Learning to Understand Figurative Language: From Similes to Metaphors to Irony

Tony Veale (TONY.VEALE@UCD.ie)
School of Computer Science and Informatics, University College Dublin
Belfield, Dublin 4, Ireland

Yanfen Hao (YANFEN.HAO@UCD.ie)
School of Computer Science and Informatics, University College Dublin
Belfield, Dublin 4, Ireland

Abstract

Simile is widely viewed as a less sophisticated conceptual device than metaphor, not least because similes are explicitly marked and are frequently more obvious about the meanings they carry. Nonetheless, this lack of sophistication makes simile an ideal basis for acquiring the category-specific knowledge required to understand metaphor. In this paper we describe a computational approach to simile and metaphor that takes the career-of-metaphor hypothesis of Bowdle and Gentner (2005) as its starting point. We describe how the category-defining knowledge required by metaphor can be acquired from exposure to explicit similes, and demonstrate that this knowledge offers a richer and more diagnostic picture of category structure than that acquired from alternate sources.

Keywords: metaphor; simile; irony; salient property; category representation.

Introduction

Figurative language can range from the sublime and the enigmatic to the banal and the obvious. Metaphor, for instance, is widely considered to be the epitome of creative expression, for metaphors often transcend the merely descriptive to yield profoundly enlightening insights; metaphors can be richly allusive, playful and challenging, and open to constant re-interpretation by new readers in new contexts (e.g., see Gibbs, 1994). Because metaphor allows us to view one concept through the prism of another, it is an inherently asymmetric device in which the meaning of a juxtaposition depends crucially on the direction of the information flow (Ortony, 1979). Similes, in contrast, seem an altogether more humble form of expression. The use of the hedge words "like" or "as" marks simile as a diffident figure of speech, which (unlike metaphor) stops short of ascribing category membership to merely draw attention to certain shared properties. So while the metaphor "drug dealers are vampires" challenges our conception of vampires and the criteria needed for membership in this category (see Glucksberg, 2001), the corresponding simile, "drug dealers are *like* vampires" merely enjoins us to look for common properties which in themselves may be insufficient to support category inclusion. This reluctance to categorize marks simile as a symmetric form of comparison.

The hedging and diffident nature of similes might also be said to signal a lack of confidence in the aptness of the equivalent categorization. Indeed, Roncero *et al*. (2006) note that similes found on the internet are far more likely than the equivalent metaphors to be accompanied by an explicit explanation, which suggests that similes are less constrained by norms of category structure, and thus less likely than metaphors to be implicitly explained by these norms. Hanks (2004) goes as far as to argue that this non-reliance on category norms makes simile a freer and more creative form of expression than metaphor, since similes can serve as dynamic "triggers for the imagination" without having to appeal either to linguistic conventions or experiential gestalts. Chiappe *et al*. (2003) demonstrate that metaphoric expressions of a relationship are preferred when the relationship is an apt one, which suggests that it should follow more obviously from the corresponding categorization. These authors also find that aptness correlates strongly with ease of comprehension, and indeed, similes can enhance both their aptness and their comprehensibility by opting for explicit self-explanation: when one says "my left tire is as bald as a bowling ball", there is no ambiguity whatsoever as to the property that is shared by topic and vehicle, even if baldness is not a literally sensible property of artifacts. Bowdle and Gentner (2005) argue, in a hypothesis they call the *career of metaphor*, that as metaphors become more conventionalized, they are more likely to be processed as categorizations than as comparisons. This suggests that increased familiarity with a particular metaphoric vehicle allows for greater competence in how the vehicle category is applied and extended to include new members (following Glucksberg, 2001).

