AAAI Technical Report FS-12-02
Artificial Intelligence of Humor

Detecting and Generating Ironic Comparisons: An Application of Creative Information Retrieval

Tony Veale
Web Science and Technology Division, KAIST, Daejeon, South Korea
tony.veale@gmail.com

Abstract

Ironic utterances promise an expected meaning that never arrives, and deliver instead a meaning that exposes the failure of our expectations. Though they can appear contextually inappropriate, ironic statements succeed when they subvert their context of use, so it is the context rather than the utterance that is shown to be incongruous. Every ironic statement thus poses two related questions: the first, "what is unexpected about my meaning?" helps us answer the second, "what is unexpected about my context of use?". Like metaphor, irony is not overtly marked, and relies instead on a listener's understanding of stereotypical norms to unpack its true meaning. In this paper we consider how irony relies upon and subverts our stereotypical knowledge of a domain, and show how this knowledge can be exploited to both recognize and generate ironic similes for a topic.

The Soul of a new cliché

Samuel Goldwyn, the co-founder of MGM studios, famously summed up Hollywood's attitude to creativity with the line "Let's have some new clichés". On the face of it, this seems like just another one of Goldwyn's many memorable misstatements (like "include me out!"): after all, it's hard to think of clichés as *new*, or as something that can be invented on demand. Yet, on closer analysis, one can find real insight in Goldwyn's remark. Clichés are considered anathema to the creative process because they represent everything that is conventional and jaded about the status quo. However, clichés become tired thru overwork, and are overworked precisely because they prove themselves so useful in so many different contexts. Few writers set out to create a new cliché, but most would like their efforts to become as much a part of the fabric of our linguistic culture as the most tenacious of clichés.

One productive form of *new cliché* is the humorously pithy comparison, as in "as durable as a chocolate teapot" or "as useful as a screen door on a submarine". Speakers recognize memorable comparisons when they hear them, and re-use them as eagerly as one retells a favorite joke. The most frequently reused comparisons can, in this way, acquire the clichéd status of a proverbial simile. When the folklorist Archer Taylor collected his corpus of proverbial similes in 1954, he observed not just a wide variety of humorous comparisons in American speech, but a wide variety of humorous forms for the same descriptive qualities, such as "durable" and "useful". Speakers are clearly drawn to popular comparisons of proven value, but are equally fond of coining their own, in the hope that their witty new descriptions are widely reused by others in turn. This constant churn of re-invention keeps our language fresh, and ensures that ironic comparisons retain their ability to challenge and to entertain, even as others – such as "crazy like a fox!" and "as clear as mud!" – acquire an idiomatic status which makes them effortlessly understood.

Stereotypes anchor a comparison in the realm of the familiar, where judgments about representativeness are made quickly and intuitively by cognitive processes that Kahneman (2011) dubs *System 1*. Conversely, similes help to perpetuate stereotypes by packaging them into such conveniently re-usable linguistic forms. However, ironic similes subvert the workings of *System 1*, and force us to engage a set of deliberative, analytic and non-intuitive processes that Kahneman dubs *System 2*. Whereas straight similes use stereotypes as exemplars of a given quality, ironic similes highlight the unexpected lack of this quality by instead using a withering counter-example, perhaps constructed from stereotypes that one intuitively associates with very different (and even antonymous) qualities. Humorously ironic comparisons often construct this counter-example by subverting the normal form of a stereotype. Thus, while we expect a vault to be the very model of a secure location, we might describe a weak and

unsafe container as being "as secure as a chocolate vault". The stereotypes "chocolate" and "vault" are combined here to produce a complex concept with internal incongruities (*secure* and *strong* versus *brittle* and *soft*) that lead to emergent qualities (such as *not secure at all*). Novel combinations such as "chocolate vault" require insights that can only come from *System 2*, to override those that are produced by *System 1*.

Nonetheless, after encountering a variety of sardonic utterances that are anchored in a construct such as "chocolate X", we might infer a general rule: If X is a stereotype of strength or solidity or durability or some related quality, then an X made of chocolate will have none of these qualities. Repeated exposure to the "chocolate X" pattern will help us acquire a *meta-cliché* that we can use, in turn, to generate novel instances of our own, such as "chocolate tank" and "chocolate castle". Of course, we cannot simply stick the prefix "chocolate" in front of an arbitrary stereotype and hope for the best; rather we need a means of identifying meaningful combinations.

