

Influence of Unlicensed Definiteness on Behaviour of Non-Existing V+NP Idioms

Lenneke van Rooij
BA Thesis
English Language and Culture
Utrecht University
13-04-2018

First reader: Dr. Loes Koring
Second reader: Prof. Dr. Martin Everaert

Table of contents

ABSTRACT	2
1. INTRODUCTION	2
1.2 SYNTACTIC FLEXIBILITY OF IDIOMS.....	3
1.3 PRONOMINALISATION.....	8
1.4 DEFINITENESS AS A POTENTIAL IDIOMATIC TRIGGER.....	10
1.5 AIM OF THE CURRENT STUDY	12
2. METHOD	13
2.1 PARTICIPANTS	13
2.2 MATERIALS.....	13
2.3 PROCEDURE	14
3. RESULTS	16
4. DISCUSSION.....	18
4.1 TYPES OF MEANINGS GIVEN FOR NON-EXISTING V+NP IDIOMS.....	18
4.2 SCORES ON PROXIMITY OF THE NOUN TO A DICTIONARY DESCRIPTION	20
4.3 SCORES ON RETENTION OF IDIOMATIC MEANING IN A PRONOMINALISED CONSTRUCTION	21
REFERENCES	23
APPENDIX.....	25

Influence of Unlicensed Definiteness on Behaviour of Non-Existing V+NP Idioms

Abstract

The current study aimed to explore a possible effect of unlicensed definiteness on idiomatic behaviour of English non-existing idioms. It has been found that definite nouns in idioms are often not licensed by the context. An online questionnaire was created in which participants were asked to produce meanings for definite and indefinite non-existing idioms and rate their properties. It was found that the definite non-idioms generated meanings that preserved less of the original VP compared to the indefinite non-idioms. The indefinite non-idioms generated more meanings with a syntactic structure similar to the structure of the non-idiom. Therefore, unlicensed definiteness can possibly be seen as a cue for idiomatic meaning.

1. Introduction

Idioms are a rather complex and extensively researched language phenomenon. They are probably best described as “phrases or sentences that involve some degree of lexical, syntactic and/or semantic idiosyncrasy” (Fazly & Stevenson, 2006, pp. 337). Many idioms remain largely unused in daily speech (Liu, 2003). Their etymology is often unclear and their semantics are often non-transparent. This means that the meaning of the idiom cannot be inferred from the individual meanings of its components and how they combine (Moreno, 2005). Furthermore, actions described in idioms often do not correspond to any historic or real world events, though this has been debated with *spill the beans*¹. Furthermore, idioms are different from literal language in that they often have limited compositionality. This means that individual components within the idiom often do not have individual meaning that contributes to the overall idiomatic meaning. Instead, the idiomatic meaning arrives from the

¹ A popular folk etymology for *spill the beans* is based on a voting system from ancient Greece. Members of secret societies used black and white beans to vote on a new member. A white bean meant a positive vote on the new member, and a black bean meant a negative vote on the new member. Sometimes it happened that the jar filled with beans was knocked over and therefore the vote on the new member was revealed prematurely (Omazić & Schmidt, 2008). It is said that this is how *spill the beans* has come to mean *give away secret information*.

phrase as a whole. Idioms such as *kick the bucket* or *bite the bullet* are called pure idioms (Fernando, 1996, cited in Liu, 2003). Their components do not individually contribute to the idiomatic meaning in a literal sense. They are therefore considered fully non-literal. Other idioms are semi-literal: they have some components that contribute to the idiomatic meaning in a literal sense and some components that do not. Lastly, some idioms are completely literal. Other than varying in figurativeness, idioms also vary in syntactic flexibility (Gibbs & Nayak, 1989; Machonis, 1985; Nunberg, Sag & Wasow, 1994). Some idioms, such as *kick the bucket*, are very syntactically frozen, which means they cannot undergo many syntactic operations without losing the idiomatic meaning. Other idioms, such as *pop the question* are much more syntactically flexible. In the following section I will look at various largely unexplained irregularities of idioms: most idioms are not able to undergo many syntactic operations without losing their idiomatic meaning, and the syntactic behaviour of idioms varies wildly from idiom to idiom. The current study aims to find which factors contribute to variation in syntactic flexibility within idioms.

1.2 Syntactic flexibility of idioms

I will continue to discuss three related issues in the domain of syntactic flexibility of idioms: how they are stored in the brain, the way in which they behave syntactically, and what factors affect syntactic flexibility. Idioms were initially thought to be represented in the mental lexicon as a single lexical entry (Bobrow & Bell, 1973). It was argued that whenever an idiom is mentally accessed through speech or writing, its idiomatic meaning is accessed immediately, much like the meaning of a single word from the lexicon. This would mean that the idiomatic meaning is accessed before the literal meanings of individual words in the idiom are considered. This claim has been found to be false however (Cutting & Bock, 1997). Cutting and Bock found that the literal meaning of idioms was active during idiom

production. This suggests that idioms are not represented in the mental lexicon as a single lexical entry, but are instead processed like literal language.

