

Incorporating Polarity in Lexical Resources

The felicity of contrast and parallel relations as well as structures such as ellipsis are in part dependent on the presence of lexical elements contrasting or sharing. Traditional dictionaries and lexical resources like WordNet code information about antonyms, yet this is often not the type of polarity found in real contrast or parallel examples. It is not yet clear how negative or positive polarity can be objectively defined and how it can be identified for individual examples. Further, many adjectives are interpreted differently depending on the context in which they are used. This ability of the context to select the sense of the adjective is parallel to meaning coercion in other adjective-noun combinations and should also be treatable in a computational linguistic framework. This suggests that using a dynamic approach like the Generative Lexicon (GL) to allow polarity coercion would be more appropriate, but GL currently doesn't even code antonym relationships, let alone polarity. Further, the use of contrastive or parallel markers like but or and play a role in the final polarity interpretation. This ability seems to be unique among function words and isn't currently handled in GL. Since the sense shifts are completely local, it is similar to other coercion processes. We conclude that this type of information, and coercion processes would perhaps be best handled within a framework like GL, but that the determination of the default polarities themselves might be extractable from already existing lexical resources like WordNet.

1 Introduction

Consider the following examples, paying particular attention to the adjectives:

- (1) a. This is an *exquisite_P* and *practical_P* idea.
- b. This is an *exquisite_P* but *impractical_N* idea.
- c. This is an *exquisite_N* but *practical_P* idea.¹
- d. This is an **exquisite_N* and *impractical_N* idea.

In (1a), *exquisite* is coordinated with *practical* to give two positive descriptions of the idea. In (1b) *exquisite* has a positive meaning, through the contrastive coordinator *but* it is contrasted with the negative descriptor *impractical*. However, in (1c) it seems that the contrastive coordinating conjunction together with the clearly positive *practical* coerces *exquisite* into a negative polarity interpretation. In (1d) we further see that somewhat surprisingly, it is not possible to coordinate *exquisite* with *impractical*, without the contrastive coordinator. This seems to be because *exquisite* has a positive polarity as its default.

In the above description we have appealed to an informal understanding of the concept of polarity, but how can polarity be objectively determined? Are some words ambiguous as to their polarity, and further how dynamic is a polarity default: can lexical items be coerced from one polarity type to another in certain contexts?

Our aim is to look at these questions from the perspective of how two formalisms, WordNet and the Generative Lexicon, would deal with them, or if not currently able to deal with them, to what extent they could be extended to deal with them. This work is relevant for computational discourse work since it seems that polarity is the major factor that licenses contrastive relations in discourse. Since contrastive relations are one of the most frequent relations, and since corpus study has shown that many contrast relations are unmarked (Marcu & Echiabi, 2001; Carlson et al. 2001), NLP applications that attempt to automatically identify contrast will need lexical resources that can deal with polarity.

¹ It seems that you can only coerce to a negative meaning if the positive adjective is clearly positive.

2 Introduction

2.1 Antonyms and Scales

Semantic polarity is usually discussed in the literature in the context of opposition. Although opposition is a readily recognized lexical relation, there is no clear-cut definition or classification of what opposites are. Put informally, opposites are two words that share the same semantic domain but differ in one particular aspect. According to this definition, one opposite of *practical* in the example above is *impractical*. Similarly, an opposite of *rich* is *poor*, and an opposite of *dumb* is *smart/clever*. In this paper, we are interested in gradable opposites only, which are shortly referred to as antonyms (Cruse: 1986). A complete classification of opposites is out of the scope of this paper².

One of the basic properties that a pair of antonyms has in common is that both adjectives share the same semantic domain that can be represented as a scale. For example, antonyms *rich* – *poor* share the same scale WEALTH, occupying opposite poles on it. Other antonyms, such as *well-off* and *needy* are also represented on this scale but in different intervals. Diagram 1 below illustrates this:

