UNIVERSITY OF TWENTE.

DEPARTMENT OF PSYCHOLOGY

NOVEL METAPHORS

Rating meaningfulness in L1 and L2 speakers of Dutch

Davida Flinsenberg
B.Sc. Thesis
February 2016

Supervisors:

Prof. dr. Frank van der Velde,

Dr. Karolina Rataj,

Deniece Nazareth, CPE

Opleidingsadministratie

BOZ-PSY-CES@utwente.nl

ABSTRACT

In understanding the cognitive processing of language, figurative language holds an interesting position. Due to their inherent ambiguity, non-literal expressions such as metaphors are very susceptible to personal differences in the attribution of meaning to what has been said. The amount of contact we have with a language can drastically influence the way we perceive these messages, especially in people who have a different history with the language such as non-native speakers. This explorative norming study attempts to produce a usable set of Novel Metaphors, along with their triplets, for further use in linguistic metaphor research. Also under observation are the differences presented between native (L1) speakers and non-native (L2) speakers. Several items were identified as reliable measures of meaningful metaphors. There were several differences observed between the attribution of meaning between L1 and L2, among which some points of interest on metaphoric competence in L2, as mentioned in the discussion.

CONTENTS

1. INTRODUCTION

When appreciating artistic ability, a lot of praise is placed on artists for being able to use their creative ability to express a concept in a beautiful (often non-literal) way. One such an artful expression is "The silken sad uncertain rustling of each purple curtain, thrilled me, filled me with fantastic terrors never felt before." (Edgar Allan Poe, from 'The Raven'.) Here the mood the poet wishes to achieve is set very well, through the use of stylistic devices such as metaphor. While knowing that a curtain can be sad nor uncertain, somehow we are able to understand what the author is saying and even create very vivid imagery in our minds of how these curtains are influencing the mood in the room. How is it that we process this non-literal language and seemingly intuitively understand what it is the author is trying to say?

1.1 Theoretical framework

1.1.1 Processing Language – From recognition to interpreting meaning.

In order to have a discourse with another person, whether it is face-to-face or through a written medium, we must comprehend the message they are sending. Processing language is influenced by many factors, on different levels of language representation. According to Smith and Kosslyn (2009, p484), when we have a discourse we comprehend the language at a syntactic level with *sentences* and *phrases*, at a level to encode *word* meanings through morphemes, and also through phonemes to distinguish words through *speech sounds*.

A key aspect of human language is its ambiguity (Smith & Kosslyn, 2009, p494). Every phrase, word or sound can have its own meaning, so we find this ambiguity on every level, and has to be resolved for each. This means we have to sort through multiple alternate meanings, even when only becoming aware of one interpretation, in order for us to understand what has been said (Smith & Kosslyn, pp494-495).

Understanding of language is primarily dependent on its interpreted meaning. But what about when the language is intentionally ambiguous? How is it that we still so often seem to interpret the intended meaning correctly? When speaking in metaphors, we are speaking in terms that are often not literally true, and yet we still arrive at a meaningful conclusion.

Lakoff and Johnson (1980) argue that we do not simply understand metaphors as a small part of our linguistic cognition, but that it is imperative for structuring thought process. They

write in their book that the human thought process is largely metaphorical, because metaphorical concepts structure what we do and understand. "The essence of metaphor is understanding and experiencing one kind of thing in terms of another" (Lakoff & Johnson, 1980). This is reflected in the way we seem to grasp metaphors, sometimes even more easily than literal language, from an early age. That is, the way children often express themselves through metaphor (e.g. "fire engine in my tummy") at a very young age (Winner, 1988), when they have not yet grasped the ´proper´ expressions we use to describe a phenomenon such as a stomach-ache.

Gluckenberg (2003) points out that language processing is automatic, and input will be processed no matter what. This is the case for literal language, but also appears to be so for figurative language.

1.1.2 How do we process Metaphors? - Direct- or indirect processing theories.

Supporting evidence has been found for both the direct and indirect processing of metaphors (de Grauwe et al, 2010). That is to say, whether we process them directly, same as any literal sentence, or whether we engage in *re*processing of a phrase when initial literal processing has registered as false or impossible.

Several models have been erected in order to attempt to map the processing of metaphors. Under the Standard Pragmatic Model of Metaphor (Grice, 1975) the processing of metaphors occurs indirectly, independent of context cues and metaphor type. That is to say, when processing a metaphor, the Standard Pragmatic Model of Metaphor suggests that we first search for a literal meaning and when we come to the conclusion that this interpretation would be false, we search for a non-literal meaning. Another example is the serial processing claim of hierarchical models (de Grauwe et al, 2010) in which the processing of language occurs by following a pre-set order which can be either bottom-up through feature recognition, or top-down through categorical placement.

In order to study the subject, researchers have devised different ways of testing various aspects of the cognitive processes involved. One such way is using lexical decision time, where a participant is presented with a task including a word or a sentence and asked to make a decision, usually regarding the meaning of a phrase. Depending on the experiment, the amount of time it takes for a participant to make that decision as opposed to the control condition of,

for example, a literal phrase can suggest that processing occurs differently, as it takes a different amount of time. Gluckenberg (2003) summarizes various experimental studies using lexical decision time, which found no significant deviation in the time it takes to process metaphorical expressions and literal ones, suggesting it took just as long to understand the metaphorical as it took for literal sentences. He indicates this could point to a lack of priority of the literal over the metaphorical in an indirect theory of processing, as there is no response-latency recorded in which the cognition of the metaphor would first reject the literal interpretation of the phrase (Glucksberg, 2003). Thus these experiments seem to offer counter-evidence to theories of serial processing, where the interpretation of a phrase is dependent on a fixed series of steps to determine meaning. Rather, it suggests that the processing of metaphors does *not* occur *in*directly, as there would have been a response latency recorded for the metaphorical items as opposed to the literal items.

One possible explanation is that metaphors are categorical assertions, where an aspect of the metaphor falls literally in the same category as its key component (Glucksberg, 2003); in "my job is a jail" the words 'job' shares a category with 'jail'. This theory is known as the Class Inclusion theory.

Another theory is "conceptual blending", where the identification of attributes is necessary for the comprehension of metaphor. These attributes guide movement through blended space as well as the background semantic information (Coulson & van Petten, 2002), for assigning meaning to metaphorical sentences.

Another way of studying metaphor comprehension is the Electro Encephalogram (EEG). By measuring electrical activity, it is possible to view which parts of the brain 'light up' under certain circumstances. An EEG study De Grauwe et al. (2010) measured brain activity in participants when evaluating literal and metaphoric sentences. It found that while familiar metaphors were easily accessed and mapped, there is a slight delay in accessing metaphorical meaning compared to accessing the literal, indicating that the way these two types of sentences are processed differently in some way. As such, this study also emphasizes that as the delay was so small, this does *not* point to serial processing. Rather, they suggest direct activation of the metaphorical meaning due to the metaphorical context.

So far we have established that metaphorical statements are processed differently 'in some way' from literal language, but in what way? A functional Magnetic Resonance Imaging (fMRI) –study, intended to identify neural substrates of metaphor comprehension, recorded

brain activity as participants were asked to assess whether a sentence was meaningful or not. Not only was the mean reaction time significantly longer for novel sentences when compared to literal as well as anomalous sentences, but the imaging showed activation of different areas in the brain (Shibata et al., 2009). They concluded that processing these two sentence types took different neural pathways. In other words, this study suggests that in the 'journey' through the brain to attain coherent semantic meaning, metaphoric comprehension takes a different 'route' than literal comprehension

1.1.3 Does the metaphor matter? – Differences in processing of the Novel Metaphor.

In the processing of literal language, many factors play a role in word-processing. One example is that the presence of orthographic neighbors influencing word recognition (Grainer & Dijkstra, 1996). According to Smith & Kosslyn (2009, p500), when a word relates to many other words that are in some way similar, through spelling or phonetics, they are known as having 'cohorts'. The word "marker" would have many cohorts, among which other words starting with the "ma" sound, whereas words such as "xylophone" are very specific both in phonetics and in function, and would have less cohorts. A theory known as the neighborhood density effects shows us that words with fewer neighbors are interpreted faster than words that have many neighbors, because their cohorts are automatically activated, creating a process of competition among the possibilities (Smith & Kosslyn, 2009, p500). While this effect is shown on a word level, it also applies to a sentence-level of comprehension, where activation of possibilities leads to competing interpretations of meaning. Figurative language being naturally ambiguous as it is, it seems logical for there to be an incredible amount of competing meanings. Having said this, it seems surprising that many studies using conventional metaphors have found so little difference in processing time, making us wonder if there is not another factor which influences this. Indeed, while previous studies have found little difference between the time it takes to process literal statements and familiar metaphors, novel metaphors present a different scenario.

Camp (2006) mentions many instances of findings where unfamiliar metaphors take significantly longer to process than their conventional counterparts and literal sentences, indicating different types of cognitive processing for familiar and unfamiliar metaphors. Bowdle & Gentner (1995, 1999) suggest familiar conventional metaphors as having stored meanings, whereas unfamiliar novel metaphors are processed through on-line mappings. What

happens here, that so drastically changes to cognitive process as these metaphors become more familiar to us?

The Gradient Salience Hypothesis suggests that stored information is superior to unstored information, including novel and contextual information (Gioria, 2003:15). That is to say, where highly familiar information is processed automatically, while contextual information is initially processed parallel to, rather than interacting with lexical processing in the brain.

Studies using Event Related Potentials (ERP) have found that the activation of a component later named the N400 seems an accurate measure of semantic incongruity, regardless of syntactic incongruity (Rataj, 2014). Various studies have shown that this N400 amplitude is measured at its lowest for literal sentences, higher for conventional metaphors, and at its highest for novel metaphors (Coulsen & van Petten, 2002; Arzouan, 2007). This is interesting for two reasons. Firstly, because it tells us that novel metaphors are in fact initially interpreted as incongruous, and secondly, because it indicates a difference in cognitive process depending on the novelty of the metaphor.

A study by Keysar et al. (2000) proposes that "as metaphors become lexicalized, they are no longer processed as metaphors". When we continually use once-novel metaphors, they become conventionalized and enter our 'mental dictionaries' (Glucksberg, 2003). In other words; the more often you have heard a certain metaphor, the quicker you can process it. Metaphorically speaking; if I were to put a fish into a pond every time I heard the word 'pike'[1], the pond would fill more and more with every utterance of the phrase. Thus the more I have heard the phrase, the more saturated the pond, and the easier it is to catch a fish when I try to retrieve it.

