Metaphor in the Age of Mechanical Production

(Or: Turning Potential Metaphors into Deliberate Metaphors)

Tony Veale

School of Computer Science,

University College Dublin,

Ireland.

Tony.Veale@gmail.com

@MetaphorMagnet

Abstract

Linguistic metaphors are most naturally viewed as the output of a language generation process, and as the input to a language understanding process. But it is just as meaningful to view the conceptual metaphors that underpin these linguistic forms as an *input* to the generation process and an *output* of the understanding process. A large repository of existing linguistic metaphors, such as a text corpus or a database of Web n-grams, can thus be viewed as an implicit source of the knowledge an agent needs to generate and understand novel or unseen linguistic metaphors. If one uses Web data as a knowledge resource for metaphor, it also makes sense to think of the algorithms and tools for manipulating this knowledge as Web services that can be called upon to generate and understand linguistic metaphors on demand. This paper argues that *potential* metaphors in Web n-grams can be used as a resource for understanding and generating novel *deliberate* metaphors. A Web service that provides this functionality on-demand is also described, allowing 3rd-party applications to exhibit a measure of their own figurative creativity.

Keywords: *Metaphor, Affect, Polarity, Lexicons, Stereotypes, Web services*

1. Introduction

Picasso famously claimed that “art is a lie that tells the truth.” Fittingly, this artful contradiction suggests a compelling reason for why speakers are so wont to use artfully suggestive forms of creative language – such as metaphor and irony – when less ambiguous and more direct forms are available. While literal language commits a speaker to a tightly fixed meaning, and offers little leeway for the listener to contribute to the joint construction of meaning, metaphorical language suggests a looser but potentially richer meaning that is amenable to collaborative elaboration by each participant in a conversation. In Picasso’s terms, a metaphor is an artifice that may be literally false (a “lie”) but it is one that can better facilitate our access to knowledge.

A metaphor *X is Y* establishes a *conceptual pact* between speaker and listener (Brennan and Clark, 1996), one that says ‘let us agree to speak of X using the language and norms of Y’ (Hanks, 2006). Suppose a speaker asserts that “*X* is a snake”. Here, the stereotype *Snake* conveys the speaker’s negative stance toward *X*, and suggests a range of talking points for *X*, such as that *X* is *charming* and *clever* but also *dangerous*, and is not to be *trusted* (Lakoff, 1987; Veale and Hao, 2008). A listener may now respond by elaborating the metaphor, even when disagreeing with the basic conceit, as in “I agree that X can be charming, but I see no reason to distrust him”. Successive elaboration thus allows the speaker and listener to arrive at a mutually acceptable construal of a metaphorical *Snake* in the context of *X*. So metaphors are flexible conceits that allow one to express a position while seeking elaboration or refutation of this position from others. Our computational models for the interpretation and elaboration of metaphors should allow speakers to exploit the same flexibility of expression when interacting with machines as they enjoy with other humans. Such a goal clearly requires a great deal of knowledge, since metaphor is a knowledge-hungry mechanism *par excellence* (Fass, 1997). However, much of the knowledge requirered for metaphor interpretation is already implicit in the large body of metaphors that are active in a community (see Martin, 1990; Mason, 2004). Existing metaphors are themselves a valuable source of knowledge for the production of new metaphors, so much so that an agent can acquire the relevant knowledge from corpora of figurative texts (Shutova, 2010; Veale, 2011).

Metaphors achieve a balance of suggestiveness and concision through the use of *dense descriptors*, familiar terms like "snake" that evoke a dense body of shared knowledge of stereotypical properties and behaviors (Fishelov, 1992). Though every concept has the potential to be used figuratively, casual metaphors tend to draw their dense descriptors from a large pool of familiar stereotypes shared by all speakers of a language (Taylor, 1954). A rich conceptual model of the lexicon is needed to allow the figurative import of these stereotypes to be inferred as needed in context. In this paper we show how a large lexicon of affective stereotypes can be mined from Web content, and further, how affective representations can be used selectively, to metaphorically highlight aspects of a given target concept in a specific metaphor. Because so many familiar stereotypes have polarizing features – think of the endearing and not-so-endearing features of babies, for instance – metaphors are ideal vehicles for conveying an affective stance toward a topic. Even stereotypes that are not used figuratively, as in the claim "Steve Jobs was a great *leader*", are likely to elicit deliberate metaphors (in the sense of Steen 2011, 2015) in response, such as "yes, a pioneer" or "such an artist!" or even "but what a tyrant!". Proper-names can also serve as dense descriptors, as when Steve Jobs is compared to the fictional inventor *Tony Stark*, or Apple is compared to *Scientology*, or Google to *Microsoft*. Our affective lexicon thus needs to be a dynamic lexicon, capable of building dense stereotype representations whenever they are needed.

