Metaphor as Sign and as Symbol

Tony Veale

Abstract. Metaphors come as second nature to users of language because they are so often the norm. We trade in them deftly, to the point of seeming indifference to, and sometimes even ignorance of, their figurative natures. But the opposite is also true, since words that are offered with the plainest of intentions can be granted a metaphorical significance by those who wish to perceive it. In this paper we contribute to the debate about *deliberate* metaphors by exploring a related concept, the *potential* metaphor. Any text that supports a non-literal interpretation is a potential metaphor, regardless of its author's avowed intentions. We build on this distinction to model the mechanical generation of metaphors as an opportunistic process, whereby potential metaphors are converted into deliberate metaphors. We argue that the distinction between potential and deliberate is mirrored in that between signs and symbols, and demonstrate how this understanding leads to a more nuanced basis for generating and interpreting metaphors on a machine.

Keywords: Signs, Symbols, Deliberate Metaphors, Potential Metaphors

1. A Clash of Signs and Symbols

The psychologist Carl Jung urges us to be wary of the profound differences between signs and symbols, especially as they relate to the interpretation of dream imagery and metaphors. "The sign is always less than the concept it represents", he tells us in (Jung, 1964:55), "while a symbol always stands for something more than its obvious and immediate meaning." Jung uses the notion of "sign" here in its conventional semiotic sense, to denote an accepted placeholder for meaning that obtains its relevance from a network of connections to other signs in an overarching system of signification. We can point to a sign in a dictionary, a taxonomy or an ontology and say with some authority that it means this but not *that*. A symbol, by contrast, is not so easily corralled into a system of mutual discrimination and signification. Rather, cultural symbols possess indefinite halos of connotative association and emotional resonance whose limits are not defined *a priori* but tested and stretched by real-world communication with others. Signs permit an

unquestioning leap from a signifier direct to its signification, while symbols encourage reasoning, inference, and collaborative elaboration in context.

Jung also cautions readers – with a claim that will seem especially apt to modern researchers of Artificial Intelligence – that symbols can never be traded for signs without a loss of meaning potential. He notes, ibid, that "no one can take a more or less rational thought, reached as a logical conclusion or by deliberate intent, and then give it 'symbolic' form." This claim also appears to form the nub of John Searle's *Chinese Room* argument (Searle, 1980), an infamous thought experiment that purports to demonstrate the insufficiency of mechanical reasoning for achieving real understanding. For no matter how intricate a system for manipulating tokens of signification may be, Searle argues that the best we can expect from these manipulations is the mere appearance of understanding, as the manipulating agent itself is never privy to the meaning of the utterances it produces from its own rules. Ironically, Searle refers to these tokens not as "signs" but as "symbols," arguing as he is against the *physical symbol system hypothesis* of Newell and Simon (1976). So while Searle is dismissive of symbol-processing, he uses the term "symbol" in the lesser sense Jung instead reserved for "sign," thus reducing his argument to a critique of mere sign manipulation. We can ask then whether machines are capable of true Jungian symbol processing, of going beyond reductive signs to find the "something more" behind them, and if not, what this might mean for their capacity for produce metaphors?

Notwithstanding Jung's injunction against trading one for the other, we communicate in signs, not symbols. It falls to us as effective speakers and writers to choose and arrange our signs so as to evoke the desired symbols in the minds of an audience. As Raymond Chandler (1944) puts it in *The Simple Art of Murder*, the task of the writer is to pick a path to "what one wants to say" from "what one knows how to say." Yet, as Orwell argues in *Politics and the English Language* (1946), the careful alignment of signs to symbols is a responsibility that many writers fail to uphold. Bemoaning a decline in written English, Orwell frets that "prose consists less and less of words chosen for the sake of their meaning, and more and more of phrases tacked together like the sections of a prefabricated henhouse." His remedy is to look past signs and return symbols to the heart of communication, to "let the meaning choose the word, and not the other way around." To avoid surrendering to signs to soon, "it is better to put off using words as long as possible and get one's meaning as clear as one can through pictures and sensations." Once the signs that best convey a desired symbol are chosen, one can "switch round and decide what impressions one's words are likely to make on another person." Orwell fulminates against calcified metaphors that have become signs without symbols, tokens of signification that can no longer evoke the vivid imagery and feelings they once stirred in audiences. It is here that Orwell's dismal diagnosis and radical prescription confronts our own interest in metaphor production by humans and machines. Where

he sees good reason to "never use a metaphor, simile, or other figure of speech that you are used to seeing in print," we spy an opportunity for the regeneration of symbolic potential in once-stale signs, an opportunity that can by exploited by humans and machines alike.

Orwell's repudiation of jaded metaphors, and his rallying cry for new metaphors to dislodge them from our political discourse, hold up a mirror to the contemporary debate about unthinking and "deliberate" metaphors in cognitive linguistics (see Steen, 2011;2015; Gibbs, 2015). In the following sections we strive to unify these viewpoints, to arrive at a computationally felicitous understanding of metaphor *potential* more generally. We begin in section 2 by reconciling the distinctions of sign/symbol and deliberate/non-deliberate with Bowdle & Gentner (2005)'s view of the career of metaphor. This leads us to consider, in section 3, how two speakers in a metaphorical discourse may operate at differing levels of deliberateness and symbolism, allowing the symbols of one to humorously trump the signs of another. In section 4 we exploit this gap to show how new metaphors can be built from old, striking fresh symbolic sparks from clichés and stale turns of phrase. In section 5 we consider two approaches to the contextualization of metaphor, before exploring, in section 6, an approach to grounding that shows how metaphor can go beyond the realm of arbitrary signs to reference the world outside. Searle considered a machine's inability to ground its abstract signs as a fundamental brake on its ability to grasp the meaning of those signs, but we argue here that a practical grounding is sufficient to achieve human-like creativity when it comes to metaphor generation by machines.