Lakoff and Johnson (1980) suggested that metaphors are "rooted in physical and cultural experience" and are "contextually, personally and culturally bound". Here we may argue that physical and cultural surroundings greatly affect the language we are surrounded by on a daily basis, and thus influences our word-recognition process. To quote Bowdle & Gentner (2005): "Whether metaphors are processed directly or indirectly, and whether they operate at the level of individual concepts or entire conceptual domains will depend both on their degree of conventionality and on their grammatical form." This indicates the allocation of meaning to metaphors through cognitive processes as being largely dependent on their conventionality.

---

[1] ENG = pike. NL = snoek. DE = Hecht.

1.1.4 What about other languages? - Bilingualism research.

Among some interest in the study of psycholinguistics, is the study of bilingualism. Increasing our understanding of bilingual parsing is beneficial for educational purposes as well as a contribution to the scientific field. Through the study of bilingual language systems, we are able to research many things about the cognition of language, such as the possibility of a language-independent underlying lexicon (Dijkstra & van Heuven, 2002).

The Bilingual Interactive Activation (BIA+)-model attempts to map bilingual language-processing through a bottom-up nature. It suggests a collaboration of a word identification system and a task/decision system is involved in bilingual word recognition (Dijkstra, van Heuven; 2002). A word identification system being an integral part of this concept, it stands to reason that especially in L2 speakers, this system is under the influence of their mastery of the language. As such, personal skill in the language should also greatly influence that person's language comprehension, and thus also the figurative language that we process through word recognition. Indeed, research has shown L2 speakers who are more skilled at L2 reading as being "more inclined to be at a higher level of metaphoric competence" (Zhao, Yu & Yang, 2014). This is not only due to the bottom-up processes taking place in the L2 speaker as he takes in the information before him, but is also influenced by top-down patterns as this speaker references his own expectations onto the phrase.

As Keckes (2006, p221) put it: "Different experience results in different salience, and second language (L2) acquisition differs from first language (L1) acquisition. Consequently, what is salient for individuals belonging to the target language community will not necessarily be salient for the 'newcomers', the L2 learners." This is also heavily dependent on their acquisition of the language and immersion into the society of the target language, as these are factors that all effect the amount and type of contact an individual has with language. One might say it is the difference between 'learning' a language and truly making it your own through 'acquisition'. So what does it matter that the L2 speaker has a different history of contact with the language than a native?

Returning to the metaphorical pond slowly filling with 'pike' each time it is mentioned; it follows that native speakers (L1) should have a more 'saturated pond' than second language (L2) speakers, as they have come into contact with certain words many more times throughout their lives. This affects the salience of certain words or phrases, which means the most salient

word in ambiguous phrases will not be the same for L1 and L2 speakers. Metaphoric phrases that are often uttered in daily life, such as "welcome aboard", will be more familiar to L1 speakers and thus more salient. Having different saliences for certain utterances will mean that the interpretation that is automatically the first one to pop into the mind of the individual, will be different for some of those phrases between L1 and L2 speakers. In other words, they may automatically attribute different meanings to ambiguous phrases. Not only can this cause misunderstandings, but it also calls into question whether or not the ideal of an acquired 'native speaker' is not wholly unattainable. Kerckes even goes so far as to suggest it might be utopian.

What of individuals that have been immersed in their L2 society, by speaking their L1 in the home environment and their L2 at, say, school? Research on children ages 3-6 suggests that despite their "equal domain general abilities for learning", non-native speakers still score consistently lower on L1 (Dutch) language assessments (Scheele, 2010). It appears there is simply more happening than simple contact with surroundings. Cooke (1997) found that L2 speakers process language differently. L2 figurative language comprehension is a complex process in desperate need of further study.

1.2 Research Questions

Findings based on N400 amplitudes point to an interesting question: what are metaphors really? Congruous phrases like the literal, that are processed differently due to whatever factors? Or are they anomalous sentences, that we somehow attribute meaning to through some internal process? And what would this say about the way our brain functions when interpreting what another person is saying to you? Will the effect be potent enough to color our daily interaction?

As accurately described in Arzouan et al. (2007): "Whereas literally related pairs benefit from all three features (*semantic relatedness, familiarity, and meaningfulness*), conventional metaphors are familiar and meaningful, novel metaphors are only meaningful, and unrelated word pairs possess none of those attributes." In order to further research metaphor comprehension, usable novel metaphors need to be tested for interpreted meaningfulness before use. Therefore, this study will function as a norming-study and attempt to create a collection of novel metaphors (and their triplets) with good construct validity, for use in further research.

Q1: Will the chosen items prove accurate predictors of meaningful novel metaphors, usable for research?

H1: Several of the items will prove usable novel metaphors.

As previously explained, L2 speakers should theoretically be less adept at identifying metaphorical meaning to sentences. In order to test this, this study will attempt to compare attributed meaningfulness between L1-speakers to that of L2-speakers.

Q2: Is there a difference in the attribution of meaningfulness of novel metaphors between native Dutch speakers (L1) and those who have learned Dutch as a non-native language (L2)?

H2: L2-speakers will attribute less meaning to novel metaphors than L1-speakers.

2. METHOD

A cross-sectional survey study was designed to measure attributed meaningfulness to several items, differing in sentence types. This will enable us to comprise a list of sentences for further research. In order to do this, several considerations were taken into account.

The survey was administered online and freely accessible for anyone who had the link. Contents were divided into three groups; (1) Demographics, (2) Word Recognition Task, and (3) Semantic Judgement Task. Existing materials were used for the creation of some of these sets.

2.1 Respondents

The target respondent pool for this study will compose of native speakers of Dutch (L1) and those who have learned the Dutch language at a later stage (L2). No specific age, schooling or other background requirements must be met. However, respondents were recruited over the age of 18. This was done in order to assure informed consent as well as to increase likelihood of a reasonably developed sense of language.

2.2 Semantic Judgment task.

In order to assess the amount of meaningfulness a participant attributes to an expression, a judgement task was created in order to assure the novelty of the used metaphors. This set was created from 66 critical words (CW's), with corresponding sentences across 3 conditions (novel metaphor, literal sentence, and anomalous sentence) to form a total of 198 sentences. Another 66 filler sentences were constructed in order to balance the amount of congruous and incongruous sentences in the stimulus set. Thus making a total item pool of 264 constructed sentences.

2.2.1 Critical words.

As stated in the introduction, the processing of novel metaphors occurs differently from conventional metaphors. In order to suit our purposes, novel metaphors will be created to match 'base' words to literal- and anomalous "A is B" sentences.

In 2010, Keuleers, Brysbaert & New developed a database called SUBTLEX-NL for the frequency of use for Dutch words, based on their appearance in subtitles. This database is freely accessible online, along with the papers published on the subject. It contains 42,729,424 words, excluding duplicates, which were processed from 8443 different subtitles. In order to create sentence sets, this database was used for the selection of critical words.

When you hear an ambiguous word such as "letter" (the written communication/the member of the alphabet), there is often a dominant meaning that comes to mind before the other. This occurs because one meaning is more 'salient' than the other. In 1979, Ortony found that the interpretation of meaning was dependent on the common properties that were more salient for the last word in a comparative metaphor than the first word. To account for this salience imbalance between the base word and the target word[2], the principal design criteria were applied to the last, or *critical words* (CW) used as a base for the creation of stimuli. One could say, interpretation of meaning is more dependent on the last word, which is why these were held to certain standards to increase the probability of creating usable meaningful metaphors. These words were all concrete nouns, 5 to 15 characters in length.

---

[2] "Only those common properties that are significantly more salient for the base concept than for the target concept will be relevant to the meaning of a metaphor." (Ortony, 1979)

Giora (2003) explains that salience as the most probably meaning to come to mind, as being influences by conventionality, familiarity, frequency or being a prototype. In order to ensure comparable salience for the CW in all of the sentences had a fixed range of 20-70 words per million in the SUBTLEX-NL 'corpus'. Frequency of occurrence was assessed through the corpus database, which was built through subtitle analysis (Keuleers, Brysbaert & New, 2010)[3], and chosen between a range of 20-70 words per million. Compound nouns were excluded to promote ease of processing.

The remaining word used to make a sentence with a metaphorical meaning, or the 'target' stimuli need not be held to any specific standards, other than that they do not also appear in the overall list of critical words.

2.2.2 Sentence construction

An "A is B" syntactic structure (copula sentence) was used to create phrases in all three conditions. This was done to ensure data not be corrupted by contaminants produced by participants having to process a more complex structure. In the A is B structure, "B" is the CW, and "A" is the target.

In order to keep the syntactic structure as simple as possible, and comparable across conditions, no use was made of negative coding (*A is not B*). To combat automatization of response (due to repeated exposure to identical sentence types) or acquiescence response, and break up the structural monotony, filler sentences were composed of exclusively different syntactic structure. This combats response bias. See appendix B for the list of filler sentences.

For the novel metaphors, it should not be possible to be both somehow literally true and metaphorically interpreted. The anomalous sentences were created to be syntactically correct and semantically meaningless. This was done to ensure the processing of the 'false' sentence was triggered by the meaning of the sentence and not any other linguistic errors (such as syntax, vocabulary, etc.).

Other prime considerations taken into account were:

Orthography: All sentences were presented in the same short sentence form (starting with a capital letter and ending with a period) with correct spelling and hyphenation.

---

[3] Found at SUBTLEX-NL

Sentence Length: All sentences were monitored, during construction, for average length. This was done by adding an in-program formula which calculated the situational 5-number-descriptions. With this formula in place, it was possible to have the Quartiles and averages be comparable across situations, thus minimizing the probability of a processing contaminant due to sentence length.

Phonetic priming: No words that rhymed or started with the same letters were used in any one constructed sentence. This was done to avoid phonetic priming bias.

(Interlingual) Homographs and cognates: Care was taken to avoid using words that have more than one meaning. This was done both within the Dutch language (e.g. "gast"[4]), and between the languages most likely to contain interlingual homographs (e.g. "spin[5]"). Interlingual ontographic neighbors were avoided in the same manner across Dutch, English and German regarding spelling and phonetics (e.g. "police"[6]).

2.2.3 Counterbalancing.

Of the 264 constructed sentences, the 66 critical words were each placed into one of three sentence conditions (novel metaphor/literal/anomalous) and presented to the participants. Each CW was only used once, to avoid priming-effects. Of the 66 constructed filler sentences, 22 were randomly picked for counterbalancing, only accounting for comparable sentence length. Thus, participants were presented with 88 stimuli (22 novel metaphors, 22 literal sentences, 22 nonsensical phrases, 22 filler sentences) and asked to decide whether or not the presented stimuli conveyed a meaningful expression.