In many ways, a metaphor resembles a *query* in that staple of the information age, *Information Retrieval* (IR). Metaphors, like queries, allow us to simultaneously express what we believe and to elicit further information that may bolster or refute our beliefs. Metaphors, like queries, are very often concise, and require unpacking and expansion to be properly understood and acted upon. An expanded IR query is successful if it leads to the retrieval of a richer set of relevant information sources. Likewise, an expanded metaphor can be considered successful if expansion produces a rich interpretation that is consonant with, and consistenty adds to, our beliefs about a topic. Of course, there are important differences between metaphors, which elicit information from other humans, and IR queries, which elicit information from search engines. For one, IR typically fails to discriminate literal from non-literal language (Veale 2004, 2011), and reduces any metaphoric query to literal keywords and key-phrases that are matched near-identically to texts (e.g. see Salton, 1968; Van Rijsbergen 1979). Yet everyday language shows that metaphor is an ideal form for expressing an information need. A query like "Steve Jobs is a good leader" can be viewed by an IR system as a request to consider all the ways in which leaders are stereotypically good, and to then consider all the metaphors that are typically used to raise these specific talking abouts about Steve Jobs.

It is unsurprising then that IR techniques prove to be useful in the robust treatment of metaphor. For instance, Kintsch (2000) tackles the metaphor-understanding problem using a staple of modern IR, the vector space model (VSM), by mapping ideas into a high-dimensional space that is defined by the texts those ideas are found in. Though a VSM does not distinguish between literal and non-literal uses of a term, it can capture the shared associations and dimensions that link the literal and non-literal meanings of the same word. In this paper we do not employ a VSM approach, but do show how other IR techniques, such as corpus-based query expansion (Vernimb, 1977; Vorhees, 1994,1998; Navigli and Velardi, 2003; Xu and Croft, 1996), can be used to understand and generate metaphors on demand. Expansion of a user's figurative information need is performed using a comprehensive lexicon of affective stereotypes that is itself acquired by harvesting many instances of figurative language usage from the Web.

With these goals in mind, the rest of the paper assumes the following structure. Section 2 provides a brief review of related work and ideas in the domains of metaphor and "creative" information retrieval. The means by which a comprehensive lexicon of affective stereotypes is acquired from the Web is then presented in section 3. We describe in section 4 how a vast collection of Web fragments rich in potential metaphors – the Google n-grams (Brants and Franz, 2006) – is used to drive the processes of metaphor comprehension and generation. These capabilities are evaluated in section 5, which considers the representational adequacy of metaphor as a proxy knowledge representation for dense descriptors that are used with an affective twist. Section 6 describes a working Web service, called *Metaphor Magnet*, that realizes this view of metaphor as a resource and a service for creativity on demand. The paper concludes in section 7 with a discussion of future work in the area of creative Web services.

2. From Potential Metaphors to Deliberate Metaphors

Metaphor has been studied within computer science for four decades, yet it remains at the periphery of NLP research (Veale *et al.*, 2015). The reasons for this marginalization are, for the most part, pragmatic ones, since metaphors can be as varied and challenging as human creativity will allow. The most success has been achieved by focusing on conventional metaphors (e.g., Martin, 1990; Mason, 2004), or on very specific domains of usage, such as figurative descriptions of mental states (e.g., Barnden, 2006). Metaphors are freshest when norms associated with one domain are newly exploited in another, to communicate meanings about very different kind of concept. Hanks (2006) refers the novel use of a linguistic norm in a new domain as an *exploitation*, and notes that metaphors lose their freshness and die as each new exploitation becomes a norm in its own right. For example, it is a norm that we catch and throw physical objects like balls; while "to catch a cold" has become a norm in the domain of infections, and thus a conventional metaphor in its own right, it would still be novel to speak of "throwing a cold" from one person to another. The norms and exploitations view of Hanks offers a lexicographer's take on flexible representations for AI/NLP, such as Wilks' (1978) *preference semantics*, later extended by Fass (1991,1997) into a *collative semantics*.