2. Signposting the Career of Metaphor

We resort to a double-standard whenever we rate the freshness of figurative language, for while metaphors are often derided for their age – e.g., Orwell favored the labels "flyblown", "worn-out" and "useless", and pleaded for stale metaphors to be consigned to the writer's "dustbin" – the plain stock of our literal lexicon never seems old no matter how often we draw from it. Indeed, when metaphors die and take their place amongst their literal kin, we no longer deride them for their staleness and age, yet as they near their end they become ever larger targets for criticism. While we use each kind of language to convey meanings, deliberate metaphors carry the additional responsibility of simultaneously proclaiming a speaker's ambition, and it is in this role as a harbinger of creative intent that stale metaphors fall short.

However, it is likely that old metaphors are not just received differently from fresh metaphors, they are processed differently too. As metaphors age, our response to them alters both aesthetically and procedurally. As argued via a hypothesis that is intriguingly named the *career of metaphor*, Bowdle and Gentner (2005) suggest that very different interpretation mechanisms

may be brought to bear on a metaphor to suit our level of familiarity with it. Specifically, the hypothesis makes space for two competing theories of metaphor to work together: the category-inclusion account of Glucksberg (1998, 2001) and the structure-mapping account of Gentner and colleagues (Gentner, 1983; Falkenhainer *et al*. (1989), Gentner *et al*. (1998)). To see how the apparent novelty of a metaphor inevitably dictates how we arrive at an interpretation, consider the old chestnut "my lawyer is a shark." Since the word "shark" has been so liberally applied to people of a ruthless bent it has long since acquired a dictionary sense that captures the uncaring nature of predatory humans. If not already dead, the metaphor is certainly stale and highly conventionalized, and Glucksberg would argue that "shark" is no longer just a signifier of the class of sharp-toothed marine predators but of the category of all things that are unstoppably cruel. It is as a signifier of the latter category that the word is used here, so that "my lawyer" may be newly included among its membership. Yet think back to a time when the metaphor was fresh and the bond between the signifier and this category was not yet set. At a point when the signifier was still alive with symbolic potential, we must surmise that another mechanism allowed us to forge a link from conscienceless lawyers to cruel predation. Gentner and Bowdle argue that this alternate mechanism is analogical reasoning. In comparing the domain of lawyers to the domain of sharks, a reader will spot certain structural similarities between the causal representations of both. Mappings for water (to human affairs, perhaps) and prey (to litigants, or cash cows) and dogged pursuit (to lawsuits, perhaps, as in "I'll see you in court!") are established, and these parallels allow the properties of sharks and their prey to be projected onto the corresponding ideas in the world of jurisprudence.

Bowdle and Gentner claim it is analogy that does the initial spadework to unearth points of overlap between source and target in novel metaphors, long before these insights are eventually stored in a category-level model. But conceptual processes other than analogy are also implicated in the shift. In "meat is murder" we find a metaphor that equates a whole industry with the most grievous of moral transgressions, but this is an equation that relies as much on metonymy as analogy. The analogy suggests parallels between two event-like structures, the industrial process of killing animals (target) and the act of killing a human being (source). While "murder" pinpoints the latter, "meat" offers only an imprecise metonymic pointer to the former. At its most symbolic, the metaphor uses "meat" to condemn not just those who kill animals but those who eat them too, as well as those who fail to object. Yet as the metaphor has become a facile slogan to suggest a moral choice, it has lost its power to shock, move and persuade. The metaphor as a whole has become a sign for a particular lifestyle and social attitude that allows us to make certain assumptions about its bearers. The word "murder" has also undergone a career shift, with its metaphorical uses often carrying a hint of ironic exaggeration, as in "shoe shopping is murder on the feet." So notice

how the career of metaphor also entails a career of metaphorical symbols, where potent symbols in the Jungian sense gradually give way to signs that conveniently allow audiences to converge more rapidly, and with much less divergent inference, from a familiar metaphor to a consensus interpretation. As metaphors age, their interpretation demands less divergent processing of symbols and more convergent processing of signs, to a point at which the use and comprehension of conventionalized metaphors becomes unthinking and far from deliberate (Steen, 2011; 2015).

3. When Symbols Trump Signs

As maturing metaphors become more convergent with age, different people may nonetheless diverge in their individual approaches to interpretation. As the many become numb to the possibilities of a metaphorical conceit, the few may remain alive to its symbolism. Though there is be little benefit in being amongst the latter from the perspective of a metaphor producer, there is a creative advantage to being a consumer of symbols in a world of signs. For in a context where metaphors are used to justify and persuade, there is value is being able to see, as Jung put it, that a calcified sign can "stand for something more than its obvious and immediate meaning." Arguments that involve an exchange of metaphors in a *winner-takes-all* contest of ideas can turn on the rejection of one metaphor in favor of another, so to dismantle an opponent's arguments we must first deliberately dismantle their metaphors.

Argument is, after all, a process of deliberation, and as argued by Veale, Feyaerts and Brône (2006), it pays to view the non-deliberate metaphors of others as quite deliberate when one seeks to obtain a humorous advantage. Those authors define *trumping* as an adversarial use of metaphor in which one speaker's conventional metaphor, cliché or platitude is undermined by bringing a *hyper*-understanding of the metaphor's symbolic origins to bear. For example, on the night when Winston Churchill lost the 1945 election, the consoling platitude "think of it as a blessing in disguise, dear" earned Churchill's wife the rebuke "well it's a bloody good disguise!" Muhammad Ali recounted the tale of being told to "buckle up" when on an airplane as it was about to depart. His metaphorical reply to the stewardess, "Superman don't need no seat-belt!" reflected Ali at his peak, yet he was resoundingly trumped by her riposte, "Superman don't need no airplane neither." In each case the respondent appears to agree with the speaker, and appears to take the speaker's metaphor at face value. Yet by bringing a deliberate analysis to bear, of the metaphor's symbolic potential to evoke a source domain rich in ideas, the respondent succeeds in turning the metaphor on its user. This divergence sits at the very heart of the debate around deliberate metaphors

(Steen, 2011;2015), though it reveals that deliberateness can apply just as much to the interpretation of metaphors as to their production and use.