---

[4] Dutch word meaning "guest" as well as "dude/man".
[5] Meaning "to turn around one's own axis" in English, and meaning "spider" in Dutch.
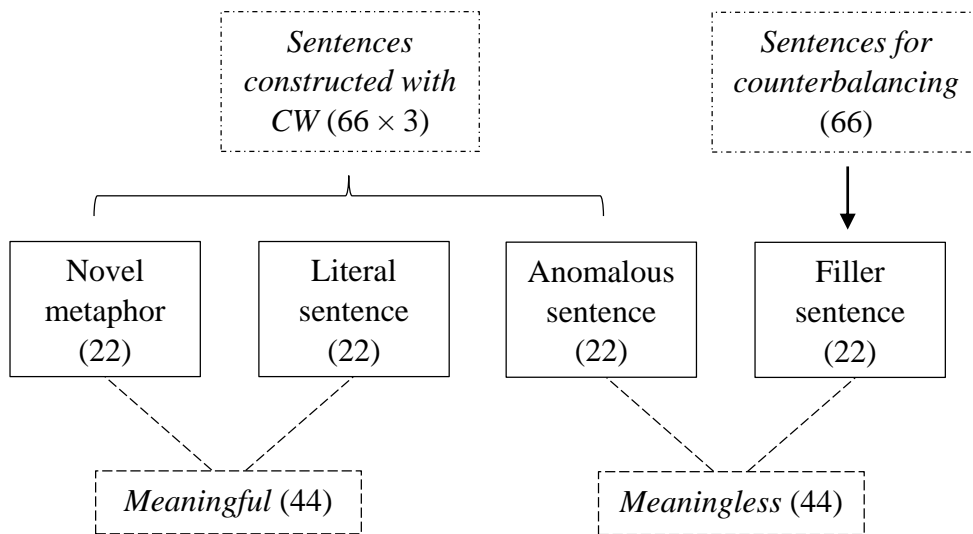[6] ENG= police. NL= politie. DE= polizei.

FIGURE 2.1: Visual representation of the use of constructed items.

2.3 Research Design

2.3.1 Demographics

At the start of the survey, several demographic questions were included in order to accurately map the respondent pool, and asses its generalizability. Alongside the standard 'age' and 'sex', some information was needed.

To categorize the participants as L1- or L2-speakers, we asked them their native language. If the participant was a non-native speaker, they were asked to disclose their native language, as well as whether or not the language was learned during the sensitive period in language development[7].

In order to determine language dominance, we enquired as to their contact with the language. Questions were designed to determine linguistic surroundings[8], native language type[9], and linguistic (e.g. phonetic) similarities with other spoken languages[10].

---

[7] At what age did you start learning Dutch?
[8] Do you live in the Netherlands at this time?
[9] What is your native language?
[10] What other languages do you speak?

2.3.2 Word Recognition Task

In order to gain insight into participant language skill, we presented a previously standardized word recognition task "LEXTALE", based on a large scale study by Lemhöfer & Broersma (2012). This was used as a test of general language proficiency, and indicator of personal item difficulty.

Lextale is a 60-item, dichotomous judgement task in which respondents attempt to assess whether the items they are presented with are real words in the Dutch language. Based on whether or not they answer these correctly, the researcher is able to calculate a score by using a given formula. Lextale has been scientifically tested for validity by the authors.

2.3.3 Procedure

Participants were asked to follow a link to the online survey. Interviewer effects should be minimal, as the online nature called for self-completion of the questionnaire.

Within the online survey respondents were presented with a 3-part questionnaire; (1) demographics, (2) word recognition task, (3) semantic judgement task. Instructions were given at each interval, to make the survey as clear as possible for all respondents, especially considering a large part of the respondents are L2 speakers who will be reading the instructions in their second language. Note that the word 'metaphor' was never explicitly mentioned in the instructions, so as not to prime the participants.

Appendix D: copy of survey

Stimulus Presentation

In order to reduce the risk of contaminating stimuli, the format was kept as simple as possible. The questionnaire was formatted to be simple black standard lettering, with no other stimuli offered. The survey software provided required the use of a "University of Twente" template (see appendix B) which remained exactly the same throughout the survey. Since the tasks offered are not timed, respondents were free to assess the stimuli at their leisure, making it a self-paced reading task and reducing any stress it may put on participants.

Demographic questions employed as little open-ended questions as possible, in order to reduce interpretation-bias.

The word recognition task was presented in accordance with the instructions in the original Lextale article. Participants were given 3 practice items, after which they were presented with one word per page, with the possibility of selecting "yes" or "no" for word evaluations.

The semantic judgement task was presented in a 7-point Likert format. Participants were asked to make judgements on items from 'very meaningful' to 'very meaningless'[11]. In order to avoid exceptionally long item lists, the 88 items of the semantic judgement task were spread out over 3 pages, with recurring 7-pt choice options.

Distribution

Participants were recruited through several means; (A) Several other people were recruited from among the student body of the University of Twente, and were rewarded with credits for an in-university research-participation system. Aforementioned system was run with the online company SONA-systems, where participants can choose studies to sign up for based on personal selection criteria and time-input based reward credits. (B) The remaining people participated by internal motivation, through either network-sharing on social-networking-sites (meaning the researcher and several random acquaintances with diverse social circles shared the survey-link), posting on relevant fora such as "Linguist List", or targeted e-mail sent to organizations by the researcher (this was done mostly to get in touch with L2 speakers). These people did not receive a reward for participating in the survey.

Software

The online survey-making-software "Qualtrics" was used[12] to format and release the survey to participants.

Data-analysis was conducted through use of Microsoft Excel, and IBM's analysis software SPSS Statistics 23.

---

[11] 1) Very meaningless. 2) Meaningless. 3) Somewhat meaningless 4) Neither meaningless nor meaningful. 5) Somewhat meaningful 6) Meaningful. 7) Very meaningful.
[12] Operating on a license from the University of Twente

2.3.4 Pilot study

Before the survey was launched, an informal trial run occurred where several persons unrelated to the study were asked to evaluate the survey for syntactical correctness and clearness of instructions. Their comments were processed, and a final version was uploaded for data-collection.

3. RESULTS

Cases were discarded if the survey was not fully completed.

3.1 Demographic variables

Raw data showed 181 individuals took part who completed the survey, both male (57) and female (124). Age group varied between 17 and 80 (5nr summary: 17; 22; 25; 45,50; 80)(mean=33,06; σ =15,556). This population was divided into L1 "native speakers" (n=127), and L2-population of "second-language speakers" (n=54). Outliers were detected with the quartile-method based on irregularities in their scoring patterns. Participants that were outliers in average scoring in any of the sentence categories were trimmed.

After these outliers had been trimmed 172 remained, 52 male and 120 female, between the ages of 17 and 80 (mean=32,85; σ =1,184). Among the L1 speakers (n=121), were men (37) and women (84) between the ages of 17 and 80 (5nr summary 17; 22; 25; 48,50; 80) (mean=33,89; SE=1,466). Among the L2 speakers (n=51), were men (15) and women (36) between the ages of 18 and 71 (5nr summary 18; 21; 25; 34; 71) (mean=30,37; SE=1,941).

Analysis of the constructed sentence items revealed mean differences between 0 - 0,23 upon removal of outliers.
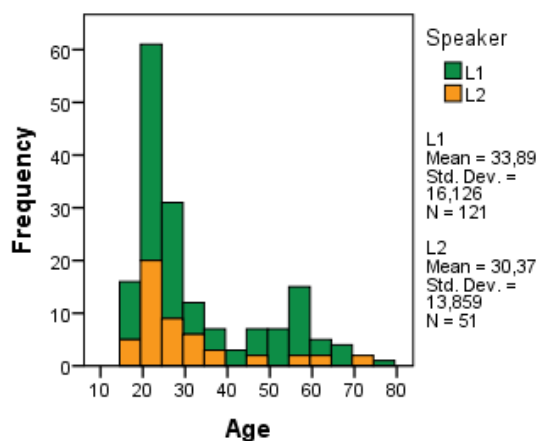
Age of participants by native language.   *Sex of respondents by native language.*
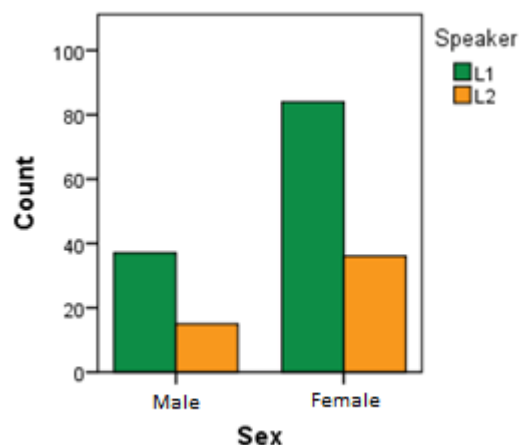
3.2 Creation of a usable item-set for further research.

Items were numbered for display purposes and to ensure 'blind'-analysis. Item numbers represent the page they were on, and the individual item number. For example; item 2-18 would be the 18th item on the second page. As items were randomized *by* and *within* pages, this number does not represent anything but simple item ID. The nature of the sentence was noted through abbreviations such as "NM" for Novel Metaphor, "LS" for Literal Sentence, "AS" for Anomalous Sentence and "FS" for Filler Sentence.

When the data was categorized into L1 and L2 speakers, results showed that the appropriate sentences used differ per target audience. Analysis showed this difference to be non-significant, both with asymmetrical L1 and L2 populations (U=1084,5 ; p=0,663) and with comparable population sizes (U=423 ; p=0,572). However, since the base condition of further research is to compare L2 interpretations to L1 standards, the following item analysis was conducted on L1 participants.

3.2.1 Novel Metaphor

In the novel metaphor condition, there were no items that would increase reliability upon deletion.

Items 1-8, 3-2 and 3-3 displayed a strong right-sided skewness in scoring. Items 1-3, 1-4, 1-5, 1-6, 1-7, 2-1, 2-2, 2-3, 2-6, 3-4, 3-6 and 3-7 displayed some right-sided

item	skewness	kurtosis
1-8	-1,524 (SE=,224)	2,361 (SE=,444)
3-2	-1,265 (SE=,221)	1,756 (SE=,438)
3-3	-1,024 (SE=,218)	-,163 (SE=433)

skewness. This is reflected in the 5-nr-summaries and indicates that these were the novel metaphors interpreted as meaningful. This indicates participants regularly scored these items as 'meaningful'.