The idea of an idiom as a single lexical entry also lead to the assumption that they are completely frozen expressions that cannot be internally modified. It is true that idioms have commonly been found to be restricted in their syntactic flexibility (Gibbs & Nayak, 1989; Machonis, 1985; Nunberg, Sag & Wasow, 1994), see (1) for example.

(1) *John kicked the bucket*

**The bucket was kicked by John*

The idiom *kick the bucket* in (1) loses its idiomatic meaning when it is passivised. Therefore, the idiomatic meaning in passive form is infelicitous. Similarly, the idiomatic meaning is lost when *kick the bucket* is topicalised, such as in (2).

(2) *John kicked the bucket*

**The bucket John kicked*

(Schenk, 1995, pp. 259, example 17)

However, it is incorrect to claim that all idioms are completely frozen expressions. To test syntactic flexibility of idioms, various syntactic operations, such as passivisation, topicalisation, quantification, modification by adjectives and relative clauses, and omission in elliptical constructions, have been applied (Machonis, 1985; Nunberg et al., 1994).² Through these processes, some idioms such as *spill the beans* have been found to be more syntactically flexible than others, such as *kick the bucket* (Mostafa, 2010). Furthermore, it was found that most idioms were able to undergo one or more of these operations without losing their

² Mostafa (2010) also discusses these modifications found using the British National Corpus and Webcorp.

idiomatic meaning. However, many questions still remain about why some idioms seem to be more syntactically flexible than others.

One possible theory for the variation in syntactic flexibility within idioms comes from Gibbs and Nayak (1989). They had participants categorise V+NP idioms and found that their syntactic behaviour was dependent on speakers' perception of how components of the idioms contributed to the overall idiomatic meaning. This means that if the individual components of the idioms were related to the overall idiomatic meaning in a literal sense, participants judged the idioms as more syntactically flexible. Based on these findings, Gibbs and Nayak introduced the Idiom Decomposition Hypothesis. Their participants were able to identify three different categories of idioms: firstly, they identified normally decomposable idioms such as *pop the question*, of which components were thought to have a literal connection to the overall idiomatic meaning. In this idiom, *the question* refers to an actual question, namely a proposal. Therefore, it has a literal connection to the idiomatic meaning *to propose*. Secondly, participants identified abnormally decomposable idioms such as *spill the beans*, of which components were thought to metaphorically link to the overall idiomatic meaning. Whereas *spill* has a close relationship to *reveal*, *the beans* does not have a literal connection to *secret or valuable information*, which is what is thought to be the meaning of *beans* in the idiom. This means that the components of the idiom contribute to the overall meaning in a metaphorical way. Gibbs and Nayak called this category abnormally decomposable, because the idiom components contributed to the overall idiomatic meaning, but not in a literal, normal sense. Lastly, participants identified non-decomposable idioms such as *kick the bucket*. Unlike the previous two categories of idioms, idioms in this category did not have components that individually contributed to the idiomatic meaning. The idiomatic meaning was only reflected in the full phrase. Later experiments revealed that the category in which each idiom was placed predicted the degree of syntactic flexibility. The more an idiom was

regarded as having separate meaningful units, the more syntactically flexible that idiom was. For example, compare *pop the question* and *chew the fat*. The former idiom was regarded as being normally decomposable, whereas the latter was regarded as being non-decomposable. Consequently, *pop the question* was later judged as being more syntactically flexible than *chew the fat*. Idiom compositionality and syntactic flexibility of idioms are therefore shown to go hand in hand. I will go into how Gibbs and Nayak tested syntactic flexibility of idioms in section 1.3.

An example of the correlation between idiom compositionality and syntactic flexibility can be found when looking at nouns in idioms. When an idiom is non-decomposable, its individual components do not contribute to the overall idiomatic meaning. This is also true for the nouns. Nouns in idioms such as *bucket* in *kick the bucket* are therefore said to be non-denoting, which means they do not have any meaning within the idiom (Fellbaum, 1993). This makes them semantically vacuous, such as *it* in sentences such as *it is raining*. Schenk (1995) explains how the noun in an idiom being semantically vacuous causes the idiom itself to be less syntactically flexible. He argues that two types of syntactic operations exist. The first type includes operations that can only be applied to constituents that carry meaning, whereas the second type includes operations that can be applied to both meaningful and meaningless constituents. Syntactic operations like topicalisation, clefting, and pronominalisation belong to the first type (see (2) for topicalisation, (3) for clefting, and (4) for pronominalisation).

(3) *John kicked the bucket*

**It was the bucket that John kicked*

(4) **Alexander spilled the beans, since he did not know they were secret*

(Schenk, 1995, pp. 262)

Schenk argues that the syntactic operations that belong to type 1 change the semantics of the expression they are applied to. He draws on an example of modification, as seen in (5). When modification is applied to a meaningless constituent, it cannot change the constituent, because it has no original meaning. Therefore these types of operations do not work on idiomatic constituents.