The deliberateness debate covers much of the same ground as the debate about cliché and stale metaphor that animated Orwell's 1946 essay. Orwell showed all the zeal of the eugenicist in his desire to cleanse English of its "huge dump of worn-out metaphors which have lost all evocative power," because, he believed, "the slovenliness of our language makes it easier for us to have foolish thoughts." Humorous gambits such as trumping certainly punish the slovenliness of lazily-used metaphors, but the humour that arises owes as much to the hackneyed metaphor as to its symbolic reimagination. To claim, as Orwell does, that one should avoid all over-exposed metaphors is to succumb to what Pullum (2008) calls "a load of Orwellian cobblers." Notice, for instance, how Pullum's denunciation of Orwell lends a special resonance to that most overused of dystopian clichés, "Orwellian." Familiar metaphors need not always be used for entirely familiar ends, and the age or familiarity or even conventionality of a metaphor offers only a statistical basis for guessing at the creativity with which it is used. Rather, the true measure of creative intent in a linguistic metaphor is the extent to which a speaker engages with the symbols behind the signs, and forces a listener to engage with them too. As Ricks (1980) argues in his defense of language's journeymen, "Instead of banishing or shunning clichés as malign, haven't we got to meet them, to create benign possibilities for and with them?" Just how benign those possibilities might be will depend on one's point of view, especially if one's goal is humour. The apotheosis of the engagement view is to be found in William Empson's sly repudiation of Orwell's proscriptive agenda. By marrying two of the most jaded metaphors in English to dismiss Orwell as "the eagle eye with the flat feet," he showed that donkeys *can* win derbies, yet only when purposeful riders take the reins (Veale, 2012).

4. Needles in a Metaphor Haystack

We can give deadbeat metaphors a new lease of life by looking for symbols where others see only signs. As a creative strategy, re-analysis applies just as much to the dead and buried as to the jaded and highly conventionalized. Dickens, for example, opens *A Christmas Carol* (1843) with a disquisition on the symbolism of the idiom "dead as a doornail." Finding little in the way of mortality about this kind of nail, he is moved to "regard a coffin nail as the deadest piece of ironmongery in the trade." A dead metaphor begets a living one, even when its symbolic origins are the stuff of just-so stories. Indeed, anything at all that merely resembles a metaphor can beget other metaphors. We define a potential linguistic metaphor as any arrangement of words that supports a metaphorical interpretation, whether or not it has been intended to do so by its author. For if metaphors can be used without

deliberation and profitably analyzed as though they were the products of deliberate creativity, nothing stops us from looking for metaphors wherever signs are used to convey meanings. Seeing potential metaphors anywhere we go permits us to be continuously inspired by the world around us.

Potential metaphors often occupy the uncertain middle ground between a deliberate figurativeness and a presumed literalness. Consider the title of George Lakoff's *Women, Fire and Dangerous Things* (1987). Nothing in this arrangement of words dictates that the title should be interpreted as a metaphor, although the book's subtitle, "What categories reveal about the mind," does suggest that the title also names a category of things. Since fire can certainly qualify as a dangerous thing, our search for a coherent reason for the title encourages us to see women as belonging to this category too. Yet there is no reason in principle why we shouldn't approach this title as we approach C.S. Lewis's *The Lion, The Witch and The Wardrobe*, namely, as a list of intriguing if seemingly mismatched elements that the book will presumably knit together into a single, satisfying narrative. Of course, there will be no doubt in reader's minds that Lakoff's title is intended to be read as a metaphor, since his book later makes this point plain. Yet at issue here is the idea that metaphors are often indeterminate, so that any arrangement of signs that permits a reasonable literal interpretation – such as e.g., this book will talk of women, it will talk of fire, and it will talk of dangerous things – is always just a potential metaphor insofar as it fails to force a non-literal reading. In contrast, the title of Gerald Durrell's *My Family And Other Animals* provides precisely this force, since its use of "other" makes us equate the members of the author's family with the animals in his zoo.

Deliberate metaphors emerge out of potential metaphors because it is the deliberate act of interpreting them as metaphors that makes them so. To build a machine that can generate metaphors of its own, we can either try to understand metaphor production from first principles, so as to build our new metaphors from the conceptual foundations up, or we can choose to build new and deliberate metaphors from the potential metaphors that reside in abundance in any body of text or any source of semiotic stimuli. For our current purposes, this abundance is provided by the Google n-grams (Brants and Franz, 2006), a vast database of snippets of web texts of between 1 and 5 tokens in length; an *n*-gram is any contiguous sequence of *n* tokens – words, numbers or punctuation marks – that is found in a text. Consider, for example, the 3-gram "romance and insanity" to which Google assigns a web frequency of 313 documents. Read as a simple coordination structure the phrase says as little as "women and fire" or "lion and witch". Pragmatically, however, readers will ponder the reasons for squeezing two ideas of opposing sentiment into a single unit, and may seek to unearth a figurative kinship that links the two. With enough knowledge of words and the world at their disposal, readers may beat a path from madness to love, and one might even repackage the resulting kinship in the following way:

It used to be that romances were enjoyed by beloved lovers. Now I say unto you that romance is insanity from which only hateful fanatics suffer.

While this framing may speak to any reader who has experienced the highs and lows of romantic attachment, it was in fact produced by a machine that draws on a rich inventory of stereotypes (e.g., of lovers and fanatics) and of familiar situations (e.g., falling in love, suffering from an affliction). These archetypes allow it to find myriad symbolic connections behind the signs "romance" and "insanity." The use of language associated with religious discourse (e.g., "I say unto you") is just one of many strategies the machine uses to package its mechanical insights as deliberate metaphors and thereby provoke an emotional and intellectual response in readers. The machine is named *MetaphorMagnet*, and the results of its deliberations can be sampled on the hour in the tweets of its Twitterbot incarnation, *@MetaphorMagnet*.