If, as the *career of metaphor* hypothesis suggests, there is an "evolutionary path …from comparison to categorization", it is consistent to also argue for an evolutionary path between simile and metaphor. Certainly, explicit similes of the form "X is as P as Y" indicate that P is a highly salient property of Y, salient enough that Y can be used to exemplify P-*ness*. If exposed to enough similes of this form, or similes with accompanying explanations (like those reported by Roncero *et al*., 2006), a cognitive agent can build a detailed conceptual picture of the features P_i that define a category Y. Since these will be the most salient and diagnostic features of Y, they can be used to build a category representation of Y that can subsequently be used to understand figurative uses of Y as categorizations rather than as comparisons. Put another way, the less sophisticated and often more explicit nature of simile means that simile is an excellent knowledge-acquisition device through which an agent can learn enough about category structure to become competent in the metaphoric uses of those structures.

Rather than viewing simile as a lesser cousin of metaphor, this view would make simile a crucial progenitor to metaphoric awareness. Though the development of figurative competence in humans is undoubtedly more complex and non-linear than this simple view presupposes, this hypothesis provides an ideal basis for training a computational agent to understand and appreciate metaphor, and in the process enrich its internal category organization. In this paper we describe a super-charged implementation of this approach, in which a computational agent is automatically exposed to a very large quantity of self-explanatory similes from the web. We describe how these similes are collected and then sense-disambiguated with respect to the lexical ontology WordNet (see Fellbaum, 1998). We then describe how these descriptions can be translated into robust computational membership functions that can be used to understand metaphors in terms of category inclusion. To demonstrate the descriptive adequacy of the conceptual picture painted by these similes, we evaluate how well each simile-derived category description predicts the overall affective rating of a category. We conclude with a consideration of irony, and offer some empirical observations from our large-scale analysis of simile.

Acquiring a Large Case-Base of Similes

As in the study reported in Roncero *et al*. (2006), we employ the *Google* search engine as a retrieval mechanism for accessing relevant web content. However, the scale of the current exploration requires that retrieval of similes be fully automated, and this automation is facilitated both by the *Google* API and its support for the wildcard term *. In essence, we consider here only partial explicit similes conforming to the pattern "*as ADJ as a|an NOUN*", in an attempt to collect all of the salient values of ADJ for a given value of NOUN. We do not expect to identify and retrieve all similes mentioned on the world-wide-web, but to gather a large, representative sample of the most commonly used.

To do this, we first extract a list of antonymous adjectives, such as "hot" or "cold", from the lexical database WordNet (Fellbaum, 1998); the intuition here is that explicit similes will tend to exploit properties that occupy an exemplary point on a scale. For every adjective ADJ on this list, we send the query "*as ADJ as* *" to Google and scan the first 200 snippets returned for different noun values for the wildcard *. From each set of snippets we can ascertain the relative frequencies of different noun values for ADJ. The complete set of nouns extracted in this way is then used to drive a second phase of the search. In this phase, the query "*as * as a NOUN*" is used to collect similes that may have lain beyond the 200-snippet horizon of the original search, or that hinge on adjectives not included in the original list. Together, both phases collect a wide-ranging series of core samples (of 200 hits each) from across the web, yielding a set of 74,704 simile instances (of 42,618 unique types) relating 3769 different adjectives to 9286 different nouns.

Simile Annotation

Many of these similes are not sufficiently well-formed for our purposes. In some cases, the noun value forms part of a larger noun phrase: it may be the modifier of a compound noun (as in "bread lover"), or the head of complex noun phrase (such as "gang of thieves"). In the former case, the compound is used if it corresponds to a compound term in WordNet and thus constitutes a single lexical unit; if not, or if the latter case, the simile is rejected. Other similes are simply too contextual or under-specified to function well in a null context, so if one must read the original document to make sense of the simile, it is rejected. More surprisingly, perhaps, a substantial number of the retrieved similes are ironic, in which the literal meaning of the simile is contrary to the meaning dictated by common sense. For instance, "as hairy as a bowling ball" (found once) is an ironic way of saying "as hair<u>less</u> as a bowling ball" (also found just once). Many ironies can only be recognized using world (as opposed to word) knowledge, such as "as sober as a Kennedy" and "as tanned as an Irishman". In addition, some similes hinge on a new, humorous sense of the adjective, as in "as fruitless as a butcher-shop" (since the latter contains no fruits) and "as pointless as a beach-ball" (since the latter has no points).