(5) **Mary spilled the well-kept beans*

(Schenk, 1995, pp. 262)

The second type of syntactic operations Schenk discusses features strictly syntactic operations. It includes operations such as verb-first operations in *yes/no* questions and object to subject operations as part of passivisation. These do not change the semantics of individual constituents, but only the syntax, which is why Schenk argues they can apply to both semantically meaningful and semantically vacuous constituents. It is now clear why non-decomposable idioms do not topicalise or pronominalise without losing the idiomatic meaning. An already meaningless element cannot suddenly gain meaning by performing any syntactic operation. However, it remains unanswered why some idioms, such as *kick the bucket* cannot passivise (see (1)), whereas others, such as *spill the beans* can (see (6)).

(6) *Pat spilled the beans*

The beans were spilled by Pat

(Nunberg et al., 1994, pp. 510, example 30)

1.3 Pronominalisation

In this section I will discuss pronominalisation of idioms, which is a syntactic operation that relies on the semantic content of the noun. If we take the idiom *kick the bucket* for example, it would look like the following in a pronominalised construction: *Mary kicked the bucket, but John did not kick it*. The noun in the main clause is referred to by an anaphoric pronoun in the subordinate clause, namely *it*. This *it* should not be confused with *it* in cleft constructions, which is considered an expletive pronoun and is semantically vacuous. The pronoun in the subordinate clause is considered a neuter personal pronoun and refers to *the bucket*. To be able to refer to the noun in an idiom such as *kick the bucket*, however, the noun must be a meaningful constituent (Schenk, 1995). If this is only possible in the literal form and not the idiomatic form, the speaker will infer a literal meaning over an idiomatic meaning for the pronominalised form. Therefore, pronominalisation of idioms can show whether the noun in the idiom has individual meaning or is non-denoting. If the noun in the idiom has individual meaning, the overall idiomatic meaning will be retained in a pronominalised construction, whereas if the noun in the idiom is non-denoting, the overall idiomatic meaning will not be retained in a pronominalised construction.

Gibbs and Nayak (1989) used pronominalisation to test syntactic flexibility of existing English idioms. They had participants rate pronominalised versions of idioms such as *pop the question*, *spill the beans* and *kick the bucket*. Participants were given these idioms in pronominalised form, along with their idiomatic meanings in pronominalised form. They were asked to rate how similar the idiomatic meaning in pronominalised form was to the pronominalised idiom on a 7-point Likert scale. An example of this with *hit the sauce* can be found in (7).

(7) *After they were divorced, Tony began to hit the sauce, but Cathy didn't begin to hit it.*

(9) a. bucket₁ MEANING: ‘bucket’

SYNTAX: -

b. bucket₂ MEANING: -

SYNTAX: [kick₂ -]

(Everaert, 2010, p. 6, example 10 & 11)

By putting restraints on the context in which these meanings are felicitous, two readings of *kick the bucket* arise. By combining (7a) with (8a), the literal reading of *kick the bucket* is formed, whereas the idiomatic reading is formed by combining (7b) with (8b). The subsenses of (7b) and (8b) have to be combined in order to receive idiomatic meaning. Their composition is identical to the literal meaning however, which suggests that idioms are potentially no different than their literal counterparts in their compositionality. Everaert therefore argues that saying some idioms are more compositional than others is therefore merely saying that some idioms are more syntactically flexible than others.

1.4 Definiteness as a potential idiomatic trigger

A factor that could potentially play a role in the degree of compositionality of idioms is the definiteness of the noun. Idioms can feature definite nouns, such as in *bite the dust*, or indefinite nouns, such as in *have a ball* (Fellbaum, 1993). For a definite noun to be felicitous however, it has to be licensed by the context, because definiteness assumes familiarity or uniqueness. When the noun is discourse-old information, it can be preceded by a definite article, which in English can be found in the form of *the* (Abbott, 2004). If the noun has not been mentioned in earlier discourse or is not uniquely identifiable, it is not felicitous to use a definite determiner and an indefinite article (*a* or *an* in English) should be used instead. Consequently, idioms that use a definite determiner often do so seemingly infelicitously. For example, *John kicked the bucket yesterday* can be uttered in conversation without previous mention of a bucket, or a uniquely identifiable bucket being present. The definite determiner

the bucket is therefore not licensed by the context. Furthermore, the definite determiner in an idiom usually cannot be changed into an indefinite determiner without losing the idiomatic meaning (Fellbaum, 1993). Mostafa (2010) found two exceptions to this, but argues that the authors deliberately used an indefinite determiner in order to refer to a general situation instead of a specific, unique one. Nonetheless, definite determiners in idioms cannot usually be changed without losing idiomatic meaning. The idiomatic meaning depends on the definiteness of the determiner. This sparks the idea that there is a specific reason for many idioms to feature unlicensed definites.

