

An Ironic Fist in a Velvet Glove:

How Often Does Irony Couch A Negative Sentiment in Positive Terms?

Tony Veale, Yanfen Hao,

School of Computer Science and Informatics, University College Dublin.

1. Introduction

Like most creative phenomena, irony is difficult to pin down in formal terms, and no single definition ever seems entirely satisfactory. To begin with then, let's consider some obvious examples of verbal irony in everyday situations:

Going to your car in the morning, you notice that one of your tyres is completely flat. A friendly neighbour chimes in with "*Looks like you've got a flat*". Marveling at his powers of observation, you reply "*Ya think?*".

When having breakfast in a greasy-spoon café, you hungrily polish off everything on your plate. Seeing your totally clean plate, your waitress quips "*Well, that must have been terrible*". "*Yes*", you reply, "*absolutely awful*".

A professor explains and re-explains Hegel's theory of the State to his class of undergraduates. "*Is it clear now*", he asks. "*Clear as mud*", a student replies.

A man goes through the entrance to a building but fails to hold the door for the woman right behind him, even though she is visibly struggling with a heavy box. She says "*Thank You!*" anyway.

A man holds the door open for the woman behind him as they both enter a building. The woman hardly seems to notice and says nothing. Nonetheless, the man says "*You're Welcome!*".

An office worker asks his company's IT guru to help him understand why a very personal email was sent to everybody in the office rather than to just its intended recipient. The IT guy replies "*Email can be such a challenge*".

These examples suggest that pretence plays a key role in irony: speakers craft utterances in spite of what has just happened, not because of it. The pretence in each case alludes to, or echoes, an expectation that has been violated (see Clark and Gerrig, 1984; Sperber and Wilson, 1992), such as the expectation that others behave in a civil fashion, speak meaningfully and with clarity, or not consume every single speck of food on their plate. This pretence may seem roundabout and illogical, but it offers a sharply effective and concise mode of communication. Irony allows a speaker to highlight the expectation that has been violated while simultaneously poking fun at, and often rebuking, the violator.

The most formulaic of the ironic rejoinders above is the cliché "*clear as mud*". At first glance, this may not even appear to be ironic: after all, the student is effectively saying "*what you have said has the clearness of mud, which offers no clearness at all*". But even in this reading, the student deliberately uses the word "*clear*" in a different

sense to that of the professor (i.e., “clear” as in “clearness” as opposed to “clarity”), an unsound but humorous strategy that logicians name the *fallacy of equivocation*. This deliberate equivocation amounts to *pragmatic insincerity* on the part of the student (see Kumon-Nakamura, Glucksberg and Brown, 1995). In conjunction with the failed expectation that the lecture is clear, this pragmatic amounts produces an ironic effect.

Of the six examples of irony above, five use pretence to appear superficially positive at the word level, if not in tone (we imagine most of the rejoinders spoken with a sarcastic tone). For instance, “*Ya think?*” seems to acknowledge the right of the addressee to hold an opinion about something so obvious as a completely flat tyre, while “*email can be such a challenge*” seems to sympathize with the addressee’s plight. The only example that is superficially negative is the exchange with the waitress, but this is also the only example where we do not imagine the words spoken with a sarcastic tone. In fact, while we imagine the positive-seeming rejoinders spoken with a sharp tone, the negative-seeming rejoinder is likely spoken with a playful, upbeat tone. In each case, the words create a pretence, while the spoken delivery of the words deliberately undercuts this pretence. This is how irony can achieve so much in so few words: it often allows us to simultaneously communicate on two different levels via two different channels.

From this cherry-picked sample we might conclude, as people generally do, that irony is mostly used to express propositions with a negative sentiment (such as rebukes, complaints and criticisms) in positive terms. This view is certainly appealing, for irony allows us to doubly-wound an addressee: first, by making them think, if only for a fleeting moment, that they are the recipients of praise; second, by withdrawing this praise, making the addressee feel silly for ever thinking it was warranted, and then delivering the underlying rebuke. The waitress example demonstrates, however, that irony sometimes uses the language of criticism to playfully hide a more positive sentiment. When we think of verbal irony, negative examples rush to mind as our archetypal experience of the phenomenon, but just how prevalent is each kind? We expect the *negative-disguised-as-positive* variety to outnumber the *positive-disguised-as-negative* variety, but by how much, and can we quantify the difference?