Given the creativity involved in these constructions, one cannot imagine a reliable automatic filter to safely identify bona-fide similes. For this reason, the filtering task was performed by a human judge, who annotated 30,991 of these simile instances (for 12,259 unique adjective/noun pairings) as bona-fide (i.e., non-ironic and meaningful in a null context; these similes relate a set of 2635 adjectives to a set of 4061 different nouns. In addition, the judge also annotated 4685 simile instances (of 2798 types) as ironic; these similes relate a set of 936 adjectives to a set of 1417 nouns. Perhaps surprisingly, ironic pairings account for over 13% of all annotated simile instances and over 20% of all annotated simile types.

Word-Sense Disambiguation

It is important to know which sense of a noun is described by a simile if an accurate conceptual picture is to be constructed. For instance, "as stiff as a zombie" might refer either to a reanimated corpse or to an alcoholic cocktail (both are senses of "zombie" in WordNet, and drinks can be "stiff" too). Sense disambiguation is especially important if we hope to derive meaningful correlations from property co-occurrences; for instance, zombies are described in web similes as exemplars of not just stiffness, but of coldness, slowness and emotionlessness. If such co-occurrences are observed often enough, a cognitive agent might usefully infer a causal relationship among pairs of properties (e.g., that coldness implies emotionless).

Disambiguation is trivial for nouns with just a single sense in WordNet. For nouns with two or more fine-grained senses that are all taxonomically close, such as "gladiator" (two senses: a boxer and a combatant), we consider each sense to be a suitable target. In some cases, the WordNet gloss for as particular sense will actually mention the adjective of the simile, and so this sense is chosen. In all other cases, we employ a strategy of mutual disambiguation to relate the noun vehicle in each simile to a specific sense in WordNet. Two similes "as A as N_1" and "as A as N_2" are mutually disambiguating if $N_1$and N_2 are synonyms in WordNet, or if some sense of $N_1$is a hypernym or hyponym of some sense of N_2 in WordNet. For instance, the adjective "scary" is used to describe both the noun "rattler" and the noun "rattlesnake" in bona-fide (non-ironic) similes; since these nouns share a sense, we can assume that the intended sense of "rattler" is that of a dangerous snake rather than a child's toy. Similarly, the adjective "brittle" is used to describe both saltines and crackers, suggesting that it is the bread sense of "cracker" rather than the hacker, firework or hillbilly senses (all in WordNet) that is intended.

These heuristics allow us to automatically disambiguate 10,378 bona-fide simile types (85% of those annotated), yielding a mapping of 2124 adjectives to 3778 different WordNet senses. Likewise, 77% (or 2164) of the simile types annotated as ironic are disambiguated automatically. A remarkable stability is observed in the alignment of noun vehicles to WordNet senses: 100% of the ironic vehicles always denote the same sense, no matter the adjective involved, while 96% of bona-fide vehicles always denote the same sense. This stability suggests two conclusions: the disambiguation process is consistent and accurate; but more intriguingly, only one coarse-grained sense of any word is likely to be sufficiently exemplary of some property to be useful as a simile vehicle.

Robust Category Representation

Each bona-fide simile contributes a different salient property to the representation of a vehicle category. In our data, one half (49%) of all bona-fide vehicle nouns occur in two or more similes, while one third occur in three or more and one fifth occur in four or more. The most frequently used figurative vehicles can have many more; "snowflake", for instance, is ascribed over 30 in our database, including: *white, pure, fresh, beautiful, natural, delicate, intricate, delicate, identifiable, fragile, light, dainty, frail, weak, sweet, precious, quiet, cold, soft, clean, detailed, fleeting, unique, singular, distinctive* and *lacy*. This is a finding compatible with the career-of-metaphor hypothesis, for as "snowflake" becomes conventionalized as a highly evocative metaphoric vehicle, its category structure should become richer and more nuanced to support figurative categorizations. But as noted in Glucksberg (2001) and Bowdle and Gentner (2005), the vehicle "snowflake" can mean different things in different metaphors: in some it stands as a symbol of purity, in others as a symbol of uniqueness, and in others still a symbol of delicacy. Either a variety of different structures should be automatically constructed from this data, or a single flexible category structure that can foreground different properties in different metaphors**.**