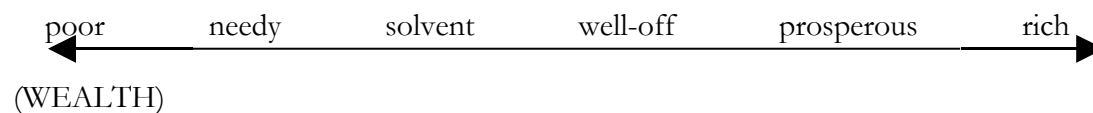


Diagram 1

A poor person clearly has less money than a person in need who has less money than someone solvent, etc. This relation can be summarized as $poor < needy < solvent < well-off < prosperous < rich$. Adjectives *poor* and *rich* appear to be a better antonym pair than *poor* – *prosperous*. This is because these two adjectives are on the (opposite) ends of the scale while *prosperous* is not³. Thus, it is important to determine the position of an adjective on the scale in relation to other opposites⁴. The exact position of the adjectives is necessary in order to identify antonym pairs. However, in our example above, the polarity of *exquisite* rather than its position on a scale plays a role. And this is exactly what the scale on Diagram 1 fails to show: the adjectives *poor* and *needy* entail negative polarity, while *solvent*, *well-off*, *prosperous* and *rich* have positive polarity.

2.2. Polar Opposition

Polarity has been widely debated in theoretical literature (Hoeksema, 1983, Cruse, 1986, Kennedy, 2001). In recent years it is also a relevant topic in computational linguistics (Hatzivassiloglou & McKeown, 1997, Kamps, Marx, Mokken & de Rijke, 2004), especially in the area of automatic identification of subjective opinion in a text.

Some adjectives like *terrible* and *ineffective* or *excellent* and *effective* have an inherently negative or positive polarity. But what about such adjectives as *simple*, *fast* and *hot*? Cruse (1986) suggested that the lack of properties can be regarded as an indication of negative polarity. That is *poor* is negative because it implies that a person has no or little money, *rich*, on the contrary, is positive because it entails possession of money⁵. Similarly, *dumb* in

² An interested reader can look for more information in Lehrer and Lehrer: 1982; Cruse: 1986; Andrew: 1993.

³ In WordNet such opposites as *poor* – *rich* are called direct antonyms and *poor* – *prosperous* are called indirect antonyms.

⁴ There are different alternatives for representing scales: the adjectives can be represented either as an ordered set of points (degrees) (Seuren: 1973; Cresswell: 1976) or as intervals (Kennedy: 2001).

⁵ Antonyms *safe* – *dangerous* seem to be a counterexample.

example (2) is negative because it indicates John's lack of intelligence, *rich* is on the other hand implies that John has something, namely, he is wealthy.⁶

(2) John is rich_P but dumb_N. (from Lakoff 1974:133)

(3) *John is rich but poor.

In (2) *poor* can be contrasted with *dumb* when predicated of John. On the contrary in (3) this is clearly not possible because *poor* and *rich* in their standard interpretations are mutually exclusive. One reason for this is that in (3) both adjectives belong to the same domain and share the scale WEALTH. In example (2), the adjectives represent different domains (*rich* entails WEALTH, *dumb* entails INTELLIGENCE). In addition, on the scale WEALTH *rich* has positive polarity, while *dumb* is negative on the scale INTELLIGENCE. Apparently, the type of positive vs. negative polarity is what licenses a contrastive relationship in (2). This means that in contrastive sentences like (2), the adjectives must represent different scales and they must entail opposite polarities on these scales. Here, rather than dealing with one domain, we are dealing with multi-dimensional scales!

Things get even more complicated in examples like (4).

- (4) a. It was a simple_N but effective_P plan.
b. It was a simple_P and effective_P plan.
c. It was a simple_P but ineffective_N plan.
d. It was a simple_N and ineffective_N plan.

In all four sentences (4a-d) there are two scales: a scale DETAIL and a scale EFFICTIVENESS. Although we can define the domains, it is not clear how the polarity is assigned. While *effective* is always (inherently) positive, the adjective *simple* shifts its polarity depending on the conjunction it is used with, either describing a lack of complexity or of sophistication. This type of polarity shift and how it is licensed by the type of rhetorical relation is discussed in the next section.

2.3 Contrast and parallelism

Polarity seems to play a central role in the rhetorical relations of contrast and parallelism: contrast is often explicitly marked with cue words such as coordinating conjunctions like *but*, *although* and *however* or numerous adverbs. Parallelism is marked with words like *also*, *as well*, or *too*. Example (4) above illustrates that coordination requires coordinated elements to be interpreted as having the same polarity, while contrast requires them to have opposite polarities.

