studies (Hubers et al., 2019). For example, decomposability is a relatively abstract characteristic, and it is possible that each participant interprets this concept differently, assigning contrasting semantic weight to the individual words. Thus, if participants are asked to judge the decomposability of an idiom, their judgement might be influenced by their interpretation of the characteristic, as well as how familiar they are with the idiom and the mental image it produces. Knowing this, it could therefore be argued that decomposability, as an idiomatic variable, is problematic as it is likely that individual subject differences guide results, as opposed to the lexical representational differences of the idioms (Nordmann et al., 2014).

Nordmann et al. (2014) further state that native speakers are unable to inhibit their knowledge of the meaning of an idiom when making decomposability judgements, and thus if a participant is highly familiar with the figurative meaning of an idiom, their perception of the phrase will be that it is semantically acceptable, and so their perception of the phrase is biased. A similar argument is made by Keysar and Bly (1995), suggesting that once the meaning of an idiom is learned, its transparency increases. In this case, a transparent idiom is one where the connection between the expression and its idiomatic meaning makes sense to native speakers. For example, knowing that a *carrot and stick approach* refers to a donkey being rewarded with a carrot and punished with a stick, the idiom becomes much more intuitive, and this can have an effect on the decomposability judgement. In other words, familiarity makes it seem like there is meaning where there is none: the more familiar a speaker is with an idiom and its figurative meaning, the harder it is to suppress this knowledge when judging decomposability. Thus, although decomposability is a relevant concept for psycholinguistic research, the way that it is measured may not be reliable because once we learn the meaning of an idiom, we lose the ability to judge how decomposable it is.

In this paper we investigate whether the native speakers' ratings (gathered by Sprenger et al. (2019) may be biased by their knowledge and fa-

miliarity of the idioms. By using Dutch idioms on English speakers, we can test if there is a relationship between decomposability rating and how well these idioms are understood. In order to look into this relationship, a new way of testing is proposed where Dutch idioms are translated into English and shown to English speakers along with 3 possible answers (one of which is correct), to see if the individual word meanings influence choosing the correct answer. As stated in the paper by Sprenger et al. (2019), if an idiom is highly decomposable, knowledge about its individual words may help a language learner to deduce the meaning. Thus, under this assumption, the pattern we should expect is that idioms with higher decomposability ratings should be easier to answer, since the meaning should be easier to work out.

2 Methods

Following from the literature presented in the previous sections, there is a discussion about the validity of the idiom decomposability metric, as decomposability judgements may be biased by familiarity of the idiom, as well as other possible factors such as imageability and idiom transparency.

In order to test whether there is a relationship between the degree of decomposability and idiom understanding by English speakers, Dutch idioms were translated into English and tested on English natives. Each translated idiom had three answer options, and the participants could select which of the three answers fits the idiom best.

2.1 Participants

The questionnaire described in the following sections was sent via social media (Facebook and Whatsapp) to participants in the personal network of the author. The participants received no financial reward, their participation was voluntary and they could end the questionnaire at any moment. The responses of 89 participants were used for the analysis out of 173 participants responding to the questionnaire. The other responses from participants were removed because: they were incomplete (54); the participant lived in the Netherlands and/or spoke Dutch (6); the participant was a non-native English speaker (English as a *second* language) (24). The participants included for the results were: monolingual native English speakers (Group 1, N=52), and multilingual native English speakers (i.e. English as their first language, but regularly speak, write, or read in another language; Group 2, N=37). The reason the participants are split into these groups is because the experiment was intended for monolingual English natives, yet having extra data from multilingual English native speakers shows interesting results. The age range of the participants for Group 1 was 16-72 (mean 42.96; 26 male), for Group 2: 21-71 (mean 41.59; 19 male).

2.2 Materials

2.2.1 Idiom selection

The idioms used in the experiment were from an idiom database by Sprenger et al. (2019), which contains 189 Dutch idioms, including their meaning, literal English transcription, familiarity ratings, and decomposability ratings. The decomposability rating for each of these idioms were gathered by asking 34 Dutch monolingual participants (age 21-26, 8 men) to read the idioms and judge to what extent the meaning of the individual words were related to the figurative meaning of the whole expression. The participants had to rate these idioms on a scale from 1 to 5. The finalized decomposability score per idiom was not the average rating made by the participants, in order to "avoid potential subject bias influencing the decomposability scores for the idioms with a low number of ratings" (Sprenger et al., 2019). Instead, an estimated decomposability score was made for each idiom by fitting a generalized additive mixed model with random effects for the idioms and participants to take into account the response bias and variation between idioms.