More recently, some success has been obtained with statistical approaches that side-step the problems of symbolic knowledge representation altogether, by working instead with latent representations that are derived from word distributions. Turney and Littman (2005) show how a statistical model of relational similarity that is constructed from Web texts can retrieve the correct answers for proportional analogies, of the kind used in SAT/GRE tests. No hand-coded knowledge is employed, yet Turney and Littman's system achieves an average human grade on a set of 376 real SAT analogies. Shutova (2010) annotates verbal metaphors in corpora (such as "to *stir* excitement", where "stir" is used metaphorically) with the corresponding conceptual metaphors identified by Lakoff and Johnson (1980). Statistical clustering techniques are then used to generalize from the annotated exemplars, allowing the system to recognize and retrieve other metaphors in the same vein (e.g. "he *swallowed* his anger"). These clusters can also be analyzed to identify literal paraphrases for a given metaphor (such as "to *provoke* excitement" or "*suppress* anger"). Shutova's approach is noteworthy for operating with Lakoff and Johnson's inventory of conceptual metaphors without actually using an explicit knowledge representation of the domains involved.

Hanks (2006) argues that metaphors exploit distributional norms, patterns of word association in one domain that are knowingly exploited in another. To understand a metaphor, one must first recognize the norm that is exploited so as to derive insight from the exploitation. Common norms in language are the preferred semantic arguments of verbs, as well as idioms, clichés and other multi-word expressions. Veale and Hao (2007a) suggest that stereotypes are conceptual norms that are a symbiotic part of many figurative expressions, since e.g. similes rely on stereotypes to illustrate the features ascribed to a topic, while stereotypes are often promulgated via proverbial similes (Taylor, 1954). They also show how stereotypical knowledge can be acquired by harvesting "Hearst" patterns (Hearst, 1992) of the form "as P as C" (e.g. "*as smooth as silk*") from the Web, and how (in Veale and Hao 2007b) how this body of stereotypes can be used in a Web-based model of metaphor generation and comprehension.

Veale (2011) builds a rich query language from these stereotypes by allowing them to be combined with a limited set of operators that turn each one into a non-literal wildcard (see Mihalcea, 2002). We consider here just two operators, @ (the stereotype operator) and **?** (the neighborhood operator), which can be combined with either a noun that denotes a stereotype (such as "knife") or an adjective that denotes a typical property of one (such as "sharp"). If **Noun** denotes a stereotype, then **@Noun** matches any adjective that denotes a stereotypical property of **Noun** (so e.g. *@knife* matches *sharp*, *cold*, etc.) while **@Adj** will match any noun that denotes a stereotype for which *Adj* is a stereotypical property (e.g. *@sharp* matches *sword*, *knife*, *pin*, etc.). The neighborhood operator **?** complements the stereotype operator @. When combined with an adjective, **?Adj** matches any property that co-occurs with, and reinforces, the property denoted by **Adj** in similes; thus, *?hot* matches *humid* because "*as hot and sultry as*" is an attested combination in observed similes; for the same reason, **?Adj** also matches *sultry*, *spicy* and *steamy*. When combined with a noun, **?Noun** matches any noun with which **Noun** is seen to form an ad-hoc set (Hanks, 2005), where ad-hoc sets are typically denoted by the coordination of bare plurals, as in "lawyers and doctors" or "pirates and thieves". Thus, *?lawyer* matches *doctor* and *engineer*, while ?pirate matches *thief* and *hacker* (among many others). The knowledge needed for the operator @ is obtained by harvesting text from the Web, while that for the operator **?** is obtained by mining ad-hoc sets from the Google n-grams (Brants and Franz, 2006).

It is worth noting that Veale (2011) does not model the affective profile of either stereotypes or their properties, so the approach does not know e.g. that *thief* is typically a negative label, or that *damp* is typically an undesirable property. Neither does the model provide a convenient means of using affective qualities in a retrieval query. So we build here on the work of Veale (2011) in several important ways. First, we enrich and enlarge the stereotype lexicon, to include more stereotypes and more adjectival properties, as well as verb-based behaviors such as *swaggering*, *cutting*, *dancing* and *crawling* (i.e. the kind of qualities that one does not typically find in "as X as Y" similes but in "Xing like a Y" similes). We determine an affective polarity for each property or behavior and for each stereotype, and show how polarized positive/negative viewpoints on a topic can be calculated on the fly. We also show how proxy representations for proper-named entities (like *Microsoft*) can be constructed on demand. Finally, we show how metaphors are retrieved from the Google n-grams, allowing a system to understand novel metaphors (like *Google is another Microsoft* or *Apple is a cult*) in terms of known metaphors, and to generate plausible metaphor expansions that better express a user's information needs (e.g., Steve Jobs was a *great leader*, Google is *too powerful*, etc.).