Koring argues that the infelicity of an unlicensed definite determiner causes the reader to consider non-literal meanings. She presented participants with short stories that featured a non-existing idiom. Participants were then asked come up with meanings for these non-existing Dutch V+NP idioms, which were divided into two conditions: idioms featuring a definite determiner, such as *clean the table*, and idioms featuring an indefinite determiner, such as *clean a table*. There were no other manipulations. The short stories were identical for both types of idioms. Based on the meanings participants provided, three categories and a leftover category were created. Meanings were coded into these categories based on how much of the original syntactic structure of the idiom was preserved. Category 1 featured meanings that preserved least of the original structure and were usually very bare syntactic structures, such as *to help* or *to be disappointed*. Category 2 featured meanings that had a V+NP structure but with a non-concrete object, such as *to offer help* or *to solve a problem*. Lastly, category 3 featured meanings that were closest to the original structure of the idiom. They had a V+NP structure with a concrete object and they were usually concrete actions that could be performed, for example *to hide something* or *to perform a task*. Koring also notes that meanings in category 3 often retained the verb from the idiom in the meaning. Therefore, this category was the least idiomatic category. Category 4, which was the leftover category,

featured meanings that could not be coded into any of the other categories. These meanings were usually considered most idiomatic. It was found that the meanings given to the non-existing idioms preserved less of the original structure of the idiom when the idiom featured a definite determiner. The fact that definite determiners in the non-idioms led to meanings that preserved less of the original VP and therefore were more non-compositional could be evidence that unlicensed definiteness could potentially be seen as a trigger for idiomatic meaning.

1.5 Aim of the current study

Keeping Koring's findings in mind, the current study aims to find how unlicensed definiteness influences syntactic flexibility and decomposability of English non-existing V+NP idioms. No research has yet been done with non-existing idioms in English. Furthermore, the question remains why some idioms allow more syntactic flexibility than others. Therefore, I want to study the behaviour of unlicensed definite nouns in non-existing idioms and the influence of definiteness on the syntactic flexibility of nouns in non-existing idioms, specifically in a pronominalised form. I expect that idioms featuring unlicensed definiteness will show stronger idiomatic behaviour than idioms featuring indefiniteness. This means that they are less decomposable and less syntactically flexible. Because the unlicensed definite nouns do not have any felicitous individual meaning, it could be argued that they do not individually contribute to the idiomatic meaning and are more like non-denoting nouns. Therefore, when comparing the meaning of the noun in definite and indefinite non-existing idioms, I expect that the unlicensed definite noun in a non-existing idiom will be further away from its literal meaning than the indefinite noun in a non-existing idiom. This would show that unlicensed definiteness leads the reader to consider stronger idiomatic meaning than indefiniteness. Furthermore, I expect that an unlicensed definite noun in a non-existing idiom will be more syntactically frozen than an indefinite noun. In line with Schenk (1995),

assuming the unlicensed definites have no individual meaning would mean they would not work in a pronominalised construction without losing the idiomatic meaning, because they cannot be reliably referred to by a pronoun. Therefore, I expect that non-existing idioms featuring indefinites would retain more idiomatic meaning in a pronominalised construction than non-existing idioms featuring unlicensed definites.

2. Method

The designed experiment aimed to find an effect of definiteness on the proximity of the noun in non-existing V+NP idioms to its literal meaning. Furthermore, it aimed to find an effect of definiteness on retention of idiomatic meaning of the non-existing V+NP idiom in a pronominalised construction. Both of these predictions were tested using an online questionnaire.

2.1 Participants

29 native speakers of English participated in the experiment. Their age ranged from 19 to 45. All participants were residing in a country with English as the first language at the time of testing. 32 participants initially completed the survey, however, in order to be able to categorise the meanings properly, 3 participants were left out of the analysis because they failed to produce a meaning for all non-existing idioms. Instead of giving a meaning for the idiom, their answers were along the lines of *it's too complicated* or *I don't know*.

2.2 Materials

For the experiment, a questionnaire was created using Qualtrics. The questionnaire consisted of 10 English short stories that featured a non-existing V+NP idiom. Each idiom had a version with a definite article and a version with an indefinite article. These two versions had an identical short story. An example of a short story with a definite idiom can be found in (10). The only manipulation between the two versions was the definite or indefinite article. All idioms featured concrete actions that could be performed. The list of idioms can be found in

Appendix A. To make sure participants did not judge both a definite and indefinite version of the same idiom, two different versions of the questionnaire were created. Both contained an equal number of definite and indefinite idioms which were evenly distributed throughout the questionnaire. Participants were assigned one of the two versions of the questionnaire randomly. 14 participants filled out one version and 15 participants filled out the other version.

(10) John had been a teacher for 10 years, but never had he been in front of a classroom as noisy as this one. He tried everything to get his students to be silent, such as banging the table and shouting, but nothing had worked so far. John bit the apple in a desperate last attempt at restoring order in the classroom.