Out of these 189 idioms, 50 idioms were chosen for this experiment, split into roughly 10 idioms per interval of decomposability score (i.e. 10 idioms with a rating from 1.0-1.9, 10 idioms with a rating from 2.0-2.9 etc.), however a perfect range of scores from 1.0 - 5.0 was not possible since there were no idioms with a rating of 1.0, or 5.0. Out of the 50 chosen Dutch idioms, the item with the lowest score (1.312) was for the idiom *He had hairs on the teeth* (meaning to dare to speak out), and the item with the highest decomposability score (4.427) was for the idiom *He sat with his nose in the books* (mean-

ing to study a lot). Included in the 50 idioms were 10 idioms that translated into existing English idioms, with a similar range in decomposability rating. These idioms were used as control items to validate the experiment, which will be explained in the sections below. To clarify, 40/50 idioms were Dutch idioms. 10/50 idioms were existing English idioms, such as *He played devil's advocate*. The full list of idioms, including control items can be seen in the table in the appendix.

2.2.2 Translation of idioms

The 50 idioms were translated by the author, who is bilingual in Dutch and English. Furthermore, the translations were checked for correctness by two Dutch native speakers who also speak English.

2.2.3 Checking whether the Dutch idioms were unfamiliar

The 40 Dutch idioms were shown to two English monolingual natives, to validate if any other idioms in the list were already known, as this could influence the experiment. It is important that the Dutch idioms for the experiment were unfamiliar. The two English natives, both university students, were given the list of translated idioms and asked to write what they thought each idiom meant, for as many idioms as they could. The participants wrote their interpretations for several idioms, which are shown the table in the appendix.

2.2.4 Multiple choice answers

In order to test whether the English participants are able to work out the meaning of the Dutch idioms, each idiom was shown along with 3 possible meanings (answers options). These answer options include one correct meaning, and two *decoy* options such that the participant must choose which of the three options *fits the idiom best*. Thus, for each idiom two decoy options were created. There were several guidelines when deciding on the decoy options:

- All 3 options must be of a similar structure, tense and length.
- Depending on the idiom and its figurative meaning: out of the three options, one option must relate to the literal word meanings, identified as a decomposable answer. One must be a nondecomposable option, where the meaning cannot be inferred from the individual word meanings. The third decoy answer option, referred to as an "abnormal" option was added so that the questionnaire did not only have two options to choose from. This answer option was intended to be more metaphoric than the decomposable answer, yet not completely abstract, similar to the IDH category *Abnormally Decomposable* defined by Gibbs Jr et al. (1989).

For a nondecomposable answer option, the answer has to be different to the individual word meanings, simulating the effect of the idiom being nondecomposable. Similarly, the decomposable answer must relate to the individual word meanings of the idiom. The abnormal answer option was included to reduce the probability of guessing the correct answer, as well as adding an extra metaphoric interpretation of the idiom in question.

The correct answers were used according to the original database by Sprenger et al. (2019), and checked by the author for correctness. Some examples are shown in table 2.1, which show the varying answer options, where the correct answers are highlighted in green.

Table 2.1: Examples of 3 answer options, correct answer highlighted in green

Est	Idiom	Decomposable answer	Abnormal answer	Nondecomposable answer
1.65	Then he fell with his nose in the butter.	To make a mess	To smell something peculiar	To be in luck
3.173	Then he got the beard in the throat.	He had a cough	His voice has deepened	He made an indecent remark
3.942	Then he went on the fist.	He had a fight	He held on tightly	He lost all of his money

As an example from table 2.1, for the idiom *Then he fell with his nose in the butter*, the options created were:

1. To make a mess
2. To smell something peculiar
3. To be in luck

The correct answer is the last option. The other two options are of a similar syntactic structure, but relate more literally to the individual words.

The challenging part of creating the two decoy answers was to make them suitable enough so that they did not give the real answer away, yet the answers could not be too extravagant so that they stand out, likewise stated in the Encyclopedia of Clinical Psychology: "excessively attractive wrong answers on a multiple choice test may seduce easily swayed examinees away from a known correct answer". (Cautin and Lilienfeld, 2015). This was especially difficult for the idioms with a high decomposability rating, since the individual words were often very obvious, such as for the Dutch idiom *He chatted the ears off his head*, the verb *chattered* makes it difficult to come up with alternative interpretations, however, this is of course the essential nature of highly decomposable idioms.

2.2.5 Control idioms

As stated previously, 10 of the 50 idioms translated into existing English idioms, for example, *Toen sloeg hij de spijker op de kop*, which translates to *He hit the nail on its head* (meaning to arrive at exactly the right answer) which were used as control items. In the questionnaire, the participants were shown the idioms along with 3 answer choices, as well as an "I know this one" option. If the participant recognized or knew the idiom, they could then select the button "I know this one" (see figure 2.1).

The purpose of the control idioms was to:

- validate whether the Dutch idioms were truly unfamiliar i.e. only the control items should be selected as "I know this one"
- validate whether the English natives were able to choose the correct interpretation for the control items.

