Psycholinguistic Variables in Lexical-Semantic Processing of Antonyms

Uštulica, Dora

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:777486

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-07-30**

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb Faculty of Humanities and Social Sciences

University of Zagreb

Faculty of Humanities and Social Sciences

Department of English

Psycholinguistic Variables in Lexical-Semantic Processing of Antonyms

Thesis advisors:

Prof. Milena Žic Fuchs, PhD

Dr. Janja Čulig Suknaić, postdoc

Zagreb, 2022

Table of contents

Table of figures

Abbreviations

RT = reaction time

NA = unavailable answer

GAM = generalized additive model

1. Introduction

This study aims to examine the role of psycholinguistic variables in the lexical-semantic processing of antonyms. It focuses on the impact of convention on processing antonym pairs, as expressed with the psycholinguistic variable of reaction time. This could help further clarify not only antonymy as a linguistic and conceptual phenomenon, but also the notion of convention and the role it plays in the processing of lexemes. The pairs of antonyms that have a higher degree of conventionalization should be processed faster, as they should be more readily available in the mental lexicon. That would suggest that different pairs of antonyms can be considered more or less prototypical, depending on their level of conventionalization and that in turn results in easier lexical access. This is an empirical study of antonyms, which should be replicable and reproducible, complying with the principles of open science.

First, the theoretical framework on antonymy, the notion of convention, and reaction time as a variable will be provided. What ensues is the central, experimental, part of the study. In it, pairs of antonyms from three categories are tested in an online experiment measuring reaction time of the participants. Pairs are divided into highly conventionalized ones, less highly conventionalized ones, and unconventional ones. The experiment, thus, aims to answer the question of the importance of convention in the lexical-semantic processing of antonyms.

Two hypotheses are posited: H1) The more conventionalized pairs of antonyms, as measured by their co-occurrence frequency in the chosen corpus are going to be processed faster than the less conventionalized ones; H2) Unconventional pairs of antonyms are going to be processed the slowest because they are usually context-dependent, although they may as well be recognized as opposites. The difference found in the processing of different categories of antonyms would then point to possibly different cognitive mechanisms involved in their processing, which would confirm the existence of different categories of antonym pairs (more or less prototypical), as they are stored in the mental lexicon. In this case, antonymy should be regarded as a conceptual category of semantic opposition, and not as just a relation between lexemes. The study ends with the discussion of the results and suggestions on possible future directions of this kind of research.

2. Theoretical framework

2.1 Antonymy

Antonyms are a ubiquitous part of everyday language, as Lyons (1977: 271) discussed in his work stating that “binary opposition is one of the most important principles governing the structure of languages.” However, there is still no consensus on the definition of antonyms in linguistics. They are most often defined simply as lexemes that are semantically opposite (Murphy et al. 2012). Murphy et al. (2012: 2) define oppositeness in the context of antonymy as “having opposed meanings in a given context”, which will be discussed in the subsequent section. Another type of oppositeness according to Murphy et al. (2012) is the logical one, which means that we cannot describe one thing with both lexemes that form the antonym pair, as they show with the example of *long* and *short.* So, if we say that something is short, it cannot at the same time be long.

However, there is another important thing to keep in mind when discussing antonyms. They are not only in the relation of oppositeness, but they also have to possess a degree of similarity. The two necessary characteristics are known as *the principle of maximal similarity* and *the principle of minimal difference* (Lyons 1977, Cruse 1986, Murphy 2003) or also as further elaborated upon *the principle of invariance* (Bianchi and Savardi 2006, 2008a) and *the principle of the degree of adequacy* (Bianchi and Savardi 2008a). That is why we do not perceive every pair of lexemes as an antonym pair, e.g. *pen* and *car* will not be considered antonyms because they do not contain enough similarities, unlike, for instance, the pair *man* and *woman.*

There are several different perspectives on antonyms, which is no wonder since they were discussed as far back as in the time of Aristotle (Murphy et al. 2013: 6). One of the most prominent perspectives on antonymy stems from structuralism, according to which antonyms make up one of the paradigmatic relations in language, which means that those words constitute a set of potentially substitutable expressions, as antonyms do (Murphy et al.2013). Saussure (1956[1916*]) claimed that lexemes get their meanings from their relations with other words, which is something that can clearly be applied to antonyms, as we can say that every lexeme in a pair of antonyms is defined by its opposite. In the structuralist perspective, antonyms can be divided into four different groups, as was done by Lehrer and Lehrer (1982) and Cruse (1986). Namely, those are complementaries, which denote pairs of two domains (e.g. *dead-alive)*; contraries, which denote a degree of some property (e.g. *fast-slow)*; reversives, which denote opposite directions (e.g. *fall-rise)* and converses, which denote opposites of transfer or relationship (e.g. *buy-sell)*.

Murphy (2003: 170) developed a theoretical model called Relation by Contrast-Lexicon Contrast (RC-LC), which she defines as “a lexical contrast set that includes only word-concepts that have all the same contextually relevant properties but one.” What makes this approach different from the previous one is the fact that it regards the mentioned relations as metalexical, meaning that those relations are about “the conceptual knowledge about words, rather than lexical or semantic representation of the words” (Murphy 2003:171). That highlights the importance of factors such as morphology, collocational relations or phonology in contrast with emphasizing only the semantic dimension of antonymy.

With the advent of cognitive linguistics in the 1970s, cognitive approaches to antonymy appeared as well. Cruse and Togia (1996) introduced the concept of the schematic ANTONYM domain, which means that the content domain is structured in terms of pairs of directionally opposed graded properties. Paradis (1997, 2001) extended this by adding the concepts of GRADABILITY, OPOSITENESS and BOUNDEDNESS. For instance, she (2001: 51) makes a distinction between scalar and non-scalar adjectives; non-scalar ones structured along a BOUNDED SCALE (e.g. *dead-alive)*, and the scalar ones, which are deemed to be ‘implicit comparatives’, along the UNBOUNDED SCALE (e.g. *short-long)*. The main idea of the cognitive linguistic approach to antonymy and the one adopted in this study is that “antonyms are not just relations between lexemes, but relations between different meaning construals” and that they are “remarkably non-arbitrary” (Croft and Cruse 2004: 40-43, 192). This position is in line with our research and the hypotheses, as convention, which will be discussed in the following section, plays a possibly important role in antonym processing. This underscores the cognitive linguistic position that antonyms are relations between meaning construals rather than just lexical relations, as our knowledge of them is construed through our relations with the language itself and the world in which language is used. That was also the conclusion reached by Čulig Suknaić (2020) who states that “we are able to conclude that antonymy does function as a conceptual prototype-based category whose members are added according to conventionalized knowledge of the language and knowledge of the world.” So, antonymy in this view can be defined as a conceptual category of semantic opposition. Her study has also confirmed that antonymy defined as a conceptual domain functions in both

English and Croatian, two typologically different languages, which additionally confirms the relevance of the conclusion.