Since the cue words for contrast or parallelism are closed class function words, they are usually not dealt with on a lexical level but on a structural or discourse level because it is assumed that their function is primarily to relate different prepositional information. Function words are treated separately from open class words because they are not considered to influence the choice of the sense of the words they co-occur with. However this generalization doesn't seem to hold for contrastive markers like *but*.

If we look at (4) we see that *simple*, which can be given a negative or positive interpretation, has its polarity determined by two factors: 1) whether or not it is

⁶ There two possible directions here. On the one hand, it can be said that some adjectives denote 'objective' values and thus do not entail any polarity. On the hand, it can be said that polarity consists of three categories: negative, positive and neutral (or objective) (Koppel & Schler, 2005).

conjoined with a contrastive or a parallel conjunction, and 2) whether or not the adjective it is conjoined with is strongly negative or positive. Contrastive marking requires opposite polarities for the conjoined adjectives, while parallel *and* requires the same polarity for conjoined adjectives. It seems quite straightforward to say that the polarity of *simple* is coerced/determined by its local context, the default polarity of the other adjective and the presence or absence of a contrast or parallel marker determines the appropriate polarity. Note also that this has nothing to do with the context of the utterance, but is much more local, having to do with what it is joined with.

However, we also saw for example (1) that this simple pattern doesn't hold for all adjectives. *Simple* seems to be more flexible than *exquisite*. In (1d), we see that conjoining *exquisite* with a negative adjective isn't sufficient to obtain its negative interpretation (*cf.* (4d)). Is *simple* a different sort of adjective than *exquisite*? It does seem that *simple* is more neutral, and allows scales to be flipped, whereas *exquisite* seems to be positive by default.

These are clearly lexical problems: the basic polarity orientation of an adjective, and what it is combined with in the local context strongly influence its final polarity interpretation, its sense, as well as limiting the contexts in which it is appropriate. How (and where!) should this information be represented in a lexical resource? Further, how can we build into our interpretation of *but* the fact that it has this sense-changing ability which seems to be otherwise lacking in other function words? In the following section we look at how WordNet and GL might deal with these questions (since unfortunately, they currently do not address them).

3 Lexical approaches

3.1 WordNet and polarity

WordNet, like traditional dictionaries, codes information about traditional lexical antonym pairs like *rich-poor*. Further, synonyms of one member of an antonym pair are considered to be indirect antonyms to the other member of the pair, i.e. *wealthy* is considered an indirect antonym of *poor* because *wealthy* is a synonym of *rich*.⁷ However, WordNet makes no reference to polarity. But it might be possible to determine the polarity of a word from the gloss. Consider the entry for *poor* we find two senses that situate *poor* on the scale of WEALTH:

- *having little money or few possessions*⁸
- *characterized by or indicating lack of money*

These definitions do mention “lack” and “little” and “few”, clues that could be used to guess negative polarity. Further if we also know that the sense of *poor* relevant for an example is on the WEALTH scale we might be more confident in our judgement if we also looked at an antonym, such as *wealth*:

- *having an abundant supply of money or possessions of value*

Abundance seems to be positive compared to lack. Thus even though WordNet doesn't specifically identify polarity, it might be possible to make fairly clear guesses at what polarity a word has. But for adjectives like *simple* where the polarity is flexible this is quite difficult. First we should determine if there is some sort of default polarity. WordNet

⁷ WordNet also tries to define antonyms for nouns and verbs, and interesting fact verbal antonyms often share an entailment, e.g. *hit-miss* both entail *aim*.

⁸ All definitions taken from the online version of WordNet 3.0:
<http://wordnet.princeton.edu/perl/webwn?s=word-you-want>

distinguishes between seven different senses of *simple*. How many are positive and how many are negative?