3. Potential Similes and Affective Models

For practical purposes, we consider just two kinds of stereotypical features: properties that can be denoted by *adjectives* (e.g., *hot* for *desert* or *refreshing* for *lemonade*) and *behaviors* that can be denoted by verbs (e.g., *cutting* for *knife* or *flying* for *bird*). We use the generic term *feature* to refer to either properties or behaviors. So if a feature *f* is stereotypical of a concept *C*, we should expect to frequently observe *f* in instances of *C* (Pasca and Van Durme, 2007), and can thus expect to see collocations of "*f*" and "*C*" in a resource like the Google n-grams (Brants & Franz, 2006). Consider these Google 3-grams from the English Web (and their recorded Google frequencies) for "cowboy":

a	lonesome	cowboy	432
a	mounted	cowboy	122
a	grizzled	cowboy	74
a	swaggering	cowboy	68

N-gram patterns allow us to find frequent ascriptions of a feature to a noun concept (like *swaggering* to *cowboys*), but frequently observed features are not always noteworthy features (e.g., see Almuhareb & Poesio 2004,2005). However, if we also observe these features used in similes – such as "*swaggering like a cowboy*" or "*as grizzled as a cowboy*" – this is evidence that speakers assume these features to be elements of a consensus knowledge representation that is shared by speakers and listeners alike. So for each hypothesis *f is stereotypical of C* derived from 3-grams like those above, we generate the corresponding simile form: we use the "like" form for verbal behaviors such as *swaggering*, and the "as-as" form for adjectival properties such as *lonesome*. We dispatch each potential simile as a phrasal query to Google. The hypothesis *f is stereotypical of C* is validated if the potential simile is found at least once on the Web.

This mining process gives us over 200,000 validated hypotheses for our stereotype lexicon. To ensure that the contents of the stereotype lexicon are of the highest quality, we manually filter these validated hypotheses. The investment of a few weeks of labor produces a reliable and reusable resource. We obtain rich descriptions for many dense descriptors, such as the stereotypical *baby*, whose 163 salient features range from *cute* and *guileless* to *crying* and *drooling*. After manual filtering, the lexicon maps 9,479 stereotypes to a set of 7,898 properties and behaviors, to comprise over 75,000 pairings.

3.1. Affective Modelling

For the purpose of affective modelling, adjectival properties and verbal behaviors are treated equally as features after being acquired from the Web via different patterns. To understand the affective uses of a feature, we employ the intuition that the features which support each other in a single simile (e.g. "as *lush and green* as a jungle" or "as *hot and humid* as a sauna") are more likely to have the same affective polarity than those that do not. To construct a support graph of mutually supportive features, we gather all Google 3-grams in which a pair of stereotypical properties or behaviors *X* and *Y* are linked via coordination, as in "*hot and spicy*" or "*kicking and screaming*". A bidirectional link between *X* and *Y* is then added to the support graph if one or more stereotypes in the lexicon contain both *X* and *Y*. If this is not the case, we ask whether both features ever support each other in Web similes, by posing the query "*as X and Y as*" to the Web. If we obtain a non-zero hit set, we link *X* to *Y* and *Y* to *X* in the graph.

We next build a reference set Ref_{neg} of typically negative words, and a set Ref_{pos} of typically positive words. Given a few seed members for Ref_{neg} (such as *sad*, *evil*, *monster*, etc.) and a few seed members for Ref_{pos} (such as *happy*, *wonderful*, *hero*), we use the neighborhood operator **?** to expand this set by suggesting neighboring words with the same polarity (e.g., “sad and *pathetic*”, “happy and *healthy*”). After three iterations in this fashion, we populate Ref_{pos} and Ref_{neg} with approx. 2000 words each.