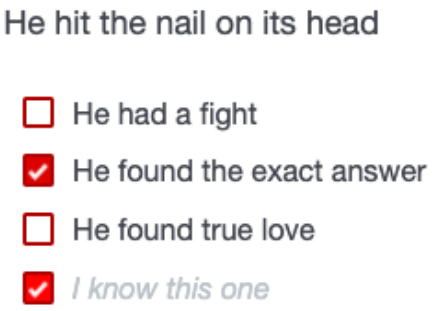

Figure 2.1: Example question from the questionnaire

If the participants choose the wrong answer for the idioms selected as "I know this one", then it shows that the data are unreliable. In other words, if the English natives choose the correct answer for the control idioms, then this show that they understand how to interpret familiar idioms in this questionnaire.

2.2.6 Anchor questions

We decided to add three anchor questions, which were presented at the start of the experiment. These three questions were in the same style as the questions to be tested, preparing the participants for the following 50 idioms, so that they would know what was expected. The three questions consisted of one metaphoric idiom (i.e. abnormally decomposable), one decomposable, English idiom (to prepare the participant that they might know the answer for some), and one nondecomposable idiom, to show that some meanings are hard to work out. These anchor questions were not used for analysis.

2.3 Design

The questionnaire was implemented using the survey software Qualtrics (Qualtrics, Provo, UT), where the participants could respond online on their smartphone, tablet or laptop. The participants could anonymously access the questionnaire via a link. This link* was sent to participants in

*https://rug.eu.qualtrics.com/jfe/form/SV_acaW9YwtIon3GPb

England and the UAE using social media (Facebook, Whatsapp).

2.4 Procedure

At the start of the questionnaire, the participants were informed about the nature of idioms and figurative meanings, and the purpose of the questionnaire: we want to investigate whether some Dutch idioms are easier to understand than others, for native English speakers.

The participants were asked to read the idioms and choose which of the three answers is the *most likely meaning of the idiom*. Furthermore, the participants were asked to select "I know this one" if they were familiar with the idiom (see figure 2.1).

After the introduction of the experiment, questions followed about the participants' age, gender, country of residence, what their native languages are, and what other languages they "regularly speak, write, or read" in. Next, the three anchor questions were shown to prepare the participants for the kind of idioms and answers to expect, followed by the 50 idioms to be tested. Each participant saw the same 50 idioms and anchor questions, but the order that the idioms were presented in was randomized. The answer options were also randomized to eliminate order bias. 5 idioms were shown per page.

3 Results

The data from the questionnaire shows the answer chosen per idiom, and whether the idiom was familiar, for each participant. The data for each participant also included their age, gender, native language(s), other language(s), and country of residence.

The questionnaire was designed to explore whether there is a relationship between the idiom decomposability rating of unfamiliar idioms and whether native English speakers could work out the figurative meaning. We assume that a participant choosing the correct answer means the participant could work out the figurative meaning. The responses from 52 monolingual native English speakers were used, referred to as Group 1. The responses from 37 multilingual native English speaking participants were used for comparison, referred to as Group 2. The participants in Group 2 speak English as their first language, and also speak one or more languages other than English.

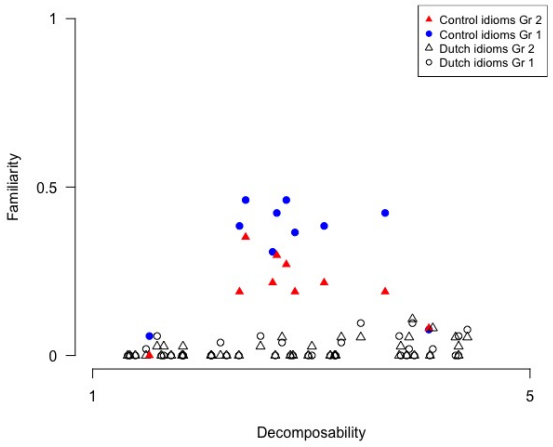

Figure 3.1: Familiarity of idioms. Calculated as number of "I know this one" selections per groups of participants

3.1 Are the Dutch idioms unfamiliar?

The first question is to find out whether the Dutch idioms in the questionnaire were truly unfamiliar for the native English speakers. By using data from the "I know this one" option, we can check which idioms were familiar to the participants. Figure 3.1 shows the familiarity on the y-axis and the decomposability rating on the x-axis. The familiarity is calculated as the number of "I know this one" selections for an idiom, per number of participants. This was calculated separately per group.

The blue points show the familiarity for the 10 control idioms for group 1 and the red triangles show the 10 control idioms for group 2. The black circles show the 40 Dutch idioms for group 1 and the black triangles for group 2. The graph shows that the Dutch idioms were mostly unfamiliar with some exceptions, where some Dutch idioms were selected as "I know this one" by 5 participants at most. Furthermore, for both groups, 2 of the 10 control idioms were selected as "I know this one" less frequently than the other 8 control idioms.