As can be seen in this section, antonymy has gone through many different definitions and conceptualizations and that is why further research into the phenomenon is needed to be able to come up with an appropriate definition, and more importantly, understanding of the phenomenon itself.

2.2 Convention

The notion of convention plays an important role in linguistics, as it "reflects a part of the complexity of language seen not only as a cognitive phenomenon or a cognitive ability, but also seen as being central to human interaction" (Žic Fuchs et al. 2013: 66). Since language is a cognitive, as well as a social phenomenon, both factors have to be taken into consideration when studying different linguistic phenomena, antonymy included. Langacker (1987: 488) defines convention as "the degree to which an expression conforms to the linguistic conventions of a language." Or rather, a linguistic expression or phenomenon deemed to be conventional is "widely shared by the members of the relevant speech community." He claims that convention is "simply contextual meaning that is schematized to some degree and established as conventional through repeated occurrence" (157). We can thus say that the level of conventionalization is reflected in the canonicity of antonyms, which is demonstrably conventionalized through their co-occurrence in corpora.

In the case of our study, the Corpus of Contemporary American English (COCA) reflects the spread of an antonym pair in the community and is thus used to signal the level of conventionalization of a certain pair. The frequency of the co-occurrence of different pairs influences the status of a given antonym pair. That is, it determines the already mentioned level of canonicity of the pair or, as Justeson and Katz claim (1991: 182), 'adjectives may be more or less antonymous rather than simply antonymous or not antonymous.' Čulig Suknaić reaches the following conclusion in her dissertation:

> [A]ntonymy is structured on the conceptual level as a prototype-based category whose members are included through conventionalized knowledge of the individual pairings and antonymic constructions. More prototypical members are more frequent in the corpora and can be found in a wider set of constructions, which can be used to express more basic opposite meanings. (2020)

In line with that conclusion, this study aims to test whether the pairs deemed to be more antonymous, or rather the pairs being more prototypical as reflected in the notion of convention based on the frequency of their co-occurrence in the given corpus, really are easier for us to process. That would point to different cognitive representations of different pairs of antonyms in the mental lexicon.

2.3 Reaction time

In order to show that some antonyms are more conventionalized and that that is an important factor in their cognitive representation, we needed to devise a test to confirm the hypotheses. We have decided to use a reaction time experiment, which is very often used in different types of psychological and psycholinguistic research. The pioneer of reaction time experiments was F.C.Donders, in as early as 1868, who used reaction time to measure behavioral responses in different experiments and who proved the existence of three types of RT experiments (Donders 1969 [1868*]), which are going to be explained below.

According to Decrochers and Thompson (2009), there are several factors influencing the processing of words. Namely, their intrinsic properties (e.g. word length), the context in which the word is found (e.g. how often it appears in a given text), and psycholinguistic variables (e.g. imageability, concreteness etc.). The focus of this study, as already explained in the previous section, is on the notion of convention. As Murphy et al. (2013: 13) claim, psycholinguistic studies provide evidence that ‘some antonyms are canonical and so presumably represented as pairs in the mind.’ Many psycholinguistic experiments proving that have been carried out: elicitation tests (e.g. Deese 1965, Charles and Miller 1989, Paradis & Willners 2006, Paradis et al. 2009, Čulig Suknaić 2020), identification tests (e.g. Herrmann et al. 1997, Gros et al. 1989, Čulig Suknaić 2020) and semantic priming tests (Becker 1980). All those tests come with certain limitations. Murphy et al. (2013) mention the choice of stimuli based on the investigator’s intuition rather than on some objective criteria and different possibly confounding factors, such as word length, as possible downsides to the experiments of the kind. We have tried to avoid those in our study, as explained in the methodology section.

As already stated, in this study, an online processing experiment is used to study the way that convention affects reaction time, or rather how long it takes the participants to react to the stimuli in the form of antonym pairs of different levels of conventionalization. Jiang explains:

> the use of RT data is based on the premise that cognitive processes take time and by observing how long it takes individuals to respond to different stimuli or perform a task in different conditions, we can ask questions about how the mind works, and infer about the cognitive processes or mechanisms involved in language processing. (2012: 2)

In the case of our study, we measure how the variation within the variable of convention affects reaction time, which can in turn inform our understanding of the cognitive processes included in the processing of antonyms.

As mentioned before, there are three types of RT experiments. Namely, simple RT experiments in which the participants simply react to different stimuli; recognition tasks (go-no go tasks), in which the participants have to answer to only one type of stimulus, while the other one serves as a distraction; and choice tasks, in which the participants have to choose the answer upon the appearance of the stimulus, usually by pressing a determined key (Baayen and Milin 2010).

There are multiple factors that may influence reaction time, such as physical and mental condition, age, gender, handedness, and general cognitive abilities of the participants (Lee and Chabris 2013). However, these factors may be controlled in the results as they mainly influence the RT average of every individual participant (Wagenmakers et al. 2004). Other factors that may influence RT, and the ones connected to the properties of lexemes, are familiarity, word length, neighborhood, concreteness, imageability, age of acquisition, spelling-sound regularity, affixation, number of meanings, number of associates and bigram frequency. They should be controlled according to the requirements of the experiment (Jiang 2012). In our study, we have focused on controlling the factors specifically important for it to get the as reliable as possible results.

3. Methodology and participants

3.1. Methodology

Pairs of antonyms used in the study were taken from the examples provided in Lyons (1977), Cruse (1986) and Jones et al. (2012) and chosen based on the search of the Contemporary Corpus of American English (COCA). Fifteen (15) pairs of highly conventionalized and fifteen (15) pairs of less highly conventionalized antonyms were chosen. They all belong to the open word class. Namely, to nouns, verbs and adjectives. The level of conventionalization in this case was based on the frequency of an antonym pair occurrence in the corpus. All the pairs considered to be highly conventionalized in this study have a frequency of above 2,000 tokens. All pairs contain only one syllable, to at least partially avoid the possibly confounding factor of word length.

The unconventional pairs of antonyms were chosen by means of ancillary antonymy. Ancillary antonymy can be defined as antonymy that is generated by a pair of antonyms in a sentence by means of association, which results in a new pair of opposites that would otherwise not necessarily be considered opposites (Jones 2004). For instance, in the following sentence- *It is meeting public need, not private greed* – provided in Jones (46)- the antonym pair *public-private* generates the pair *need-greed.*

The table below contains all the conventionalized pairs together with their frequencies and also all the chosen unconventional pairs.

