

Metaphorically opposed A conceptual metaphor theory-based approach to the classification of antonymy pairs

Gustav Styrbjørn Johannessen
Aarhus University
gustavstyrbjoern@yahoo.dk

Abstract

Theories of antonymy have traditionally been based on semantics, with some studies offering a cognitive approach as an alternative. However, neither approach has been able to account for the use of antonyms in discourse. This theoretical paper suggests a new cognitive approach based on the framework of cognitive metaphor theory, which could potentially account for discourse observations. The paper suggests methods to investigate the cognitive reality of this claim, and discusses its implications for human cognition.

1. The opposite of a conclusion

This paper aims to suggest a new approach to a topic in cognitive semantics, namely the nature of antonymy. Antonyms have been extensively described from many approaches, as they are widely agreed to be the most robust type of semantic relationship between lexemes. Antonymy is both the most easily identified semantic relationship, and it is extremely common in natural language, both with and without speakers' explicit acknowledgement (Murphy 2003: 169). Its relevance to the study of human cognition is evident: Arguably, the nature of semantic relations, including antonyms, can indicate how humans conceptualize the world around them (Stocker 2015). However, previous theories of antonymy have in some respects fallen short when confronted with actual language data (Lobanova et al. 2010). This paper presents the outline of an alternative approach based on Conceptual Metaphor Theory (CMT), a theoretical framework which has been demonstrated to underlie a number of linguistic mechanisms (Gibbs 2017). The paper argues that the concept of opposition is a fundamental part of the human conceptualisation of the world, making antonymy a linguistic manifestation of such conceptualisations. In this way, a theory of antonymy based on CMT could provide a framework to better explain the phenomenon as we see it in actual language use. As a framework, CMT is a particularly appropriate candidate for such a theory, since it already describes a set of well-established mechanisms which link conceptualisation to linguistic expression, which is yet to be applied in existing theories of antonymy.

The aim of this paper is to propose an approach to antonymy whose psychological reality can be empirically tested. Previous studies have found that the established categorization of antonyms into for example gradable antonyms, non-gradable antonyms, and converses is not validated by actual

discourse data (as discussed in Lobanova et al. 2010: 43). The purpose of this paper is to provide an alternative cognitive model of antonymy in the hopes that this will better align with the observed use of antonymy in language. This first section of the paper is an introduction to the aims of the paper. Section 2 contains brief accounts of relevant previous research in the fields of antonymy research and CMT. Section 3 discusses the claim of this paper. Finally, Section 4 contains a short summary and conclusion. The overall aim of this paper is to propose a potential new direction for future research on antonymy, exemplified with an outline of a theoretical model.

2. Existing literature

This section will serve as an overview of previous research relevant to the claim of this paper. The materials cited can broadly be divided into two separate bodies of research: Antonymy and CMT.

2.1. Antonymy Research

Antonymy has been studied with many different theoretical approaches. In this section I will briefly review three of these: The theoretical-semantic approach, the discourse-analytic approach, and the cognitive-semantic approach. For a more thorough summary of the history of antonymy research, see Jones et al. (2012: chapter 1).

In the theoretical-semantic approach, antonymy pairs have traditionally been sorted into a number of subcategories depending on semantic criteria, going back as far as the observations of Aristotle. The exact number and nature of the categories has varied depending on the aims of different researchers. Arguably the most commonly used set of categories is as follows: *Gradable antonyms*, also called *true antonyms* (antonyms which are gradable on a scale, such as HAPPY/SAD), *non-gradable antonyms*, also called *opposites* (mutually exclusive antonyms, such as DEAD/ALIVE, where something is always described by one of the terms, but never both), and *converses*, also called *multiple incompatibles* (antonyms which are, as the name suggests, sets of multiple mutually incompatible terms). However, Lobanova et al. note that "psycholinguistic studies of antonyms suggest these distinctions [the three aforementioned categories] do not play a role in the way antonyms are represented in the mental lexicon" (2010: 21). This conclusion is based on numerous studies described in the same paper (Deese 1964, 1965; Charles and Miller 1989; Kučera and Francis 1967; Justeson and Katz 1991; Fellbaum 1998), which explore antonymy in various ways. Through psycholinguistic experiments, the studies find a certain set of antonyms which they label as *canonical*, but which are not adequately explained by purely semantic criteria. This suggests that the purely semantic established definitions, while terminologically useful, do not accurately represent the cognitive landscape of the human language faculties.