For group 1, 8/10 control idioms had a mean familiarity of 0.401 (SD=0.0518), and for group 2 this was lower with a mean familiarity of 0.240 (SD=0.0603). As an example from the data, the control idiom *He rubbed salt in the wound* was selected as "I know this one" by 22/52 of the monolingual English speakers, so the familiarity is 0.423 for group 1. The Dutch translated idiom *He had a hole in his hand* was selected as "I know this one" by 2 participants of group 1. The reason why these participants might have selected "I know this one" for this Dutch idiom is because it is similar to the English idiom *money burns a hole in your pocket*. Group 2 shows similar results to Group 1 where, in general, the 8/10 control idioms (in red) are selected as "I know this one" more frequently than the Dutch idioms. However, when the idioms were selected as "I know this one", it is unclear whether this is because the participants really knew the Dutch idiom, or because the idiom was similar to an English idiom, or because the participant selected this option unintentionally.

In order to assess whether selecting "I know this one" meant the participant was indeed familiar with the idiom, we can use the mean accuracy of each idiom. As previously mentioned, a participant choosing the correct answer out of the three options meant a score of 1, and a wrong answer meant a score of 0, per response. When we take the total score of each idiom over the number of participants we have the mean accuracy. For example, for Group 1, 51/52 participants chose the correct answer for the control idiom *He hit the nail on its head*, so the accuracy for this idiom is 0.981. 24/52 participants selected "I know this one" for this idiom, so the familiarity is 0.462. When we plot these means for all idioms, per group of participants, we can see the results in figure 3.3. The circles are the 50 idioms for group 1, where the blue circles are the 10 control items. The triangles are the data for group 2, where the 10 red triangles are the control items. The y-axis is the accuracy, and the x-axis is the frequency of "I know this one" selections for the idiom over the number of participants, referred to as familiarity. This plot shows that although the control items were selected as "I know this one" by less than half of the participants, due to their overall high accuracy, it can be argued that they are actually *familiar* idioms. Interestingly, group 2 selected "I know this one" for fewer control items than group 1.

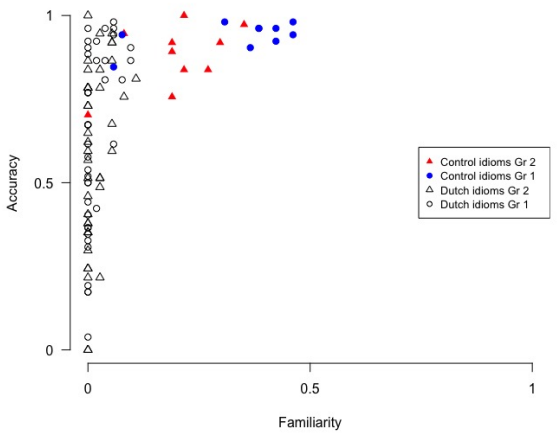

Figure 3.2: The accuracy (mean score) for each idiom vs. familiarity. Familiarity is calculated by number of "I know this one" selections per number of participants for that group.

Following from these results, we can argue that in general, the contorl idioms were familiar, whereas the Dutch idioms mostly unfamiliar to the participants. Thus, for the analysis we removed the control idioms for all participants. Further, we removed the cases where any Dutch idioms were selected as "I know this one, leaving only the data for *unfamiliar* idioms. For example, the Dutch idiom *He sat with the hands in the hair* was selected as "I know this one" by 3/52 participants for group 1. As these cases are removed, the mean score i.e. accuracy, is now divided by 49 instead of 52 for that idiom.

After removing these data, the mean number of unfamiliar idioms per participant for group 1 was 39, ranging between 34 and 40 idioms. For group 2 the mean number of unfamiliar idioms left per participant for analysis is 39, ranging from 33 to 40.

3.2 Relationship between decomposability rating and accuracy for unfamiliar Dutch idioms

For the remaining, unfamiliar idioms, we can plot the accuracy against the idiom decomposability rating of each idiom. The graph in figure 3.3 shows

the 40 Dutch idioms and their decomposability rating. The accuracy is thus adjusted for when "I know this one" was selected by any of the participants. The blue points are the data for group 1, and the red points are the data for group 2. The two dashed lines are the linear regression trends for group 1 (blue), and group 2 (red). The data shows a sparse spread, though a positive trend. However, the linear regression trends do not take into account the variation between participants. Furthermore, when we look at the raw data we see that the variation between idiom accuracy is large, despite having similar decomposability ratings. For example, table 3.1 shows 4 idioms with similar decomposability scores, yet varied accuracy, one showing 51/52 participants from group 1 choosing the correct answer, and one where only 2/52 chose the correct answer.

As stated by Sprenger et al. (2019), "it is important to realize that there is no such thing as an average native speaker: they differ with respect to socio-economic backgrounds, education, personality, and age". Knowing this, and the variability between idioms, it is clear that there are many sources of random variability.