An approach to linguistic creativity called CIR, *Creative Information Retrieval,* offers such a means (Veale, 2011). We describe in this paper how CIR allows us to harvest readymade phrases from a corpus such as the Google n-grams, and allows us to identify the most likely qualities of the complex concepts denoted by those phrases. In the next section we show how the stereotypical knowledge exploited by CIR is acquired from web-harvested similes, and consider the signs, and the likelihood, that a web-harvested simile is in fact ironic. We then describe how this knowledge is operationalized in an information-retrieval guise as a set of powerful query operators, which can be used to retrieve similes, metaphors, analogies and other pithy descriptions from text corpora. Finally, we explore how this knowledge can be applied to a corpus of creative similes that has been annotated for irony, to automatically learn a collection of meta-clichés than can then be used to generate new examples of ironic similes.

Stereotypes, Similes and Irony

Similes – even ironic similes – exploit our mutual knowledge of stereotypical norms to communicate their meaning, even if these norms are ultimately subverted. As such, similes are also an ideal source of stereotypical knowledge for a *tabula rasa* agent such as a computer, for as Dickens notes in *A Christmas Carol*, "the wisdom of our ancestors is in the simile". There is, however, an important caveat: such agents are liable to make the worst mistakes possible if they misread a speaker's intention to be ironic.

Veale & Hao (2007a) explored the use of similes as a source of stereotypical norms, and harvested tens of thousands of simile bodies (of the form "as X as a Y") from the web. Lacking an automatic procedure for filtering straight similes from ironic similes, Veale & Hao filtered their simile-set manually, to obtain a collection of over 12,000 unique straight simile bodies (such as "as hot as an oven") and almost 3000 unique ironic simile bodies (such as "as subtle as a sledgehammer"). Veale & Hao (2007b) then used the straight similes as a source of stereotypical norms (e.g. ovens are hot, jungles are humid, snow is soft) in a system for comprehending and generating metaphors.

The unused ironic similes were later used as a test-set for an irony detection algorithm in Hao & Veale (2010). A simple question was posed of each simile in the combined set (straight + ironic): can a simile "as X as Y" be plausibly recast in a form that is highly likely to be non-ironic, or can it more plausibly be recast in a form that is highly likely to be ironic? For the idealized non-ironic form, the pattern "X Zs such as Xs" was chosen, since irony is very infrequently observed in this construction. Thus, the simile "as hot as a sauna" is recast as "hot Zs such as saunas". For the idealized ironic form, the pattern "about as X as Y" was chosen, since "about" has been hypothesized to be a subtle marker of irony (Moon, 2008). Thus, the simile "as muscular as a paper-clip" is recast as "about as muscular as a paper-clip". To test the plausibility of each recasting, simple web frequency for each form was calculated using the search engine *Google*. Though Veale & Hao describe a number of ancillary heuristics in their methodology, this simple recasting is the backbone of their approach, which achieves promising results overall: 87% of ironic similes are recognized with **.63** precision, while 89% of straight similes are recognized with **.97** precision. The F-score for classification of ironic similes is **.73**, for straight similes it is **.93**, and for irony/straight classification overall it is **.88**.

Veale (2012) used the "about" construction to harvest a collection of irony-rich similes, and extended the form of the search to accept complex similes in which the vehicle is a multiword phrase (such as "a wet rag" or "a chocolate teapot"). This collection of 20,299 web similes was then hand-annotated for irony, so that it could be used as a gold-set for future irony studies. The collection lends further support to the role of "about" and other markers of imprecision (such as "not exactly X" and "almost as X as") in marking irony, since 76% of the newly harvested similes (such as "about as modern as a top-hatted chimneysweep") are marked as ironic. Just 14% of these 20,299 "about" simile types use a vehicle with a single content-word, and a mere 3% (i.e., 676 types) are also found in the original harvesting process of Veale & Hao (2007a). So the "about" marker not only signals a speaker's intention to be playful and ironic, it also signals a speaker's intention to be humorous and perhaps creative. This finding suggests that when computers aim to be ironic in their descriptions, the use of the "about" construction will help ensure that their efforts are more readily received as well-formed irony.

Creative Information Retrieval

Ironic utterances can be viewed as congruous statements in incongruous contexts. As Sperber & Wilson put it in their echoic-mention theory of irony, an ironic utterance echoes one that would plausibly be made in a very different context, thus drawing an audience's attention to a failure of expectations in the current context. One can ironically echo an utterance that is explicitly associated with another context (an *explicit* echo) or one can allude to a general expectation arising from another context (an *implicit* echo). Fludernik (2007) generalizes this notion of an implicit echo to include any pointed allusion to accepted beliefs, received wisdom, or stereotypical norms. In any case, to process irony a computer needs the ability to retrieve either the appropriate utterance or the appropriate norm. Veale (2011) describes a platform for the retrieval of creative language that can be used to achieve each of these ends.

