A Computational Exploration of Creative Similes

Tony Veale

1. Introduction

Irony is a curious form of double-speak in which a speaker implies the opposite of what is said (Giora, 1995), or expresses a sentiment in direct opposition to what is actually believed (Grice, 1978). Intriguingly, an ironic speaker does this in the hope that the audience will actually see past this artifice to comprehend the speaker's actual meaning. On the surface, this seems a most irrational, round-about and risky way to communicate meanings (Sperber and Wilson, 1992). But on closer analysis, irony reveals itself to be anything but round-about: it is, in fact, a very compact way of saying or doing two useful things at once. Irony can be used to divide an audience into those who "get it" and those who don't; it can be used to soften a criticism with humour, or more often, to salt a wound by cloaking it in an apparent compliment that is quickly dashed; and most concisely of all, it can echo a viewpoint that is advanced by another while simultaneously undermining that viewpoint (see again Sperber and Wilson, 1992). This compression of function and viewpoint is most obviously apparent in ironic similes, as in the comparison "*you are about as tough as a marshmallow cardigan*" (an example from the web-corpus we describe in section 3): this simile integrates the expectation that the audience ("you") is believed to be "tough" with a comparison that utterly undermines this expectation.

There is something appealingly democratic and unpretentious about similes. Not only are they pervasive in language, they are at home in any register of speech and any genre of text, from tabloid newspapers to romantic poetry (Fishlov, 1992). Conveniently, most languages provide a wealth of pre-fabricated similes that are as well-known to native speakers as the adjectival features they serve to exemplify (e.g., "*as strong as an ox*", "*as sober as a judge*", etc.; see Taylor, 1954; Norrick, 1986; Moon, 2008). Such formulaic similes allow us to quickly identify the key stereotypes of a language and culture, and to recognize those which are shared by different language cultures, such as English and Chinese (Veale *et. al,* 2008). But just as importantly, languages like English make it easy for speakers to mint their own similes on the fly, by imposing low barriers to creation.

In contrast to metaphors, similes are always marked, allowing their audience to immediately and unambiguously construe them as comparisons (Taylor, 1954; Hanks, 2004). The syntactic form of explicit *as*-similes – "*A Topic is as Ground as a Vehicle*" – provides a ready-made infrastructure that authors can populate with their own bespoke vehicles (Fishlov, 1992; Moon, 2008), while the ability to explicitly state the grounds of a comparison allows an author to use vehicles that are neither obvious or entirely to the point. Unlike metaphors, which often employ coherent systems of mappings to support the pretence that the topic really is a member of the vehicle category (see Glucksberg, 2001), similes can be as wildly colorful and incongruous as an author wants, as long as the ground is effectively communicated. Hanks (2004) thus argues that similes provide a freer and more creative means of expression than metaphor, since similes can serve as dynamic "triggers for the imagination" without having to appeal either to underlying schemata or to experiential gestalts (Lakoff and Johnson, 1980). Creative similes are often used as glosses to illustrate or tie together elements of an argument, much as canned jokes are often invoked in conversation to illustrate a key point (Oring, 2003). As such, creative similes do not have to be as rigorously constructed as metaphors, and Roncero *et al.* (2006) have shown that similes found on the web are much more likely to be accompanied by explanations than comparable metaphors that convey the same message.

Consider this example from Jerry Seinfeld, in his New York Times remembrance of the late comic George Carlin (see Seinfeld, 2008):

> "[George] was like a train hobo with a chicken bone. When he was done there was nothing left for anybody."

The image of a "train hobo with a chicken bone" is visually striking, but needs an explicit explanation to help draw out its meaning potential. Note that this additional explanation only conveys a small part of the simile's meaning. The image itself is rich in descriptive resonance, and one sees Carlin as just as leery, disheveled, acid-tongued and mocking of social conventions as the hobo on the train. But without the support structure of the adjunct "nothing left for anybody", we would fail to see the relevance of the chicken-bone: Carlin's comedy is driven by a visceral hunger that leads him to thoroughly exhaust the humor potential of his targets. In other words, the range of possible similarities is just too large, so the simile needs a support structure to direct us toward the desired meaning.