We test the relationship between accuracy and decomposability rating by fitting a generalized linear mixed-effects model with random effects for participants and idioms, to account for the participants' response bias and the variation between idioms. First we analysed the data for group 1. The dependent variable is the score of each idiom by each participant (1 or 0), which is a binomial predictor, hence we fitted a logistic generalized linear mixed effect model. As fixed effect we included decomposability rating. We included random intercepts for idioms and participants. To test whether

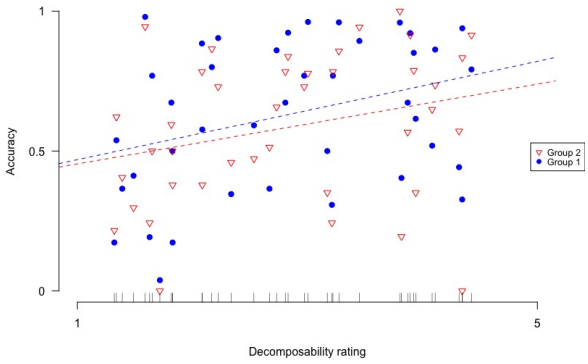

Figure 3.3: Accuracy vs. decomposability rating for group 1(blue) and group 2(red). Lines are linear models for corresponding groups.

decomposability rating contributed significantly to the explanation of this data we compared a model with decomposability as a fixed effect to a model without decomposability as fixed effect. ANOVA between these two models shows that the decomposability rating does significantly affect accuracy ($\chi^2(1)= 4.2771$, $p = 0.03865$) for monolingual native English speakers.

The same procedure was used to analyse the data for group 2, but in this case the decomposability rating does not significantly affect accuracy for multilingual native English speakers ($\chi^2(1)= 2.4947$, $p > 0.1$).

Table 3.1: Frequency of chosen answers, monolingual English native, N=52

Decomposability Rating	Translation	Answers	Frequency Monolingual - English
1.589	The monkey came out of the sleeve.	He revealed his secret	51
		He rolled up his sleeves	0
		He arrived late to the party	1
1.628	He joined for bacon and beans.	He was not playing for real	10
		He received little reward	39
		His work was below average	3
1.65	He fell with his nose in the butter.	He was in luck	40
		He made a mess	7
		He smelled something peculiar	5
1.717	He added butter to the fish	He paid straight away	2
		He cooked delicious food	2
		He exaggerated the story	48

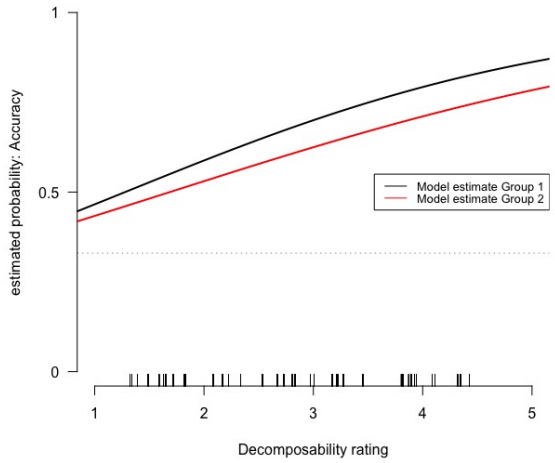

Figure 3.4: Generalized linear mixed effect model estimates for accuracy probability vs. Decomposability, Group 1(black) and 2(red)

When we plot what the model estimates, we can see the results for both groups in figure 3.4. The estimates have been transformed from the logit scale to the probability scale. The graph shows us that when an idiom has a low decomposability rating (i.e. nondecomposable), then the estimated probability of choosing the correct answer is around 0.45, and increases gradually as the decomposability increases. The horizontal broken line is the chance performance line, which is 33% as there was 1 correct answer out of the 3 answer options per idiom. The rug show the decomposability ratings of the idioms.

4 Discussion

The aim of this paper was to investigate whether the decomposability ratings made by Dutch natives in a study by Sprenger et al. (2019) may be biased, due to their familiarity with the idiom. In order to investigate this, we tested the relationship between the decomposability ratings of Dutch idioms, and if English speakers can work out their meanings in a multiple choice style questionnaire. As stated in the paper by Sprenger et al. (2019), if an idiom is highly decomposable, knowledge about its individual words may help a language learner to deduce the meaning. Thus, under this assumption, the pattern we should expect is that idioms with higher decomposability ratings should have a higher accuracy, since the meaning should be easier to work out.

The results show that for monolingual English speakers (i.e. group 1), the decomposability rating does have an effect on accuracy. In other words, the decomposability rating has shown to affect choosing the correct answer in the multiple choice questionnaire. This result agrees with current literature that the meaning of decomposable idioms should be easier to work out, since the individual words contribute to the overall figurative meaning. However, there are various points of discussion concerning this experiment.