If we label enough terms in the support graph with a discrete *pos* or *neg* sign, we can reliably interpolate a non-discrete *pos*/*neg* score for every feature in the graph. Let *N*(*f*) denote the set of neighboring terms to a feature f in the support graph. Now, we define:

(2.1) $$N_{pos}(f) = N(f) \cap Ref_{pos}$$

(2.2) $$N_{neg}(f) = N(f) \cap Ref_{neg}$$

We assign non-discrete positive and negative affect scores (from 0 to 1) to f as follows:

(2.3) $$pos(f) = \frac{\|N_{pos}(f)\|}{\|N_{pos}(f) \cup N_{neg}(f)\|}$$

(2.4) $$neg(f) = 1 - pos(f)$$

where ||.|| denotes the cardinality of a set. We can think of *pos*(*f*) as an estimate of the probability that *f* is going to be used in a positive description of a target concept, and *neg*(*f*) as an estimate of the probability that *f* will be used in a negative description.

If a term *S* denotes a stereotypical idea that is described in the lexicon with the set of typical features (adjectival properties and verbal behaviors) denoted *typical*(*S*), then:

(2.5) $$pos(S) = \frac{\sum_{f \in typical(S)} pos(f)}{\|typical(S)\|}$$

(2.6) $$neg(S) = 1 - pos(S)$$

That is, we calculate the mean affect of the properties and behaviors of *S*, as represented in the lexicon via *typical*(*S*). Note that (3.5) and (3.6) are simply gross defaults. One can always use (3.3) and (3.4) to separate the features of *typical*(*S*) into subsets which are more negative than positive (i.e., to put a negative spin on *S*) or into subsets which are more positive than negative (i.e., to put a positive spin on *S*). Thus, we define:

(2.7) $$typical_{pos}(S) = \{f \mid f \in typical(S) \quad \wedge \quad pos(f) > neg(f)\}$$

(2.8) $$typical_{neg}(S) = \{f \mid f \in typical(S) \quad \wedge \quad neg(f) > pos(f)\}$$

For instance, a positive spin on the stereotype *baby* highlights features such as *smiling*, *adorable* and *cute*, while the negative spin focuses on features such as *crying*, *wailing* and *sniveling*. This ability to place a positive or a negative filter on the representation of a stereotypical concept is key to generating affective metaphors on demand.

4. Metaphor Interpretation as Metaphor Expansion

The category-inclusion view of metaphor (see Glucksberg and Keysar, 1990) sees metaphors of the form "X is a Y", such as "my job is a jail", not as identity statements but as categorization statements. The "jail" of "my job is a jail" does not denote a literal jail, but the category of oppressive, jail-like situations, and so the metaphor identifies the referent of "job" as yet another member of that category. The word "jail" serves as a convenient, if oblique, label for this category, and though we may not be able to name the category more directly, we can assume it will impart many of the same features to its members as the category *jail* does to its own members. If a computational system is to appreciate the features that are projected from a source S onto a target concept T in a metaphor, it matters little if we cannot precisely identify a literal mediating category. What matters is that the system can identify a set of feature-rich intermediate categories that seem apt for both S and T, so that it can reason about the features that are projected from T to S. Following Kintsch (2000), it also matters little whether these intermediate categories are metaphorical or literal. So perhaps the set of mediating categories for "my job is a jail" is the set {*hell, trap, cage, nightmare*, …}, all of which share features with *jail* and all of which have features that can aptly be applied to certain kinds of jobs. Expanding on the category-inclusion view, we model metaphor interpretation as a process of expansion, in which an agent searches for the set of mediating categories that are attested for both the target concept T and for the source concept S, and whose stereotypical properties can meaningfully be projected onto the target concept T. In this section we present a set-theoretic view of the expansion process, showing how it uses corpus-attested associations and categorizations to arrive at a feature-rich interpretation.

4.1. Metaphor Expansion

The Google n-grams database is a rich source of established metaphors of the copula form *Target is Source*, such as “politicians are crooks”, “Apple is a cult”, “racism is a disease” and “Steve Jobs is a god”. Let *src*(*T*) denote the set of stereotypes that are used to describe *T* (i.e., potential source concepts for *T*) in the Google n-grams in explicit copula metaphors and categorization frames. To find potential metaphors for proper-named entities such as “Donald Trump”, we focus on n-grams of the form *stereotype First* [*Middle*] *Last*, such as “*billionaire* Donald Trump.” Thus, for example:

src(racism) = {*problem, disease, joke, sin, poison, crime, ideology, weapon*}

src(Hitler) = {*monster, criminal, tyrant, idiot, madman, vegetarian, racist, ...*}

Following Kintsch (1998), we do not discriminate literal from non-literal assertions (e.g. “racism is a problem” versus “racism is a disease”). So we remain agnostic on literality, assuming each element of *src*(*T*) is a potential metaphor that may or may not be deliberate. What matters is that each can be framed as a deliberate metaphor for *T*.