Because the explanation provides just a small part of the overall meaning, the potency of Seinfeld's simile does not seem diminished by its addition. This suggests that the explanations observed by Roncero *et al.* (2006) are used as a form of support structure, or *scaffolding* (see Veale and Keane, 1992), for creative similes, allowing speakers to choose comparisons primarily on their basis of their visual and affective resonance without fear of miscommunication. Nonetheless, explanations can greatly diminish the potency of jokes (Oring, 2003), so when a simile is humorously used as a joke, we hypothesize that support structures, if any, will be far more subtle. Ironic comparisons, for instance, would be utterly undermined if accompanied by an explicit explanation, since as Grice (1978) notes, "*to announce [irony] as pretence would be to spoil the effect*" (p. 125). Nonetheless, irony always runs a risk of being misdiagnosed (Sperber and Wilson, 1992), and so requires that great care is exercised in its construction. Grice (1978) further notes that when "*speaking ironically ... a tone suitable to such a feeling or attitude seems to be mandatory*" (p. 125). So when ironic comparisons are creatively minted on the fly, in conversationally-styled texts, we hypothesize that some form of lexicalized support structure will often be used in place of an ironical tone, to subtly direct the audience toward the desired meaning. For instance, the *indeed*-construction "*a [characterization], indeed*!" commonly acts as a scaffolding for ironic observations about situations or events that fail to behave as advertised, such as the remark "an officer and a gentleman, indeed" discussed in Sperber and Wilson (1992). Likewise, the construction "a fine [X]!" is commonly used to express a negative evaluation whether X has a positive or negative sentiment, as in "a fine romance!", "a fine holiday!" and even "a fine mess". These examples can also be followed by a use of the "indeed" marker, or "I must say", to compound the effect.

Speakers rarely have as much time as writers to rework and polish their outputs, so it is intuitive to believe that they often use some kind of support structure to ensure that their most creative (and risky) efforts achieve successful communication. Moon (2008) has noted that the marker "about" has a special role in signaling irony, and goes as far as to argue that the *about*-form of similes, "*about as ADJ as NOUN*", always conveys an ironic meaning. But Moon's analysis is based on relatively formulaic similes, of the kind one expects to find in common usage, and so this claim is based on a very small sample set. If one looks at the much larger space of creative similes that speakers mint on the fly, as we do in this current work, then it becomes clear that "about" does not always signal irony, but more general-

ly signals an attempt by a speaker to be imprecise, humorous and creative. By signaling creative intent, speakers ask for – and generally receive – additional support for their desired interpretations. In this current work we seek robust empirical support for this claim, and support more generally for our intuition that structures like the "*about as*" form act as scaffolding or support structures for creative utterances that are at risk of being misunderstood.

This investigation is computational in nature, and views the problems of identifying and analyzing the meaning of creative comparisons from the perspective of a naïve computer. As argued in Veale (2006), such a computational perspective – which Jackendoff (1987) has dubbed the *hypothesis of computational sufficiency* – forces us to address vexing issues of under-specification and over-generation in our theoretical accounts of cognitive/linguistic phenomena. In this paper we ask: what kinds of knowledge must a computational agent possess so as to recognize the ironic intent of a humorous comparison, and to what degree is this intent telegraphed to the audience by the use of specific markers or support structures? To this end, we use automated means to collect two very large corpora of similes, one that favors the conventional and one that favors the creative use of similes. In section 2, we describe how simple similes with one-word vehicles, such as "*as cunning as a fox*", can be harvested from the internet, while in section 3 we explicitly seek out more complex similes prefixed with the putative support marker, "about". In section 4 we present a statistical analysis of these corpora, to tease apart their similarities and differences and thus reveal the extent to which creative comparisons exploit the stock imagery of more formulaic similes. In section 5 we turn to a consideration of irony and the affective signature of similes marked by "about". Finally, we conclude in section 6 with a discussion of our results.