4.1 Does choosing the correct answer mean the participant has actually understood the idiom?

Firstly, a clear point of discussion is the fact that for the analysis of the results the accuracy per idiom is used as the dependent variable. The accuracy is the mean score of the idiom by the participants. If a participant chose the correct answer out of the three options, then that instance received a score of 1, and otherwise a score of 0. A participant could guess the answer randomly, or choose the correct answer because the two decoy options were not likely. Put differently, there was no "I don't know" option, therefore the participant could quickly choose an answer and move to the next idiom. A possible improvement for future testing could be to give each idiom more than 3 answer options, to reduce the guessing chance. However, this would make the test more time consuming.

Another problem is that because of the multiple choice style of experimenting, the participants were exposed to the correct answer instead of having to work out the meaning. It could be the case that a participant always chose the most literal answer option out of the three options. As the decomposable idioms had two decoy answers which were created to be more abstract, always choosing the most literal option would result in a higher score for decomposable idioms by default. An improvement on this experiment would be to have an open

format questionnaire, where the participant can fill in their own meaning. A different approach could be to present only one of the three meanings to a participant at random, and ask how likely it is that this particular answer is the meaning of the idiom.

4.2 Familiarity and "I know this one"

The second discussion point for this experiment is that from the results of the "I know this one" option, we do not know when a participant would have chosen this option. A participant could have selected this option when they thought they knew it, because it was similar to an English idiom, or because they actually knew the Dutch idiom. It might also be the case that the participant only selected that option if they had heard of the idiom before, but were not familiar with the meaning of the idiom. When analysing the results, we assumed that any instances where "I know this one" was selected meant that the participant was *familiar* with the idiom and its meaning. Furthermore, some participants did not use this option at all, which could mean that they really weren't familiar with any idioms, or even that they did not understand the instructions properly.

4.3 Similarity of Dutch and English idioms and variation between results

A related point is that some idioms are very similar to English idioms. So even though the participants might not have been 100% familiar with the Dutch idiom, it might be so similar that the answer was already known and easier to choose the correct answer. For example, the Dutch idiom *The monkey came out of the sleeve* is similar to *To pull a rabbit out of the hat* in meaning and structure, and arguably presents a similar mental image. As can be seen in table 3.1, the meaning for this idiom was correctly identified by 51/52 participants from group 1 despite having a low decomposability rating (1.589). The low rating would suggest that the meaning is not easy to work out, yet this result shows otherwise. However, when we compare this result to the accuracy for the idiom *He added butter to the fish* the opposite result is achieved, where 2/52 participants chose the correct meaning, despite having a similar (even slightly greater) decomposability rating of 1.717.

4.4 Native English: Monolingual vs. Multilingual

The graph in Figure 3.1 also shows that multilingual English natives' (group 2) familiarity of the control idioms was less than for group 1. The graph in 3.2 shows that the accuracy of these control idioms was also generally lower. This suggests that the multilingual speakers were less familiar with these control idioms, even though they were existing English idioms. This is interesting as this suggests that if someone is bilingual or multilingual, then they would know fewer English idioms, similarly explained by Al-Lami (2006): "non-native speakers of English often lack the ability to understand and produce English idiomatic expressions appropriately and in an adequate amount." This adequate amount of idiomatic knowledge is a distinctive characteristic for native speakers compared to non-native speakers, and thus in the case of this experiment, for multilingual *native* English speakers.

Since some of the Dutch idioms were similar to English idioms, monolingual native English speakers could have an advantage because they should know more idioms, and thus should be better at working out figurative meanings because of the similarities between Dutch and English. For this reason, the multilingual native English speakers might actually be more representative for this experiment because of their limited knowledge of English idioms. However, we do not know their actual English proficiency, and perhaps their second language also contained similar idioms to Dutch. The focus was on monolingual English natives as we could control that the participants all spoke the same (and only) language. In contrast, the participants in group 2 were native English speakers, but also spoke one or more of the following languages: Arabic, French, Gujarati, Hindi, Irish, Italian, Japanese, Malayalam, Mandarin, Marathi, Norwegian, Portuguese, Punjabi, Russian, Sinhala, Spanish, Swedish, Tamil, or Urdu. A follow-up study could be to experiment on native monolingual Arabic speakers to minimize the chance that some idioms are similar to Dutch idioms.

5 Conclusion

In this paper we examined whether there is a relationship between the decomposability ratings, and whether speakers of a different language are able to work out the meaning. Since decomposability ratings are made by native speakers, it could be the case that these ratings are biased, and do not reflect the true decomposability of the idiom.

Specifically, Dutch idioms were translated into English and tested on native English speakers. We expected that the greater the decomposability rating, the more decomposable the idiom is and thus the easier it is to work out the meaning. The results show that this is the general trend for monolingual English natives. The results further show that the familiarity of English idioms is lower for multilingual English natives, than monolingual English natives, though the control idioms were only a small part of the experiment, and due to the limited number of participants this cannot be a conclusive result.