The name *MetaphorMagnet* derives from the machine's main generation strategy: searching for potential metaphors in web n-grams with the copula form "A is B" or "A is a B" and the coordinated form "A and B." It always attempts to interpret these forms as deliberate metaphors, using the specific associations of A and B to explore how B might partially mirror A. While most n-grams satisfying these forms will yield nothing, the database of n-grams is so large that millions of new metaphors will still be generated. It hardly matters if the n-grams that do yield successful analyses were never meant to understood as metaphors, since the strategy looks to the n-grams for linguistic inspiration, not corpus evidence. Inevitably, our magnet will extract needles from its haystack that turn out to be the verbal expression of metaphor schemas in the mold of Lakoff and Johnson (1980), such as "Life is a Journey," "Time is Money" and "Argument is War" to say nothing of famous lines from poetry and song such as "love is a battlefield." These n-grams are interpreted using exactly the same symbolic search mechanisms as "meat is murder" and "love is a game" and "romance and insanity."

Metaphor Magnet relies on a logical representation of norms and beliefs that old-school AI researchers describe as "symbolic AI" (e.g., Newell and Simon, 1976). This serves as a logical basis for most computational models of analogy or metaphor, from Falkenhainer *et al.* (1989) to Hofstadter *et al*. (1995) to Veale and O'Donoghue (2000) to Barnden (2006). For instance, the propositions *lovers enjoy romances*, *lovers are beloved*, *romances are sweet*, *fanatics suffer-from insanity* and *fanatics are hateful* are just some of the many propositions that Metaphor Magnet can reach from the signifiers "romance" and "insanity" in its divergent search to unify these two terms. To imbue its deliberate metaphors with semantic tension, Metaphor Magnet is especially drawn to semantic oppositions, such as *hateful* versus *beloved*. In this case it hypotheses that *if* romance were truly a kind of insanity, then lovers would be more like fanatics, so that lovers would be more hateful

and fanatics would be more beloved. The machine knows nothing of how jealousy and rejection can sour love into hate, or of how fanatical zeal can inspire love for those who would kill what they hate. It is enough that it can frame the opposition in such a fashion as to nudge human readers to bring their own experiences of the world to bear on the final textual rendering.

Metaphor Magnet's forced analysis of potential metaphors as deliberate metaphors is only as insightful as the representations that underpin it, so it is convenient that figurative language often gives back as much as it takes. As shown in Veale (2012), the machine acquires many of its associations and norms from the low-hanging figurative fruit of the web, such as similes with explicit ground terms. For example, when the query "as hateful as a" is posed to the Google search engine, the retrieved texts offer "toad", "mass shooting" and "Russian autocracy" as simile completions. Automating this process so that it can operate on a massive scale, Metaphor Magnet obtains a wealth of symbolic associations that it would never find in a dictionary.

5. Metaphor In The Moment

The three challenges faced by any user of metaphor are, broadly speaking, knowing *what* to say, knowing *how* to say it, and knowing *when* to say it. As discussed earlier, Raymond Chandler saw writers as explorers who must go from the *what* to the *how* in as clean and elegant a manner as possible, and Metaphor Magnet conducts a similar search using mechanical means. Yet the machine still falls short with regard to the *when*. Metaphor Magnet generates and stores millions of vivid metaphors during its sweeping passes over the Google n-grams, so that it can later pluck any of these metaphors from its database whenever it is called upon to serve up a new metaphor. Its incarnation as a Twitterbot, @MetaphorMagnet, avails of this abundance to tweet a randomly plucked metaphor every hour on the hour, and while each is well-formed and internally meaningful, and perhaps thought-provoking too, most fail to speak to the moment. @MetaphorMagnet's metaphors are deliberate in construction but they are very far from deliberate in delivery.

Most metaphors are created as a response to what others have said or done, so a maker of metaphors that lacks a model of context might still rely on an interlocutor to tacitly provide some context by what it says or does. For this reason we have built an interactive version of Metaphor Magnet, as a public web service that both humans and machines can avail of. Humans use a web browser[1] to query the service and to interact with its suggestions, while machines do much the same without the intermediary of a browser, obtaining XML-structured data directly. Let's consider the service from the perspective of a human interlocutor. A user can enter individual terms, such

[1] The website may be accessed via the URL http://ngrams.ucd.ie/metaphor-magnet-acl/

as “love” or “life” or “war” or “religion”, and receive a variety of related metaphors in return. Entering the term “politics”, for instance, the user is presented with a flurry of related metaphors that the system has acquired or invented during its sweep through the Google n-grams. These range from “challenging sport” to “convoluted mess” and include many more besides. For single terms such as “politics” – these are assumed to designate a target concept – the system first examines any copula metaphors in the n-grams of the form “politics is [a/an] X” such as “politics is myth” (frequency 3948), “politics is war” (freq. 867), “politics is religion” (freq. 116), “politics is rubbish” (freq. 96), “politics is poison” (freq. 47), “politics is a joke” (freq. 99) and “politics is a disease” (freq. 46). Using its symbolic models of the stereotypical qualities of those source concepts, the service composes new metaphors of its own that accentuate many of the same patterns of qualities in the target. Users are thus presented with a mix of the commonplace and the novel, and can choose to further explore any that pique their interest.