2. Corpus I: Simple Comparisons

2.1. Compiling Lists of Simple Similes

To compile a collection of conventional similes, one can look to authoritative sources such as printed dictionaries, or exploit the syntactic frame of the *as*-simile to identify matching instances in large text corpora. Norrick (1986), for instance, uses the former approach, and bases his analysis on 366 similes listed in the 1970 edition of *The Oxford Dictionary of Prov-*

erbs. Moon (2008) uses a hybrid approach, and compiles a collection of 377 similes from multiple sources, one of which is the *Bank of English* corpus. But the pervasiveness and ease of creation of similes means that one is likely to find a much wider range of different similes in the collected texts of the world-wide-web (Roncero *et al*., 2006). The syntactic marking of similes means that most of these similes can be harvested automatically, using a simple process of pattern-matching. Thus, when we pose the query "*as * as **" to the Google search-engine, the wildcard elements are bound by Google to the corresponding elements of a comparison.

Google returns a large number of snippets from online documents that contain matching phrases, such as "*as hot as an oven*" or "*as strong as an ox*". In these snippets, we are likely to see the same combination of ground and vehicle recurs in many different contexts. This combination of ground and vehicle is the semantic core of a simile, the part that transcends context to be reused in the description of many different topics. The relationship of this core combination to the topic, will in many cases, be entirely contingent and subjective; most similes are used, after all, to communicate information about a topic that is not fully understood or appreciated by an audience, and so for purposes of corpus construction, the topic has very little bearing on the semantics of the simile. For instance, the simile "*my boss is as cunning as a fox*" tells us nothing at all about bosses *per se*, but does tell us that foxes are either stereotypically cunning (if the simile is non-ironically *straight*) or stereotypically naïve (if the simile is ironic). We are primarily interested therefore in the collection of simile *types* – the context-transcending reusable combination of a specific ground with a specific vehicle – rather than of simile *instances* – the contextually-tied application of a ground and vehicle to a specific topic.

To ensure that we acquire the widest range of simile types with the widest range of adjectival grounds, we need to seed our queries with specific adjectives. For example, to ensure that we find similes for *strength*, we need to use the queries "as *strong* as *" and "as *weak* as *". To automate the harvesting process, we use the lexical resource WordNet (Fellbaum, 1998) as a source of adjectives for these queries. In particular, we use WordNet as an inventory of antonymous adjective pairs, such as "strong" and "weak", since these often define the gradable properties for which similes are used to indicate extreme values. In all, we generate over 2000 queries of the "as * as *" form, in which the ground position (the first wildcard *) is successively bound to a different adjective. For tractability, we cannot consider every document returned by Google for these queries. Ra-

ther, we consider just the first 200 snippets returned for each, allowing us to harvest a corpus of simile types by taking a wide-ranging series of different core-samples from across the full breadth of the web. While the core-sample for each adjective is just 200 snippets deep, this is sufficient for a frequency analysis to reveal the most culturally entrenched English similes. For instance, in the query "as *strong* as *", the * matches "horse" 27 times, "bull" 19 times, "gorilla" 12 times, and "Viking" just once.

2.2. Annotating the Data

When we consider only those simile instances with a single-term vehicle, as listed in a conventional lexical resource like WordNet, the above processes harvest 74,704 instances of the "as * as *" pattern, 42,618 of which are unique. In all, these instances relate 3769 different adjectival grounds to 9286 different noun vehicles. However, while each of these instances is a legitimate instance of a comparison, not all qualify as similes. As defined by Ortony (1979), the difference between comparisons and similes is best characterized in terms of salience: a simile uses a vehicle for which a given ground property is especially salient to highlight this property in a topic. Simple comparisons, on the other hand, merely point out correlations and commonalities between two things, regardless of whether those commonalities are particularly salient in the vehicle. If a doctor states that a tumour is "as big as a tennis-ball", this is certainly cause for alarm, but it is not a simile, since bigness is not a salient property of tennis-balls.

Since there is no automatic way of separating similes from simple comparisons, human judges are used to annotate all those instances where the ground is obviously a salient property of the vehicle (the bona-fide or *straight* cases) or where a property that is diametrically opposed to the ground is salient of the vehicle (the ironic cases). The extensive grey area between these positions – where the ground is neither strongly associated with, nor strongly opposed to, the vehicle – is not always clear cut, and instances like "*as cuddly as a bear*" might fall into either category in one context or another. The human judges thus perform a conservative separation, discarding those instances that might lean both ways. Those that are not discarded are annotated as either straight or ironic. In all, 30,991 instances are identified as straight (non-ironic) similes; of these instances, 12,259 are unique simile types, that is, unique pairings of a ground property to a vehicle. A smaller body of 4685 instances are annotated as ironic simi-

les, such as "*as hairy as a bowling-ball*", and of these, 2798 form unique types.