Though there is a relationship between the decomposability rating and accuracy, the bigger picture of this paper was to investigate whether these ratings were perhaps biased, since Nordmann et al. (2014) state that native speakers are unable to inhibit their knowledge of the meaning of an idiom when making decomposability judgements. If a participant is highly familiar with the figurative meaning of an idiom, their perception of the phrase will be that it is semantically acceptable, and so their perception of the phrase is biased. The results show that in general the idioms that have higher ratings were correctly answered more often, showing that these ratings do suggest the true decomposability of the idioms. There are of course many exceptions where the meanings of decomposable idioms were not correctly identified by the majority, and conversely where the meanings of nondecomposable idioms were correctly identified by the majority.

This study suggests that, in general, the ratings gathered by Sprenger et al. (2019) do reflect the degree of decomposability of the Dutch idioms. However, since there are various cases where decomposable idioms were incorrectly answered and vice versa, there is an ongoing challenge to create a reliable metric for idiom decomposability.

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.

Al-Lami, M. (2006). Assessing efl students' cultural awareness of the english language. *Unpublished Dissertation, University of Baghdad: Baghdad.*

Cain, K., Oakhill, J., and Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of experimental child psychology*, 90(1):65–87.

Cautin, R. L. and Lilienfeld, S. O. (2015). *The Encyclopedia of Clinical Psychology, 5 Volume Set.* John Wiley & Sons.

Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of language*, pages 22–42.

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

Gibbs Jr, R. W. and Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive psychology*, 21(1):100–138.

Gibbs Jr, R. W., Nayak, N. P., and Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of memory and language*, 28(5):576–593.

Hamblin, J. L. and Gibbs, 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.

Hubers, F., Cucchiarini, C., Strik, H., and Dijkstra, T. (2019). Normative data of dutch idiomatic expressions: Subjective judgments you can bank on. *Frontiers in psychology*, 10:1075.

Keysar, B. and Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, 34:89–89.

Nordmann, E., Cleland, A. A., and 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.

Paivio, A., Yuille, J. C., and Smythe, P. C. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 20(4):362.

Pawley, A. and Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191:225.

Searle, J. R. (1975). Indirect speech acts. In *Speech acts*, pages 59–82. Brill.

Sprenger, S., la Roi, A., and Van Rij, J. (2019). The development of idiom knowledge across the lifespan. *Frontiers in Communication*, 4:29.

Swinney, D. A. and Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of verbal learning and verbal behavior*, 18(5):523–534.

Tabossi, P., Fanari, R., and Wolf, K. (2008). Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2):313.

Tabossi, P., Wolf, K., and Koterle, S. (2009). Idiom syntax: Idiosyncratic or principled? *Journal of Memory and Language*, 61(1):77–96.

Titone, D. A. and Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5):1126.

A Appendix

Example questions from the questionnaire.

He held his face in the crease

- [ ] He did not age
- [ ] He showed no emotion
- [ ] He was very upset
- [ ] *I know this one*

He crawled under the wool

- [ ] He went to bed
- [ ] He was hiding
- [ ] He was cold
- [ ] *I know this one*

He was a hero on socks

- [ ] He won the game with ease
- [ ] He was a coward
- [ ] He acted like a super hero
- [ ] *I know this one*

He added water to the wine

- [ ] He avoided getting drunk
- [ ] He reached a compromise
- [ ] He understated the story
- [ ] *I know this one*

He did not let the cheese get eaten off his bread

- [ ] He ate every last bite
- [ ] He stood up for himself
- [ ] He saved his money
- [ ] *I know this one*