Creativity often arises from a process of *appropriation*, in which something is wrenched from its normative context of use and given new meaning in a new setting. In this vein, Duchamp popularized the notion of an artistic *ready-made* when his *Fountain* – a signed urinal – was presented with some controversy at a Dada exhibition in 1917. We normally think of readymades as physical objects whose artistic merit derives wholly from their selection by an artist, but language is also rich in linguistic readymades. Just think of how many movies, songs, novels, poems, etc. allusively borrow utterances and phrases from each other. For example, the movie *The Usual Suspects* takes its name from a famous quote from the movie *Casablanca*, while the novel *All The King's Men* takes it title from a famous nursery rhyme; this title has, in turn, inspired the title of Woodward and Bernstein's book *All The President's Men*.

The well-formed phrases that one can extract from the Google database of n-grams can be viewed as a vast collection of linguistic readymades, insofar as each is a phrase that is wrenched from its original context of use on the web. To find the right readymade for a given task, one needs a set of retrieval tools that go beyond the normal range of query operators that are employed in standard IR. Veale (2011) describes a set of *non-literal* query operators that allow a user to search for n-grams on the basis of metaphorical similarity rather than literal string similarity. These operators are denoted **@**, **?** and ^ and defined thus:

@ is the stereotype operator, where @*noun* will match any stereotypical property of *noun* and @*adj* matches any noun denoting a concept that has the stereotypical property *adj*. For example, @*razor* will thus match the words "sharp", "straight", "clean" and "smooth", while @sharp will match any of the words "razor", "knife", "sword", and so on.

? is the neighborhood operator, where ?*noun* will match any other noun that is frequently clustered with *noun*, and ?*adj* will match any other adjective that is frequently used with *adj* in a complex description. The neighborhood for nouns is determined by gathering matches for the pattern "Xs and Ys" (a coordination of bare plurals) in the Google n-grams (such as "angels and demons", "doctors and nurses", "cups and saucers", etc.), while the neighborhood for adjectives is determined by gathering matches for the pattern "as X and Y as" (a multi-property simile) on the web. For example, ?*disaster* will match any of the words "tragedy", "catastrophe", "calamity", "misfortune", "hardship", "plague", "famine" and so on, while ?*tragic* will match any of the words "sad", "shocking", "terrible", "unfortunate", "ridiculous" and so on. All matches for ?*noun* are ranked by the WordNet similarity of each match to *noun*, while matches for ?*adj* are ranked by the co-occurrence frequency of each match with *adj* on the web

^ is the category operator, where ^*class* will match any member of the pre-defined category *class*. Users can give names to their own ad-hoc categories, or designate names for the results of previous retrievals (for instance, a user might populate the category ^*sharptool* with words that match the CIR query @*sharp* & ^*tool*). The ^ operator can also be used to refer to WordNet categories; e.g., ^*person* matches any noun denoting a kind of person in WordNet.

To these three operators we add a fourth:

- is the antonym operator, where –P will match any antonym of the adjective P. Antonyms are derived from the lexical resource WordNet. This, for example, -*soft* matches the word "hard", and –*strong* matches the word "weak".

Importantly, these non-literal operators can be used in complex combinations. For instance, ?@adj will match any noun in the neighborhood of a noun / concept that exhibits the stereotypical property *adj*, while @-*adj* will match any noun/concept that is a stereotypical holder of any property denoted by an antonym of *adj*. Likewise, @?adj will match any noun / concept with a stereotypical property that is like *adj*, while @@*noun* will match any other noun / concept that shares at least one stereotypical property with *noun*.

Veale (2011) describes how non-literal operators, when used to augment an IR system for retrieving Google n-grams, can provide a generic foundation for retrieval-oriented linguistic creativity. For instance, CIR can be used to turn the readymade phrases of the Google n-grams into vehicles for creative comparison. For a topic X and a property P, straight similes of the form "***X** is as **P** as **S***" are easily generated by retrieving values for $\mathbf{S} \in (\mathbf{@P} \cap \mathbf{??X})$. Likewise, the query "***?P @P***" will retrieve corpus-attested elaborations of stereotypes in **@P** to suggest similes of the form "***X** is as **P** as* **(*?P @P*)**". The simile "*as cold as a fish*" can thus be elaborated to yield "*as cold as a wet fish*", "*as cold as a dead haddock*", "*as cold as a wet January*", "*as cold as a frozen corpse*", and "*as cold as a heartless*

robot". Complex queries can retrieve more elaborate vehicles for creative similes, so "***@P @P***" (which retrieves "*robot fish*" and "*snow storm*" for *cold*) and "**?P @P @P**" (e.g. "*creamy chocolate mousse*" for *rich*) each retrieve n-grams that blend two different but overlapping stereotypes.

By using the antonym operator **-**, ironic similes can also be generated for the *P-ness* of a topic X using the pattern "***X** is as **P** as* (**@-P ∩ ??X**)". In effect, (**@-P ∩ ??X**) finds counter-examples of *P-ness* that are comparable to X, rather than the stereotypical examples that we expect in the normative, non-ironic use of *as*-similes. However, adjectives can be ambiguous, and *–P* may not be an appropriate antonym for the intended sense of *P*.