HIGHLY CONVENTIONALIZED		LESS CONVENTIONALIZED		UNCONVENTIONAL
black-white	34707	near-far	1017	sun-moon
good-bad	16337	wet-dry	968	book-film
left-right	16212	hard-soft	921	car-bus
come-go	15629	fast-slow	873	sea-land
day-night	10016	laugh-cry	786	faith-doubt
life-death	7799	thick-thin	652	food-drink
high-low	7329	push-pull	578	debt-cash
east-west	5684	sit-stand	520	dress-pants
front-back	4182	tall-short	486	walk-run
long-short	4159	thin-fat	343	greed-need
big-small	3844	well-ill	219	red-blue
boy-girl	3715	pass-fail	199	spoon-fork
poor-rich	3690	lean-fat	181	sun-rain
give-take	3306	dim-bright	74	heart-brain
love-hate	2827	break-fix	34	ice-sand

The survey and the experiment were compiled in PsyToolkit (Stoet 2010, 2017), a free toolkit for running cognitive-psychological experiments and surveys. The survey and the code for the experiment can be accessed in the Open Science Framework repository via the link provided in the references. The first part of the study consists of a survey of participants' basic demographic information. The participants had to answer the questions on their gender, age, education, possible language disorders, and their mother tongue, along with the country they come from. The introductory part also includes a brief description of the experiment, the informed consent checkbox, author's email for any further information, statement on data anonymization and the instructions for the experiment.

In the experiment we measure the reaction time of participants. The task presented to the participants is to press the correct key in order to decide whether the pair presented represents a pair of opposites or not. Pairs of synonyms are used as a control group. The software measures the time that passes from the appearance of the word pair on the screen to the participant pressing the key. The stimuli containing the instructions appearing at the beginning of the training session and the experiment, and the fixation point appearing prior to every word pair for 500 ms were created in Inkscape (Harrington et al. 2004-2005), an open-source graphics editor. Our experiment falls into the category of a decision RT task, as was explained in the previous section. Namely, it is a combination of simple and choice task. Participants are presented one stimulus at a time and then have to choose the right answer by pressing the correct key. To be precise, after the appearance of the word pair on the screen, participants have two seconds to press the right key. The key used in the situation in which the opposites are presented is the right arrow, and the key used in the situation in which there are no opposites presented is the left arrow. The keys were chosen to exclude the possible effect of right- or left-handedness of the participants. After pressing the key, the participants see the message saying if they got it right or wrong. If it took them too long to press the key, they see the message that they were too slow. The experiment takes around three minutes to finish.

The anonymized results of every participant are then saved in the form of Excel sheets for further analysis. Incorrect responses are excluded from the analysis. Outliers, considered to be the responses outside the range of 200 and 2000 ms are excluded as well, as it is assumed that an individual is highly unlikely to process lexemes faster or slower than the given processing speed values. Those results are more likely to be the result of accident or distraction (Jiang 2012).

3.2 Participants

Fifty participants took part in the study. The mean age of participants was 38, age range being between 19 and 73. Thirty-two participants identified themselves as female and eighteen participants as male. As for the level of education, for twenty-two participants the highest level of education reached was high school, for twenty it was the bachelor's degree, for six it was the master's degree and for two it was the PhD. They were all native speakers of English, which was required for the experiment. Forty-six were from the UK and four were from the USA. When it comes to possible language disorders, one participant reported having some sort of language disorders, two didn't know and all the rest reported not having any.

4. Results

There is still no consensus on the best way to analyze reaction time data, as they have their specific features, which will be discussed in this section, and consequently different researchers choose to analyze reaction time data differently. So, before starting the analysis some of the methodological concerns regarding the analyses of RT data are going to be presented.

The first problematic point is that the distributions of RTs are often positively skewed, resulting in non-normal distributions (Baayen and Milin 2010). This problem may be solved by doing non-linear transformations of data, transforming the data logarithmically. However, recently it has been shown that that may not be really useful if one only wants to see whether RT covariates with variables such as frequency (Schramm and Rouder 2019). The second problem is that individual responses are not statistically independent, as a trial-by-trial sequential correlation is present (Baayen and Milin 2010). That is why it is best to account for the individual differences in the model as well.

When it comes to the distribution of the data, nor the mean, nor the standard deviation are robust measures because the distribution is often skewed. That is why many researchers choose to report the median value as a central tendency parameter and interquartile range for estimating dispersion. This is also not a perfect way to analyze data, since the median is a biased estimator when the data is skewed, resulting in underestimation (Whelan 2008). One of the proposed solutions to this is to analyze central tendencies of the whole distribution to better understand the differences between different categories (Whelan 2008).

As can be seen, there are many possible obstacles one can stumble upon when analyzing RT data. However, led by the position articulated in one of the papers by the statistician George Box (1976), which could be paraphrased as saying that all statistical models are wrong but some are useful, we have analyzed the results of the experiment with the intention of getting the most out of them, while always being aware of their possible limitations.

All the results were converted into .csv format. They were imported into RStudio (version 1.4.17) and analyzed with R (version 4.1.0). The analysis was done primarily with the tidyverse (Wickham et al. 2019), patchwork: The Composer of Plots (Lin Pedersen 2020), mgcv (Wood 2011) and lme4 (Bates et al. 2015) packages. The code used for the analysis can be accessed in either HTML or R Markdown form via the Open Science Framework link provided in the references.

4.1 Descriptive statistics

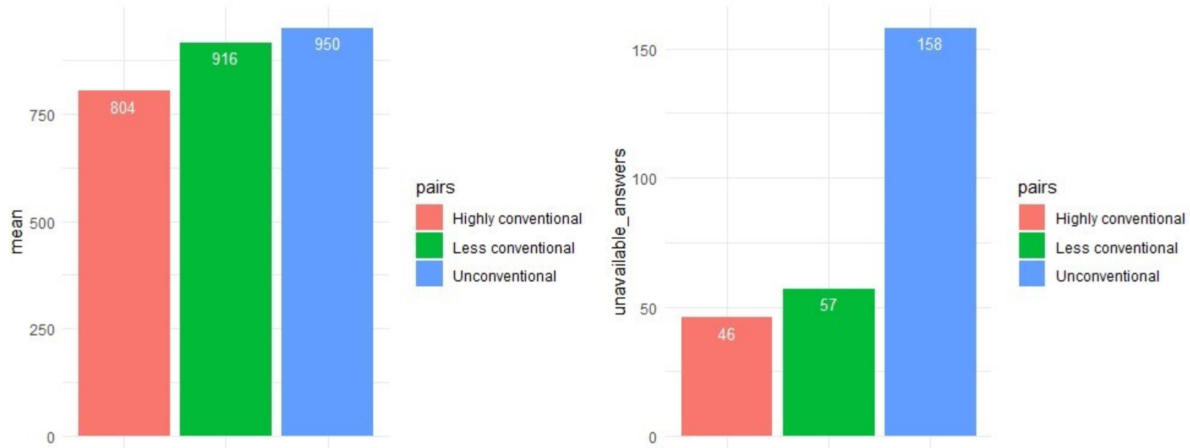

Figure 1: RT mean (left) and number of NAs (right)