With this in mind, it is useful to consider antonymy from a perspective other than the purely theoretical-semantic. One such candidate is through discourse analysis: Categorising antonymy pairs by the different functions they can have in conversations. These functions are discussed by Jones (2007), who uses discourse functions rather than semantic criteria when investigating the differences in the nature of antonymy in children versus in adults. Based on his data, Jones divides

discourse functions into nine discursive categories, two of which are by far more common than the rest. These are *ancillary antonymy* (in which a pair of antonyms is used to highlight contrastive features of a corresponding pair of terms or phrases, e.g. "The worst episode of *Frasier* beats the best episode of *Friends*") and *coordinated antonymy* (in which a pair of antonyms is used to denote opposing ends of a spectrum, usually in order to refer to everything within that spectrum, e.g. "We serve everything from pumpkin spice latte to regular black coffee").

The significance of discursive functions is further explored by Lobanova (2012:21-24), who uses a corpus-driven approach to establish functionally distinct uses of antonymy naturally occurring in speech. Among other things, the study finds that noun-noun pairs are represented far more in collections of antonyms identified by algorithms than when identified based on semantic criteria, again casting doubt on the psychological reality of the established semantic categories. A relevant question here is whether the established semantic categories are even compatible with the findings of these more recent usage-based studies. For the purposes of this paper, I take the view that both types of categorisation (those based on semantic features and those based on usage patterns) shed light on aspects of the nature of antonymy, but that for such a categorisation to have meaningful implications about human cognition, a theoretical model must correlate with observable data, such as in the psycholinguistic studies discussed by Lobanova et al. (2010:21-23).

Paradis and Willners (2011) posit a framework of antonymy which, broadly speaking, bases the phenomenon in cognitive semantics, and thus attempts to establish a more psychologically accurate theory of antonyms. The study works within the framework of Lexical Meaning as Ontologies and Construals (ibid: 2), which views conceptualisation as a process arising from structures and construals (with different subcategories under each of these terms). Here, *structures* describe how we conceptualise the constituent elements of the world, and *construals* describe how we conceptualise the relations between them (Paradis 2005). It concludes that "antonyms are meanings that are used in binary opposition through a construal of comparison" (2011: 383). This construal is a mental 'border' along which the semantic features of any given antonymy pair are compared. For example, in the antonymy pair 'TALL/SHORT', the construal is the border of *height*, along which the semantic features of the two terms are compared. Within this framework, an antonymy pair is, cognitively speaking, two lexical items whose relationship is defined by a specific type of construal: a comparison of features. Paradis and Willners provide a cognitive framework for the conceptualisation of antonymy, but do not link this to discursive functions. As an alternative, this paper suggests a theory that might be able to connect the cognitive and discursive findings from the papers described so far.

2.2. Conceptual Metaphor Theory

The seminal work of CMT is Lakoff & Johnson (1980), in which the theory and its base implications are first presented. The basic claims of the theory are as follows: Rather than just a rhetorical flourish, metaphor is a crucial mechanism in conceptualisation which helps humans to understand concepts by comparing them to other concepts. Typically, the concept that we use in order to understand (called the source domain) is more integral to the human experience than the

concept being understood (called the target domain). This means that certain concepts, mostly those directly involving the physical experience of having a human body, are employed to understand more abstract ideas. CMT further claims that the source domain used in a metaphor hides some aspects of the target domain while highlighting others. The theory views metaphors in language as an expression of the conceptual metaphors in our minds, and the same conceptual metaphor can, and usually does, have multiple linguistic expressions.