2.3. Simple Elaborations

Taylor (1954) notes that speakers sometimes elaborate existing similes to create new and more emphatic variations. For instance, the conventional simile "*as cunning as a fox*" is sometimes elaborated into "*as cunning as an* educated *fox*" or "*as cunning as an* old *fox*". In effect, the existing simile acts as a recognizable support structure that a speaker can exploit to achieve minor-level creativity. To quantify the extent to which this happens, and thereby determine the relative productivity of a simile-elaboration strategy, we generate a query of the form "as <*GROUND*> as a * <*VEHICLE*>" for every simple simile type in the corpus of 12,259 straight types harvested above. This finds over 5,700 elaborations of conventional similes on the web that mostly add perceptual information to aid visualization; thus, we find "*as white as a* frightened *ghost*"; "*as dangerous as a* ravening *wolf*"; "*as green as a* pickled *toad*" or "*an* Irish *meadow*"; "*as dry as a* stale *biscuit*" or "*a* stiff *martini*". However, not all the basic simile types yield attested elaborations, and these 5,700 extended types derive from just 700 adjectival grounds, that is, less that 20% of the set of 3769 adjectival grounds in our corpus of simple straight similes. Elaboration is a productive strategy, but clearly not an widely used one.

2.4. Subversive Elaborations

While these elaborated forms add just a single word to an existing simile, this additional word can sometimes alter its meaning in quite a dramatic fashion. We find that 2% of these elaborations (or 109 simile types) subvert an original simile to achieve an ironic effect, as in "*as dangerous as a* toothless *wolf*", "*as accurate as a* blind *archer*" and "*as lethal as a* toy *gun*". The majority of subversions – 93% – undermine a simile with a positive evaluation to produce a newer variant with a distinctly negative attitude. To communicate this critical viewpoint, subversions ask us to imagine broken, dysfunctional or pathetic instances of concepts whose stereotypical guise is far more impressive. But as these figures suggest, subversion of existing similes is a little used strategy for generating an ironic effect. For-

tunately, since the “about” form of similes appears quite commonly on the web, this promises to yield a much richer vein of creative comparisons.

3. Corpus II: Complex Comparisons

3.1. Compiling Lists of Complex Similes

Unlike metaphors, similes are hedged assertions, since a topic is merely stated to be approximately similar to, and not absolutely identical to, a given vehicle. Indeed, some similes are doubly-hedged, as if to indicate to their audience that the similarity on display is even more approximate. We see double-hedging in the following simile from Raymond Chandler, who uses the marker “about” to emphasize the wildly approximate nature of his comparison: *“[Moose Molloy] looked about as inconspicuous as a tarantula on a slice of angel food”*. The “about” marker seems to telegraph an author’s intention to use an inventive vehicle which exhibits an inexact *ballpark* similarity to the topic. Because the most culturally-entrenched similes are the most frequently reused, the simple query pattern “as * as *” is implicitly biased toward the retrieval of these most common types. This bias is reinforced by our efficiency-driven cut-off of 200 snippets per query, since many one-off originals are likely to fall outside this threshold. However, we now rerun our two-phase harvesting process with the doubly-hedged query “*about* as * as *”, so we are more likely to retrieve one-off similes of the kind that exhibit creativity.

Fishlov (1992) argues that excessive vehicle length is an attention-grabbing characteristic of creative similes, so we now extract all syntactically well-formed vehicles, whether they comprise one word or many, from the returned snippets. The extracted instances thus run the gamut from the short and punchy to the long and overwrought; “*about as pervasive as air*” is typical of the short variety, while “*about as difficult as finding work as a school teacher after a child-abuse conviction*” typifies the longer variety. In all, this second sweep of the harvester yields 45,021 instances of the “about” construction. Most of these instances occur just once overall, and this second harvesting sweep yields almost as many unique types (38,294) as instances, suggesting that 85% of these instances are bespoke one-offs. When hand-annotated for the salience profile that we expect from similes, we find that 20,299 of these types (53%) are more than mere comparisons,

and use vehicles for which the stated ground is either very salient or ironically opposed.