Modeling Stereotypical Behavior

The effectiveness of CIR depends crucially on the matching ability of its operators **@**, **?**, **-** and ^, and on the range of **@** in particular, since only this operator can match properties (adjectives) to stereotypes (nouns) and vice versa (**?**, **-** and ^ each match words that reside in the same syntactic categories). The knowledge provided by @ is a model of common-sense stereotypical norms, which is precisely the kind of knowledge that is subverted by irony. However, the simile pattern "as X as Y" that Veale & Hao first used to harvest stereotypical associations from the web is limited to the retrieval of adjectival properties. It cannot harvest stereotypical behaviors, such as the fact that babies drool and cry, that politicians lie, or that dogs bark.

To rectify this limitation and to simultaneously enlarge the stereotypical model underpinning CIR, we use a bottom-up, evidence-driven approach to identifying stereotypical properties and behaviors for a given topic. We first harvest all 3-word phrases from the Google 3-grams that match the pattern <DET PROPERTY NOUN>. Here PROPERTY can match either an adjective in WordNet or an inflected verb that denotes a behavior, such as *swaggering* or *armored*. For adjectival cases, we generate the *as*-simile "as ADJ as a NOUN", while for verbal behaviors we generate the *like*-simile "BEHAVIOR like a NOUN". We then use the resulting simile as a web query to see how many times the simile occurs on the web. That is, *as*-similes are generated for adjectival properties and *like*-similes are generated for verbal behaviors, so e.g. the 3-gram "a *reckless* cowboy" yields the query "*as* reckless *as a* cowboy" and the 3-gram "a *swaggering* cowboy" yields the query "swaggering *like a* cowboy".

We next consider only the queries that return a non-zero result set; these correspond to stereotypical associations that have been attested by usage on the web. Nonetheless, this set still contains a great deal of noise, and we find many under-specified behaviors such as "walking like a drunk" or "talking like a baby". We could use statistical techniques here to separate the most discriminating behaviors (e.g., "*staggering* like a drunk" and "*babbling* like a baby") from the least informative (e.g., "*walking* like a sailor"). Yet, since this task only needs to be done once to create a precise and highly reusable lexico-conceptual resource, we do it manually to achieve the best results.

It takes a matter of weeks to perform manual filtering, but the stereotype model that results from this effort is significantly larger than that produced by Veale & Hao (2007a). It contains 9,479 different stereotypes overall, and ascribes to each of these a selection of 7,898 different properties and behaviors. In all, the new resource contains over 75,000 unique noun-to-property associations. This compares very favorably with the 12,000+ associations in Veale & Hao's original resource. Consider the term *baby*: the new resource suggests these 163 associations as being stereotypical of babies:

{*delicate, squalling, weeping, baptized, adopted, startled, attentive, blessed, teeny, rocked, adorable, whining, bundled, toothless, placid, expected, rescued, treasured, new, sleepy, indulged, slumbering, weaned, pure, supple, helpless, small, sleeping, animated, vulnerable, wailing, cradled, kicking, soft, rested, bellowing, blameless, grinning, screaming, orphaned, mute, cherished, reliant, thriving, loveable, guileless, sniveling, inexperienced, harmless, dribbling, unthreatening, nursed, angelic, bawling, beaming, tame, naked, spoiled, scared, weak, squirming, blubbering, contented, smiling, wiggling, mewling, blubbing, sniffling, overtired, dimpled, loving, dear, tired, powerless, bewildered, peaceful, distressed, naive, wee, soiled, sucking, fussy, gurgling, vaccinated, heartwarming, pouting, constipated, drooling, quiet, wiggly, lovable, bare, weaning, suckling, cute, bald, whimpering, tender, pampered, incontinent, fleshy, charming, dependent, artless, fussing, flabby, babbling, warm, giddy, crawling, snoozing, hairless, cuddled, sweet, sobbing, squealing, wrapped, tiny, cooing, swaddled, laughing, toddling, fragile, innocent, moaning, gentle, terrified, precious, cranky, giggling, confused, pink, cuddly, fat, ignorant, snoring, young, howling, screeching, shrieking, trusting, shivering, napping, resting, frightened, fresh, loved, demanding, chubby, adored, appealing, happy, relaxed, bumbling, wriggly, rocking, wriggling, conceived, clean, content, smooth, crying, submissive*}

Strategies of Ironic Subversion

The use of antonym-based counter-examples is just one strategy that is available to the ironic speaker. We can express this strategy $S_{antonym}$ in CIR terms as follows:

$$S_{antonym}(P) \leftarrow \quad \textbf{?}\text{-}P \quad \textbf{@}\text{-}P$$

In other words, a set of ironic expressions for a property *P* can be found by retrieving all 2-gram phrases where the

first word is an adjective that reinforces the idea of *–P* (any antonym of *P*) and the second word is a noun denoting a concept for which any element of *–P* is stereotypical. Thus, for example, the phrase "soggy pillow" is retrieved for the property *hard*, since *soft & hard* are antonyms in WordNet.