An example of a conceptual metaphor is ARGUMENT IS WAR, in which the abstract concept 'argument' is conceptualised in terms of the concrete experience of physical fights. Lakoff and Johnson list multiple examples of expressions of this metaphor, such as "Your claims are *indefensible*" and "I've never *won* an argument with him" (ibid: 4). In this case, the source domain highlights the competitive aspects of an argument and hides others, such as the potential goal of learning from each other's perspective (since that is a part of the concept 'argument', but not 'war').

One part of the framework posits the existence of 'orientational metaphors', a type of metaphor often used for abstract concepts such as emotions, values, or morals. According to Lakoff and Johnson, "[...] most of them have to do with spatial orientation: up-down, in-out, front-back, on-off, deep-shallow, central-peripheral" (1980:14). An example of orientational metaphor in action can be seen in the expression "to *rise above* something", in reference to a person who chooses not to engage in conflict in reaction to someone else's actions, insults or differing opinions. This expression is built on the orientational metaphor RATIONAL IS UP.

3. The claim

When Lakoff and Johnson define the concept of orientational metaphors (1980: 14), all of their examples are pairs of antonyms. Based on this observation, this paper's central claim can be stated as follows: *Antonymy pairs are cognitively framed in terms of a metaphorical experiential basis. More specifically, a comparatively small number of antonymy pairs - those described as orientational metaphors - constitute the source domains for our conceptualisation of all other antonymy pairs.* For example, the antonymy pair UP/DOWN, which is described by Lakoff and Johnson as the source domain in some orientational metaphors (1980: 14), is the source domain for the antonymy pair HAPPY/SAD, where UP is the orientational metaphor for HAPPY, and DOWN is the orientational metaphor for SAD. This paper proposes that antonymy pairs can be sorted into two different categories: Intrinsic antonyms and non-intrinsic antonyms. Intrinsic antonyms make up the source domains of orientational metaphors. Their constituent concepts are directly experienced in the physical world, and so is their intrinsic oppositeness. Examples include UP/DOWN or INSIDE/OUTSIDE. These pairs are intrinsically antonymous. In contrast, non-intrinsic antonyms are not *intrinsically* antonymous, but are nevertheless viewed by speakers as indisputably antonymous. Examples include HAPPY/SAD and FIRE/WATER. Concepts found in this category may have more than one antonym each. For example, HAPPY can be the antonym of SAD, but also of ANGRY. In contrast, UP has only one antonym: DOWN. Thus, this paper claims that we perceive e.g. HAPPY and SAD as antonyms not because the concepts they denote are

intrinsically opposite, but because we conceptualise them via UP and DOWN, respectively, which *are* intrinsically opposite.

The most useful approach to testing this claim would be a multi-faceted research project collecting many different types of data. Different types of data could include brain scans (such as magneto- and electroencephalography), conversation analysis, psycholinguistic laboratory experiments, language user surveys or studies of patients with aphasia. Brain scans, psycholinguistic experiments, language user surveys and studies of aphasia may shed light on the neurological and psychological underpinnings of antonymy. Neurologically, the two categories of antonymy pairs may elicit different neurological responses in speakers and listeners. Psycholinguistic experiments may reveal whether the categories are differently affected by phenomena such as priming, or whether they influence reaction times. For example, the claim would be supported if non-intrinsic pairs are observed to have a higher reaction time, or if the effect of priming is observed to be stronger in intrinsic pairs (in both cases due to intrinsic pairs being closer connected in meaning). Conversation analysis could establish whether the two categories have different functions in conversation, such as one category preceding longer hesitations than the other, or being used more often to create contextual antonymies. The claim would similarly be supported if studies of patients with aphasia were to reveal the possibility of losing either use of understanding of some or all non-intrinsic antonymy pairs, but not the intrinsic. The proposed claim could be used as a theoretical basis to account for the discourse observations of Lobanova et al. (2010), which is something Lobanova et al. observe that existing frameworks fail to do. In doing so, it would provide an alternative to Jones (2002) and Paradis & Willners (2011). This makes the claim a potential basis for future research, and a help to more accurately determine the cognitive role of metaphor and opposition. If supported, the theory would also support the CMT framework as a whole.