3.2. Annotating the Data

Interestingly, just 14% of these 20,299 simile types involve a vehicle with just one content-word, and a mere 3% of "about" simile types (676 types) are found in the original harvesting process of simple similes. In other words, the overlap in simile types found using both harvesting processes – single-hedged ("*as * as **") and double-hedged ("*about as * as **") – is negligible, on the order of 3 to 4%. Clearly, the addition of an "about" marker causes the second web sweep to harvest an almost completely different set of similes. We thus see a clear quantitative and qualitative separation between similes that are marked with "about" from more conventional similes. The "about" similes are typically longer, with a mean size of three words per vehicle, excluding initial determiners. They are also more heavily inclined toward the ironic. Hand-annotating for straight or ironic descriptions, we find that only 4797 unique simile-types (or just 24%) employ a vehicle for which the ground is both salient and apt, while 15,502 simile-types (76%) are ironic, as in "*about as modern as a top-hatted chimneysweep*.

The "about" form thus seems to be syntactic scaffolding that allows an author to telegraph an attempt to coin an unconventional, creative and potentially "spurious" (in the sense of Oring, 2003) simile. We can only speculate why the word "about" is semantically suitable to this role, but it does seem likely that the semantics of "about" allows it to act an implicit negation marker, in the sense of Giora (1995). Perhaps the non-spatial meanings of "about" – *imprecise*, *approximate* and *not quite* – impart a diluted sense of negation that alerts an audience to the possibility that all is not as it should be within the apparent logic of the simile.

4. Comparing Corpora

While most simple similes are formulaic evergreens, we find that 12% of "about" similes are topical and largely perishable, making use of well known names from the current cultural climate, such as "*Karl Rove*" and "*Paris Hilton*". Though there is just a 3% overlap between the longer

"about" similes and the shorter, more conventional figures of speech, this number significantly underestimates the role of conventional imagery in the construction of creative similes. On closer analysis, we find that 62% of the "about" similes use at least one stock image (such as *library*) drawn from the inventory of conventional vehicles. The longer similes do not use these stereotypes in isolation, or even to exemplify the same grounds, but combine them in novel ways to create memorable images, such as "*as lost as Paris Hilton in a library*". For instance, our first corpus of simple similes contains both "*as quiet as a cat*" and "*as noisy as a blender*", while our second corpus of "about" similes contains a simile that combines both of these to achieve an emergent, ironic effect: "*about as soothing as a cat in a blender*".

As in this example, a substantial number of "about" similes – 30% – use a vehicle that is a composite structure of two or more concepts linked by a preposition. The combination above employs two stock images with contrary properties – the stealthy cat and the loud blender – to evoke a visceral feeling of unease and disgust that stands in ironic opposition to the stereotype of calm relaxation that the simile initially promises. Notice how the simile cleverly plays each stock image against type: the cat, which might be considered soothing in normal circumstances, is placed in a cruel situation that prompts us to feel its suffering; and the blender, which is stereotypically loud and jarring, is ironically put forward as an exemplar of the very opposite. So while the longer "about" similes achieve more imaginative and creative effects than their conventionalized brethren, they are not completely distinct. They frequently draw upon the same conventional imagery, but in combinations that are designed to subvert stereotypical properties and create a heightened sense of perception and affect.

5. Empirical Analysis: Irony and Affect

5.1. Quantifying Attitude

A critical attitude is typical of irony, and creative "about" similes should be no different in this respect than simple similes with short, single-word vehicles. However, while some adjectives are uniformly critical in any context, such as "dull", "unattractive" and "stupid", most adjectives (such as "fragile", "tough" and "controversial") occupy a usage-sensitive middle ground between clearly-positive and clearly-negative. Lacking specific

knowledge of a speaker's views on a topic, or indeed of the topic itself, the quantification of a simile's positive or negative affect is too subjective to be meaningfully performed by a small group of human annotators. To achieve as much consistency as possible in the rating of attitudes, we turn to Whissel's (1989) *dictionary of affect*, an inventory of over 8000 English words with pleasantness scores that are statistically derived from human ratings. These scores range from 1.0 (most unpleasant) to 3.0 (most pleasant), with a mean score of 1.84 and a standard deviation of 0.44. For our purposes, we assume that the ground of a simile is negative if it possesses a pleasantness score less than one standard deviation below the mean (<= 1.36), and positive if it possesses a pleasantness score greater than one standard deviation above the mean (>= 2.28).