Since we so often seek to impress with irony, our goal is not merely to communicate an implicit negation, but to communicate an implicit negation in the most imaginative, memorable and quotable words we can muster. A vivid juxtaposition of ideas can help us to achieve this goal. We can thus use the following variant of the $S_{antonym}$ strategy:

S_{combo}(P) ← **@**-*P* **@**-*P*

For example, since a *wall* is typically *hard*, and a good basis for ironic descriptions of softness, the phrases "brick wall", "stone wall", "steel wall", "titanium wall", "oak wall", "granite wall" etc. are retrieved by S_{combo}(soft). Likewise, phrases like "marshmallow bunny" and "snow baby"– which can seem decidedly odd, and thus fresh and imaginative when considered out of their original context – are retrieved as ironic descriptions of hardness. In this way, strategies like S_{combo} embody Fishlov's (1992) view of poetic similes, by providing more elaborate and more vivid mental images than a single stereotype alone could do. A computer that uses a database of readymade phrases to suggest possible word/idea combinations for creative descriptions is thus tapping into the collective imagination of many different speakers at once.

We use the term "strategy" to denote a high-level approach to generating ironic descriptions, where each such approach is represented as a CIR query for retrieving matching phrases from a corpus such as the Google n-grams. For instance, we can define another strategy S_{group} that is defined via the following CIR query:

S_{group}(P) ← (^*group* ∩ **@***P*) "of" **@**-*P*

The CIR query for S_{group} matches any 3-gram phrase in which: the first word denotes both a group (like family or army) *and* a concept for which the given property *P* is stereotypical; the second word is the string literal "of"; and the third word is a noun that denotes a concept for which the antonym of the given property *P* is stereotypical. Under normal circumstances, any word matching (^*group* ∩ @P) would serve as an excellent example of *P-ness*, denoting a mass of *P* stereotypes in a coherent grouping. However, S_{group} finds phrases that subvert this stereotypical group, by populating the group with counter-examples of *P-ness*. Thus, "army of dreamers", "army of civilians" and "army of irregulars" are all retrieved from the Google 3-grams as ironic vehicles for the property *disciplined*, while the 3-grams "army of cowards", "army of babies", "army of ants", "army of cripples", "army of kittens", "army of girls" and "army of worms" are retrieved for *strong*. Notice the way in which S_{group} creates descriptions that suggest *P-ness* right up to the last word, whereupon a final ironic reversal of meaning is delivered. S_{group} can thus be viewed as a more sophisticated ironic version of the adolescent strategy for generating sarcasm, where "Not!" is placed at the end of an otherwise affirmative utterance.

To use the language of Attardo and Raskin's GTVH, each strategy is effectively a *logical mechanism* that causes an incongruous combination of ideas to be juxtaposed, and to be subsequently resolved as an attempt at irony. In $S_{antonym}$ and S_{group} the incongruity derives from lexical knowledge of antonyms and the properties they denote. For example, consider the following strategy, dubbed $S_{material}$

$S_{material}$(P) ← (^*material* ∩ **@**-*P*) **@***P*

The category ^*material* is populated with elements of the WordNet category ^*substance* for which 3-gram matches can be found for the CIR query "made of ^*substance*", such as "chocolate", "concrete", "steel", etc. Thus, though *brick* is a stereotypically *hard* object, $S_{material}$(hard) is populated with the 2-gram phrases "silk brick", "snow brick" and "water brick". To provide a garden-path effect, each of these vehicles can be re-formulated so that the ironic kicker is found at the end of the phrase, as in "a brick made of silk", "a brick made of snow" and "a brick made of water".

As shown in $S_{antonym}$, S_{combo}, S_{group} and $S_{material}$, each CIR query acts as a high-level rule in which the key elements are matched non-literally against a large corpus of phrasal n-grams. But, as in all high-level rules, each rule may give rise to unexpected, low-level exceptions. The more generic the strategy, the more exceptions it brooks. Consider this somewhat under-constrained strategy:

S_{proxy}(P) ← **?**-*P* **@?**-*P*

This strategy is a variant of $S_{antonym}$ that uses neighbors of antonyms of P (that is, **?**-*P*) rather than just antonyms of *P*, and evokes stereotypes of those neighborly oppositions (via **@?**-P). For instance, this strategy retrieves the 2-gram "meek lamb" as an ironic vehicle for the property *savage*. However, given the wide range of neighbors in **?**-*P*, this strategy will significantly over-generate. Though it is well meaning, we can expect just a small proportion of the phrases suggested by S_{proxy} to achieve an ironic effect.