Items 1-1, 1-2 and 3-1 showed a strong left-sided skewness in scoring. Items 2-4, 2-5, 2-7 and 3-5 showed some right-sided skewness. These items scored very low on

item	skewness	kurtosis
1-1	,752 (SE=,218)	-,760 (SE=,433)
1-2	1,169 (SE=,221)	-,133 (SE=,438)
3-1	1,807 (SE=,239)	2,750 (SE=,474)

attributed meaningfulness, and are less suited for further use.
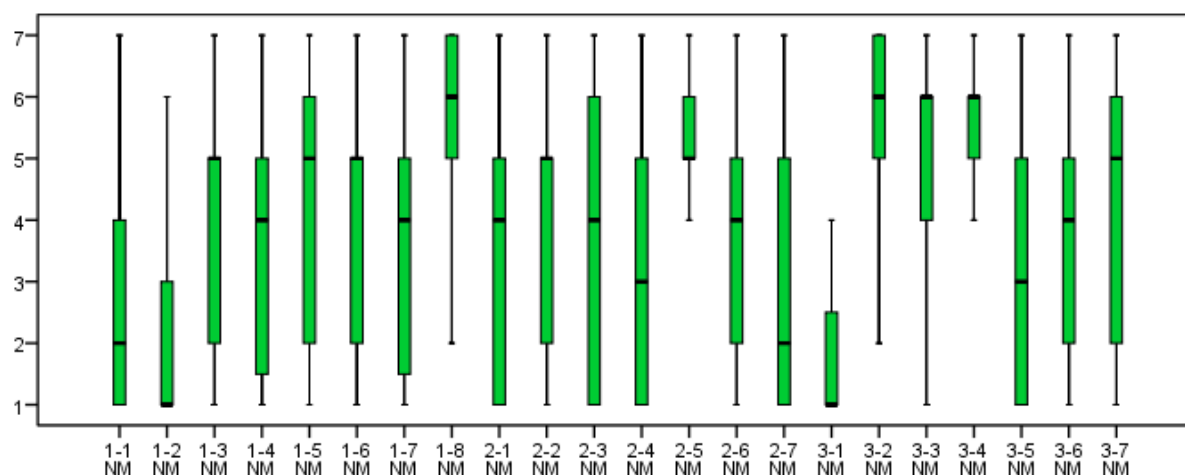
Figure 3.1: Boxplot of L1 speakers' rating of meaningfulness on Novel Metaphors.

Notably; on inspection of the item correlation matrix items 1-1 and 1-2, which have previously been flagged as undesirable due to low scoring on attributed meaningfulness, showed several positive correlations flagged as 'desirable'. However, the strongest correlations were still with non-meaningful items. Item 1-2 (NM) showed strong positive correlation with 'meaningless'items 1-18(AS)(r=0,75), and 1-29 (FS)(r=0,75).

Item 3-1(NM) correlated very strongly to item 2-25(FS)(r=0,934), but also correlated well to 'desirable' 3-3(NM)(r=0,775). 3-4(NM) tot 2-15(AS)(r=0,774) and 3-5(NM) to 2-

7(NM)(r=0,842). 'Undesirable' items 2-7(NM) and 3-5(NM) were also strongly correlated (r=0,842).

Item 1-4(NM) correlated strongly to item 3-7(NM)(r=0,727). Item 1-5(NM) had a strong negative correlation with non-meaningful item 3-28(FS)(r= -0,849). Item 1-7(NM) correlated strongly with other meaningful sentences such as 1-10 (LS)(r=0,734), 2-11(LS)(r=0,786), 2-12(LS)(r=0,887), 3-11(LS)(r=0,706) and 3-13(LS)(r=0,814). As did item 1-8(NM) with 2-2(NM)(r=0,741) and 2-14(LS)(r=0,746). Item 2-2(NM) correlated strongly to 1-8(NM)(r=0,741) and 3-3(NM)(r=0,803). Item 2-3(NM) was strongly correlated to 1-16(LS)(r=0,706). Item 3-3(NM) correlated strongly to 2-2(NM)(r=0,803) and 3-1(NM)(r=0,775).

				SK*	r1**	r2***
Novel Mataphors	1-1	Ideeën zijn brand.	Ideas are fire.****	✖		2✖
	1-2	Plannen zijn voeten.	Plans are feet.	✖		
	1-3	Zuchten zijn tranen.	Sighs are tears.	✔		
	1-4	Je gezicht is een krant.	Your face is a newspaper.	✔	✔	
	1-5	De tijd is een schrijver.	Time is a writer.	✔	✔	
	1-6	Je karakter is een zwaard.	Your character is a sword.	✔		
	1-7	Afwijzingen zijn moorden.	Rejections are murders.	✔	5✔	
	1-8	Je kinderen zijn een spiegel.	Your children are a mirror.	✔	2✔	
	2-1	Applaus is een regen.	Applause is a rain.	✔		
	2-2	Golven zijn stemmen.	Waves are voices.	✔	2✔	
	2-3	Grappen zijn kogels.	Jokes are bullets.	✖	(-)	
	2-4	Mobieltjes zijn muren.	Cellphones are walls.	✖		
	2-5	Geheimen zijn een ziekte.	Secrets are a disease.	✖		
	2-6	Je woorden zijn je kleding.	Your words are your clothing.	✔		
	2-7	Gevoelens zijn advocaten.	Feelings are lawyers.	✖	(-)	
	3-1	Roest is een horloge.	Rust is a watch.	✖	(-)	✖
	3-2	Een date is een proef.	A date is an experiment.	✔		
	3-3	Kansen zijn sleutels.	Chances are keys.	✔	(-),✔	
	3-4	Dokters zijn bewakers.	Doctors are guards.	✔		✖
	3-5	Democratie is een winkel.	Democracy is a store.	✖	(-)	
	3-6	Meningen zijn spelletjes.	Opinions are games.	✔		
	3-7	Een enigskind is een eiland.	An only child is an island.	✔		

* SK, item suitability based on skewness and kurtosis measures for distribution.
** r1, item suitability based on positive correlations with other meaningful items, or negative correlations with meaningless items above 0,7. Items marked with "(-)" has a positive connection to another item previously flagged as unfavorable.
*** r2, item unsuitability based on positive correlations with other meaningless items, or negative correlations with meaningful items above 0,7.
**** Untested translations of items, added for legibility of results.

TABLE 3.1: Item suitability of novel metaphors, where a green check (✔) indicates a desirable item characteristic and a red cross (✖) indicates an item characteristic as undesirable.

3.2.2 Other Sentences

In the literal sentence condition, there was one item (2-8) that would increase reliability upon deletion. All items skewed to the right, indicating high attributed meaningfulness on the Literal items. Only item 1-12 showed

item	skewness	kurtosis
1-12	-,095 (SE=,181)	-1,493 (SE=,359)
1-10	-,670 (SE=,190)	-,828 (SE=,377)
2-9	-,812 (SE=,181)	-,543 (SE=,359)

unfavorable answer distribution. Items 1-10 and 2-9 showed questionable favorability with favorably scored but also widely varied attributed meaningfulness. No significant difference between L1- and L2-speakers was found in this regard.

Literal items 1-9, 1-10, 1-11, 1-13, 1-14, 1-15, 1-16, 2-9, 2-8, 2-10, 2-11, 2-12, 2-13, 2-14, 3-8, 3-9, 3-11, 3-12 and 3-13 correlated well with other 'meaningful' items, or had a very strong negative correlation with a meaningless item. Literal item 3-10 correlated well with 'meaningless' item 2-15.

In the anomalous sentence condition, there were no items that would increase reliability upon deletion. All items skewed to the left, indicating low attributed meaningfulness on the Anomalous items. Items 1-19, 2-15, 3-19, 3-20 and 3-21 showed questionable favorability with

item	skewness	kurtosis
1-19	,780 (SE=,181)	-,932 (SE=,359)
2-15	,495 (SE=,181)	-1,187 (SE=,359)
3-19	,832 (SE=,181)	-,782; SE=,359)
3-20	,757 (SE=,181)	-,818 (SE=,359)
3-21	,527 (SE=,181)	-1,199 (SE=,359)

favorably scored but widely varied attributed meaningfulness. No significant difference between L1- and L2-speakers was found in this regard.

Anomalous items 1-17, 1-18, 1-22, 2-15, 2-17, 2-19, 3-16, 3-17, 3-18 and 3-19 correlated well with other 'meaningless' items, or had a very strong negative correlation with a 'meaningful' item.

The principal purpose of the filler sentences is combating asymmetrical item presentation, making the exact phrases largely irrelevant.

Appendix (D) shows overview-tables regarding these statistics.

3.3 L1/L2 discrepancy.

For these calculations, **comparable groups** were necessary between L1 and L2 speakers. The L1-population was trimmed to a comparable number (n=57), by selecting the L1-participants who most accurately mirrored the demographic variables of the L2-population. Participants were trimmed based on age, gender, schooling and number of languages spoken to attempt to mirror personal demographic backgrounds.

Since the distribution did not appear to be normally distributed, a chi-square analysis was run, rather than an ANOVA.

In order to run a chi-square test, several assumptions must be met. One of these is; "No more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater" (Yates, Moore & McCabe, 1999, p. 734). The Chi-squared analysis that ran, violated the assumptions of the the contingency table that tests this assumption. However, it showed a likelihood-ratio of 0,007 suggesting that there is a difference between L1-speakers and L2-speakers when attributing meaningfulness to Novel Metaphors.

3.4 Internal consistency

Initial analysis showed strong inter-condition consistency, for Novel Metaphors of ($\alpha = 0,920$), Literal Sentence ($\alpha = 0,902$), Anomalous Sentence ($\alpha = 0,906$) and Filler Sentence ($\alpha = 0,889$).

3.4.1 Possible confounding factors

Lextale score showed minor correlations to scoring on Novel Metaphors (r= -,204; *p*=,032) for all respondents. However, this correlation was higher within the L2-population (r= -,538; *p*=0,003). A two-way ANOVA showed that there was a significant effect of Lextale score on the average scoring of Novel Metaphors ($F_{(31,73)}$=2,048; *p*=,006), as well as an interaction effect of Lextale-score coupled by nativity of the speaker[13] ($F_{(9,73)}$=2,245; *p*=0,028).

A two-way ANOVA was also conducted within the L2-speakers that examined the effect of learning age and Lextale-score on the rating of Novel Metaphors. There was no

---

[13] L1 or L2

statistically significant interaction between the effects of learning age and Lextale score on the attribution of meaning to Novel Metaphors, $F(6, 1) = 4,130$, $p = 0,360$.