2.3 Procedure

The questionnaire was sent to the participants online. Participants were first asked about their age, country of residence, and native speaker status. After this, they were presented with a short instruction. First, an explanation of idioms was given including examples of four common existing English idioms with varying syntactic structures of both the idioms and their idiomatic meanings. This was done to prevent any bias in the syntactic structures of participants' meanings for the non-existing idioms. Participants were then informed that they would be shown 10 short stories that were told by a sibling who likes to make up non-existing idioms. They were also informed they would be asked to think of a possible meaning for the non-existing idioms. It was specified that there was no right or wrong meaning for the idioms. This was done to prevent participants from hesitating too much or failing to produce a meaning. Participants were then told they would be rating properties of the idiom on a scale

from 1 to 7 in two questions following each idiom. Lastly, they were instructed to fill out the questionnaire in a quiet room without distractions.

After having read the general instruction, participants were shown the short stories featuring the non-existing V+NP idioms. The idioms were underlined in each short story. Participants were then asked to produce a meaning for the non-existing idiom they were presented with. This was done for multiple reasons: firstly, to make sure that participants did not skip reading the short story, which would cause them not to consider the infelicity of the definite article. It was also done to make sure that participants produced a possible meaning before rating the properties. This would allow for the property scores for the idiom to be more clearly motivated by the choice of meaning. Lastly, the syntactic structure of the produced meanings could possibly also show an effect of definiteness, with definite idioms producing a meaning that preserves less of the original syntactic structure of the idiom. After the participants produced a meaning for the idiom, an Oxford dictionary definition of the noun in the idiom was given. This definition was shortened and simplified in order to prevent it from being too specific or complicated. An example of a definition used in the experiment can be found in (11) for *the/an apple*.

(11) "A round piece of fruit that grows from a tree and is typically red or green."

Participants were asked to rate to what extent this definition corresponded to the meaning of the definite or indefinite noun in the idiom on a 7-point Likert scale. The low end of the scale was labelled *not at all*, the middle was labelled *somewhat*, and the high end of the scale was labelled *entirely*. Next, participants were shown the idiom in a pronominalised construction, such as in the definite version in (12) and the indefinite version in (13).

(12) *John bit the apple, but his students did not bite it.*

(13) *John bit an apple, but his students did not bite one.*

Participants were asked how similar the meaning that was given to the non-existing idiom earlier was to the meaning of the same idiom in the pronominalised construction on a 7-point Likert scale. The low end of the scale was labelled *not similar at all*, the middle was labelled *somewhat similar*, and the high end of the scale was labelled *very similar*. Afterwards, results were collected and analysed using SPSS.

3. Results

The idiom meanings participants gave were coded into three categories, based on how much of the original syntactic structure of the non-existing idiom was retained in the given meaning. These categories were based on the categories established by Koring (submitted). Category 1 featured meanings that preserved least of the original syntactic structure of the non-existing idiom and therefore were considered to be very idiomatic. Meanings in category 1 had a very bare syntactic structure usually consisting of a bare verb or a verb and adverbial and no object (*relax, do well, fall asleep*). Category 2 featured meanings that had a V+NP structure including an object, but the object was non-concrete (*focused his anger, take a risk, killed some time*). This category was considered less idiomatic than category 1. Category 3 featured meanings that also had a V+NP structure, but used concrete objects (*closed her eyes, keep the money, stamped his feet*) The meanings in this category were concrete actions, like the original non-existing idiom stimuli, and were considered least idiomatic. Category 4 was created for any meanings that did not fit into the other categories. This category featured answers with varying syntactic structures and often existing idioms that could not be categorised in the other categories (*think better of it, throw in the towel*). Answers in category 4 were not a unified class of answers, but instead consisted of varying syntactic structures that

did not correspond to the other three categories. They were therefore considered most abstract. A distribution of answers per category per condition can be found in figure 1.

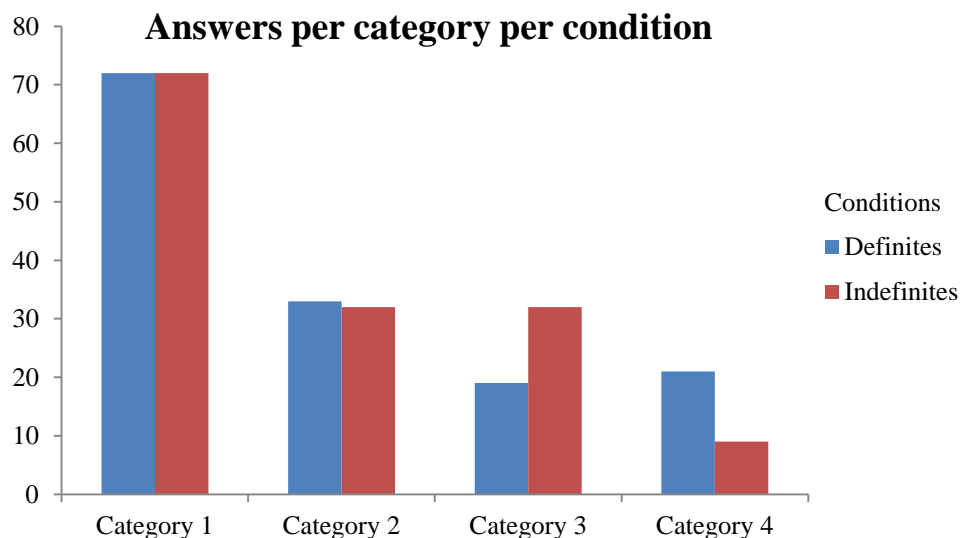


Figure 1. Distribution of idiom meanings to categories for the Definite and Indefinite non-existing idioms.