Moreover, words in unexpected combinations often have specific nuances that we cannot predict from high-level rules alone, and specific combinations may not give rise to the predicted ironic effect. To address this problem, high-level strategies – such as S_{proxy} and others considered above – should be implemented via lower-level tactics that are grounded in attested examples of irony, and which specifically employ words that have a proven record of communicating an ironic meaning.

Subversive Tactics

A computer can learn specific tactics for each strategy by considering specific instances of ironic descriptions. So, given an ability to detect irony in figurative comparisons, a computer can observe which ironic comparisons instantiate which strategies, and can learn word-specific tactics for each as new examples are encountered. To begin with, Veale's (2012) corpus of annotated *about-as*-similes provides a rich set of initial observations on which to build.

Consider this ironic simile from Veale's (2012) corpus, describing the level of password protection in MS Word:

"about as *secure* as a cardboard bank vault"

Both *bank* and *vault* are stereotypically associated with the ground property *secure*, so "bank vault" is a compound term with the same stereotypical association. The original simile thus exhibits the following general pattern:

"about as *secure* as a *cardboard @secure*"

Since the simile is annotated as ironic, its ironic effect must emerge from its specific use of the word "cardboard". That is, *cardboard* must possess some quality that subverts the secureness of an otherwise very secure container. This case appears to be a specific use of the $S_{material}$ strategy, yet *insecure* is not a stereotypical property of *cardboard*, so there is no semantic incongruity between *cardboard* and *secure* (as there is between *concrete* and *soft*, say, or *marshmallow* and *hard*). In such a context, when used to describe a container that should be secure, *cardboard* prompts the emergent inference that it is not secure at all.

This specific use of words suggests this tactic for irony:

$T_{cardboard}$(secure) ← "cardboard" ***@secure***

Tactics, labeled with a T, are more specific than strategies (labeled with an S). Note how the tactic above contains a literal content-word (*cardboard*), and pertains to a specific property (*secure*) rather than to a generic property *P*. Tactics implement a high-level strategy, but are bound to specific words and derive from attested examples of irony. It is the strategy, when applied to the specific example, that allows the computer to understand the workings of the example and to generate the corresponding tactic. So when used to retrieve phrases from the Google 2-grams, the tactic $T_{cardboard}$(secure) above suggests these other ironic descriptions of a secure location: "cardboard fortress", "cardboard bank", "cardboard jail" and "cardboard prison".

The opposition found in ironic descriptions tends to be pragmatic rather than semantic, and relies on experiential knowledge of the world. This kind of opposition is unlikely to be found in lexico-semantic resources like WordNet. For instance, consider this simile which is marked as ironic:

"about as *sharp* as a rubber carving knife"

Since carving knives are stereotypically sharp, the *rubber* modifier must be subverting this quality to produce an ironic counter-example. Our knowledge of the world tells us that rubber objects are unlikely to be very sharp, and even those that look pointy are probably too flexible to do any real harm. This example provides the following tactic:

T_{rubber}(sharp) ← "rubber" ***@sharp***

Phrases retrieved using CIR for this tactic include: "rubber sword", "rubber shark", "rubber pencil", "rubber arrow", "rubber dart", "rubber tooth" and "rubber dagger". Similar modifiers like "plastic" can subvert multiple properties across different stereotypes. Consider these examples:

"about as *sharp* as a plastic knife"

"about as *pretty* as a plastic flower"

which yield the following two tactics:

$T_{plastic}$(sharp) ← "plastic" ***@sharp***

$T_{plastic}$(pretty) ← "plastic" ***@pretty***

Even in cases such as these, where there is no obvious semantic incongruity between the features of the words in the description and the property that is ironically conveyed, we should expect a stereotypical model of the world to help a computer appreciate the logic of an ironic pairing. Consider another attested simile that is marked as ironic:

"about as *threatening* as a wet puppy"

As humans who have experience of real puppies, we know them to be unthreatening and an unlikely source of fear. Yet we cannot realistically expect our models of the world to enumerate all of the properties that our stereotypes do *not* typically possess. To a computer, there must be some property that puppies *do* typically possess that makes them unthreatening, a property that is even more salient when they are *wet*. Since *threatening* and *harmless* are antonyms, the qualities of stereotypical puppies that make them unthreatening are those that make them more *harmless*, such as being *soft and gentle*. Both *soft* and *gentle* are prominent members of **?***harmless*, and our web-derived knowledge-base contains 16 stereotypes that are both *soft* and *harmless*, and 10 that are both *soft* and *gentle*. The following two tactics can thus be inferred:

$T_{wet+soft}$(threatening) ← "wet" ***@soft*** (.94)