One question of particular interest, should this theory be empirically supported, is the nature and number of non-metaphorical antonymy pairs - that is, the antonymy pairs that form the metaphorical basis of other antonymy pairs. Other than Lakoff and Johnson's original orientational metaphors (1980: 14), I have been unable to think of many true such pairs, and the only ones that come to mind are directional antonyms, i.e. UP/DOWN, NORTH/SOUTH. Is it likely that directions are the only true example of opposition that occurs in nature? And, if so, are all antonyms in human language metaphorical extensions of that one opposition? One could argue that an opposition such as ALIVE/DEAD occurs non-metaphorically in nature, but the fact that one term denotes simply the absence of the other could mean that the terms actually cover related, mutually exclusive, but not technically *opposite* concepts - that is, not opposite without the involvement of human conceptualisation. This question has far-reaching implications for human cognition, but requires a solid theoretical foundation to investigate, which this theory could begin to provide.

3.2. *Points of criticism*

Following the presented theory, there are a number of factors which would be likely to complicate any empirical testing and which should be kept in mind. First, one major complicating factor is the effect of negation morphemes. In psycholinguistic experiments, for example, antonymy pairs where

one term is identical to the other term, but with a negation morpheme added (such as HAPPY/UNHAPPY), it is likely that HAPPY would elicit UNHAPPY as its antonym in speakers of English far more consistently than other properties of the words allow this model to predict; and it is to be expected that the effect would be even stronger the other way around, with UNHAPPY eliciting HAPPY almost without exception. In the testing of any model based on this paper's claim, this should be considered a source of error. In experiments involving researcher-picked stimuli, the issue can be partially addressed by avoiding the use of these words when picking out stimuli - but if so, that discrimination of stimuli will itself be a source of error to take into account when drawing conclusions.

Second, the theory can be argued to be incompatible with other well-established approaches. One example of this is the approach of Jones et al. (2012: 8-10), in turn adapted from Murphy (2003). In this view, antonymy is primarily viewed as arising from context, with some antonymy pairs (known as *canonical antonyms*) becoming universally accepted within one or more languages by way of repetition. According to this framework, a pair becomes canonical when it is used often enough in a context that construes it as antonymous for speakers to register and internalise it as such. Contextual antonymy is in itself an extremely common and well-documented phenomenon and has been shown to affect the judgement of language users in antonymy research (Van de Weijer et al. 2014). For this reason, studies of antonymy tend to benefit from clearly delineating the distinction between canonical and contextual antonyms. The view taken here is that this paper's claim is in fact compatible with the approach of Jones et al. (2012), which defines the criteria of how contextual antonymies arise, but not the cognitive mechanisms behind that process. However, contextual antonymy does give rise to another apparent issue with this claim. Through contextual antonymy, practically anything can - with a little linguistic innovation - be construed as the antonym of anything else. This seems to contradict the idea that antonyms stem from orientational metaphors, since there may not always be an obvious orientational metaphor for every instance of contextual antonymy. This apparent inconsistency can, however, be explained by considering the individual features of each lexical item in this context. In the example sentence "*I'd rather die than colour my hair*", 'to die' is contextually made antonymous with 'to colour my hair', which is not commonly the case, and which brings no obvious orientational metaphor to mind. However, the two terms are compared on only one of their features, namely their agreeableness to the speaker (from the speaker's perspective, simply 'goodness'), and it's evident from the context that the former has more of that than the latter. Murphy (2003: 170) describes that antonymy arises exactly from the comparison between lexical items which "have all the same contextually relevant properties but one". In this case, that one contextually relevant property is 'goodness'. To see how this fits with this paper's claim, we now no longer have to identify an orientational metaphor that maps onto 'die' and 'colour my hair', but instead onto 'good' and 'bad' - and as discussed in Lakoff & Johnson (1980:16-17), the orientational metaphor UP/DOWN does exactly that.