We opt for the second course, by describing each category structure with a single mathematical membership function that converts the available property-based evidence for category inclusion into a score in the range 0 … 1. Consider the function of Figure 1:

```
(define Snowflake.0 (arg0)
        (*      (%sim  arg0  Snowflake.0)
                (combine     (attr pure arg0)
                             (attr unique arg0)
                             (attr delicate arg0)
                             ...
                )
        )
)
```

Figure 1: A partial view of the membership function the category Snowflake.0

Note that the function is named Snowflake.0 to represent a particular WordNet sense of the word "snowflake", while the single argument arg_0 is bound to any candidate entity we wish to consider for membership. The function *%sim* returns a WordNet-based measure (in the range 0…1) of taxonomic similarity between two terms, e.g., as determined by link distance to a common hypernym. The function *attr* measures the salience (also in the 0…1 range) of a property to an entity; this association can be based on their relative frequency of co-occurrence in the annotated simile database, or on their relative frequency of co-occurrence in a large text corpus, or on a mixture of both these factors. For instance, *attr* can be implemented using either the Jacquard coefficient or the Dice coefficient (see Cimiano *et al*., 2005). Finally, the function *combine* implements a simple probabilistic *or*, in which different pieces of evidence are naïvely assumed to be statistically independent.

$$(\text{combine } e_0\ e_1) = e_0 + e_1\,(1 - e_0)$$
$$(\text{combine } e_0\ e_1 \ldots e_n) = (\text{combine } e_0\ (\text{combine } e_1 \ldots e_n))$$

The more evidence that can be combined for a particular member arg_0, the higher its assigned membership score. In effect, each function represents a radially structured category (see Lakoff, 1987) in which the most prototypical members are assigned a score closer to 1.0 and the least typical members are assigned scores closer to 0.

For a category as property-rich as Snowflake, only a few properties need be adduced to obtain a reasonable membership score. As such, the category can mean different things in different figurative contexts, depending on the features that are known about arg_0. Indeed, if we use a text corpus to represent a particular context or domain of discourse, the syntagmatic use of a term in this corpus will determine its membership scores. For instance, the function of Figure 2 is automatically constructed for Gladiator:

```
(define Gladiator.0 (arg0)
      (*     (%sim arg0 Gladiator.0)
             (combine   (attr manly arg0)
                        (attr violent arg0)
                        (attr competitive arg0)
             )
      )
)
```

Figure 2: An automatically generated membership function for Gladiator.0

Now consider the category Athlete.0, to which the property *competitive* is also ascribed in the simile database. This in itself is sufficient evidence for an instance of Athlete.0 to also be considered a member of the category Gladiator.0, and since Athlete.0 and Gladiator.0 are taxonomically similar in WordNet, this single property yields a middling membership score. However, if in a corpus the term "athlete" is repeatedly modified by the adjectives "violent" or "manly" or both, this categorization of athletes as gladiators will become all the more appropriate, to produce an altogether higher membership score.

Empirical Evaluation

A membership function like that of Figures 1 and 2 is automatically generated for each of the 3778 disambiguated WordNet senses in our simile database. But how accurate are these simile-derived functions? Furthermore, how we can be sure that similes are the most insightful source of the world knowledge needed to compose these functions?