A notable difference between conditions can be found in category 3 and 4. More indefinites (32) than definites (19) fell in category 3. The opposite was found for category 4. This category has 21 definites and 9 indefinites. Therefore, the indefinite condition yielded more answers that fell in category 3 compared to the definite condition, whereas the opposite was found in category 4, where more definites (21) were found than indefinites (9).

The data for the dictionary questions and the pronominalisation questions was based on a 7-point Likert scale. Data for both questions was not normally distributed. A Wilcoxon signed-rank test was done for the dictionary questions. It did not show a significant effect of definiteness on the scores for proximity of the noun in the idiom to the dictionary description ($T = 96$, $p > .05$, $r = -.04$). Means for the dictionary description question were 2.98 for the definites and 2.91 for the indefinites. Another Wilcoxon signed-rank test was done for the pronominalisation questions. No significant effect of definiteness was found on the scores for retention of the idiomatic meaning in a pronominalised construction ($T = 206$, $p > .05$, $r = -$

.02). Means for the pronominalisation question were 3.33 for the definite condition and 3.40 for the indefinite condition.

4. Discussion

The current study aimed to find an effect of unlicensed definiteness on compositionality and syntactic flexibility of non-existing English idioms. It was expected that unlicensed definite nouns in idioms would be semantically vacuous and therefore the definite idioms would be less compositional. Furthermore, in accordance with Schenk (1995), a semantically vacuous element cannot undergo syntactic operations that change the semantics of the element they are applied to, such as in pronominalisation. Therefore, it was expected that definite idioms would retain less of the idiomatic meaning in a pronominalised construction than indefinite idioms. I will first discuss the qualitative results from the coded meanings, after which I will go into the quantitative data from the dictionary and pronominalisation questions.

4.1 Types of meanings given for non-existing V+NP idioms

The qualitative data from the meanings given by the participants led to some interesting findings in category 3 and 4. Category 3, which was considered least idiomatic in the sense that the given meaning reflected most of the original VP, featured more indefinites than definites. This shows that the indefinite non-idioms tended to generate meanings that were less idiomatic than the definite non-idioms. Category 4 also showed an interesting finding. More definites than indefinites fell into this category. Since this category was considered most abstract and featured meanings that could not easily be categorised, it can be assumed that definite non-idioms generated more abstract meanings than indefinite non-idioms. This is in line with the findings by Koring, who found that meanings given to definite non-idioms preserved less of the original VP of the idiom than meanings given to indefinite non-idioms. Furthermore, the findings of the current study show an effect for meanings given to indefinite

idioms as well. They were more frequent in category 3, which means they preserved more of the original VP than the meanings given to definite idioms.

Interestingly, many non-existing idioms lead to meanings that featured existing idioms with the same verb. For example, *bite the/an apple* led to *bite the bullet*, *light the/a match* led to *light up*, and *throw the/a ball* led to *throw in the towel*. Koring found similar retention of the idiom verb in the meaning in category 3, but did not find this led to generation of an existing idiom. The reason for this could be that Koring's non-existing idiom stimuli were mixed with existing idiom stimuli, which could have prevented participants from giving the non-existing idioms meanings that included an existing idiom. Furthermore, it is important to note that answers featuring existing idioms were given as much for definite and indefinite conditions (15 times for definites and 16 times for indefinites). Therefore, there was no effect of definiteness on the frequency of given answers featuring existing idioms.

It was quite difficult to code the meanings into the different categories for multiple reasons. Firstly, the categories in which meanings were coded did not completely straightforwardly apply to the data. This is due to the fact that Koring created the categories based on the types of meanings she found, whereas I applied her categories to my data. The most idiomatic category, category 1, was easy to code. However, category 2 and 3 often had some overlap where the noun could not be considered completely non-concrete but also not completely concrete (see *complete a task* for example). Consequently, it was difficult to code such meanings, which could have potentially influenced results slightly. Furthermore, participants' meanings often featured multiple meanings per idiom or an explanation of how the idiom could be interpreted to be transparent. For example, one answer to the idiom *sing a verse* explained how it takes a reasonably long time to sing a verse and therefore the idiom meant that it took a long time. It was very interesting to see what participants came up with, but it often led to the meaning being placed in the leftover category because it did not fit into

the other categories. I suspect that the task being complicated and participants not being able to ask questions during online completion could have contributed to the wide variety in interpretations of the task at hand.