In the table above two bar plots are presented. The first one represents the distribution of mean RTs per different experimental categories. Namely, highly conventional pairs, less conventional pairs and unconventional pairs. All the values were rounded. Highly conventional pairs were on average processed the fastest by the participants (M = 804, SD = 240, Mdn = 757, range = 324 - 1949, IQR = 256). Then less highly conventional pairs (M = 916, SD = 280, Mdn = 878, range = 273 - 1967, IQR = 347). Finally, unconventional pairs took participants the longest to process ($M = 950$, $SD = 287$, $Mdn = 923$, range = 271 - 1845, $IQR = 371$). What can be observed right away is that the means in all cases are greater than the medians and that there are data points that exceed two standard deviations, which tells us that the data is right skewed. We can also see that interquartile ranges are ranked in the same way as means, from the lowest for the highly conventionalized pairs to the highest for the unconventional pairs. That tells us that the dispersion in the middle half of our distribution, which is especially important in skewed distributions, is the greater the less conventional the pairs are. We have performed two-sample T-tests to determine whether the differences between the means are in this case statistically significant. The mean difference between highly conventional and less highly conventional pairs turned out to be statistically significant ($p = 0.03$, $d = 0.43$, $t = -2.15$), just like the difference between highly conventional and unconventional pairs means ($p = 0.0069$, $d = 0.55$, $t = -2.76$). The difference between the less highly conventionalized pairs and unconventional pairs is not statistically significant and the magnitude of difference is small ($p = 0.55$, $d = 0.12$, $t = -0.59$). However, that does not mean that the difference between these two groups is insignificant. That is why we also observed the unavailable answers for each category.

In the bar plot on the right-hand side of the graphic we can see the distribution of unavailable answers per each category. Those were the answers that either took the participants too long to answer or they did not give the correct answer. In the case of our study, it would be wrong to simply ignore those answers because they are informative. We can see right away that the difference between them is significant. 17.62% of those answers can be found in the group of highly conventional pairs, 21.84% in the group of less conventional pairs and 60.54% in the group of unconventional pairs. Two sample T-tests were also performed on them. The difference between the highly conventionalized and less conventionalized pairs was not significant ($p = 0.12$, $d = 0.32$, $t = -1.58$). However, tests for highly conventionalized and less conventionalized pairs paired with the unconventional pairs were statistically significant, with large magnitude of difference effects ($p < 0.001$, $d = 1.57$, $t = -7.84$; $p < 0.001$, $d = 1.57$, $t = -7.83$).

What we can conclude upon analyzing the distributions of data is that when it comes to reaction time highly conventionalized pairs stand out, as they are supposedly easier to access and subsequently to process. The difference between the less conventional and unconventional pairs is not significant, which points out that, when unconventional pairs are recognized as antonyms, they are equally easy, or rather, equally hard, to access as the less conventional pairs. That underscores the fact that, when it comes to lexical access, the degree of conventionalization should be high for it to matter. If we observe the number of unavailable answers, the conclusion is opposite. Namely, there is no significant difference between the highly and less highly conventionalized pairs, which means that the latter are recognized as antonyms, just not as easily. Unconventional pairs are sometimes not recognized as antonyms, or would take much longer to process, specifically because they are not conventional. Another observation to be made is that unconventional antonyms are still very often recognized as antonyms by the majority of participants (in our sample they appeared 750 times, but there were only 158 unavailable answers for them). That tells us that antonyms, as we discussed them in the theoretical part of the paper on antonymy, should primarily be regarded as the relation of conceptual opposition and not as merely lexical opposition.

After performing the initial analysis on the whole sample and per category, we have decided on the further path of analysis. Namely, outlier values exceeding 2 SDs will not be removed, as in the case of our research they affect both the results and the assumptions made. Unavailable answers are going to be excluded from the analysis after we proved that the difference in means is significant even without them being replaced by means per categories or by regression imputation. Instead, the number of unavailable answers is going to be included in the mixed-effects model to account for that factor. Finally, we have decided not to perform a non-linear transformation on our data, as it wouldn't significantly affect the analysis.

4.2 By-item plots

In order to get a better picture of our data we have made kernel density graphs for every category of antonym pairs. They show the distribution of every item across participants. To each of them a vertical line, which marks the mean of the distribution, was added.

4.2.1 Highly conventionalized pairs

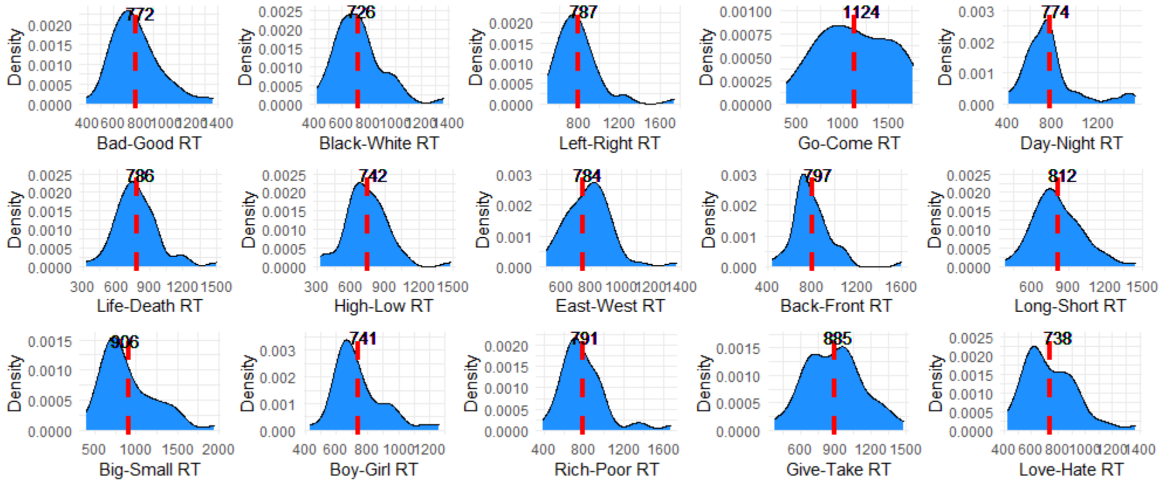

Figure 2: Highly conventional pairs - kernel density graphs

If we look at the graphs above we can see that almost all of them are positively skewed, which was expected, since the means tend to be greater than the medians in this type of data because of the outliers. There are a couple of cases here that are interesting to observe. Firstly, the pair which has the highest RT, *go-come,* is dispersed across the RT distribution, which means that for some participants it was rather easy to process and for some it was rather hard, which was probably the case because it was presented in the unconventional order, as *go-come* and not as *come-go*. Bimodal distributions can be seen in pairs *love-hate, give-take, boy-girl* and *big-small.* So, there was a significant number of people who had longer RTs than the observed mean, or rather, the distribution of RTs is dispersed with significant individual differences. The examples which have a distribution similar to the normal distribution are *bad-good, high-low,* and *long-short*. They were on average similarly difficult to process for all the participants.