In cases where users enter copula metaphors of their own – such as “life is a game” – the service proceeds in the same way, though it promotes the constructed metaphors that focalize many of the same aspects of the target as the user’s own input. A promoted result is granted a greater presence on screen. Consider the case of “meat is murder,” which prompts the service to offer up the related metaphors “burning sin,” “threatening virus,” “perilous evil” and “alarming outbreak” in response to the user’s metaphor. If a user now clicks on “burning sin” to drill down into this possibility – it is, after all, the closest to the user’s own, and the one that offers the clearest moral judgment – the service will display its own interpretation of this metaphor as the set of qualities it believes will be focalized in the target. Metaphors are easiest to appreciate in the flesh, and the system’s own constructions are easiest to appreciate when they are rendered in polished linguistic forms. Since the service cannot squeeze the full potential of a metaphor into a single sentence, it offers the user a machine-generated poem instead. Poetry serves many purposes, but the service uses it as a summarization device, to distill the many rendering possibilities of a metaphor in a coherent form. So unlike most machine-generated poetry, which emphasizes meter and rhyme over meaning and symbolism, Metaphor Magnet’s poems are generated in blank verse so as to showcase all of the rendering strategies at its disposal. Users need only click a link to have any metaphor woven into a poem. The following is one such poem, as generated for “meat is a burning sin”:

No Sin Burns More Terribly

Terrify me with your edible fruit
By perverted religions are meats punished,
    and meaty love do these religions promote
The most despicable racist is not more blatantly criminal

Reward me with the perishable fruit of your sin
Let your tempting meat excite me
Did ever a crime cause a more terribly heinous sin?
How you pollute me so terribly, like an ugly sin
Does any sin burn more terribly than this meat?
You degrade me with your terrible immorality
Organized sins do disorganized religions sometimes proscribe
Even if you were a feared criminal wouldn't you want to
commit this sin of beloved angels?
O Meat, you menace me with your criminal seduction

The poem draws together diverse strands of the same metaphor (i.e., meat is a burning sin) and related metaphors that focalize the same qualities (e.g. meat is forbidden fruit). We see metaphors of crime, pollution and religion in the lines above, framed as similes, superlatives and rhetorical questions. Notice how this poem, and others like it[2], are tacitly influenced by context in two different ways. Firstly, and principally, the poem is a response to the user's initial metaphor, and is thus systematically shaped by that stimulus. But secondly, and more subtly, each of the various metaphors that comprise the poem offers a tacit context for the interpretation of each of the others.

5.1 Metaphors In The News

Metaphors allow us to speak about events in ways that significantly depart from the norm, and so, when responding to the same event, we can expect a deliberate metaphor and a literal or conventional exposition to share a deep but not a superficial similarity. While each may speak to many of the same underlying topics using different words, statistical topic modeling allows a machine to identify the latent topics that are present in a text and thereby quantify the deep similarity between a metaphor and the literal rendering of an event. In the case of news events and their more-or-less literal headlines, a context-sensitive Metaphor Magnet can rank and choose apt metaphors using topical similarity to incoming headlines in an evolving news context. Latent Topic Analysis, or LDA (Blei *et al*. 2003), builds a fixed number of topics that best explain the observable similarity between texts in a data set. Each of the resulting topics is a bin of weighted words, so that the texts that contain these words can be assigned a probability of having been generated by the corresponding topic. Each text is thus assigned a graded membership in every topic, and the array of membership probabilities across all topics for a given text serves as a compressed representation that allows similarity judgments to be made on the basis of simple vector algebra.

[2] Generate a new variant of the poem for yourself at: *http://ngrams.ucd.ie/metaphor-magnet-acl/p?source=burning:sin&target=meat*

An LDA model corresponds to a highly compressed semantic space of fixed dimensionality – n topics yields n dimensions – in which each text is represented by an n-dimensional vector of topic probabilities for all topics. The angle between any two vectors is indicative of the differences between vectors and thus of the differences between texts. The cosine of this angle is a useful measure of the deep similarity of two texts since the cosine of 0 degrees (no angle, so a tight fit) is 1 and the cosine of 180 degrees (the largest possible angle between vectors) is -1. We assume this similarity to be a *deep* similarity since the compression of a nuanced data set into, say, a space of 150 dimensions necessarily entails a high degree of generalization. This allows an LDA model to take texts that speak to the same topics using different words and to map them into the same locality in a space. Since we wish to measure the angle between the LDA vectors of news headlines and the vectors of metaphors, a single topic model must be constructed for a composite dataset that unites a large corpus of news headlines with a large collection of metaphors. For the former we harvest several years of news stories from the websites of mainstream media outlets such as CNN; for the latter we ask Metaphor Magnet to generate 10 million metaphors from the Google n-grams. A model with 150 topics – which represents an empirical trade-off between generality and specificity – is built from the joint dataset.

Each of these 10 million metaphors is assigned a 150-dimension vector. These serve as the fixed stars in the model's semantic firmament, to which the newly arriving headlines of breaking news stories can be compared. So as headlines arrive over Twitter, from *@CNNbrk*, *@BBCbreaking*, *@WSJ*, *@FOXnews*, *@Reuters, @nytimesworld*, and @AP, they are mapped onto topic vectors by the model and compared to vectors of known metaphors to identify the nearest matches. The model is recomputed at regular intervals to incorporate new stories, so it can keep pace with the cultural Zeitgeist. For instance, the *#MeToo* movement is at present dominating the news, with nightly reports of sexual harassment in workplaces as diverse as the media, film and TV, and government. One story of peak interest concerns senate candidate Roy Moore, who stands accused by multiple women of predatory behavior against adolescent girls. Though the offenses date back many years, the *#MeToo* movement has given them a new public airing. A news-sensitive version of Metaphor Magnet, in the guise of a Twitterbot named @MetaphorMirror, responds in the following way to this headline (see Veale *et al*., 2017):

@Buzzfeed: "*He did not perpetrate sexual misconduct with me...but I now know for sure he is a liar," said a woman who claims she dated Roy Moore when she was 17 and he was 34*

To some oppressors, every victim is a harmless child.
To others, every victim is an overwhelming fire.

The LDA-mediated mapping between headline and metaphor often seems gnostic in its imagery. What, for instance, is the symbolism of *fire* above? Does it refer to the tide of anger that has swept the public sphere since the shocking revelations concerning Hollywood producer Harvey Weinstein? Or does it represent the righteous fire of judgment and condemnation? The model refuses to tell, limited as it is to its 150 numeric dimensions. We are on somewhat surer ground however in seeing the relevance of "oppression" "victim" and "child", since the unfolding Moore saga has given the model a diversity of headlines that nudge new Moore headlines into this dark corner of the LDA space. So far from diminishing the allusive charm of metaphor, the use of a reductive mathematical model actually heightens the mystery of interpretation, encouraging audiences to bring their own insights to bear when exploring the symbolism of even a machine-crafted metaphor.