If similes are indeed a good place to mine the most salient properties of categories, we should expect the set of properties for each category to accurately predict how the category is perceived as a whole. For instance, humans – unlike computers – do not generally adopt a dispassionate view of ideas, but rather tend to associate certain positive or negative feelings, or affective values, with particular ideas. Unsavoury activities, people and substances generally possess a negative affect, while pleasant activities and people possess a positive affect. Whissell (1989) reduces the notion of affect to a single numeric dimension, to produce a *dictionary of affect* that associates a numeric value in the range 1.0 (most unpleasant) to 3.0 (most pleasant) with over 8000 words in a range of syntactic categories (including adjectives, verbs and nouns). So to the extent that the adjectival properties yielded by processing similes paint an accurate picture of each noun vehicle, we should be able to predict the affective rating of each vehicle via a weighted average of the affective ratings of the adjectival properties ascribed to these vehicles (i.e., where the affect of each adjective contributes to the estimated affect of a noun in proportion to its frequency of co-occurrence with that noun in our simile data). More specifically, we should expect that ratings estimated via these simile-derived properties should correlate well with the independent ratings contained in Whissell's dictionary.

To determine whether similes do offer the clearest perspective on a category's most salient properties, we calculate and compare this correlation using the following data sets:

A. Adjectives derived from annotated bona-fide (non-ironic) similes only.

B. Adjectives derived from all annotated similes (both ironic and non-ironic).

C. Adjectives derived from ironic similes only.

D. All adjectives used to modify the given vehicle noun in a large corpus. We use over 2-gigabytes of text from the online encyclopaedia Wikipedia as our corpus.

E. All adjectives used to describe the given vehicle noun in any of the WordNet text glosses for that noun. For instance, WordNet defines Espresso as "strong black coffee made …" so this gloss yields the properties *strong* and *black* for Espresso.

Predictions of affective rating were made from each of these data sources and then correlated with the ratings reported in Whissell's dictionary of affect using a two-tailed Pearson test ($p < 0.01$). As expected, property sets derived from bona-fide similes only (A) yielded the best correlation (+0.514) while properties derived from ironic similes only (C) yielded the worst (-0.243); a middling correlation coefficient of 0.347 was found for all similes together, demonstrating the fact that bona-fide similes outnumber ironic similes by a ratio of 4 to 1. A weaker correlation of 0.15 was found using the corpus-derived adjectival modifiers for each noun (D); while this data provides far richer property sets for each noun vehicle (e.g., far richer than those offered by the simile database), these properties merely reflect potential rather than intrinsic properties of each noun and so do not reveal what is most salient about a vehicle category. More surprisingly, perhaps, property sets derived from WordNet glosses (E) are also poorly predictive, yielding a correlation with Whissell's affect ratings of just 0.278.

While it is true that these WordNet-derived properties are not sense-specific, so that properties from all senses of a noun are conflated into a single property set for that noun, this should not have dramatic effects on predictions of affective rating. If one sense of a word acquires a negative connotation, it is generally believed that "bad meanings drive out the good" so that the word as a whole becomes tainted. Rather, it may be that the adjectival properties used to form noun definitions in WordNet are simply not the most salient properties of those nouns. To test this hypothesis, we conducted a second experiment wherein we automatically generated similes for each of the 63,935 unique adjective-noun associations extracted from WordNet glosses, e.g., "as strong as espresso", "as Swiss as Emmenthal" and "as lively as a Tarantella", and counted how many of these manufactured similes can be found on the web, again using *Google*'s API

We find that only 3.6% of these artificial similes have attested uses on the web. From this meagre result we can conclude that: a) few nouns are considered sufficiently exemplary of some property to serve as a meaningful vehicle in a figure of speech; b) the properties used to define categories in general purpose resources like WordNet are not always the properties that best reflect how humans think of, and use, these categories.