Let *srcTypical*(*T*) denote the aggregation of all properties ascribable to a target concept *T* via the attested source concepts in *src*(*T*):

$$srcTypical(T) = \bigcup_{M \in src(T)} typical(M) \qquad (3.1)$$

Let us denote a negative spin on a topic T as **-**T, and a positive spin as +T. We can thus formulate positive and negative variations of *srcTypical* for these special cases, in (4.2):

$$srcTypical(+T) = \bigcup_{M \in src(T)} \text{typical}_{\text{pos}}(M) \qquad srcTypical(-T) = \bigcup_{M \in src(T)} \text{typical}_{\text{neg}}(M) \qquad (3.2)$$

So (4.1) and (4.2) offer a feature representation for topic *T* as viewed through the prism of metaphor. This is useful when the source *S* in the metaphor *T is S* is not a stereotype in our lexicon, as happens if one describes *Rasputin as Karl Rove* (George W. Bush’s mesmeric political advisor) or *Apple as Scientology*. When the set *typical*(S) is empty, *srcTypical*(*S*) may not be, so *srcTypical*(*S*) can still act as a proxy representation for *S*.

The set of features that are evoked by a source concept S that can be meaningfully projected onto to a topic *T*, as attested by our n-grams corpus, is given by (4.3):

$$(3.3)\qquad salient(T,S) = \begin{array}{c} srcTypical(T) \bigcup typical(T) \\ \bigcap \\ srcTypical(S) \bigcup typical(S) \end{array}$$

The more of S's stereotypical features that are salient in a description of T, the more apt the choice of S as a metaphor for T. We can quantify this aptness using (4.4):

$$(4.4)\qquad aptness(M;T,S) = \frac{\| salient(T,S) \bigcap typical(M) \|}{\| typical(M) \|}$$

We can now construct an interpretation for the metaphor *T is S* by considering not just {*S*}, but the stereotypes in *src*(*T*) that are apt for *T* in the context of *T is S*, as well as the stereotypes that are commonly used to describe *S* – that is, *src*(*S*) – that are apt for *T*:

$$(4.5)\qquad interpretation(T,S) = \left\{ M \middle| M \in src(T) \bigcup src(S) \bigcup \{S\} \quad \wedge \quad aptness(M;T,S) > 0 \right\}$$

In effect, the interpretation of *T is S* is itself a set of apt metaphors for *T* that expand upon *S*. The elements $M_i \in interpretation(T,S)$ can now be sorted by $aptness(M_i;\ T,\ S)$ to produce a ranked list of interpretations (M_1, M_2 … M_n). For any given interpretation M_i, the salient features of M_i are given by:

$$(4.6)\qquad salient(M_i;T,S) = typical(M_i) \bigcap salient(T,S)$$

Some metaphors, and many similes of the form "X is as F as Y", explicitly direct our focus to one dimension or quality of a target topic. To model this explicit focus, we employ the following variants of (4.6):

$$(4.6.1)\qquad salient(M_i;+f;T,S) = typical(M_i) \bigcap salient(T,S) \bigcap N_{pos}(f)$$

$$(4.6.2)\qquad salient(M_i;+f;T,S) = typical(M_i) \bigcap salient(T,S) \bigcap N_{pos}(f)$$

(4.6.3) $$salient(M_i;-f;T,S) = typical(M_i) \cap salient(T,S) \cap N_{neg}(f)$$

Thus, for any viewpoint M_i in *interpretation*(T, S), the set *salient*(M_i; T, S) identifies the features of M_i that T is likely to exhibit when it behaves like M_i. Moreover, we can use the support graph N (and its sub-graphs N_{pos} and N_{neg}) to focus on just those features that are both salient to the metaphor and of explicit interest to the metaphor-maker.

4.2. Metaphor in Action: A Worked Example

Consider the metaphor "*Google is another Microsoft*". We can expect the most salient aspects of *Microsoft* to be those that underpin our common metaphors for *Microsoft*, i.e., the stereotypes in *src*(Microsoft). These stereotypes and their associated features will provide the major talking points for any interpretation of the metaphor.