4.2.2 Less highly conventionalized pairs

The graphs below represent less highly conventionalized pairs. We can also make a couple of observation. The pairs that have a distribution similar to the normal one are *cry-laugh, lean-fat, fix-break,* and *ill-well.* They are the pairs that on average had longer processing times. So, we can say that they were on average equally difficult to process for the participants. The pair that took the longest to process, *dim-bright*, does have a distribution similar to the normal one but there was a smaller number of participants for whom it was even harder to process, so it can be considered bimodal. If we look at the other pairs with bimodal distributions, this type of distribution can be seen in the examples *near-far* and *tall-short.* These examples were then also considerably harder to process for a number of participants.

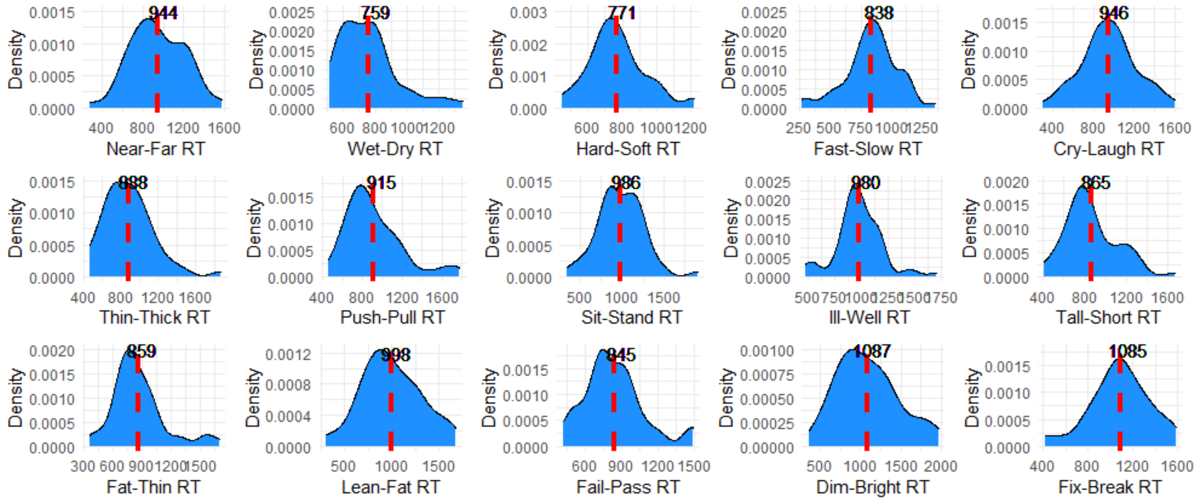

Figure 3: Less conventional pairs – kernel density graphs

4.2.3 Unconventional pairs

In the case of unconventional pairs, we can see that the majority of graphs are actually similar to normal distributions. So, in their case, there were not that many individual processing differences among the participants. There are no graphs with bimodal distributions and the means and medians of this group of pairs are close and that is why there are not many examples that are significantly right skewed. The only pairs that have a somewhat more distributed dispersion are *car-bus, need-greed* and *dress-pants*.

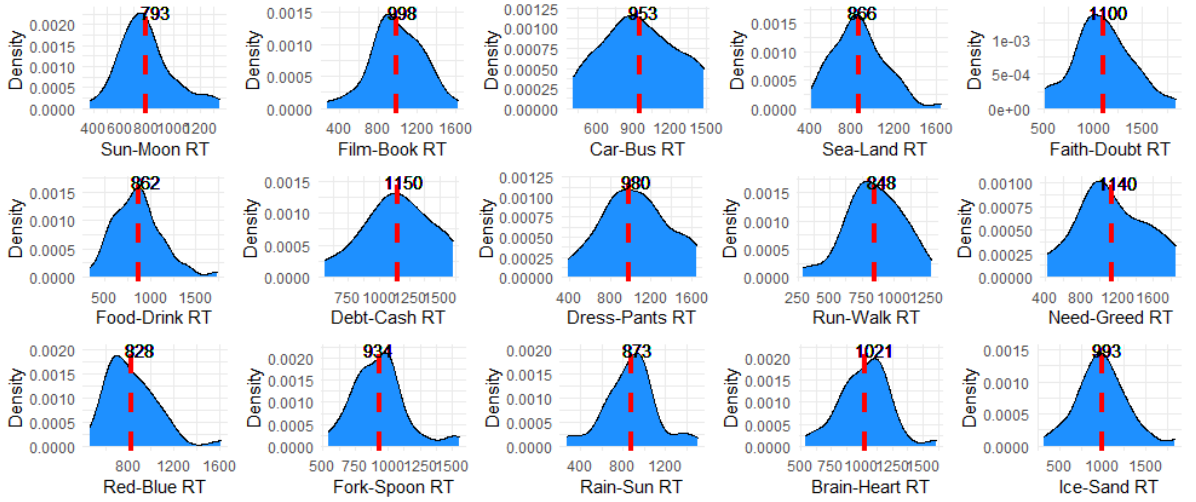

Figure 4: Unconventional pairs – kernal density graphs

Upon observing these kernel density graphs, we can conclude that data like these are dominated by individual differences, which should ideally be taken into account when statistically modelling it. We can also conclude that, although the outliers that skew the graphs are present in the majority of them, there is still a number of normal distribution graphs that can be observed and a number of bimodal ones, so it was not necessary to log-transform the data. The outliers can in this case also be informative. For instance, we have observed that they were less present in the group of unconventional pairs than in the other two, which tells us that their processing is more averaged. On the other hand, they were prevalent in the group of highly conventionalized pairs.

4.3 Inferential statistics

To generalize our conclusions to the population level we have fitted several generalized additive and mixed-effects models. In generalized additive models we have looked at the correlation between frequencies and RTs. RT is modelled on the y-axis as the dependent variable and frequency is modelled on the x-axis as the predictor variable. In mixed-effects models we have added the number of unavailable answers as a random variable to the model.