$T_{wet+gentle}$(threatening) ← "wet" ***@gentle*** (.91)

$T_{wet+soft}$(threatening) and $T_{wet+gentle}$(threatening) are specific tactics that implement the generic strategy S_{proxy}. Numbers in parentheses indicate the matching scores of *soft* for **?***harmless* and of *gentle* for **?***harmless* respectively. Since these tactics rely on unsafe inferences (e.g. that soft

things are mostly harmless), the scores provide confidence levels for the corresponding tactics. Phrases retrieved for $T_{wet+soft}$(threatening) include "wet diaper", "wet poodle" and "wet blanket", while $T_{wet+gentle}$(threatening) retrieves phrases such as "wet breeze" and "wet sheep" using CIR.

In this example, *wet* and *threatening* are not antonyms, yet *wet* helps to undermine the potential of the ironic counter-example to pose a threat. Other similes annotated as ironic, such as "about as threatening as a sleeping poodle" and "about as threatening as a dead hamster", allow a computer to infer that *sleeping* and *dead* can be just as effective in subverting the property *threatening*. The tactics that result from these examples retrieve other ironic vehicles, such as "sleeping lamb" and "dead goldfish".

Some properties that are commonly subverted by irony are so generic that they are only noticed when they are absent. Consider the property *useful*: all man-made objects are designed to be useful, but their usefulness derives from a wide-range of other, context-specific properties. When we complete the ironic simile "about as useful as …" we often create mental images of dysfunctional objects that can no longer serve their intended purpose, whatever that may be. Consider this attested example from the web:

"about as *useful* as a foam hammer"

Foam is typically neither useful nor useless. However, hammers need to be *hard* to function well, and foam is stereotypically *soft*, so an opposition between *hard* and *soft* can be detected here. This gives rise to the following tactic:

$T_{foam+hard}$(useful) ← "foam" @*hard*

Hardness does not imply usefulness any more than softness implies uselessness, but the attested example of a foam hammer gives us a reusable pattern from which to build new ironic examples. Phrases retrieved from Google 2-grams for this tactic include "foam wall", "foam knife" and the enigmatic "foam tombstone". We detect a similar opposition in the following example from the web:

"about as *useful* as a rubber crutch"

Crutches are stereotypically *rigid*, while rubber is typically *flexible*. This opposition of *rigid* and *flexible* yields the following tactic for generating uselessly *un*-rigid objects:

$T_{rubber+rigid}$(useful) ← "rubber" @*rigid*

Phrases retrieved by this tactic include "rubber sword" (which is also an ironic counter-example for sharpness), "rubber tripod", "rubber clamp" and "rubber poker".

Conclusions: Computer Says 'No' (Implicitly)

Ironic similes subvert the norms of descriptive language. Norms have highly salient exemplars – shared stereotypes – on which speakers can draw to create a vivid description, but ironic speakers instead construct their own *counter*-examples, often identifying exceptional cases where the standard inferences do not hold. For instance, bank vaults are stereotypically secure, yet a vault made of cardboard is anything but secure. Of course, a cardboard bank vault is a thought experiment of sorts: no sensible person would ever construct such an object to function as a real bank vault, yet the workings of language and thought allow us to at least contemplate the possibility of such an exceptional object (Gendler, 2000). The result may be pragmatically silly on one level – the level of real-world functionality – but it is semantically valid, and has pragmatic value as a conveyer of meaning if not as a container of money. It takes imagination to create these humorous follies, but as with jokes, there are reusable patterns that one can exploit.

For while irony subverts norms, ironic descriptions obey their own higher-level norms for constructing counter-examples. We have used the term *strategy* throughout to refer to these high-level norms of ironic description, and have outlined a variety of strategies – such as S_{group} and S_{proxy} – in this paper. Strategies are generic: they show how one can construct a counter-example for a property *P* from stereotypes for *P* or from antonyms of *P*, or from stereotypes of (the neighbors of) antonyms of P. But for irony to be humorously clever, it must do more than state an obvious negation: it must create the circumstances for this negation to emerge from an unlikely pairing of ideas. Emergent properties are of course difficult to predict, and they emerge from specific contexts, not from generic rules.

One can hone one's sense of irony by deconstructing the irony of others, and specific tactics for implementing a particular strategy for irony can be acquired by observing how others use words to achieve similar ironic ends. In this paper we have provided the technical foundations for uniting these strategic elements into a single computational system. These foundations comprise: a large, nuanced knowledge-representation of word-concepts and their stereotypical properties, acquired from a large-scale analysis of similes on the web; a set of non-literal query operators for retrieving phrases from a large corpus of linguistic readymades (such as the Google n-grams); a large corpus of annotated similes, harvested from the web using the "about as …" pattern; tools for detecting irony in similes harvested from the web; and automatic tools for deriving specific tactics for irony from these attested cases.