Using these numeric criteria, we can quantify the balance – or imbalance – of attitudes in different kinds of simile. In the most conventional straight similes, we see that a positive attitude is conveyed twice as often as a negative attitude (67% versus 33%). In contrast, simple ironic similes convey a negative attitude six times more often than a positive attitude (86% versus 14%). Turning to the more creative "about" similes, we see that straight "about" similes communicate a negative attitude a little more often than a positive attitude (56% versus 44%), but that ironic "about" similes carry a negative affect in almost 9 out of 10 cases (89% versus 11%). Simple similes are thus more likely to impart a positive view of a topic, while longer "about" similes are more likely overall (whether straight or ironic) to impart a negative view of a topic.

5.2. Irony and Affect

This difference is exacerbated by the strong preference for irony with the "about" form. Recall from section 4 that 76% of "about" simile types are ironic, while just 18% of the shorter, more conventional similes are ironic. Overall then, 83% of "about" similes impart a negative view of a topic, since 12% of "about" similes are non-ironic with a negative ground, and 71% ironically use a positive ground to impart a negative property. Tables 1 and 2 give an overview of the breakdown between irony and affect in each case.

Table 1. Total breakdown of similes with one-word vehicles. All cells sum to a total of 100%.

	Straight	Ironic
Positive Ground	55%	16%
Negative Ground	26%	3%

Table 2. Total breakdown of similes with similes with the "about" support-structire. All cells sum to a total of 100%.

	Straight	Ironic
Positive Ground	9%	71%
Negative Ground	12%	8%

The reliance of similes on familiar and evocative stereotypes in which particular properties are not just salient, but highly concentrated, means that similes have an exaggerated effect when attributing those properties to a topic. A positive description via simile is thus more likely to be seen as flattering than a non-figurative attribution of the same grounds, and a negative description is likely to be seen as more cutting. For example, it is less wounding to be described as "very ugly" than "*as ugly as a warthog*". This is in part because stereotypes represent extreme points of reference, and partly because stereotypes often have other unstated but resonant properties that are implicitly evoked (e.g., our corpus also attributes "dirty" to warthogs). When a stereotype-based vehicle is used to attribute just a single property to a topic, these other resonant properties will also be primed. The description "*as ugly as a warthog*" is thus a compact way of implying "as ugly *and dirty* as a warthog". There is a sardonic humour then in negative descriptions that are communicated via simile, but the precise degree of humour, and its effect, will depend both on the ingenuity of the simile and on the quality of the delivery.

As shown in Table 2, we can see that 83% of "about" similes have this potential for sardonic humour, either by directly describing a topic in negative terms (12%) or by indirectly implying a critical perspective via irony

(71%). In contrast, Table 1 shows that simple similes can be used for sardonic purposes in just 42% of cases (16% are ironically positive and 26% are non-ironically negative). These numbers suggest not just that irony is widely used in simile, but they also begin to explain why it is used. Table 1 shows that negativity is under-represented in simple similes, and that straight conventional similes communicate a positive description more than twice as often as a negative description (55% versus 26%). Irony provides a necessary corrective to this imbalance, allowing negative descriptions to be crafted from positive grounds. In simple similes, the balance is almost restored, with positive outweighing negative by 58% to 42%. Table 2 shows that "about" similes more than correct the remaining imbalance by choosing to employ their increased length and ingenuity in the service of negativity and ridicule.

6. Discussion

6.1. Ironic Interactions

Irony is a most vexing form of communication because – superficially, at least – it uses imagination and ingenuity to artfully disguise the expression of a negative sentiment. Consider this extract from an online discussion of the rules of baseball (Schwarz, 2003):

> "[B]aseball's rules structure has remained remarkably steady for more than 100 years. While basketball fiddles with 3-point lines and football puts its pass-interference, overtime and ref-upstairs rules in a Cuisinart each offseason, baseball rules remain as suggestible as a glacier."