To justify the use of errors as a random variable we have fitted a linear regression model to show the correlation between the number of unavailable answers and RT. As can be seen in the graph below, there is a clear positive correlation between the two. That was confirmed in the summary of the model ($p < 0.01$, adjusted R-squared = 0.51, Pearson's r = 0.71). This tells us that those pairs that had more unavailable answers were also the ones that took participants longer to process, which means that accuracy and lexical access are also connected. That also means that if we were to transform the data of the unavailable answers they would have to be transformed carefully, as those answers would on average take longer than the mean of the group, which is often used to replace them.

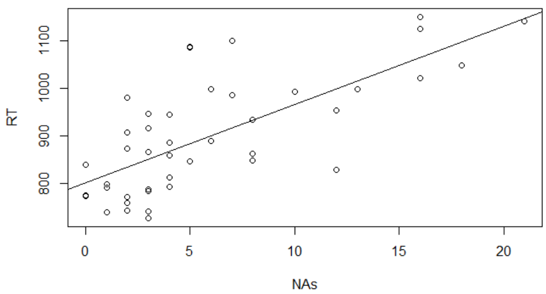

Figure 5: Linear regression model of RT and NAs

4.3.1 Highly conventional pairs

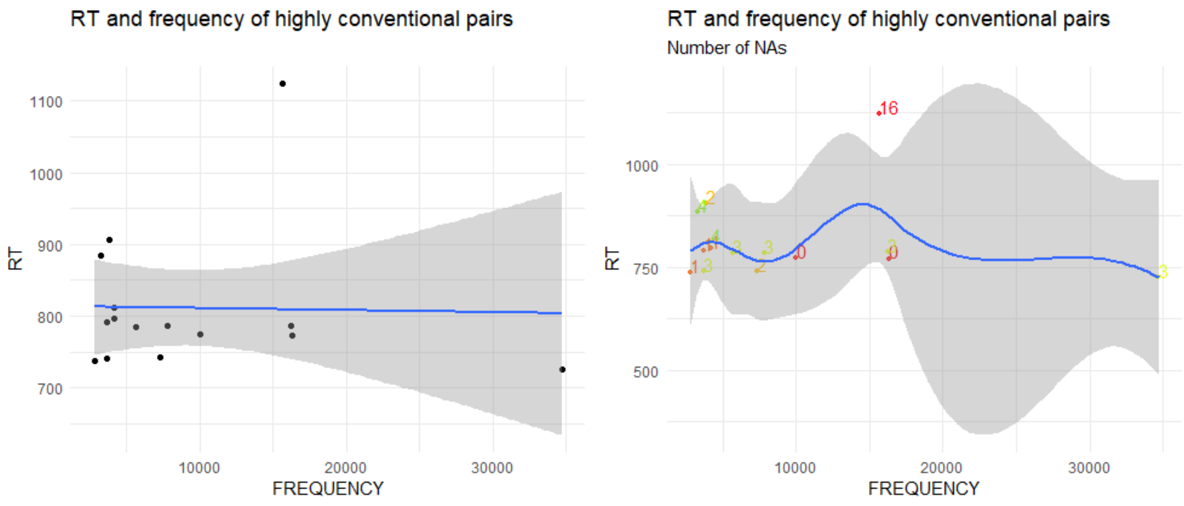

Figure 6: GAM of highly conventional pairs (left) and mixed-effects model of highly conventional pairs (right)

The first model that can be seen above on the left-hand side is the generalized additive model (GAM) of the highly conventional group of pairs. As can be seen, the regression line is flat and we can immediately conclude that the correlation between frequency and RTs in this case is not significant, which was also confirmed by performing the summary of the model (p = 0.93, adjusted R-squared = -0.076, deviance explained = 0.686%, Pearson's r = -0.025). To the right of the GAM we can see a mixed-effects model, in which we included the errors made per every pair. We can see that the only outlier with the highest RT is the one which has 16 unavailable answers and that is the pair *come-go*. To sum up, frequency is not in correlation with RT in the highly frequent data. That also confirms that the cutoff value of 2000 tokens chosen for the highly conventional pairs was a good one, as participants did not seem to perceive the differences between the pairs within this group.

4.3.2 Less conventional pairs

The first model that can be seen below on the left-hand side is the GAM of the less highly conventionalized group of pairs. As can be seen, the regression line in this case is not flat, so we can immediately conclude that correlation between frequency and RT is in this case present. The summary of the model has confirmed that there is a strong negative correlation between the two (p = 0.007, adjusted R-squared = 0.39, deviance explained =

43%, Pearson's $r = -0.66$). So, in the case of less highly conventionalized pairs frequency is in a strong negative correlation with RT. To the right of the GAM we can see a mixed-effects model, with the unavailable answers made per every pair just like in the previous example. In this case there is much more variation. There are no particular outliers, meaning that there were not great differences in accuracy in this group.

We can conclude that when it comes to the pairs that do not have a high, or rather a high enough, frequency, differences between different pairs do matter. The range between the frequencies of the pairs of highly conventionalized antonyms is greater (2,827-34,307; the difference being 31,480 tokens) than that of the less highly conventionalized pairs (34-1,017; the difference being 983 tokens) but no correlation was found. It seems that after a certain level the difference does not matter, as we simply do not perceive it.

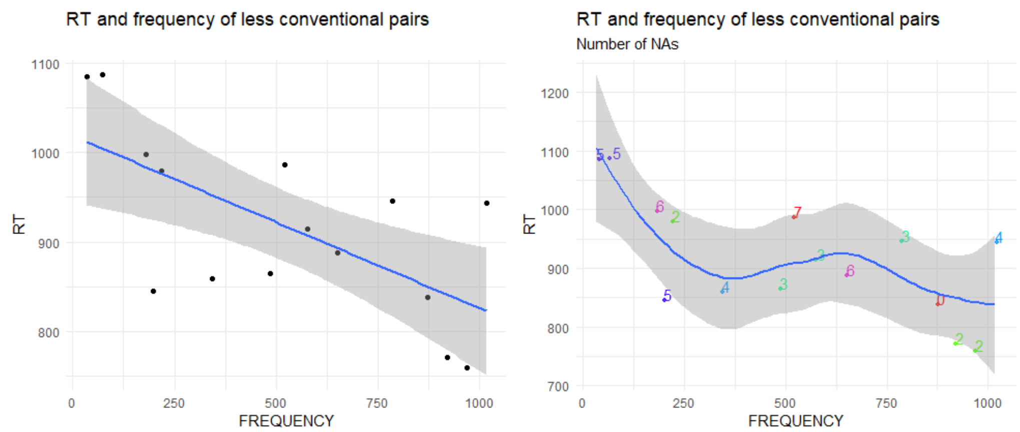

Figure 7: GAM of less conventional pairs (left) and mixed-effects model of less conventional pairs (right)