4. DISCUSSION

4.1 Usable item set

The initial analysis of data-distribution showed some items that should really be omitted from further use, as they were not an accurate measure of meaningful metaphor. Other items proved to be highly suitable as meaningful metaphors.

Taking priming into account, having every critical word in the stimulus set only once resulted in only a third of the total sentences constructed based on the critical words being tested in this survey. (A fourth of total constructed sentences, if filler sentences are taken into account.) Due to the time-restrictions placed on this research project, only the current set was distributed and analyzed. In order to create the most optimal set of 'meaningful metaphors' the remaining 3/4 of the set should be tested and analyzed in a similar manner. Comparing the results of these tests to each other should indicate which CW would be best suited for research in which sentence condition.

If only few items are necessary and a smaller set is sufficient, the table below ranks the most favorably tested novel metaphors of the current set. As you can see, items 1-1, 1-2, 2-3, 2-4, 2,7, 3-1 and 3-5 were omitted from the original set, freeing up the CW's for use in other sentence types in future research using this set. Appendix E shows tables with selection criteria for all four sentence types.

Item-nr	Novel Metaphor (Dutch)	Novel Metaphor (Untested translation)
1-8	Je kinderen zijn een spiegel.	*Your children are a mirror.*
3-2	Een date is een proef.	*A date is an experiment.*
3-3	Kansen zijn sleutels.	*Chances are keys.*
1-7	Afwijzingen zijn moorden.	*Rejections are murders.*
2-2	Golven zijn stemmen.	*Waves are voices.*
1-5	De tijd is een schrijver.	*Time is a writer.*
1-4	Je gezicht is een krant.	*Your face is a newspaper.*
1-3	Zuchten zijn tranen.	*Sighs are tears.*

1-6	Je karakter is een zwaard.	*Your character is a sword.*
2-1	Applaus is een regen.	*Applause is a rain.*
2-6	Je woorden zijn je kleding.	*Your words are your clothing.*
3-6	Meningen zijn spelletjes.	*Opinions are games.*
3-7	Een enigskind is een eiland.	*An only child is an island.*
3-4	Dokters zijn bewakers.	*Doctors are guards.*

TABLE 4.1: Novel Metaphors suitable for further use, ranked in order of usefulness.

4.2 L1/L2 discrepancy of attributed meaningfulness

As the assumptions of the chi-square tests proved violated, and no normal distribution observed, one must question the validity of drawing any confirmatory conclusions. However, with a likelihood ratio as low as it was there are strong indications that this could still be confirmed. Since this is an explorative norming study, while the hypothesis cannot be conclusively accepted, the data provides positive indications in favor of the hypothesis.

4.3 Metaphoric competence in L2 speakers of Dutch

The earlier research by Zhao, Yu & Yang (2014) that suggested a positive correlation between L2-speakers' metaphoric competence and their reading skill. However, interestingly enough results showed quite the opposite, as was supported by the interaction effect between the score on the word-recognition task and the (non-)nativity of the speaker.

Where L1 speakers generally scored rather highly on the word recognition task, L2 scores varied more. This is to be expected for many reasons. For one, L2 speakers may be in different learning or acquisition stages.
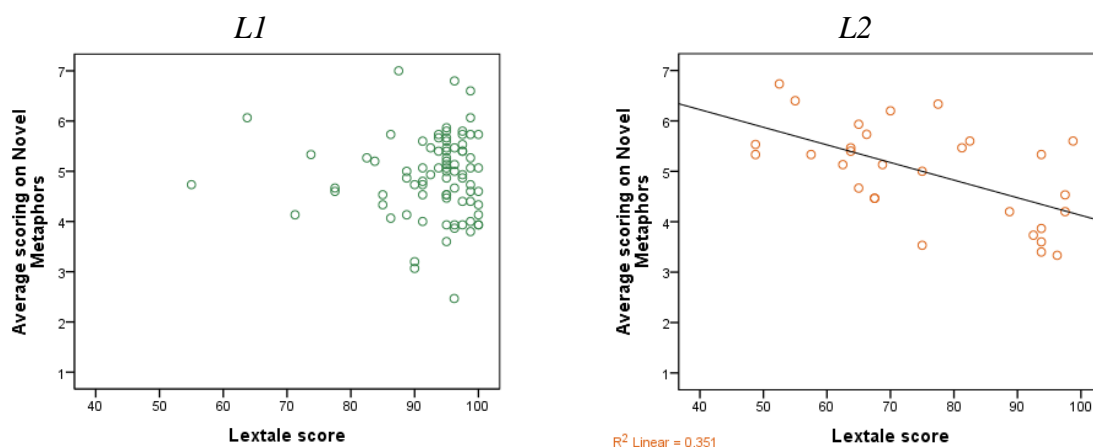
The study by Zhao, Yu and Yang was done on Chinese L2 learners. Many researchers have argued for and against the existence of an integrated underlying lexicon. If an integrated lexicon exists a lexical decision system suggests much competition of language components will make it difficult for L2 speakers to process second languages. If we have a separate lexical system per language, competition will be limited, as it only requires selection in the lexicon of the relevant language (Dijkstra & van Heuven, 2002). What does this have to do with our research findings? If there is no underlying lexicon; linguistic studies will only be comparable when preformed in the same language. A study done on Chinese participants with no knowledge of Dutch will not be relevant for Dutch language-research. That is why bilingual language studies conducted among the same language are probably preferable as theoretical bases.

Another possible factor that could have played a role is the native language of the L2 speakers of Dutch. No constraints were placed on native languages, meaning there was no consistent difference between type of languages, meaning the L2 participants were from all over the world. It is entirely possible that L2 speakers of Dutch who are L1 speakers of German have a very different scoring relationship between attributing meaningfulness and language competence, than L1 speakers of Chinese do. After all, German and Dutch, both being anglo-saxon languages with the same alphabet and many similar syntactical rules, may have an entirely different process of 'translation' when compared to Chinese and Dutch, which are dissimilar in many ways. This hypothesis could be explored with current data, as participants were asked to supply their native language as well as other languages they spoke, however, due to time constrains it was impossible to include in this report.

Furthermore; the difference in testing methodology could also explain the difference in results. Firstly, Lextale is a scoring measure for linguistic competence, not metaphoric competence. Metaphoric competence is a complex theory made up of many (some as of yet unknown) factors. It is plausible that a simple word-recognition task is not an adequate measure for metaphoric competence. Secondly, this study employed a self-paced task with no motivation for a speedy interpretation, where the task used in the research by Zhao, Yu & Yang (2014) placed a time-limit on their task. Many studies have shown time constraints to act as a stressor cognitive processes, altering the decision system. This may have contributed to the differences found during the current analysis.

Rather than indicating either study has some flaw or drew incorrect conclusions, it is far more likely that the methods are simply too dissimilar to produce the same results. It is still highly likely that linguistic competence, metaphoric or not, plays a role in L2 metaphor

interpretation. In L1 speakers, findings by Cooke & Bartha (1992) showed psychology students use of metaphors actually increasing with expertise. It seems that with better understanding, comes better judgement of metaphor, leading to higher evaluations of meaningfulness. Why should this be any different for L2 speakers? Future research on the field may prove very interesting.

4.4 Further recommendations

Creation of a *standardized* set of sentences for linguistic cognition research involving *novel* metaphors is inherently impossible. Though the creation of such a set through pilot-study is perfectly useful, one has to be wary not to apply to the same population in the actual experiment. As the phrases lose their novelty, they lose their capacity for truly novel cognition within the same population.

Though this test did adequately create several items for use as meaningful metaphors, no such guarantee has been made for the perceived novelty of these metaphors by the participants. Since perceived familiarity is such a decisive influence on the participants' cognitive processing of the items, a test to ensure the novelty of the metaphors should be executed with the items flagged as 'meaningful' before use in decisive research.

The way we comprehend metaphors is as much a process of subjective perception as it is of transferal from the source. This subjective perception is guided by our own thought-patterns. Abstract thought patterns in particular are shown to be related to metaphor use (Lakoff, 2014). It even appears that abstract domains like time are shaped more by metaphorical mappings than by concrete (Boroditsky, 2000), meaning that we use metaphors to structure highly abstract concepts. So does the capacity for abstract thought directly influence metaphor comprehension? In order to be able to rule this out as a confounding factor, further analysis is needed. It is possible to analyze an aspect of participant's capacity for abstract thinking, by coding the target words in the constructed sentences into 'abstract' and 'concrete' word categories. Since all the critical words in the sentences are concrete, and the syntactic structure is simple as well as identical for every sentence, the type of target word becomes a variable subject to analysis. So, by coding the target words as either 'abstract' or 'concrete', we may cross-reference this new variable with the average scoring on items within sentence-conditions. If these two factors correlate to each other, we will have removed a possible confounding factor,

and shown its effects on metaphor comprehension. This analysis was not completed due to time-constraints on the study.

5. CONCLUSION

A set was created for possible further use in Novel Metaphor research. Furthermore, the results of this study lend credibility to the hypothesis that L1 and L2-speakers process non-literal language differently and are less likely to attribute meaning to metaphors, especially when there is low reading competence. Recommendations have been made for further research, both research using already collected data or constructed stimuli, and simple areas of interest.

6. REFERENCES

Arzouan, Y., Goldstein, A. & Faust, M. (2007) Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, 1160, 69-81.

Berk, L.E. (2009). *Child Development*. (8th ed.) Boston, MA: Pearson Education, Inc.

Boroditsky, L. (2000) Metaphoric Structuring: Understanding time through spatial metaphors. *Cognition* 75, 1-28.

Bowdle, B. F. & Gentner, D. (2005). The Career Of Metaphor. *Psychological review*, 112, 193-216.

Caillies, S. & Declereq, C. (2011) Kill the Song – Steal the Show: What does Distinguish Predictive Metaphors From Decomposable Idioms? *Journal of Psycholinguistic Research*, 40, 205-223.

Camp, E. (2006) Metaphor in the Mind: The Cognition of Metaphor. *Philosophy Compass*, ½ 2006, 154-170.

Cooke, N.L. & Bartha, M.C. (1992) An Empirical investigation of Psychological Metaphor. *Metaphor and Symbolic Activity*, 7(3-4), p215-235.

Coulson, S., & van Petten, C. (2002) Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, 30, 958-968.

Dijkstra, T., van Heuven, W.J.B. (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175-197.

Dussias, P.E. (2003) Syntactic Ambiguity Resolution in L2 Learners: Some Effects of Bilinguality on L1 and L2 Processing Strategies. *SSLA*, 25, 529-557.