4. The opposite of an introduction

The human ability to use opposition as a tool of conceptualisation might stem from the metaphorical extension of oppositions found in the way humans experience the world. Antonymy, being the linguistic expression of that ability, is suggested in this paper to arise from orientational metaphors. This paper presents the claim that there are two types of antonym pairs: Intrinsic and non-intrinsic. While intrinsic antonym pairs are directly experienced in the physical world, non-intrinsic antonym pairs are not intrinsically opposite, but are considered opposites due to the metaphors through which they are conceptualized. The present paper seeks to communicate this idea, and to provide useful context for any further studies, including a summary of relevant previous research as well as a discussion of how to test this claim. Investigations of this claim could potentially shed light on aspects of human cognition and play a role in semantics as a novel framework for studying antonymy, complementing a number of existing approaches.

Future investigations on the topic might include a multi-faceted research project such as that described in Section 3, in order to test the claim. Another approach would be through language user surveys, which could investigate whether the proposed distinction between intrinsic and non-intrinsic antonymies affects speakers' attitudes to antonyms differently.

References

- Charles, Walter G. & George A. Miller. 1989. Contexts of Antonymous Adjectives. *Applied Psycholinguistics* 10(3): 357–375.
- Deese, James. 1964. The Associative Structure of Some Common English Adjectives. *Journal of Verbal Learning and Verbal Behavior* 3: 347–357.
- Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Gibbs, Raymond W. Jr. 2009. Why Do Some People Dislike Conceptual Metaphor Theory? *Cognitive Semiotics* 5(1): 14–36.
- Gibbs, Raymond W. Jr. 2011. Evaluating Conceptual Metaphor Theory. *Discourse Processes* 48(8): 529–562.
- Gibbs, Raymond W. Jr. 2017. *Metaphor Wars*. Cambridge: Cambridge University Press.
- Jones, Steven. 2002. *Antonymy: A corpus-based perspective*. London: Routledge.
- Jones, Steven. 2007. 'Opposites' in discourse: A comparison of antonym use across four domains. *Journal of Pragmatics* 39: 1105–1119.
- Jones, Steven, M. Lynne Murphy, Carita Paradis & Caroline Willners. 2012. *Antonyms in English: Construals, Constructions and Canonicity*. Cambridge: Cambridge University Press
- Justeson, John S. & Slava M. Katz. 1991. Co-occurrences of Antonymous Adjectives and their Contexts. *Computational Linguistics* 17: 119.
- Kövecses, Zoltán. 2008. Conceptual metaphor theory: Some criticisms and alternative proposals. *Annual Review of Cognitive Linguistics* 6(1): 168–184
- Kövecses, Zoltán. 2011. Methodological issues in conceptual metaphor theory. In: Schmid, Hans-Jörg & Sandra Handl (eds.), *Windows to the Mind: Metaphor, Metonymy and Conceptual Blending*. Berlin: De Gruyter Mouton. 23–39.
- Kučera, Henry & W. Nelson Francis 1967. *Computational Analysis of Present-day American English*. Providence: Brown University Press.

- Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago & London: The University of Chicago Press.
- Lobanova, Anna. 2012. *The Anatomy of Antonymy: a Corpus-driven Approach*. Groningen: University of Groningen dissertation.
- Lobanova, Anna, Tom van der Kleij & Jennifer Spenader. 2010. Defining Antonymy: A Corpus-Based Study of Opposites by Lexico-Syntactic Patterns. *International Journal of Lexicography* 23(1): 19–53.
- Murphy, Lynne. 2003. *Semantic Relations and the Lexicon*. Cambridge: Cambridge University Press.
- Paradis, Carita. 2005. Ontologies and construals in lexical semantics. *Axiomathes* 15: 541-573.
- Paradis, Carita & Caroline Willners. 2011. Antonymy: from conventionalization to meaning-making. *Review of Cognitive Linguistics* 9(2): 367-391.
- Stocker, Kurt. 2015. Toward an Embodied Cognitive Semantics. *Cognitive Semantics* 1(2): 178-212.
- Van de Weijer, Joost, Carita Paradis, Caroline Willners & Magnus Lindgren. 2014. Antonym canonicity: Temporal and contextual manipulations. *Brain and Language* 128: 1-8.