4.3.3 All pairs combined

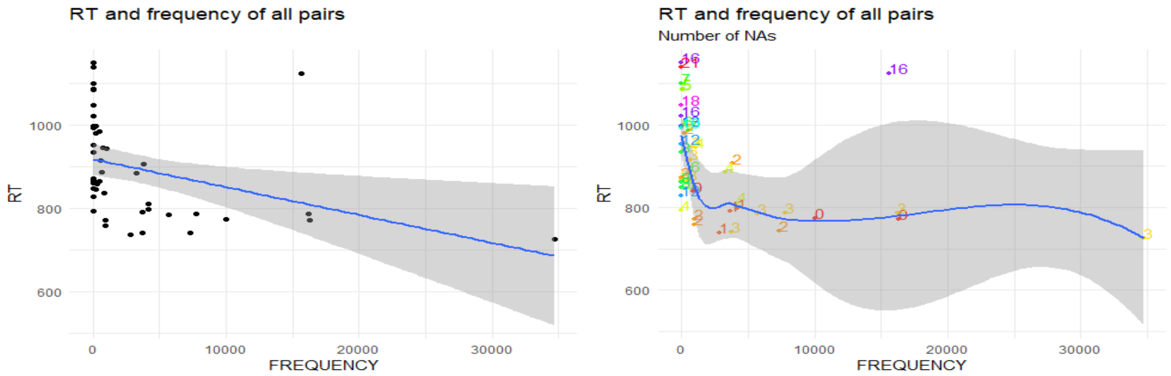

Figure 8: GAM of all pairs (left) and mixed-effects model of all pairs (right)

The first model that can be seen above on the left-hand side is the GAM of all pairs combined. The frequencies in the case of the unconventional pairs were set to zero, and that is why data points are concentrated around it. As can be seen, the regression line in this case is also not flat, so we can conclude that once again this is a case of negative correlation. The summary of the model has confirmed that ($p = 0.01$, adjusted R-squared = 0.11, deviance explained = 12.8%, Pearson's r = - 0.36). In this case correlation is not as strong as in the previous case. It is a moderate one. Still, we can say that frequency, or rather convention as we have modelled it, does indeed affect reaction time. If we look at the mixed-effects model, we can see that the data points per category with the highest number of unavailable answers are indeed placed on the highest points of the y-axis, indicating high RTs. Those are the following pairs: *need-greed* (21), *dress-pants* (18), *brain-heart* (16), *debt-cash* (16), *come-go (16), faith-doubt* (7) and *dim-bright* (5).

To conclude this section of the research paper, we can say that the degree of conventionalization, which we have modelled as frequencies derived from the corpus analysis, does in fact influence reaction time. That can be confirmed both by looking at the models of the highly conventionalized pairs, in which no correlation between the two was found, and which only underscores the fact that those pairs are really highly conventionalized (or rather, prototypical) and at the model of less conventionalized and all pairs combined where we have observed that frequency does indeed influence lexical-semantic processing. However, there are multiple other factors that should be considered and they are going to be discussed in the following section.

5. Discussion

Having made observations of the data obtained in our experiment, we can say that both our hypotheses have been confirmed. Namely, the pairs that we listed as highly conventionalized were indeed processed the fastest and the pairs that we listed as unconventional were processed the slowest. What's more, we have confirmed that the difference is also statistically significant. We have also shown that there is correlation between frequency and reaction time. The combination of p-values and Pearson's *r* has confirmed the statistical significance, even though the R-squared adjusted values were not very high. However, "language is complex and humans are messy" (Winter 2020:77), so models in linguistics can hardly account for all the variance present. What has to be kept in mind is that we chose a particular way of analyzing the data, which may not be a perfect way to analyze this kind of data with such a high level of micro variations, both by participant and by item. What would probably fit better in this case would be a multilevel analysis, such as a Bayesian one, which would take into consideration those variations as well.

Apart from the type of analyses we decided to implement, some other factors should be considered as well. As we discussed it in the theoretical part of the paper, in the section on reaction time, there are multiple factors that may influence reaction time and which could be accounted for. In the case of our experiment age is one of those factors. Since the range of participants' age was quite big (19-73), it would be good to include it in the model as a random variable. Also, when it comes to the individual features, in the survey prior to the experiment we asked participants whether they had any known language disorders. We could then also include those results in the model, although the number was not significant in this case.

Another thing to consider are the pairs used in the experiment. Word frequencies, which we used to mark the level of conventionalization of the antonym pairs, can be a complicated factor as well. Word frequencies are highly correlated with a number of other word features, such as word length, age at which the word was acquired, and similarity to other words. As a result, these features can act as confounding factors (Brysbaert et al. 2018). Brysbaert et al. point out "that the analyses of the reaction times obtained in megastudies suggest that all of these potential confounds have an independent effect on word processing (e.g., Brysbaert et al. 2016)." Another challenge that is encountered when it comes to word frequency is that the lexemes of low frequencies may indeed be processed as fast and as accurately as the ones with high frequencies. We could also see such examples in our study, e.g. *hard-soft* from the group of less highly conventionalized pairs is comparable to *high-low* from the group of highly conventionalized antonyms. Brysbaert et. al (2016) have found that word frequencies account for 30-40% of the variance found in lexical processing. Some of the explanations that Brysbaert et al. (2018) offer to explain that is that the words of lower frequencies are often morphologically connected to the ones of higher frequency, some are rarely used but people are still familiar with them and the number of encounters that we have with different words may not be that important as not all encounters are equally influential. For instance, in our study the example *come-go* has a rather high frequency but it took significantly longer to process than all the other pairs in the group, which points out the fact that, although some lexemes or pairs of lexemes may appear more often in the corpora, it does not mean that that is going to have an effect on the lexical access. That example also shows that some other factors, such as the order in which the pairs are presented should also be taken into account. Furthermore, we could observe that some of the examples from the group of unconventional pairs had less mistakes and shorter RT than some from the group of less highly conventionalized (e.g. *sun-moon* as opposed to *fix-break,* or as opposed to *come-go*). To try to counteract that effect, Brysbaert et al. (2016) introduced the variable of word prevalence, which was found to be especially helpful for the lexemes of low frequencies. So, in the case of our study we could introduce a questionnaire prior to the experiment in which we would test participants' knowledge of particular pairs, which would also enable a sort of data triangulation. The same thing could be done for different psycholinguistic variables, such as concreteness, imageability and valence, all of which could also influence processing times.

Finally, since reaction times are only indirect indicators of what may be going on inside the human mind, other types of tests should be used to study this phenomenon. Some of the often-used types of tests for this kind of research are studies using EEGs and fMRIs, which have their own specificities, and which will not be discussed here but could definitely be useful in further studying the phenomenon of antonymy and its correlation with convention.