Table A.1: Idioms used for questionnaire

Item	Dutch Idiom	Control	Meaning	Translated Idiom	Est
1	Toen had hij haren op de tanden.		To have a sharp tongue/to dare to speak out	He had hairs on the teeth	1.321
2	Toen had hij boter op zijn hoofd.		To be guilty oneself too	He had butter on his head	1.339
3	Toen keek hij de kat uit de boom.		To wait and see	He watched the cat out the tree	1.39
4	Toen kocht hij de kat in de zak.		To buy something that turns out to be a waste of money	He bought the cat in the bag	1.488
5	Toen had hij een vinger in de pap	Yes	To have a finger in the pie/to have influence on something	He had a finger in the porridge	1.519
6	Toen kwam de aap uit de mouw.		Finally the truth came out	Then the monkey came out of the sleeve	1.589
7	Toen deed hij voor spek en bonen mee.		To count for nothing/to not play for real	He joined for bacon and beans	1.628
8	Toen viel hij met zijn neus in de boter.		To be in luck	He fell with his nose in the butter	1.65
9	Toen deed hij boter bij de vis.		To pay cash on the nail/to deal with something directly	Then he added butter to the fish	1.717
10	Toen koos hij eieren voor zijn geld.		To make the best of a bad job/to be forced to content oneself with less than one desires	He chose eggs for his money	1.817
11	Toen was de kogel door de kerk.		The decision was finally made	The bullet was through the church	1.826
12	Toen was hij een held op sokken.		To be a coward	He was a hero on socks	1.828
13	Toen was het hek van de dam.		There's no stopping it now/now things are getting out of hand	The gate was off the dam	2.083
14	Toen kreeg hij het deksel op zijn neus.		To be rebuffed/to get the door slammed in one's face/to be punished for being too greedy	He got the lid on his nose	2.086
15	Toen had hij een gat in zijn hand.		To spend money like water/to throw one's money about/to spend too much money	He had a hole in his hand	2.169
16	Toen gooide hij roet in het eten.		To make it difficult to achieve something	He threw soot in the food	2.224
17	Toen had hij een bord voor zijn kop.		To be insusceptible to critique that is expressed in guarded terms	He had a plate in front of his head	2.337
18	Toen stak hij zijn kop in het zand.	Yes	To bury one's head in the sand/to not want to face a problem	He stuck his head in the sand	2.343
19	Toen sloeg hij de spijker op de kop.	Yes	To find exactly the right answer.	He hit the nail on its head	2.399
20	Toen ging hij met de kippen op stok.		To go to bed early	He went with the chickens to roost	2.535
21	Toen hield hij het hoofd boven water.	Yes	To keep one's head above the water/to manage financially	He held his head above water	2.646
22	Toen deed hij water bij de wijn.		To make a compromise	He added water to the wine	2.67
23	Toen wreef hij zout in de wond.	Yes	To rub salt in(to) a wound/to increase someone's suffering	He rubbed salt in the wound	2.684
24	Toen schreeuwde hij zich de longen uit het lijf.		To shout very loud, Shout at the top of your lungs	He screamed the lungs out of his body	2.732
25	Toen speelde hij advocaat van de duivel.	Yes	To be the devil's advocate/to be someone who prefers to highlight the bad or negative sides of a situation	He played devil's advocate	2.771
26	Toen had hij een appeltje voor de dorst.		To have a nest egg/to have a buffer	He had an apple for the thirst	2.808
27	Toen nam hij geen blad voor de mond.		To speak plainly/not hold back (mince) words	He took no leaf in front of the mouth	2.832
28	Toen zag hij door de bomen het bos niet meer.	Yes	To lose sight of the big picture	He couldn't see the forest because of the trees	2.85
29	Toen liet hij zich niet de kaas van het brood eten.		To be able to stand up for oneself	He did not let the cheese get eaten off his bread	2.972
30	Toen legde hij de vinger op de zere plek.		To get someone where it hurts	He laid the finger on the sore spot	3.005
31	Toen stond hij met zijn rug tegen de muur.	Yes	To be with one's back to the wall/to have no way out	He had his back against the wall	3.117
32	Toen kreeg hij de baard in de keel.		His voice is breaking/to get a deeper voice during puberty.	He got the beard in the throat	3.173
33	Toen viel hij met de deur in huis.		To be straight to the point	He fell with the door in the house	3.213
34	Toen lachte hij als een boer met kiespijn.		To laugh on the wrong side of one's face/to laugh sourly	He laughed like a farmer with toothache	3.222
35	Toen haalde hij de druk van de ketel.		To blow off some steam/the pressure is gone	He took the pressure off the kettle	3.274
36	Toen kletste hij hem de oren van het hoofd.		To talk the back legs off a donkey/to talk a lot	He chattered the ears off his head	3.452
37	Toen schreeuwde hij het van de daken.	Yes	To shout something from the rooftops/to announce something widely	He shouted it from the rooftops	3.675
38	Toen zat hij met de handen in het haar.		To be at one's wit's end/to be desperate	He sat with the hands in the hair	3.804
39	Toen was hij vel over been.		To be very skinny	He was skin over leg	3.818
40	Toen hield hij zijn gezicht in de plooi.		To keep a straight face/to not show one's emotions	He held his face in the crease	3.872
41	Toen voegde hij de daad bij het woord.		To suit the action to the word/ To carry out an expressed intention immediately	He joined the deed with the word	3.894
42	Toen klopte zijn hart in zijn keel.		To have one's heart beat in one's throat/to be afraid	His heart was beating in his throat	3.925
43	Toen ging hij op de vuist.		To have a scrap/to fight	He went on the fist	3.942
44	Toen bleef hij op de achtergrond.	Yes	To keep a low profile/to not show oneself/to make sure one does not get attention	He stayed on the background	4.075
45	Toen hield hij hem buiten de deur.		To keep someone or something out	He held him outside the door	4.085
46	Toen zette hij hem op straat.		To throw someone out into the street/to dismiss someone	He put him on the street	4.111
47	Toen liep hem het water in de mond.		His mouth is watering/that makes his mouth water/	The water flowed in his mouth	4.319
48	Toen kroop hij onder de wol.		To go to bed	He crawled under the wool	4.346
49	Toen was hij het gesprek van de dag.		To be the talk of the town/to be the subject that is most frequently discussed at a certain moment	He was the conversation of the day	4.347
50	Toen zat hij met zijn neus in de boeken.		To be buried in one's books/to have one's nose in one's books	He sat with his nose in the books	4.427