6. Metaphors on the Ground

As much as we need to heed Jung's distinction of symbols versus signs, our computer implementations tend to view these ideas as interchangeable. For instance, it is convenient to view signs as *lexical* tokens that correspond to words in a natural language like English, and to view symbols as *semantic* signs that derive their usefulness from their logical connections to others. So we do exactly what Jung cautions against, and trade symbols for signs in our representations. Searle's arguments about the insufficiency of syntactic manipulation of tokens of any kind still holds water unless we can show how our tokens can relate to the world outside their closed logical systems. To do this, we must call on signs that already possess an external reference.

Symbols in the Jungian sense are richly evocative: they both denote and connote. Consider just one connotative dimension of a symbol, its ability to vividly suggest a colour in the mind's eye. A red rose, blood and gore each suggests a similar shade of red, but in doing so each also brings its own aesthetic and affective overtones to our appreciation of the colour. The red of roses is different than the red of wine or the red of a cardinal's biretta, and so, conversely, a specific colour tone may be suggestive of the symbol with which it is associated in our imaginations. So we should do more than associate the symbols *rose*, *blood*, *wine*, *cardinal* or even *anger* with a sign like *red*: we should use colour symbols that are richly evocative of the right shades and tints of the colour red. We can, for instance, use the RGB colour coding system of a TV or computer monitor. With a numeric value for each of the three additive colour components Red, Green and Blue, RGB codes can specify millions of colors and shades with a six-digit hexadecimal code such a D43800 (the red of paprika) or F6E7B8 (the yellow of parmesan). It is true that we are trading one kind of sign for another here, for what else is

a token like D43800 but a sign that only makes sense within a code system? However, these tokens have purchase in the external world that other signs do not: they can be used to paint the associated colours on a screen, or they can map the colours in a camera image to the relevant signs and symbols. They link the symbols behind our words to the world of human perception, and in this sense they can be said to provide a grounding for our symbols.

This being the case, we can expect a metaphor maker that is grounded in this way to generate more compelling colour metaphors than one that is not. Consider the challenge of naming a specific colour with a vivid metaphor. Paint makers rely on colour metaphors (their catalogues are full of them) to sell paints that, in perceptual terms, have the same shades as a competitor's. Yet can a metaphor machine use RGB-grounded symbols to create colour metaphors that are deemed to be just as creative and descriptively apt as those created by an embodied human? This is a task we set ourselves here. Metaphor Magnet will seek out potential colour metaphors in the Google n-grams and transform these, when possible, into deliberate colour metaphors that can be compared in a blind test with the products of human ingenuity. To ground the system's symbols, we first build a colour lexicon that maps from lexical signs, such as "rose," to the signs for the corresponding colour term, such as "red", as well as to the most conventional RGB code, such as F19CBB (rose-red). We begin by collating n-grams of the form "X-colour" such as "wine-red" and "sky-blue" in which an archetype is used to suggest a specific colour. We then use the web-site ColourLovers.com to find the most appropriate RGB code in each case. The resulting lexicon provides over 1000 mappings of archetypes to colours and RGB colour codes.

We consider the lexical signs associated with RGB codes in this way to be the names of colour archetypes. To identify potential colour metaphors in an n-grams database, we can simply harvest all 2-grams XY where both X and Y name a colour archetype, such as "chocolate sky" or "paper tiger." While past users of "paper tiger" may not have used it to name a colour, it has the potential to be used in this way, perhaps to name a blend of paper-white and tiger-orange. Likewise, "chocolate sky" can be taken as the name of a blue-brown blend of chocolate-brown and sky-blue. We are guided by n-gram frequency when harvesting potential metaphors, since this is a good indicator of phrasal well-formedness. While "paper tiger" has a frequency of 25,690 in Google's database, "tiger paper" has a frequency of just 100. As we cannot be sure that any given 2-gram is a valid English noun-phrase, we accept "tiger paper" and reject its less frequent inverse, "tiger paper."

To turn these objets trouvés into colour metaphors, we must assign an RGB code for a specific colour to each phrasal form. But what is the colour of a "paper tiger," a "midnight sun" or an "alien brain"? We make a rather simplifying assumption and calculate the midpoint in RGB space between each of the two component colours, to yield a 50:50 colour blend. Thus, "banana curry" denotes a mix of 50% banana-yellow (FFE135) and 50%

curry-brown (DA9E19). The resulting blend in such cases is identifiably similar to each of its colour ingredients, yet in other cases the blend is not recognizable as a variant of either. We thus impose an additional constraint. By adopting an analogical colour scheme (see Pentak, 2010), in which both component colours and their 50:50 blend must reside in somewhat adjacent areas on the colour wheel, we ensure that the blend is an intuitive one. It is worth noting that analogical colour schemes owe their name to the colour changes that one can observe in nature. For instance, when leaves change colour in Autumn, their hues slide across adjacent areas of the RGB wheel. We now have a large trove of colour metaphors with specific RGB codes.

To evaluate these machine-generated colours, we compared them (see Veale and Alnajjar, 2016) to colour names chosen by humans for much the same RGB codes. The website *ColourLovers.com* is frequented by humans who live up to the site's name, and who suggest their own names for specific RGB values. Other users are invited to comment on these names, and to vote for them too (the votes are called "loves", by analogy to Facebook's "likes"). We downloaded the top 100,000 colour codes from the site[3], ranked from most to least *loves*; the mean number of *loves* per colour code is 13, while each code has at least one *love* and just one human-assigned name since the site does not permit multiple names for the same RGB code. For the purposes of a comparison we need to ensure alignment between the named codes on the website and the named codes generated by Metaphor Magnet. We mapped each RGB code into the CIE LAB colour space, as distances between points in this space are more intuitively mapped to differences in colour as perceived by humans. We then used the *Delta E CIE76* distance function to measure the Euclidean distance between two colours in this space. When comparing a colour code generated by Metaphor Magnet to a colour code from ColourLovers.com, we allowed the colours to differ by as much as 14 according to our distance function and still be considered the same colour perceptually. We chose the threshold 14 empirically, after experiments on 141 named HTML colours.