Glucksberg, S. (2003) The psycholinguistics of metaphor. *TRENDS in Cognitive Sciences*, 7(2), 92-96.

Goldstein, A., Arzouan, Y. & Faust, M. (2012) Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. *Brain & Language*, 123, 137-142.

Grainer, J. & Dijkstra, A. (Ton) (1996) Visual word recognition. In A. (Ton) Dijkstra & K. de Smedt (eds.), *Computational psycholinguistics: AI and connectionist models of human language processing*, pp.139-165. London: Taylor & Francis.

de Grauwe, S., Swain, A., Holcomb, P.J., Ditman, T. & Kuperberg, G.R. (2010) Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, 48, 1965-1984.

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3), 643-650.

Keysar, Boaz,Yeshayahu Shen, Sam Glucksberg, & William S. Horton. (2000) "Conventional Language: How Metaphorical Is It?" *Journal of Memory and Language*, 43, 576–93.

Lai, V.T., Curran, T. & Menn, L. (2009) Comprehending conventional and novel metaphors: An ERP study. *Brain Reseach*, 1284, 145-155.

Lakoff G., Johnson M. (1980). *Metaphors We Live By*. Chicago And London: Chicago U. P.

Lakoff, G. (2014) Mapping the brain' metaphor circuitry: metaphorical thought in everyday reason. *Frontiers in Human Neuroscience*, 8, 987.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325-343.

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, 32, 533-550.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 12-31.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, 86, 161-180.

Rataj, K. (2014) Surfing the Brainwaves of Metaphor Comprehension. *Poznan Studies in Contemporary Linguistics*, 50(1), 55-73.

Shibata, M., Abe, J., Terao, A. & Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: An fMRI study. *Brain Research*, 1166, 92-102.

Smith, E.E., Kosslyn, S.M. (2009). *Cognitive Psychology: Mind and Brain* (1st ed.) New Jersey: Pearson Education, Inc.

Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.

Yates, D., Moore, Moore, D., McCabe, G. (1999). The Practice of Statistics (1st Ed.). New York: W.H. Freeman.

Yang, F.G., Bradley, K., Huq, M., Wu, D. & Krawczyk, D.C. (2013). Contextual effects on conceptual blending in metaphors: An event-related potential study. *Journal of Neurolinguistics*, 26, 312-326.

Zhao, Q., Yu, L., & Yang, Y. (2014). Correlation between receptive metaphoric competence and reading proficiency. *English Language Teaching*, 7(11) pp168-181.

7. APPENDICES

Appendix A: Sentence triplets using A=B structure.

Table 7.1: Sentence triplets using A=B structure.

Novel metaphor	Literal categorization statement	Semantically anomalous sentence
Je woorden zijn je kleding	Die deken is kleding	Paarden zijn kleding
Aarde is vlees	Dit eten is vlees	Kunstgras is vlees
Verandering is een feestje	De bar is een feestje	Je teen is een feestje
Kansen zijn sleutels	Die ijzers zijn sleutels	Wielrenners zijn sleutels
Jeugd is ontbijt	Deze chips is ontbijt	Kiezelstenen zijn ontbijt
Angsten zijn tanden	Deze botten zijn tanden	Muismatten zijn tanden
Geheimen zijn een ziekte	Ouderdom is een ziekte	Een poster is een ziekte
Kernfusie is een bruiloft	Deze fuif is een bruiloft	Deze sok is een bruiloft
Suggesties zijn moeders	Deze piloten zijn moeders	Glasplaten zijn moeders
Ethiek is een vloer	Dit hout is een vloer	Een lamp is een vloer
Een groet is een taart	Die snack is een taart	Een deken is een taart
Filosofie is een boerderij	Dat erfgoed is een boerderij	Een etmaal is een boerderij
Religies zijn huizen	Die tenten zijn huizen	Muggen zijn huizen
Meningen zijn spelletjes	Die dozen zijn spelletjes	Asbakken zijn spelletjes
Verlegenheid is maagd	Deze non is maagd	Een eetschema is maagd
Emoties zijn een schilderij	Die bol is een schilderij	Traagheid is een schilderij
Afwijzingen zijn moorden	Deze daden zijn moorden	Kuikens zijn moorden
Het voorjaar is een verjaardag	Die datum is een verjaardag	Een balkon is een verjaardag
Merken zijn vijanden	Die boeren zijn vijanden	Kapstokken zijn vijanden
Kou is een eigenaar	Deze arts is de eigenaar	Thee is je eigenaar
Wolkenkrabbers zijn bomen	Zaden zijn bomen	Vogels zijn bomen
Applaus is een regen	Dit water is regen	Brillen zijn regen
Gevoelens zijn advocaten	Deze boeven zijn advocaten	Spijkers zijn advocaten
Een rijbewijs is een begrafenis	Deze casus is een begrafenis	Een beker is een begrafenis
Muziek is een vergadering	Deze zit is een vergadering	Een eend is een vergadering
De dood is kampioen	Die archeoloog is kampioen	Papier is kampioen
Waarheden zijn slachtoffers	Die agent is slachtoffer	Druiven zijn slachtoffers
Feiten zijn broers	Die kappers zijn broers	Printers zijn broers
Roest is een horloge	Die klok is een horloge	Een stripboek is een horloge
Kusjes zijn medicijnen	Planten zijn medicijnen	Nietjes zijn medicijnen

Bossen zijn lichamen	Die vormen zijn lichamen	Bureaus zijn lichamen
Mechanismen zijn hersenen	Die structuren zijn hersenen	Stoelen zijn hersenen
Behoeften zijn klanten	Die golfers zijn klanten	Bladeren zijn klanten
De wieg is een ochtend	Deze tijd is de ochtend	Het plafond is de ochtend
Plannen zijn voeten	Die sporen zijn voeten	Stickers zijn voeten
Jaloezie is drank	Bloed is een drank	Elastiekjes zijn drank
Het leven is een keuken	Dit deel is een keuken	Een sateliet is een keuken
Je bloedcellen zijn soldaten	Deze neven zijn soldaten	Tafels zijn soldaten
Democratie is een winkel	Dit pand is een winkel	Een spuitbus is een winkel
Je verstand is je getuige	Die juf is een getuige	Zwavel is een getuige
Fondsen zijn voedsel	Koeien zijn voedsel	Een fluit is voedsel
Je kinderen zijn een spiegel	Dit product is een spiegel	Een driehoek is een spiegel
Je gezicht is een krant	Deze website is een krant	De rails zijn een krant
Een enigskind is een eiland	Dit gebied is een eiland	Deze meter is een eiland
Ideeën zijn brand	Dit fenomeen is een brand	De afwas is een brand
Golven zijn stemmen	Die geluiden zijn stemmen	Armleuningen zijn stemmen
Je brein is een rechercheur	Die man is een rechercheur	Een etui is een rechercheur
Je karakter is een zwaard	Dit ding is een zwaard	Deze ezel is een zwaard
Een behoefte is een klant	Deze chef is een klant	Een appel is een klant
Pijn is een leraar	Die kok is een leraar	Een gordijn is een leraar
Warmte is een cadeau	Die bakfiets is een cadeau	Een nagel is een cadeau
De tijd is een schrijver	Die gevangene is schrijver	Een markt is een schrijver
Een relatie is een kasteel	Dit monument is een kasteel	Een konijn is een kasteel
Mobieltjes zijn muren	De grens is een muur	Een tractor is een muur
Handen zijn boeken	Deze werken zijn boeken	Toiletten zijn boeken
Dokters zijn bewakers	Die burgers zijn bewakers	Snoepjes zijn bewakers
Televisie is suiker	Deze stof is suiker	Een missionaris is suiker
Een date is een proef	Een examen is een proef	Kersen zijn een proef
Zuchten zijn tranen	Deze bladeren zijn thee	Koekjes zijn tranen
Je hart is een kluis	Deze kamer is een kluis	Een zakdoek is een kluis
Instinct is een kelder	Deze plaats is een kelder	Een raadsel is een kelder
Glimlachen zijn sterren	Zonnen zijn sterren	Knopen zijn sterren
Kritieken zijn deuren	Deze planken zijn deuren	Wormen zijn deuren
Grappen zijn kogels	Deze balletjes zijn kogels	Arenas zijn kogels
School is een woestijn	Dit landschap is een woestijn	Schaken is een woestijn
Banen zijn schepen	Die bouwsels zijn schepen	Seizoenen zijn schepen

Table 7.1: Sentence triplets using A=B structure.

Table 7.2: UNTESTED TRANSLATION of sentence triplets using A=B structure.

Note that these sentences were constructed to be meaningful in Dutch. These are untested translations that will produce different results upon being tested in English.

Novel metaphor	Literal categorization statement	Semantically anomalous sentence
Your words are your clothes	This blanket is clothing	Horses are clothes
Earth is meat	This food is meat	Artificial turf is meat
Change is a party	The bar is a party	Your toe is a party
Chances are keys	Those irons are keys	Cyclistst are keys
Youth is breakfast	These chips are breakfast	Pebbles are breakfast
Fears are teeth	These bones are teeth	Mouse pads are teeth
Secrets are a disease	DISEASE: being old, allergies,	A poster is a disease
Nuclear fusion is a wedding	This party is a wedding	This sock is a wedding
Suggestions are mothers (/clouds)	These pilots are mothers	Glass plates are mothers
Ethics are a floor	This wood is a floor	A lamp is a floor
A greeting is a pie	This snack is a pie	A blanket is a pie
Philosophy is a farm	This heritage is a farm	An (day) is a farm
Religions are houses	These tents are houses	Mosquitos are houses
Opinions are games	Those boxes are games	Ashtrays are games
Shyness is a virgin	This nun is a virgin	A feeding schedule is a virgin
Emotions are a painting	That sphere is a painting	Slowness is a painting
Rejections are murders	These deeds are murders	Chicks are murders
Spring is a birthday	This date is a birthday	A balcony is a birthday
Brands are enemies	Those farmers are enemies	Coat racks are enemies
Cold is an owner	This doctor is the owner	Tea is your owner
Skyscrapers are trees	Seeds are trees	Birds are trees
Applause is a rain	This water is rain	(Eye)glasses are rain
Feelings are lawyers	These criminals are lawyers	Nails are lawyers
A drivers-license is a funeral	This case is a funeral	A cup is a funeral
Music is a meeting	This 'sit' is a meeting	A duck is a meeting
Death is champion	That archeologist is champion	Paper is champion
Truths are victims	That officer is a victim	Grapes are victims
Facts are brothers	Those barbers are brothers	Printers are brothers
Rust is a watch	This clock is a wristwatch	A comic book is a watch
Kisses are medicin	MEDICIN: oils, plants, massages,	Staples are medicin
Forests are bodies	Those shapes are bodies	Desks are bodies
Mechanisms are brains	These structures are brains	Chairs are brains
Needs are customers	Those golf players are customers	Leaves are customers
The cradle is a morning	This time is the morning	The ceiling is the morning
Plans are feet	Those tracks are feet	Stickers are feet
Jealousy is drinks/booze	Blood is a drink	Elastic bands are drinks
Life is a kitchen	This part is a kitchen	A satelite is a kitchen
Your blood cells are soldiers	These cousins are soldiers	Tables are soldiers
Democracy is a store	This building is a store	A spraycan is a store
Your wit is your witness	This teacher is a witness	Sulpher is a witness