In fact, the properties ascribed to noun concepts in WordNet glosses are, overall, no more diagnostic of these concepts than adjectival properties used to modify the corresponding nouns in free text. To see this, consider again data set D, the set of all adjective:noun collocations in the text of Wikipedia. For each unique collocation type (such as "timeless myth"), we generated the corresponding simile (e.g., "as timeless as a myth"), and used to Google to search for all 568,400 of these similes on the web. Interestingly, we find that 5% (or 28,400) have attested uses in web-accessible documents. Of course, since Wikipedia is a reasonably authoritative encyclopedia, we should expect that the properties one can glean from it will paint a somewhat accurate picture of each noun concept. The result of the affect prediction task (a correlation coefficient of just 0.15 for data set D) means that these more salient properties are more heavily disguised by a large body of unsalient properties than is the case in WordNet glosses. Nonetheless, the simile-generation task (5% versus 3%) suggest that Wikipedia is as good a source of property knowledge for figurative processing as WordNet.

Concluding Remarks

Of course, the truth is most likely to lie somewhere between these two alternatives. The space of potential similes is doubtless much larger than that currently found on the web, and many of the similes generated from WordNet are probably quite meaningful and apt. However, even the WordNet-based similes that can be found on the web are of a different character to those that populate our database of annotated web-similes, and only 9% of the web-attested WordNet similes (or 0.32% overall) also reside in this database. Thus, most (> 90%) of the web-attested WordNet similes must lie outside the 200-hit horizon of the acquisition process described in section 2, and so are less frequent (or used in less authoritative web pages) than our acquired similes. What then makes for a good simile?

In "A Christmas Carol", Dickens asks a similar question before concluding that "… the wisdom of our ancestors is in the simile". Similes do not always convey truths that are universally true, or indeed, even literally true (e.g., bowling balls are not literally bald). Rather, similes hinge on properties that are possessed by prototypical or stereotypical members of a category, even if most members of the category do not also possess them. As a source of knowledge, similes combine received wisdom, prejudice and over-simplifying idealism in equal measure. As such, similes reveal knowledge that is pragmatically useful but of a kind that one is unlikely to ever acquire from a dictionary (or from WordNet). Furthermore, while similes are, in principle, reversible (at least from a conceptual perspective), it is rarely pragmatically sensible to do so. If a simile is to be a useful descriptive device, the vehicle category should be better understood than the topic. So although a simpler rhetorical device than metaphor, we have much to learn about language and its underlying conceptual space by a comprehensive study of real similes in the wild, not least about the recurring vehicle categories that signpost this space.

The knowledge acquired from basic similes allows a cognitive agent to gradually develop a more sophisticated understanding of irony. For instance, if the agent knows that P is a salient property of V, then the simile "as not-P as V" must be ironic. Likewise, if the agent learns from overlapping simile descriptions that P_1 often implies P_2 (e.g., that *dead* implies *cold*, or that *cold* implies *stiff*), then the simile "as not-P_2 as V" is also likely to be seen as ironic (though more subtly so) if the agent knows that P_1 is a property of V. We expect the sense-disambiguated and annotated database of similes described here to prove especially helpful in developing a model of ironic implicature. For readers who wish to see this simile data for themselves, it can be viewed at *http://afflatus.ucd.ie/sardonicus/tree.jsp*.

References

Bowdle, B. F. and Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, 112, 193-216.

Chiappe, D., Kennedy, J. M. and Smykowski, T. (2003). Reversibility, aptness, and the conventionality of metaphors and similes. *Metaphor & Symbol*, 18, 85-105.

Cimiano, P., Hotho, A., and Staab, S. (2005). Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. Journal of AI Research, 24: 305-339

Fellbaum, C. (ed.). (1998). *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.

Gibbs, R. W. (1994). *The poetics of mind*. Cambridge, UK: Cambridge University Press.

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press.

Hanks, P. (2004). The syntagmatics of metaphor. *Int. Journal of Lexicography*, 17(3).

Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago University Press.

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

Roncero, C., Kennedy, J. M., and Smyth, R. (2006). Similes on the internet have explanations. Psychonomic Bulletin and Review, 13(1), 74-77.

Whissell, C. (1989). The dictionary of affect in language. In R. Plutchnik & H. Kellerman (Eds.) *Emotion: Theory and research*. New York, Harcourt Brace, 113-131.