We chose 2587 of ColourLover.com's named colours for a comparison. The mean number of "loves" for this name/colour set is 2.188. Each colour can be aligned to one generated by Metaphor Magnet within the tolerances of the distance function, allowing us to present both names (machine versus human) to human subjects in an empirical test. We used the crowd-sourcing platform *CrowdFlower.com* for our experiments. A swatch of each colour and a choice of names, one human-generated *and* one machine-generated, was put before human judges, who were asked to take a moment to imagine the colour being used. The ordering of the names was randomly decided on a case-by-case basis, so that the human-generated name was listed first in 50% of cases, and the machine-generated name was listed first in the other

[3] http://www.colourlovers.com/colors/most-loved/all-time/meta

50% of cases. In all cases, judges were *not* told of the origin of any name. Each judge was paid a small sum to answer the following 4 questions:

1. *Which name is more descriptive of the colour shown?*
2. *Which name do you prefer for this colour?*
3. Which name seems the most creative for this colour?
4. *Why did you answer these questions they way you did?*

The fourth question is a source of qualitative data that may later yield insights into the factors that shape the appreciation of colour metaphors. Judges were timed on their responses, and those that spent less than 10 seconds presenting their answers for any colour were classed as *scammers* and dismissed. We also required that each question be answered by 5 non-scamming judges to be trusted, and in this way we obtained 12,608 trusted judgments for evaluation. Another 5,040 untrusted judgments were ignored.

The experiment was terminated after its budget of $220 was exhausted, at which point 940 judges had been paid to contribute to the task and 1578 of the 2587 colours had received five trusted judgments for each question. It is on the judgments for these 1578 colours that we based our evaluation. Tallying individual judgments per question, we see that 70.4% for *most descriptive name* (Q1) favored the machine; that 70.2% of judgments for *most preferred name* (Q2) favored the machine; and 69.1% of judgments for *most creative name* (Q3) favored the machine. Similarly, when we tally the majority judgment for each question under each colour – the choice picked by three or more judges – we see that for just 354 (23%) of the 1578 colours, a majority of judges deemed the human-assigned name to be more descriptive than that assigned by the machine. The results for the next two questions, Q2: *which name do you prefer?* and Q3:*which name is most creative?*, are in line with those of the first question. Only for 355 colours does a majority of the five human judges for a given colour prefer the human-assigned name over that assigned by the machine, and only for 357 colours does a majority of judges consider the human-assigned name to be more creative than the machine-assigned name. This consistent breakdown of approx. 3-to-1 in favour of the machine suggests that machine-generated colour metaphors can be more than competitive with human creations when the machine grounds its symbols in ways that we humans take for granted.

7. Summary and Conclusions

While we communicate with discrete signs, our most effective metaphors – and certainly our most deliberate – exploit cultural symbols with diffuse halos of shared sentiment and connotation. Symbols add to the elasticity of our interactions because there is no obvious limit to their evocative power.

In this paper we have presented a variety of systems and implementations that attempt to capture the resonance of symbols in a machine dedicated to the task of deliberate metaphor generation. Sitting at the head of this family of systems is Metaphor Magnet, a public web service and a Twitter *bot* that pursues an opportunistic approach to metaphor generation. Distinguishing between potential and deliberate metaphors, our metaphor machine trawls a very large corpus of textual scraps for the web equivalent of *objets trouvés* – phrases with the symbolic potential to be deliberately read as metaphors. As described in detail in Veale (2015), the system prizes phrases that evoke a tension between different cultural norms or stereotypical expectations. Its interpretations of the metaphors suggested by these phrases are packaged in a diversity of ways that further exploit the conventions of topic and genre. But while a single packaging strategy reveals just one facet of a metaphor, a poem allows a metaphor system to compress many mutually-reinforcing renderings and perspectives into a single generative artefact. As described in Veale (2013), a functional poem is an internally consistent structure that compresses and suggestively summarizes a space of figurative possibilities. As shown here, we view these machine-generated poems not as aesthetic artefacts in their own right, but as convenient viewfinders akin to linguistic kaleidoscopes. Users peer, and perhaps ponder, before twisting the lens to generate another swirl of words and another batch of interlinked metaphors.

Yet for all that, readers have some justification for thinking that we only pay lip service to Jung's distinction between signs and symbols. Where we promised symbols we have offered old-school AI representations composed entirely of, well, more signs. Searle's Chinese Room argument has not been vanquished by a use of terminology that, while useful from a philosophical and design perspective, is much more aspirational than it is accomplished. Searle asserts that computationalists can never escape the world of signs, no matter what linguistic contortions they might attempt. Yet a way out of the Chinese room is suggested by a pair of Metaphor Magnet variations that we have discussed here. The first is the use of signs with external reference, such as RGB codes that can be used to create and analyze real images and to mediate between those images and the linguistic processes of metaphor. These codes are signs, yet each has the capacity to be turned into something more, a vivid colour with its own visual and emotional resonances. When a machine builds its metaphors on a foundation of "grounded" signs, they too can evoke the same vividness and resonance as human-crafted metaphors.