1. *having few parts; not complex or complicated or involved*
2. *easy and not involved or complicated*
3. *apart from anything else; without additions or modifications*
4. *exhibiting childlike simplicity and credulity*
5. *lacking mental capacity and subtlety*
6. *(botany) of leaf shapes; of leaves having no divisions or subdivisions*
7. *unornamented*

Senses 1, 2, 3, 5, 6 and 7 are all clearly described as showing something lacking, e.g. not complex”, “uncomplicated”, “without additions”, “lacking mental capacity”, “unsubdivided” or “unornamented”, but in contrast to the case of *poor*, it is positive to lack things for sense 1, 2 and 3, while it is negative to lack mental capacity in 5, and it seems to be neither negative nor positive to lack something in 6 and 7. It’s also unclear if 4 is positive or negative, since childlike simplicity could be both. This could be a reflection of the ability of *simple* to be both negative or positive, or it could be evidence that apart from simple examples like *poor*, there will be no automatic and simple way to code WordNet entries for polarity.

Because WordNet is a static resource (a so-called enumerated lexicon), where each sense is listed beforehand, the case of positive polarity items being coerced into negative polarity interpretations would only be treatable if the potential negative polarity is already coded as one of the senses of the lexical item. Because example (4) refers to a plan, and compares the plan on both the EFFECTIVENESS and the DETAIL scale, it almost seems that we will have to either have double entries for the way *simple* can be interpreted in relation to plans. Because this would lead to enumeration of a large number of entries that perhaps follow some predictable pattern it might be more efficient to deal with the polarity coercion issue by rules, such as those developed in the generative lexicon (GL) approach (Pustovjesky 1995).

3.2 Generative Lexicon

Even though adjectives have been discussed in the literature in the framework of GL (Pustejovsky & Boguraev: 1993; Bouillon: 1996, 1999), the problem of polarity coercion has not been addressed. GL is a dynamic theory that provides a gracious way of dealing with ambiguity and context-dependence of adjectives exemplified in (6) (from Pustejovsky, 1995a:73).

- (6) a. a fast plane: a plane that is inherently fast.
b. a fast book: one that can be read in a short time.

Besides ambiguity affected by the noun *fast* modifies, we want to know whether its polarity in (6a-b) stays the same. According to the discussion in Section 2.2, the adjective *fast* entails positive polarity by default since lack of speed is something negative. In fact, *fast* in (6a) is positive because it means that the plane needs less time to get from point A to point B. However, the polarity of *fast* in (6b) is not clear because the same explanation is not applicable since the fact that a person needs less time to read a book does not

imply that it has positive polarity. It is not clear how the polarity can be assigned in this case.

Another case is the adjective *simple*. In the example below, simple understanding implies some mental inability (to understand complex concepts):

- (7) a. a simple understanding
b. a simple plan

In (7b) the same approach works in the opposite direction: the absence of complexity in a plan makes it easy (or simple) to fulfill. This kind of polarity shift causes problems since the same reasoning (absence of complexity) leads to different polarity. Now let's go back to our example (4), repeated here for convenience in (8):

- (8) a. It was a simple_N but effective_P plan.
b. It was a simple_P but ineffective_N plan.

Contrastive constructions like (8a,b) represent a clear example of shift in polarity, and the coercion stems both from the polarity of the coordinated adjective and from the presence of the contrastive conjunction *but*. Although coercion in (6a-b) has been discussed in the framework of GL, the shift in polarity shown in the sentences (7) and (8) has not. GL could possibly be modified to capture this type of coercion.

Besides dealing with polarity, another problem with the GL approach is that it attempts to interpret sentences from the content words alone, where the sense of the lexical entries is clearly influenced by the type of coordinating conjunction. While the restriction of the realm of the interpretations to content words alone is quite defensible, since e.g. it is unlikely that closed class items like prepositions, frequency adverbs, articles and helping verbs do not effect this lexical information, it seems that contrastive marking *does* effect the interpretation of the lexemes. In this short report we've only looked at the contrast between *and* and *but*, but we suspect that words such as *however* and *although* will have a similar effect. Perhaps these words then do need to be incorporated into GL so that they can affect polarity coercion. Of course, this cannot be handled with WordNet or any other static resource, but would have to be handled in the discourse relations level.