To avoid cherry-picking, we aim to automatically harvest a very large sample of ironic remarks from the texts of the world-wide-web. For this goal, we require a formulaic format in which irony is frequently expressed in a self-contained manner: we do not want to harvest whole exchanges between multiple speakers, or arbitrary co-texts for the ironies so that we can later understand their meanings. The simile-phrasing of the rejoinder “*clear as mud*” provides the key. We can easily harvest self-contained similes from the Web (e.g., see Veale *et al.* 2008, who show that the method can be applied cross-lingually), and we can predispose the harvester to seek out ironic uses of the simile pattern by using the hedge-marker “about”. The imprecision afforded by “about” often prefigures the use of irony, as if speakers wish to telegraph their creative intent and give their audience a fighting chance at interpretation. To this end, we collect two very large corpora of similes: in section 2, we describe how simple similes with one-word vehicles, such as “*as cunning as a fox*”, are harvested from the Web, while in section 3 we explicitly seek out more complex similes prefixed with imprecision marker, “about”. In section 4 we present a statistical analysis of these corpora, to tease apart their similarities and differences. In section 5 we turn to a consideration of irony, to quantify the degree to

which our large corpus of ironic similes favours the *negative-disguised-as-positive* profile. In the process, we also statistically determine the effectiveness of “about” as a marker of irony.

2. Corpus I: Simple Comparisons

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 Proverbs*. 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 queries “*as * as **” and “*about 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 recur 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. Rather, 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.

2.1. 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 similes, such as “*as hairy as a bowling-ball*”, and of these, 2798 form unique types.

2.2. 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 than 20% of the set of 3769 adjectival grounds in our corpus of simple straight similes. Elaboration is a productive strategy, but clearly not a widely used one.

3. Corpus II: Complex Comparisons

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.

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 the disguise of true sentiment

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.

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 I and II give an overview of the breakdown between irony and affect in each case.

Table I: *Total breakdown of similes with one word vehicles. All cells sum to 100%.*

	<u>Straight</u>	<u>Ironic</u>
<u>Positive Ground</u>	55%	16%
<u>Negative Ground</u>	26%	3%

Table II: *Total breakdown of similes with “about” marker. All cells sum to 100%.*

	<u>Straight</u>	<u>Ironic</u>
<u>Positive Ground</u>	9%	71%
<u>Negative Ground</u>	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 II, 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 I 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 I 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 II 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. Conclusions

Since over 20% of “about” similes are non-ironic, it is not correct 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 interpretation 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.

Irony does not always use a positive expression to disguise a negative sentiment, and “about” does not always signal irony in a simile. We knew this before we began, of course, but through the analysis of a large trove of Web-similes, we can now quantify the degree to which these generalizations often hold true. In most cases, an “about”-marked simile really is an ironic fist in a velvet glove. It remains to be seen in future analyses whether the particular distribution we observe here for similes accurately reflects the behavior of other framing devices for verbal irony.

References

- Clark , H. H. and Gerrig, R. J.** (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General* 113(1):121-126
- Fellbaum, C.** (ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Fishlov, D.** (1992). Poetic and Non-Poetic Simile: Structure, Semantics, Rhetoric. *Poetics Today*, 14(1), 1-23.
- Giora, R.** (1995). On Irony and Negation. *Discourse Processes* **19**, pp 239-264.
- Kumon-Nakamura, S., Glucksberg, S. and Brown, M.** (1995) How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General* 124(1):3-21.
- Moon, R.** (2008). Conventionalized as-similes in English: A problem case. *International Journal of Corpus Linguistics* **13**(1), 3-37.
- Norrick, N.** (1986). Stock Similes. *Journal of Literary Semantics* **XV**(1), 39-52.
- Oring, E.** (2003). *Engaging Humor*. University of Illinois Press.
- Ortony, A.** (1979). The role of similarity in similes and metaphors, in *Metaphor and Thought*, edited by A. Ortony. Cambridge, MA: Cambridge University Press.
- Roncero, C., Kennedy, J. M., and Smyth, R.** (2006). Similes on the internet have explanations. *Psychonomic Bulletin and Review*, 13(1), 74-77.
- Sperber, D. and Wilson, D.** (1992). On verbal irony. *Lingua* **87**, pp 53-76.
- Taylor, A.** (1954). Proverbial Comparisons and Similes from California. *Folklore Studies* **3**. Berkeley: University of California Press.
- Veale, T. and Keane, M. T.** (1992). Conceptual Scaffolding: A spatially founded meaning representation for metaphor. *Computational Intelligence* **8**(3), pp 494-519.
- Veale, T., Hao, Y. and Li, G.** (2008). Multilingual Harvesting of Cross-Cultural Stereotypes. *In proc. of the 46th Ann. Meeting of the Assoc. of Computational Linguistics*.
- 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.