We believe that the topic-modeling approach behind *@MetaphorMirror* and its timely mapping of metaphors to news headlines has equal promise as a way out of Searle's Chinese room. Metaphor has always been viewed as a reconciliation of two distinct realms of experience, and when viewed from the perspective of statistical topic-modeling, we can see that those realms differ in more than mere content; they also differ in the number and meaning of the dimensions that structure them. The LDA approach we have

presented here takes both of these realms, represented using signs, and maps each into the same mathematical space so that the same dimensions are used to characterize them both. While this third space functions rather like the *blend* space of Fauconnier and Turner's (2002) conceptual blending theory, it is ultimately built from continuous numbers rather than discrete signs, much like the vector-space models of Kintsch (2000). It is capable of subtleties that a traditional AI representation is not. Moreover, if we rebuild the topic model at regular intervals, it can adapt to nuances in the news cycle in ways that only reveal themselves in the shrewdness of its mappings. That these nuances resist explicit codification as a system of discrete signs is precisely the point of Jung's sign vs. symbol distinction.

When *@BuzzFeedNews* tweets the four-word headline "Mood going into 2018" with a link to a skater tumbling on the ice, what are we to make of those four words? Ideally, our computational model should incorporate information from this video too, yet regardless of its content, we can be sure that the word "mood" means much more here than its dictionary entry. Rather, it represents the fractious political and cultural divisions that are so often mentioned in the same context as "the public mood" in news stories. We conclude this paper then by revealing @MetaphorMirror's implicit sense of this "mood", as encoded in the dimensions of its topic model:

Enemies nurture hate.
Heroes inspire the love that creates the jealousy that inspires hate.
Who is worse?

7. References

Barnden, John A.
2006 Artificial Intelligence, figurative language and cognitive linguistics. In: G. Kristiansen, M. Achard, R. Dirven, and F. J. Ruiz de Mendoza Ibanez (Eds.), *Cognitive Linguistics: Current Application and Future Perspectives*, 431-459. Berlin: Mouton.

Blei, David. M., Ng, Andrew. Y. and Michael I. Jordan
2003 Latent Dirichlet Allocation. J. of Mach. Lear. Res. 3:993–1022.

Bowdle, Brian F. and Dedre Gentner
2005 The Career of Metaphor. *Psychological Review*, 112(1):193-216.

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

Chandler, Raymond
1944 The Simple Art Of Murder. *The Atlantic Monthly*, Dec. issue.

Dickens, Charles
1843 *A Christmas Carol*. Middlesex, UK: Puffin Books (1984 reprint).

Falkenhainer, Brian, Forbus, Kenneth D. and Dedre Gentner.
1989 Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, **41**:1-63.

Fauconnier, Gilles and Mark Turner
2002 *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.
Gentner, Dedre
1983 Structure-mapping: A Theoretical Framework. *Cognitive Science* 7(2):155–170.
Gentner, Dedre, Falkenhainer, Brian and Janet Skorstad
1989 Metaphor: The Good, The Bad and the Ugly. In Y. Wilks (ed.), *Theoretical Issues in Natural Language Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Gibbs, Raymond W.
2015 Does deliberate metaphor theory have a future? *Journal of Pragmatics*, vol. 90, December 2015, pp 73–76.
Glucksberg, Sam
1998 Understanding metaphors. Curr. Dir. Psychol. Sci. 7, 39--43.
2001 *Understanding Figurative Language: From Metaphors to Idioms*. Oxford University Press.
Hofstadter, Douglas R. and the Fluid Analogies Research Group
1995 *Fluid Concepts and Creative Analogies. Computer Models of the Fundamental Mechanisms of Thought*. Basic Books.
Jung, Carl G. (with M-L von Franz, J. L. Henderson, J. Jacobi and A. Jaffe).
1964 Man and his Symbols. Ferguson Press.
Kintsch, Walter
2000 Metaphor comprehension: A computational theory. Psychonomic Bulletin Review, 7(2):257-266.
Lakoff, George and Mark Johnson
1980 *Metaphors We Live By*. University of Chicago Press.
Lakoff, George
1987 Women, Fire, and Dangerous Things. University of Chicago Press.
Newell, Allen and Simon, Herbert. A.
1976 Computer Science as Empirical Inquiry. Symbols and Search. *Communications of the ACM, **19**(3): 113–126.*
Orwell, George
1946 Politics and the English language. *Horizon* 13(76), April issue.
Pullum, Geoff
2008 A load of old Orwellian cobblers from Fisk. *Language Log,* August 31. http://languagelog.ldc.upenn.edu/nll/?p=551
Ricks, Christopher B.
1980 Clichés. In Leonard Michaels and Christopher B. Ricks (Eds.), *The State of the Language*. University of California Press.
Searle, John R.
1980 Minds, Brains and Programs. Behavioral and Brain Sciences 3(3): 417–57.
Steen, Gerard
2011 The contemporary theory of metaphor -- now new and improved! *Review of Cognitive Linguistics*, vol. 9, 26–64.
2015 Developing, testing and interpreting deliberate metaphor theory. *Journal of Pragmatics*, vol. 90, December 2015, pp 67–72.

Veale, Tony and Diarmuid O'Donoghue
2000 Cognitive Linguistics 11(3/4): 253-281 .
Veale, Tony, Feyaerts, Kurt and Geert Brône
2006 The Cognitive Mechanisms of Adversarial Humor. Humor: The International Journal of Humor Research, 19-3:305-338.
Veale, Tony
2012 *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity*. London, UK: Bloomsbury Academic.
2013 Less Rhyme, More Reason: Knowledge-based Poetry Generation with Feeling, Insight and Wit. *Proc. ICCC 2013, the 4th Int. Conf. on Computational Creativity. Sydney, Australia, June.*
2015 Unnatural Selection: Seeing Human Intelligence in Artificial Creations. *Journal of General Artificial Intelligence*, 6(1):5-20.
Veale, Tony and Khalid Alnajjar
2016 Grounded for life: creative symbol-grounding for lexical invention. *Connection Science*, 28(2):139-154.
Veale, Tony, Chen, Hanyang and Guofu Li.
2017 I Read The News Today, Oh Boy! Making Metaphors Topical, Timely and Humorously Personal. Proc. of *HCII 2017 Distributed, Ambient and Pervasive Interactions, Vancouver, Canada.*