6. Conclusion

In this study we were interested in the correlation of convention, which we modelled as the degree of frequency of different pairs, and the lexical-semantic processing of antonyms. As it turns out, the level of conventionalization does indeed affect processing, so we can say that both of our hypotheses have been confirmed. Antonyms deemed to be highly conventionalized were on average processed the fastest and unconventional antonyms were processed the slowest. The same effect was observed in the number of unavailable answers per each category. These differences possibly point out that different pairs of antonyms may be represented differently in our mental lexicon, as they are easier for us to access. We can say that this is just another confirmation of the importance of the cognitive linguistic meaning construal view of antonymy, which has also been confirmed by the mere fact that pairs of unconventional antonyms were also recognized as antonyms by the majority of participants. That underscores the fact that antonyms are not just lexical but also conceptual opposites, for processing of which we use both the knowledge of the language and the knowledge of the world.

Since we have also demonstrated that the relation between the three categories of antonyms and conventionalization may be complicated to study and interpret, further studies taking into account some of the features discussed in the previous section should be carried out to gain additional confirmation of the effects in question.

To conclude, we can say that an interplay between convention and the lexical-semantic processing of antonymy exists, once again confirming that language is a complex social and cognitive phenomenon to be further studied.

Abstract

In this study we have aimed to examine the role of psycholinguistic variables in the lexical-semantic processing of antonyms. The focus of the study was on the impact of convention on processing antonym pairs, as expressed with the psycholinguistic variable of reaction time. The aim was to further clarify antonymy, not only as a lexical, but also as a conceptual phenomenon. We used a reaction time experiment to test processing of three different categories of antonyms. Namely, those classified, upon doing a corpus analysis to determine their respective frequencies, as highly conventionalized, less highly conventionalized and unconventional. Our experiment has confirmed that there is indeed correlation between convention, as expressed by frequency, and reaction time. We take that as just another confirmation of antonymy being not only a lexical but a conceptual phenomenon too, in which both the knowledge of the language and the knowledge of the world play an important part. Finally, we have laid out some of the ways in which this lexical and conceptual phenomenon could be further studied to get more reliable results to account for the observed effects.

Key words: *antonymy, convention, reaction time*

References

Baayen, R.H. and Milin P. 2010. Analyzing Reaction Times. *International Journal of Psychological Research, 3 (2), 12-28.*

Bates D., Machler M., Bolker B., Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statstical Software*, 67 (1), 1-48.

Becker, Curtis A. 1980. Semantic context effects in visual word recognition: an analysis of semantic strategies. *Memory and Cognition* 8, 493-512.

Bianchi I. & Savardi U. 2006. Oppositeness in visually perceived forms. *Gestalt Theory*, 4, pp. 354-374.

Bianchi I. & Savardi U. 2008a. *The perception of Contraries*. Aracne, Roma.

Box, George E.P. 1976. Science and Statistics. *Journal of the American Statistical Association* , 71(356), 791-799.

Brysbaert, M., & Ellis, A. W. 2016. Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology* 30, 1240–1263.

Brysbaert M, Mandera P, Keuleers E. 2018. The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science* 27(1). 45-50. doi:10.1177/0963721417727521

Croft, William and D.Alan Cruse. 2004. *Cognitive linguistics.* Cambridge University Press.

Cruse, D.Alan. 1986. *Lexical semantics*. Cambridge University Press.

Cruse, D.Alan and Pagona Togia. 1996. Towards a cognitive model of antonymy. *Journal of Lexicology* I, 113-41.

Čulig Suknaić, Janja. 2020. Antonymy as a Conceptual Category of Semantic Opposition in English and Croatian. PhD diss., University of Zagreb.

Deese, James.1964. The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior* 3, 347-57.

Descrochers, Allan and Glenn L. Thompson. 2009. Subjective frequency and imageability ratings for 3,600 nouns. *Behavior Research Methods* 41 (2): 546–557. https://doi.org/10.3758/BRM.41.2.546.

Donders, Franciscus. 1969. On the speed of mental processes. *Acta Psychologica*, 30 , 412-431. (Translated by W. G. Koster)

Harrington, B. et al. 2004-2005. *Inkscape*.

Jiang, Nan. 2012. *Conducting Reaction Time Research in Second Language Studies*. Routledge.

Jones, Steven. 2004. *Antonymy. A corpus-based perspective.* Taylor & Francis e-Library.

Jones, Steven, M.Lynne Murphy, Carita Paradis, and Caroline Wilners. 2012. *Antonyms in English:construals, constructions and canonicity.* Cambridge University Press.

Lee, J.J. and C.F.Chabris. 2013. General cognitive ability and the psychological refractory period: Individual differences in the minds bottleneck. *Psychological Science* 24 (7): 1226-1233.

Lehrer, Adrienne and Keith Lehrer. 1982. Antonymy. *Linguistics and Philosophy* 5, 483-501.

Lyons, John. 1977. *Semantics.* Cambridge University Press.

Murphy, M.Lynne. 2003. *Semantic relations and the lexicon.* Cambridge University Press.

Paradis, Carita. 1997. *Degree modifiers of adjectives in spoken British English.* Lund University Press.

---------------- 2001. Adjectives and boundedness. *Cognitive Linguistics* 12, 47-66.

Paradis, Carita and Caroline Willners. 2006. Antonyms and negation: the boundedness hypothesis. *Journal of Pragmatics* 38, 1051-80.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

RStudio Team. 2021. RStudio: Integrated Developmen Environment for R. RStudio, PBC, Boston, MA URL: http://www.rstudio.com/.

Saussure, Ferdinand de. 1956. *Course in general linguistics.* London:Peter Owen.

Schramm, Pele & Rouder, Jeffrey. 2019. Are Reaction Time Transformations Really Beneficial?.10.31234/osf.io/9ksa6.

Stoet, Gijsbert. 2010. PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104.

------------------- 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.

Wagenmakers E., S. Farrell, and R.Ratcliff. 2004. Estimation and interpretation of 1/fα noise in human cognition. *Psychonomic Bulletin & Review* 11, 579-615.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.

Winter, Bodo. 2020. *Statistics for Linguists: An Introduction Using R.* Routledge: New York.

Whelan, Robert. 2008. Effective analysis of reaction time data. *The Psychological Record* 58, 475-482.

Wood SN. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society* (B), 73(1), 3-36.

Žic Fuchs, Milena, Vlatko Broz and Nina Tuđman Vuković. 2013. Communication technologies and their influence on language: the notion of convention revisited. *Jezikoslovlje* 14, 65- 84.

Corpus

COCA (Corpus of Contemporary American English) https://corpus.byu.edu/COCA/

Data and code repository