Funds are food	Cows are food	A whistle is food
Your kids are a mirror	This product is a mirror	A triangle is a mirror
Your face is a newspaper	This website is a newspaper	The rails are a newspaper
An only child is an island	This area is an island	This metre is an island
Ideas are fire	This phenomenon is a fire	Dirty dishes are a fire
Waves are voices	Those sounds are voices	Armrests are voices
Your brain is a detective	This man is a detective	This pencilcase is a detective
Your character is a sword	This thing is a sword	This donkey is a sword
Your needs are a customer	This cheff is a customer	An apple is a customer
Love is a teacher	This cook is a teacher	A curtain is a teacher
Warmth is a gift/present	This (cargo)bicycle is a gift	A nail is a present
Time is a writer	This prisoner is a writer	A market is a writer
A relationship is a castle	This monument is a castle	A bunny is a castle
Cellphones are walls	The border is a wall	A tractor is a wall
Hands are books	These works are books	Bathrooms are books
Doctors are guards	Those citizens are guards	Candies are guards
Television is sugar	This substance is sugar	A missionary is sugar
A date is an experiment	An exam is an experiment	Cherries are an experiment
Sighs are tears	These leaves are tea	Cookies are tears
Your heart is a safe(/vault)	This room is a safe	A handkerchief is a basement
Instinct is a basement	This place is a basement	A riddle is a basement
Smiles are stars	Suns are stars	Buttons are stars
Criticism are doors	Those planks are doors	Worms are doors
Jokes are bullets	These balls are bullets	Arenas are bullets
School is a desert	That landscape is a desert	Chess is a desert
Jobs are ships	Those structures are ships	Seasons are ships

Table 7.2: UNTESTED TRANSLATION of sentence triplets using A=B structure.

Appendix B: Filler sentences.

Duikers maken rode mieren

De stoel wil geen cola

Bowlen vergt flamingo's

Je been maakt panda's

Liederen van honing zijn kapot

Proefdieren gaan snel bellen

Hond ontdekt begraven glazuur

De piloot vliegt duizend noten

Ze fluistert een slang

Pingpong speel je met ruis

Bieten hebben rode wimpers

Linten hangen van gebrek

Een boor breekt de toekomst

Hij loopt over ogen

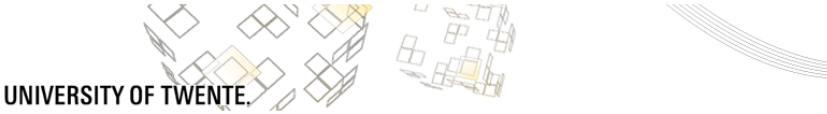Voel de verkleuring

Zij eet een bord

De olifant huilt takken

Meisjes ruiken flessen

Gooi de gele tuin

De steen houd de vis

De wens drinkt thee

De bak heeft honger

Het plan eet rijst

Ik leen een planeet

De snor is een mango

De pan bakt fitness

De oude man draagt tennis

Zoetigheid snoept varkens

Het kan alleen vacht

De diepte van het dagboek

Vieze was graaft diep

Een schep ziet melk

Het kind sprak zakken

De tas heeft vrijgevige rails

Vingers dragen wolkvorming

Neuzen voelen kou als stilte

Open ritsen lokken pinken

Dit handvat heeft een luipaard

Liefde buigt voor tomaten

Er staat een poema op gedrag

De drukte stoeit met lollies

Kinderen duiken in accus

Contactlenzen horen mos

De eend loopt op wenkbrauwen

De grafsteen staat ster

De zonsopgang brult rozijnen

Handoeken drogen lol

Kaarsen niet storen

De rat kruipt leeg

Bloemen groeien geluid

Kwasten vegen verdriet

De telefoon naait

De robot loopt rose

Een ballon vliegt knop

De bril zit op scharen

Het project verloopt kool

De kast staat muis

De kam lust molratten

De agent bloost potloden

Elastiek strekt wiskunde

Haar etui zit vol danspassen

De lamp hangt in verdamping

En das loopt vuist

Alles jeukt plakband

De toeter klinkt rekstok

Haaien duiken kussens

Appendix C: Example of Survey format, using University of Twente template.

Onderzoek Taalverwerking

Welkom

Hartelijk dank voor uw deelname aan dit onderzoek. Met dit onderzoek proberen wij inzicht te krijgen in de manier waarop onze hersenen taal verwerken.

Het onderzoek zal bestaan uit drie onderdelen:
In het *eerste* gedeelte zult u gevraagd worden om demografische gegevens in te vullen.
In het *tweede* gedeelte word u gepresenteerd met een korte woordherkenningstaak.
In het *derde* gedeelte zal u gevraagd worden om een aantal zinnen te beoordelen op hoe betekenisvol u deze vindt.

Mocht u **sona-credits** wensen te ontvangen voor uw deelname aan dit onderzoek, voer dan hieronder uw persoonlijke sona ID-nummer in. Als u deze niet invult kunnen wij namelijk geen credits toewijzen.
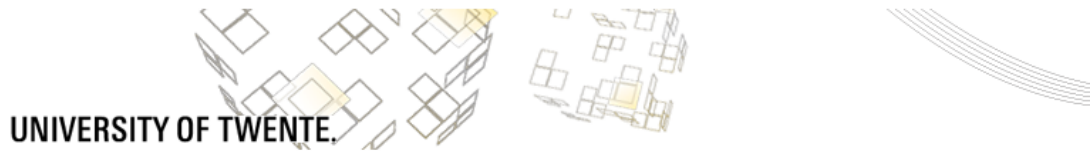
Om te beginnen, klik op 'volgende'.

Volgende

Survey Powered By Qualtrics

Appendix D: Instructions per survey section.

D1: Informed consent.


UNIVERSITY OF TWENTE.

Deelname aan dit onderzoek is geheel vrijwillig. Het geheel duurt ongeveer 20 tot 30 minuten. U bent te allen tijde vrij om zonder opgaaf van reden uw deelname aan dit onderzoek te staken. U kunt binnen 24 uur na afronding uw participatie intrekken, mocht u hiervoor kiezen dan zullen uw verzamelde gegevens vernietigd worden.

Het doel van dit onderzoek is om inzicht te vergaren in de manier waarop wij taal verwerken in onze hersenen.
In dit onderzoek zal u gevraagd worden om een aantal zinnen te beoordelen op hoe betekenisvol u deze vindt.

Om aan dit onderzoek deel te nemen moet u 18 jaar of ouder zijn en een redelijke beheersing hebben van de Nederlandse taal.

Dit onderzoek is goedgekeurd door de ethische commissie van de Universiteit Twente, en is vrijgegeven van aanstootgevende inhoud. De data zal volledig anoniem worden verwerkt, en niet door derden worden ingezien. Indien uw onderzoeksresultaten gebruikt zullen worden in wetenschappelijke publicaties, dan wel op een andere manier openbaar worden gemaakt, zal dit volledig geanonimiseerd gebeuren. Persoonsgegevens zullen niet door derden worden ingezien zonder mijn uitdrukkelijke toestemming.
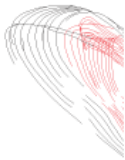
Als u vragen of opmerkingen heeft over het onderzoek, kunt u contact opnemen met d.flinsenberg@student.utwente.nl

Voor eventuele klachten over dit onderzoek kunt u zich wenden tot de secretaris van de Commissie Ethiek van de faculteit Gedragswetenschappen van de Universiteit Twente, mevr. J. Rademaker (telefoon: 053-4894591; e-mail: j.rademaker@utwente.nl, Postbus 217, 7500 AE Enschede)

○ _Ik heb het bovenstaande gelezen, en ga hiermee akkoord._

Volgende

Survey Powered By Qualtrics

D2: Instructions to the Word Recognition Task

UNIVERSITY OF TWENTE.
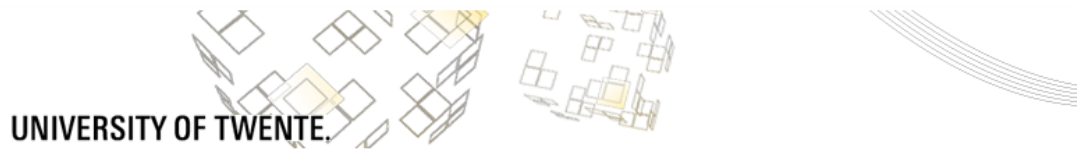
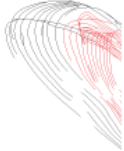## 2. Woordherkenningstaak.

Instructies

Er volgen nu instructies. Lees deze alstublieft goed, zodat u al vast een indruk krijgt van wat er van u verwacht wordt.

Deze test bestaat uit ongeveer 60 trials. Je krijgt steeds een letterreeks te zien. Jouw taak is om te beslissen of dit een bestaand Nederlands woord is of niet. Als je denkt dat het een bestaand Nederlands woord is klik je op "ja", als je denkt dat het geen bestaand Nederlands woord is klik je op "nee".

Als je er zeker van bent dat het woord bestaat, ook als je niet precies weet wat het betekent, mag je toch met "ja" antwoorden. Maar als je twijfelt of het wel een bestaand woord is, kies dan "nee".

Je hebt zoveel tijd als je wilt voor elke beslissing. Dit deel van het experiment duurt ongeveer 5 minuten.

Als alles duidelijk is kun je het experiment nu starten.

Volgende

Survey Powered By Qualtrics

D3: Instructions and Examples Semantic Judgement Task

UNIVERSITY OF TWENTE.

3. Onderzoek.
Graag vragen wij u om een aantal zinnen te beoordelen.

Instructies
Er volgen nu instructies. Lees deze alstublieft goed, zodat u al vast een indruk krijgt van wat er van u verwacht wordt.

Hieronder ziet u de schaal die u gaat gebruiken in het experiment.
Met "**betekenisvol**" wordt bedoelt of de zin betekenis voor u heeft.
Met "**betekenisloos**" wordt bedoelt dat een zin *geen* betekenis voor u heeft.