4.2 Scores on proximity of the noun to a dictionary description

Scores on the dictionary question did not show an effect of definiteness on proximity of the noun to its literal meaning. Nouns in both conditions had scores on the low-medium part of the scale, which means that the meaning of the nouns in the non-idioms were moderately similar to the dictionary description. This shows that participants could not give a conclusive rating to the nouns in the non-idioms. A reason for this could be the nature of the instruction participants received. The data for the dictionary question had a very wide variance for both conditions (3.79 for the definite condition, and 3.45 for the indefinite condition). This could mean that the instructions for this question were interpreted in multiple ways that resulted in different scores. Various participants gave a score of 7 to all nouns, whereas several participants gave a score of 1 to all nouns. It is possible that participants who gave a maximum score on likeness to the literal meaning for all nouns interpreted the instruction for the dictionary question as rating the dictionary description for the general, prototypical noun, instead of its particular meaning within the idiom. Because the dictionary meanings presented to the participants were very descriptive, this interpretation resulted in maximum scores. Conversely, those who gave a minimum score to all nouns likely interpreted all nouns in the idioms to have a figurative meaning, potentially brought on by the general instruction, in which it was explained that idioms had figurative meanings. The dictionary question was designed to test how literal the meaning of the noun in the idiom was. However, it did not manage to generate any conclusive results. Therefore, the task needs to be modified or replaced with a different task in any future studies in order to rule out different interpretations of the instruction.

Because of the wide variance, no other conclusions can be drawn about the individual meaning of the definite and indefinite nouns in non-existing V+NP idioms. It is possible that the nouns are non-denoting when the meaning participants gave for the non-existing idiom did not specify an object, such as the meanings in category 1, however this cannot reliably be inferred from the current design of the dictionary question. The rating for the noun does not give enough information to make any assumptions about the semantic vacuity of the noun, because a low rating on the dictionary question can point either to a more figurative meaning of the noun, or no semantic meaning at all.

4.3 Scores on retention of idiomatic meaning in a pronominalised construction

The pronominalisation question was designed not only to provide information on the syntactic flexibility of the non-existing idioms, but also to provide information on the semantic vacuity of the definite or indefinite noun in non-existing V+NP idioms. In accordance with Schenk (1995), pronominalisation is only possible when the pronominalised constituent is semantically meaningful. Because the definites in idioms are unlicensed, it was assumed that they did not have any individual meaning and therefore would not be able to be pronominalised. However, no effect of condition was found on retention of idiomatic meaning in a pronominalised construction. Therefore, it cannot readily be concluded from the results of the current study that unlicensed definiteness in idioms leads to the noun being non-denoting or less syntactically flexible.

Scores for both conditions were on the medium part of the scale. This means that participants judged the meaning they gave to the non-existing idioms to be somewhat similar to the meaning of that idiom within a pronominalised construction. Again, this does not provide any conclusive findings on retention of idiomatic meaning in a pronominalised construction. This is very different from Gibbs and Nayak's findings, who found that compositionality influenced the idiom's ability to retain idiomatic meaning in a

pronominalised construction. A reason for this discrepancy could be that the pronominalisation ratings participants gave in the current study were based on the meaning of the non-existing idiom that they had produced earlier. Conversely, the pronominalisation ratings in Gibbs and Nayak's experiment were based on the meaning of existing idioms that participants were already familiar with. Therefore, their participants were able to rely on a predetermined meaning, instead of having to provide the meaning for the idiom themselves, which may have complicated the task too much. An alternative method of testing pronominalisation of non-existing idioms could be to present the participants with the intended meaning of the non-existing idiom. It could then be seen whether there is a difference between definite and indefinite idioms based on them having the same meaning. This ensures that the only difference between the two types of idioms is the definite or indefinite determiner and that their idiomatic meanings remain identical as a baseline. Alternatively, different meanings could be presented for the same idioms to see if pronominalisation ratings differ when the meaning has a syntactic structure identical to the structure of the idiom or when the syntactic structures of the meaning and the idiom are different. This design ensures that the types of the idiom meanings presented to the participants can be regulated, instead of being left to participants to produce.

References

- Abbott, B. (2004). Definiteness and Indefiniteness. In L. R. Horn & G. Ward (eds.), *The Handbook of Pragmatics* (pp. 122-149). Malden, MA: Blackwell.
- Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343-346.
- Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57-71.
- Everaert, M. (2010). The Lexical Encoding of Idioms. In M. Rappaport Hovav, I. Sichel & E. Doron (eds.), *Syntax, Lexical Semantics and Event Structure* (pp. 76-98). New York, NY: Oxford University Press.
- Fazly, A., & Stevenson, S. (2006). Automatically Constructing a Lexicon of Verb Phrase Idiomatic Combinations. In *11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 337-344).
- Fellbaum, C. (1993). The determiner in English idioms. In C. Cacciari, & P. Tabossi (eds.), *Idioms: Processing, structure and interpretation* (pp. 271-296). Hillsdale, NJ: Erlbaum.
- Gibbs, R. W., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, *21*(1), 100-138.
- Koring, L. (submitted). Producing figurative meanings: the case of idioms.
- Liu, D. (2003). The Most Frequently Used Spoken American English Idioms: A Corpus Analysis and Its Implications. *Tesol Quarterly*, *37*(4), 671-700.
- Machonis, P. A. (1985). Transformations of Verb Phrase Idioms: Passivization, Particle Movement, Dative Shift. *American Speech*, *60*(4), 291-308. doi:10.2307/454907