While one can try to analyze the underlined simile (our marking) in isolation, it is clear that the take-home message is consolidated over the entire paragraph. Note how the ground of the simile, "suggestible", contrasts sharply with the property "steady" that is highlighted in the first sentence, and note how the second sentence uses "While" to establish a contrast between baseball and the more changeable games of basketball and football. Moreover, the extreme changeability of football is conveyed metaphorically, via the exaggerated claim that the football rulebook is shredded in a food processor at the end of each season. Though the irony can be localized

to the last clause, the irony is primed and supported by the paragraph as a whole, through a variety of interacting support structures.

Computational analysis reveals that this superficially novel simile is a simple variation on a well-worn comparison. For while Google identifies just one documentary source for the novel combination "*as suggestible as a glacier*" (i.e., Schwarz, 2003), "glacier" is a commonly used vehicle that occurs in a range of established similes. For instance, our test-set from section 5 contains 20 non-ironic similes with "glacier" as a vehicle, highlighting the properties *cold, cool, strong, fresh, impressive, unstoppable, pure, gradual, slow, slick, relentless, unwieldy, irresistible, frozen, frosty, implacable, impenetrable, unforgiving, forceful* and *implacable*. Glaciers are also used ironically in our test-set, to highlight the lack of the following properties: *mobile, erotic, excitable, speedy* and, of course, *suggestible*. Using the web query "*as steady and * as*" to find co-descriptors that are lexically primed by "steady", we find that these properties of "glacier" are primed: *strong, slow, cool, cold, implacable* and *unstoppable*. It follows that when an agent (whether a human or a computer) has already acquired a rich feature description of a vehicle from similes that were previously encountered and classified as non-ironic, it can choose to ignore the explicit ground in a new simile if it is not lexically primed by its context, and rely instead on those features of the vehicle that are primed. In this case, the features *slow* (to change) and *implacable* (in the face of change) are most appropriate to the topic of baseball rules. In other words, the figurative familiarity of the term "glacier" is itself a support structure for creative variation, so the irony of "*as suggestible as a glacier*" does not need the additional support of an "about" marker.

6.2. Support Structures for Irony

Since over 20% of "about" similes are non-ironic, it is incorrect to assume that "about" always signals the presence of irony. Our corpus analysis, the largest of its kind for similes, shows that the "about" form is more nuanced than a simple marker, but that it acts as a scaffolding structure for creative similes, priming an audience to view comparisons with positive grounds as ironically critical and comparisons with negative grounds as plainly critical. We employ the term *scaffolding* in the sense of Veale and Keane (1992), to mean a structure that allows immediate but superficial interpretation of a figurative utterance, and on which a deeper and more insightful interpreta-

tion can gradually be elaborated. In other words, the "about" form allows an audience to quickly construct a basic and mostly accurate interpretation of a speaker's intent without having to fully understand the meaning of the vehicle. All that is required is that the audience can determine the intended evaluative affect – positive or negative – of the simile's ground: if correctly ascertained as positive, then the simile has close to a 90% chance of being ironic and critical; if ascertained as negative, the simile has just a 40% of being ironic and is 60% likely to mean what it overtly says.

In a very real sense then, the "about" form appears to be a support structure for humorous linguistic creativity. Consider that creative similes of an obviously poetic bent (e.g., the kind analyzed by Fishlov, 1992) are typically crafted off-line, where they can be reworked and polished until they fully cohere with their narrative surroundings. They frequently give rise to complex mappings and associations, which encourage close-reading and deep analysis from their audience. In contrast, similes of a humorous bent are often generated spontaneously in fast-moving interactive situations, and genres of text that are rich in humorous comparisons (such as dialogue-heavy novels, comic narratives, and the online texts from which we harvest our corpora of similes in section 4) are typically designed to mimic the free, fast-paced flow of everyday conversation. In such time-constrained conditions, it is useful to be able to be to telegraph the basic meaning of a comparison, to minimize both the risk of information loss (if the comic conceit fails, or falls flat) and the risk of complete misinterpretation (if ironic intent is not recognized, or wrongly assumed where it is not intended). 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, suggesting that simile authors feel a need to cue readers as to the proper interpretation of their creative efforts. Explanations robs jokes of their potency, so we can expect humorous similes to eschew explicit explanations. The "about" marker is a more subtle cue than an explanation, but it is a cue nonetheless, one that signals a playfulness on the part of the author and one that licenses the audience to seek out a playful, and perhaps even ironic, interpretation when one is available.