5. Conclusions and Future Work

In this paper, we tried to show that the idea of polarity plays an important role in licensing parallel and contrastive relations. We also showed that polarity is not as straightforward as some examples might suggest: some adjectives seem to allow polarity coercion, switching between negative and positive senses depending on the local context. Neither WordNet nor GL deal with polarity explicitly, and while we might be able to derive approximate polarity values for some WordNet entries based on their glosses, it isn't clear how to deal with adjectives with flexible polarity such as *simple*. GL doesn't offer any way to determine or code polarity yet, but its ability to deal with sense changes due to coercion might make it a more suitable framework to deal with coerced adjective senses.

In future work we intend to look more closely at natural examples to see what kind of adjectives are similar to *simple*. Further we want to see if it is feasible to could automatically guess polarity on the basis of WordNet glosses, using the intuitions of naïve subjects as a gold standard. We hope that this work will allow us to eventually reach our goal, which is automatic recognition of contrasted or parallel elements in natural texts.

References

- Andrew, J.M.: 1993, 'The Conceptual Basis of Antonymy and Synonymy in Adjectives', *Journal of Memory and Language*, **32**, 301-319.
- Bouillon, P.: 1996, 'Mental state adjectives: the perspective of generative lexicon', *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*, Copenhagen, Denmark, 143-148.
- Bouillon, P.: 1999, 'The adjective "vieux": the point of view of "Generative Lexicon"' in E. Viegas (ed.), *Breadth and depth of semantic lexicons*, Dordrecht : Kluwer, 148-166.
- Carlson, L., D. Marcu, & M. Okurowsky (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In Proceedings of the 2nd SigDial Workshop on Discourse and Dialogue.
- Cruse, D.A.: 1986, *Lexical Semantics*, Cambridge University Press.
- Cresswell, M. J.: 1976, 'The Semantics of Degree', in B. Partee (ed.), *Montague Grammar*, Academic Press, New York.
- Hatzivassiloglou, V., & K. R. McKeown: 1997, 'Predicting the semantic orientation of adjectives', *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*. Madrid, ES: Association for Computational Linguistics, 174-181.
- Hoeksema, J.: 1983, 'Negative Polarity and the Comparative', *Natural Language and Linguistic Theory*, **1**, 403-434.
- Kamps, J., M., Maarten, Mokken, R. J., & M. de Rijke: 2004, 'Using WordNet to measure semantic orientation of adjectives', *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, vol. IV, 1115-1118.
- Kennedy, C.: 2001, 'Polar Opposition and the Ontology of 'Degrees'', *Linguistics and Philosophy*, **24**, 33-70.
- Klein, E.: 1980, 'A Semantics for Positive and Comparative Adjectives', *Linguistics and Philosophy*, **4**(1), 1-45.
- Koppel, M., & J. Schler: 2006, The importance of neutral examples for learning sentiment. *Computational Intelligence*, **22**(2), 100-116.
- Lakoff, R.: 1971, 'IFs, And's and But's about Conjunction', *Studies in Linguistic Semantics*, New York: Holt, Reinhart and Winston, 114-149.
- Lehrer, A., & K. Lehrer: 1982, 'Antonymy', *Linguistics and Philosophy* **5**, 483-501.
- Marcu, D. & Echihiabi, A. (2001). An Unsupervised Approach to Recognizing Discourse Relations. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania.
- Sapir, E.: 1944, 'Grading: a Study in Semantics', *Philosophy of Science* **11**, 93-116.
- Seuren, P. A. M.: 1973, 'The Comparative', in F. Kiefer and N. Ruwet (eds.), *Generative Grammar in Europe*, Riedel, Dordrecht.
- Pustejovsky, J.: 1995, *The Generative Lexicon*. MIT Press, Cambridge MA.
- Pustejovsky, J.: 1995a, 'Linguistic Constraints on Type Coercion', in P. Saint-Dizier & E. Viegas (eds.), *Computational Lexical Semantics*, Cambridge University Press.
- Pustejovsky, J. & B. Boguraev: 1993, 'Lexical Knowledge Representation and Natural Language Processing', *Artificial Intelligence*, **63**, 170-210.

Anna Lobanova and Jennifer Spenader
University of Groningen, Artificial Intelligence Department
Grote Kruisstraat 2/1,
9712 TS Groningen
The Netherlands

a.lobanova@ai.rug.nl
j.spenader@ai.rug.nl

Title: Incorporating Polarity in Lexical Resources

Normal presentation preferred