- Moreno, R. E. V. (2005). Idioms, Transparency and Pragmatic Inference. *UCL Working Papers in Linguistics*, 17, 389–425.
- Mostafa, M. (2010). Variation in V+the+N idioms. *English Today*, 26(4), 37-43.
<https://doi.org/10.1017/S0266078410000325>
- Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, 70(3), 491-538.
- Omazić, M., & Schmidt, G. (2008). Polysemy in phraseology. In M. Álvarez de la Granja (ed.), *Lenguaje figurado y motivación. Una perspectiva desde la fraseología* (pp. 97-109). Frankfurt: Peter Lang.
- Schenk, A. (1995). The Syntactic Behavior of Idioms. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives* (pp. 253-271). Hillsdale, NJ: Erlbaum.

Appendix

Appendix A: List of non-existing V+NP idioms used in the experiment

1. Bite the/an apple
2. Snap the/a branch
3. Sing the/a verse
4. Wear the/a beanie
5. Hit the/a cat
6. Light the/a match
7. Break the/a bottle
8. Throw the/a ball
9. Jump the/a ledge
10. Clean the/a table



Faculty of Humanities

Version September 2014

PLAGIARISM RULES AWARENESS STATEMENT

Fraud and Plagiarism

Scientific integrity is the foundation of academic life. Utrecht University considers any form of scientific deception to be an extremely serious infraction. Utrecht University therefore expects every student to be aware of, and to abide by, the norms and values regarding scientific integrity.

The most important forms of deception that affect this integrity are fraud and plagiarism. Plagiarism is the copying of another person's work without proper acknowledgement, and it is a form of fraud. The following is a detailed explanation of what is considered to be fraud and plagiarism, with a few concrete examples. Please note that this is not a comprehensive list!

If fraud or plagiarism is detected, the study programme's Examination Committee may decide to impose sanctions. The most serious sanction that the committee can impose is to submit a request to the Executive Board of the University to expel the student from the study programme.

Plagiarism

Plagiarism is the copying of another person's documents, ideas or lines of thought and presenting it as one's own work. You must always accurately indicate from whom you obtained ideas and insights, and you must constantly be aware of the difference between citing, paraphrasing and plagiarising. Students and staff must be very careful in citing sources; this concerns not only printed sources, but also information obtained from the Internet.

The following issues will always be considered to be plagiarism:

cutting and pasting text from digital sources, such as an encyclopaedia or digital periodicals, without quotation marks and footnotes;
cutting and pasting text from the Internet without quotation marks and footnotes;

copying printed materials, such as books, magazines or encyclopaedias, without quotation marks or footnotes;

including a translation of one of the sources named above without quotation marks or footnotes;

paraphrasing (parts of) the texts listed above without proper references: paraphrasing must be marked as such, by expressly mentioning the original author in the text or in a footnote, so that you do not give the impression that it is your own idea;

copying sound, video or test materials from others without references, and presenting it as one's own work;

submitting work done previously by the student without reference to the original paper, and presenting it as original work done in the context of the course, without the express permission of the course lecturer;

copying the work of another student and presenting it as one's own work. If this is done with the consent of the other student, then he or she is also complicit in the plagiarism;

when one of the authors of a group paper commits plagiarism, then the other co-authors are also complicit in plagiarism if they could or should have known that the person was committing plagiarism;

submitting papers acquired from a commercial institution, such as an Internet site with

summaries or papers, that were written by another person, whether or not that other person received payment for the work.

The rules for plagiarism also apply to rough drafts of papers or (parts of) theses sent to a lecturer for feedback, to the extent that submitting rough drafts for feedback is mentioned in the course handbook or the thesis regulations.

The Education and Examination Regulations (Article 5.15) describe the formal procedure in case of suspicion of fraud and/or plagiarism, and the sanctions that can be imposed.

Ignorance of these rules is not an excuse. Each individual is responsible for their own behaviour. Utrecht University assumes that each student or staff member knows what fraud and plagiarism

entail. For its part, Utrecht University works to ensure that students are informed of the principles of scientific practice, which are taught as early as possible in the curriculum, and that students are informed of the institution's criteria for fraud and plagiarism, so that every student knows which norms they must abide by.

I hereby declare that I have read and understood the above.

Name: Lenneke van Rooij

Student number: 5556627

Date and signature: 16-04-18



Submit this form to your supervisor when you begin writing your Bachelor's final paper or your Master's thesis.

Failure to submit or sign this form does not mean that no sanctions can be imposed if it appears that plagiarism has been committed in the paper.