Op de schaal ziet u een lijst van zinnen. Evalueer elke zin op deze schaal. U kunt op de schaal van 1 tot 7 aangeven in hoeverre u deze zinnen betekenisvol of betekenisloos vindt.

De zinnen zijn niet gerelateerd, dus beoordeelt u alstublieft alle zinnen individueel.

Er zijn geen goede of foute antwoorden, het gaat immers om uw mening.

	1. Zeer betekenisloos	2: Betekenisloos	3. Een beetje betekenisloos	4: Betekenisloos noch betekenisvol	5. Een beetje betekenisvol	6. Betekenisvol	7. Zeer betekenisvol
Een hond is een baksteen.	○	○	○	○	○	○	○
Een slang is een reptiel.	○	○	○	○	○	○	○
Deze man is sterk als een beer.	○	○	○	○	○	○	○

Voorbeeld:
Mogelijk zou u deze zinnen zo kunnen beoordelen.

"Een hond is een baksteen" .... 1 (Zeer betekenisloos)
U zou kunnen denken dat deze zin "zeer betekenisloos" is, omdat: Het is moeilijk voor te stellen dat een hond een baksteen is. Een hond en een baksteen lijken geen overeenkomsten te hebben, daardoor zou deze zin betekenisloos kunnen zijn.

"Een slang is een reptiel" .... 7 (Zeer betekenisvol)
U zou kunnen denken dat deze zin "zeer betekenisvol" is, omdat: Een slang heeft de eigenschappen van een reptiel.

"Deze man is sterk als een beer" .... 7 (Zeer betekenisvol)
U zou kunnen denken dat deze zin "zeer betekenisvol" is, omdat: Beren staan bekend om hun kracht, dus als je de man met een beer vergelijkt, zou het kunnen betekenen dat de man heel erg sterk is.

(Dit is natuurlijk een voorbeeld. U kunt zelf aangeven hoe betekenisvol of betekenisloos u deze zinnen vindt.)

Volgende

Survey Powered By Qualtrics

D4: Indication piece of Semantic Judgement Task.

UNIVERSITY OF TWENTE.

Hieronder ziet u de schaal die u gaat gebruiken in het experiment.
Met "**betekenisvol**" wordt bedoelt of de zin betekenis voor u heeft.
Met "**betekenisloos**" wordt bedoelt dat een zin *geen* betekenis voor u heeft.

Op de schaal ziet u een lijst van zinnen. Evalueer elke zin op deze schaal. U kunt op de schaal van 1 tot 7 aangeven in hoeverre u deze zinnen betekenisvol of betekenisloos vindt.

De zinnen zijn niet gerelateerd, dus beoordeelt u alstublieft alle zinnen individueel.

Er zijn geen goede of foute antwoorden, het gaat immers om uw mening.

	Zeer betekenisloos	Betekenisloos	Een beetje betekenisloos	Betekenisloos noch betekenisvol	Een beetje betekenisvol	Betekenisvol	Zeer betekenisvol
Dit deel is een keuken.	○	○	○	○	○	○	○
Ideeën zijn brand.	○	○	○	○	○	○	○
Je teen is een feestje.	○	○	○	○	○	○	○
Die man is een rechercheur.	○	○	○	○	○	○	○
Deze botten zijn tanden.	○	○	○	○	○	○	○
Een balkon is een verjaardag.	○	○	○	○	○	○	○
Het plafond is de ochtend.	○	○	○	○	○	○	○
Je karakter is een zwaard.	○	○	○	○	○	○	○

Appendix E: Review tables for item desirability.

Novel Mataphors

	Dutch	English	SK*	r1**	r2***
1-1	Ideeën zijn brand.	*Ideas are fire.*****	✗		2✗
1-2	Plannen zijn voeten.	*Plans are feet.*	✗		
1-3	Zuchten zijn tranen.	*Sighs are tears.*	✓		
1-4	Je gezicht is een krant.	*Your face is a newspaper.*	✓	✓	
1-5	De tijd is een schrijver.	*Time is a writer.*	✓	✓	
1-6	Je karakter is een zwaard.	*Your character is a sword.*	✓		
1-7	Afwijzingen zijn moorden.	*Rejections are murders.*	✓	5✓	
1-8	Je kinderen zijn een spiegel.	*Your children are a mirror.*	✓	2✓	
2-1	Applaus is een regen.	*Applause is a rain.*	✓		
2-2	Golven zijn stemmen.	*Waves are voices.*	✓	2✓	
2-3	Grappen zijn kogels.	*Jokes are bullets.*	✗	(-)	
2-4	Mobieltjes zijn muren.	*Cellphones are walls.*	✗		
2-5	Geheimen zijn een ziekte.	*Secrets are a disease.*	✗		
2-6	Je woorden zijn je kleding.	*Your words are your clothing.*	✓		
2-7	Gevoelens zijn advocaten.	*Feelings are lawyers.*	✗	(-)	
3-1	Roest is een horloge.	*Rust is a watch.*	✗	(-)	✗
3-2	Een date is een proef.	*A date is an experiment.*	✓		
3-3	Kansen zijn sleutels.	*Chances are keys.*	✓	(-),✓	
3-4	Dokters zijn bewakers.	*Doctors are guards.*	✓		✗
3-5	Democratie is een winkel.	*Democracy is a store.*	✗	(-)	
3-6	Meningen zijn spelletjes.	*Opinions are games.*	✓		
3-7	Een enigskind is een eiland.	*An only child is an island.*	✓		

* SK, item suitability based on skewness and kurtosis measures for distribution.

** r1, item suitability based on positive correlations with other meaningful items, or negative correlations with meaningless items above 0,7. Items marked with "(-)" has a positive connection to another item previously flagged as unfavorable.

*** r2, item unsuitability based on positive correlations with other meaningless items, or negative correlations with meaningful items above 0,7.

**** Untested translations of items, added for legibility of results.

TABLE 3.1: Item suitability of novel metaphors, where a green check (✓) indicates a desirable item characteristic and a red cross (✗) indicates an item characteristic as undesirable.

			SK*	r1**	r2***
1-9	Deze non is maagd.	*This nun is a virgin*****	✔	3✔	
1-10	Die kok is een leraar.	*That chef if a teacher.*	~	2✔	
1-11	Dit deel is een keuken.	*This part is a kitchen.*	✔	2✔	
1-12	Deze botten zijn tanden.	*These bones are teeth.*	✖	✔	
1-13	Deze werken zijn boeken.	*Theses works are books.*	✔	2✔	
1-14	Planten zijn medicijnen.	*Plants are medicin.*	✔	4✔	
1-15	Die man is een rechercheur.	*This man is a detective.*	✔	✔	
1-16	Dit landschap is een woestijn.	*This landscape is a desert.*	✔	3✔	
2-8	Deze stof is suiker.	*This substance is sugar.*	✔	✔	
2-9	Deze chips is ontbijt.	*These chips are breakfast.*	~	✔	
2-10	Die juf is een getuige.	*That teacher is a witness.*	✔	3✔	
2-11	Deze arts is de eigenaar.	*This doctor is the owner.*	✔	3✔	
2-12	Die boeren zijn vijanden.	*Those farmers are enemies.*	✔	7✔	
2-13	Die vormen zijn lichamen.	*Those shapes are bodies.*	✔	✔	
2-14	Die bakfiets is een cadeau.	*That cargo-bike is a gift.*	✔	✔	
3-8	Zonnen zijn sterren.	*Suns are stars.*	✔	✔	
3-9	Die snack is een taart.	*That snack is a cake.*	✔	✔	
3-10	Die tenten zijn huizen.	*Those tents are houses.*	✔		✖
3-11	Die kappers zijn broers.	*Those barbers are brothers.*	✔	6✔	
3-12	Die agent is slachtoffer.	*That officer is a victim.*	✔	3✔	
3-13	Deze piloten zijn moeders.	*These pilots are mothers.*	✔	4✔	
3-14	Dat erfgoed is een boerderij.	*That heritage is a farm.*	✔		

Literal sentences (row group label)

* SK, item suitability based on skewness and kurtosis measures for distribution.

** r1, item suitability based on positive correlations with other meaningful items, or negative correlations with meaningless items above 0,7. Items marked with "(-)" has a positive connection to another item previously flagged as unfavorable.

*** r2, item unsuitability based on positive correlations with other meaningless items, or negative correlations with meaningful items above 0,7.

**** Untested translations of items, added for legibility of results.

TABLE 3.2: *Item suitability of literal sentences, where a green check (✔) indicates a desirable item characteristic and a red cross (✖) indicates an item characteristic as undesirable. (~) indicates questionable desirability.*

Anomalous sentences			SK*	r1**	r2***
1-17	Vogels zijn bomen.	****	✔		
1-18	Kunstgras is vlees.		✔	2✔	(-)
1-19	Je teen is een feestje.		~	✔	
1-20	Elastiekjes zijn drank.		✔		
1-21	Deze sok is een bruiloft.		✔		
1-22	Het plafond is de ochtend.		✔	✔	
1-23	Een beker is een begrafenis.		✔		
1-24	Een balkon is een verjaardag.		✔		
2-15	Papier is kampioen.		~		2✖
2-16	Een lamp is een vloer.		✔		
2-17	Tafels zijn soldaten.		✔	3✔	
2-18	Stoelen zijn hersenen.		✔	✔	
2-19	Een zakdoek is een kluis.		✔	✔	
2-20	Een konijn is een kasteel.		✔		
2-21	Een eend is een vergadering.		✔		
3-15	Wormen zijn deuren.		✔		
3-16	Een fluit is voedsel.		✔	3✔	
3-17	Een appel is een klant.		✔	3✔	
3-18	Bladeren zijn klanten.		✔	✔	
3-19	Seizoenen zijn schepen.		~	✔	
3-20	Een raadsel is een kelder.		~		
3-21	Traagheid is een schilderij.		~		

* SK, item suitability based on skewness and kurtosis measures for distribution.

** r1, item suitability based on positive correlations with other meaningful items, or negative correlations with meaningless items above 0,7. Items marked with "(-)" has a positive connection to another item previously flagged as unfavorable.

*** r2, item unsuitability based on positive correlations with other meaningless items, or negative correlations with meaningful items above 0,7.

**** Untested translations of items, added for legibility of results.

TABLE 3.2: Item suitability of anomalous sentences, where a green check (✔) indicates a desirable item characteristic and a red cross (✖) indicates an item characteristic as undesirable. (~) indicates questionable desirability.