The CIR elements of this foundation have been evaluated in Veale (2011/2012). The complementary layers of manually-defined strategies and automatically-extracted tactics have now been implemented upon this foundation, to yield the various tactical patterns and their retrieved instances we have described here. Thus far, yields are low but promising: of the 15,554 *about-as*-similes that are

annotated as ironic in the Veale (2012) data set, just over 10% yield a reusable tactic that can be applied – via creative information retrieval – to an n-gram database to retrieve further ironic suggestions. In all, 1,694 tactics are acquired from the annotated similes, such as the following:

$T_{unweaned+gentle}$(savage) ← "unweaned" *@gentle*

This tactic, derived from the attested ironic simile "about as savage as an unweaned pup", serves a dual purpose: it offers an explanatory interpretation of the irony in similes of this kind, by understanding the vehicle as an exemplar of youthful gentleness rather than of savagery; and it allows a computer to exploit this understanding to compose novel similes of its own, albeit with a similar logic. Nonetheless, the search for readymade phrases that can satisfy a particular tactic can yield some truly unexpected word pairings, as when "foam tombstone" is retrieved as an ironic exemplar of *hard*, or "rubber tripod" for *useful*.

The next steps in this work include an evaluation of the outputs of the system by real users who demand real irony. We shall also need to improve the tactical yield of the system, to extract more tactics from our existing corpus of annotated similes (and thereby identify more of the logical mechanisms/strategies that shape these similes), to acquire more examples of creatively ironic similes from the web, and to extend the reach of the tactics that are successfully extracted. These are related goals. For instance, given that *strong*, *secure* and *durable* are related properties that are mutually reinforcing (e.g. strong containers tend to be secure, secure containers tend to be both durable and strong, and so on), it should be possible to generalize a tactic like $T_{cardboard}$ (secure) to create related tactics such as $T_{cardboard}$ (strong). Indeed, once a modifier like "cardboard" is seen to have tactical value in ironic similes for the property *secure*, our system can then generate targeted web queries using the words in **?***secure*, such as:

"about as *strong* as a cardboard *"
"about as *durable* as a cardboard *"
"about as *reliable* as a cardboard *"
"about as *trustworthy* as a cardboard *"

Since the inspiration for these patterns is an attested ironic simile, we can expect that most matches for these patterns on the web will be ironic also. For instance, we find the following matches for "about as *reliable* as a cardboard *" via the Google API: "cardboard boat", "cardboard bridge", "cardboard shelter", "cardboard oven glove", "cardboard umbrella", "cardboard raft" and "cardboard lifebelt". By using its understanding of known examples to perform a targeted exploration of the web in this fashion, the system can acquire, and actively learn from, a constantly growing corpus of ironic examples, and thereby develop and steadily hone its own ironic sensibility over time.

Acknowledgements

This research was supported by the WCU (World Class University) program under the National Research Foundation of Korea (Ministry of Education, Science and Technology of Korea, Project No: R31-30007).

References

Attardo, S. and Raskin, V. 1991. Script theory revis(it)ed: joke similarity and joke representational model. *Humor: International Journal of Humor Research,* **4**(3):293-347.

Brants, Y. and Franz, A. 2006. Web 1T 5-gram Version 1. *Linguistic Data Consortium*.

Fellbaum, F. (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.

Fludernik, M. 2007. Interfaces of Language: The Case of Irony. In: T. Honegger, E.-M. Orth & S. Schwabe (eds.), *Irony Revisited*, 11-26. Würzburg: Könighausen & Neumann.

Gendler, T. S. 2000. *Thought Experiment: On the Powers and Limits of Imaginary Cases*. Garland Publishing.

Hao, Y. and Veale, Y. 2010. An Ironic Fist in a Velvet Glove: Creative Mis-Representation in the Construction of Ironic Similes. *Minds and Machines*, **20**(4):483-488.

Kahneman, D. 2011. *Thinking Fast and Slow*. Penguin books.

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

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

Veale, T. and Hao, Y. 2007a. Making Lexical Ontologies Functional and Context-Sensitive. *Proceedings of ACL 2007, the 45th Annual Meeting of the Association of Computational Linguistics*, 57–64.

Veale, T. and Hao, Y. 2007b. Comprehending and Generating Apt Metaphors: A Web-driven, Case-based Approach to Figurative Language. *Proceedings of AAAI- 2007*, *the 22nd AAAI conference on Artificial Intelligence.*

Veale, T. 2011. Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity. *Proceedings of ACL'2011, the 49th Annual Meeting of the Association of Computational Linguistics.*

Veale, T. 2012. *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity*. London, UK: Bloomsbury Academic.

Wilson, D., and Sperber, D. 1992. On Verbal Irony. *Lingua* 87: 53-76.