6.3. Conclusions

We conclude by noting that the presence of "about" does not make a simile humorous, nor does its absence undo any potential a simile may have for

humour. Though we can identify structural and semantic features of similes that contribute to their humorousness, but we cannot identify structural or semantic features that are sufficient to make a simile humorous. Ultimately, humour is not semantically or structurally determined, but arises from the pragmatic effects of an utterance's use in a given context. Nonetheless, structural properties – like the presence of "about" – can encourage an audience to collude with the author in constructing a humorous interpretation. The "about" form is unlikely to be the only construction that supports and primes a humorous interpretation in this way, though it does seem to be one of the simplest and most direct, at least for similes. Further investigation is needed to see whether other linguistic markers of equal utility can be identified, both for predicting creative intent and for automatically harvesting potentially creative texts from the web.

Acknowledgments

The author gratefully acknowledges the contribution of Yanfen Hao to the compilation of the large web corpora described in this chapter.

References

Fellbaum, Christiane
1998 *WordNet: An Electronic Lexical Database*. Cambridge, Massachusetts: MIT Press.

Grice, Herbert Paul
1978 Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics: Vol. 9*. Pragmatics. New York: Academic Press.

Fishlov, David
1992 Poetic and Non-Poetic Simile: Structure, Semantics, Rhetoric. *Poetics Today*, 14(1): 1-23.

Hanks, Patrick
2004 The syntagmatics of metaphor and idiom. *International Journal of Lexicography*, 17(3): 245-273.

Giora, Rachel
1995 On Irony and Negation. *Discourse Processes* **19**: 239-264.

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

Jackendoff, Ray
1987 *Consciousness and the Computational Mind*. Cambridge, Massachusetts: MIT Press.

Lakoff, George and Mark Johnson
1980 *Metaphors We Live By.* Chicago, Illinois: University of Chicago Press.

Moon, Rosamund
2008 Conventionalized as-similes in English: A problem case. *International Journal of Corpus Linguistics* **13**(1(): 3-37.

Norrick, Neal
1986 Stock Similes. *Journal of Literary Semantics* **XV**(1): 39-52

Oring, Elliott
2003 *Engaging Humor*. Chicago, Illinois: University of Illinois Press.

Ortony, Andrew
1979 The role of similarity in similes and metaphors. *Metaphor and Thought,* (Ed.) A. Ortony. Cambridge, Massachusetts: Cambridge University Press. 342-354.

Roncero, Carlos, Kennedy, John M., and Ron Smyth
2006 Similes on the internet have explanations. *Psychonomic Bulletin and Review*, 13(1): 74-77.

Seinfeld, Jerry
2008 Dying is Hard. Comedy is Harder. *New York Times Op-Ed Contributor,* June 24 edition.

Sperber, Dan and Deirdre Wilson
1992 On verbal irony. *Lingua* **87**: 53-76.

Schwartz, Alan
2003 The history of rule changes. *ESPN.com, February 4th* posting.

Taylor, Archer
1954 Proverbial Comparisons and Similes from California. *Folklore Studies* **3**. Berkeley, California: University of California Press.

Veale, Tony and Mark T. Keane
1992 Conceptual Scaffolding: A spatially founded meaning representation for metaphor. *Computational Intelligence* **8**(3): 494-519.

Veale, Tony
2006 Computability as a test on linguistic theories. In G. Kristiansen, M. Achard, R. Dirven and F. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives, Applications of Cognitive Linguistics*. The Hague: Mouton de Gruyter: 461 – 484.

Veale, Tony, Hao, Yanfen and Guofu Li
2008 Multilingual Harvesting of Cross-Cultural Stereotypes. *In proceedings of the 46th Annual Meeting of the Association of Computational Linguistics (ACL):* 888 – 895.

Whissell, Cynthia
1986 The dictionary of affect in language. In R. Plutchnik & H. Kellerman (Eds.) *Emotion: Theory and research*. New York: Harcourt Brace: 113-131.