Google n-grams yield the following sources, 57 for Microsoft and 50 for Google:

src(Microsoft) = {*king, master, threat, bully, giant, leader, monopoly, dinosaur ...*}

src(Google) = {*king, engine, threat, brand, giant, leader, celebrity, religion ...*}

The following features are aggregrated for each:

srcTypical(Microsoft) = {*trusted, menacing, ruling, threatening, overbearing, admired, commanding, ...*}

srcTypical(Google) = {*trusted,admired,reigning, ruling, crowned, shining, determined, lurking, ...}*

Now, the salient features highlighted by *Google is another Microsoft* are given by:

salient(Google, Microsoft) = {*celebrated, menacing, trusted, challenging, established, threatening, admired, respected, ...*}

So, applying (4.4), we obtain:

interpretation(Google, Microsoft) = {*king, criminal, master, leader, bully, threatening, giant, threat, monopoly, pioneer, dinosaur, ...*}

Suppose we focus on the metaphorical expansion "Google is king", since *king* is the most highly ranked element of the interpretation (using (4.4), we calculate *aptness*(*king*, Google, Microsoft) = 0.48). Now,

salient(*king*; Google, Microsoft) = {*celebrated, overbearing, admired, arrogant, respected, ruling, commanding, revered, ...*}

We should note that these properties and behaviours are already implicit in our perception of Google, insofar as they are highly salient aspects of the stereotypical concepts to which Google is frequently compared. The metaphor "*Google is another Microsoft*" – and its potential expansion, "Google is king" – simply crystalizes this set of features, from perhaps different comparisons, into a single act of figurative ideation.

The metaphor "*Google is another Microsoft*" is vague and lacks an affective stance. So suppose a user instead inputs the metaphor "Google is **-Microsoft**", where **-** is used to explicily impart a negative spin (+ can likewise impart a positive spin). In this case, *srcTypical*(-T) is estimated relative to $\textit{typical}_{neg}$(T) as specified in (4.2), so that:

srcTypical(**-**Microsoft) = {*menacing, threatening, twisted, raging, feared, sinister, lurking, domineering, overbearing, ...*}

salient(Google, **-**Microsoft) = {*menacing, bullying, roaring, dreaded...*}

It follows then that

interpretation(Google, **-**Microsoft) = {*criminal, giant, threat, bully, evil, devil, ...}*

In contrast, one may impart a positive spin on *Microsoft* to view *Google* positively too, in line with how a technology investor (as opposed to a technology user) might think:

interpretation(Google, +Microsoft) = {*king, master, leader, pioneer, partner, ...}*

To focus on a specific dimension of a target concept, one can use a simile with an explicit ground, such as "Google is as **powerful** as Microsoft". To identify the sub-set of properties that are salient to this ground, we use the variants of *salient* in (4.6.1) to (4.6.3). The negative consequences of being as powerful a king as Microsoft are thus:

salient(*king*; -powerful; Google, Microsoft) = {*overbearing, arrogant, pompous, ...*}

Just as a few simple concepts can yield a wide range of options for the creative speaker, so too can these concepts yield a wide range of options for a creative system.

5. Empirical Evaluation

The affective stereotype lexicon is the cornerstone of our computational approach to metaphor, and must reliably assign polarity scores both to stereotypes and the features they exemplify. Our affect model is simple in that it relies principally on *pos*/*neg* affect scores, but as demonstrated above, users can articulate their own expressive moods to suit their needs: for example, one can express disdain for excessive power by using the term **-powerful**, or express admiration for guile with the terms **+cunning** or **+devious**.

5.1. The Affect of Stereotypes and Properties

The polarity scores assigned to a feature f in (3.3) and (3.4) do not rely on any prior classification of f, such as whether f is in Ref_{pos} or Ref_{neg}. That is, Ref_{pos} and Ref_{neg} are not used as training data, and (3.3) and (3.4) receive no error feedback. We expect that for $f \in Ref_{pos}$ that $pos(f) > neg(f)$, and for $f \in Ref_{neg}$ that $neg(f) > pos(f)$, but (3.3) and (3.4) do not iterate until this is so. Measuring the extent to which these simple intuitions are validated offers a good evaluation of our graph-based calculation of polarity scores.

Just five features in Ref_{pos} (approx. **0.4**% of the 1,314 properties and behaviors in Ref_{pos}) are given a positive affect score of less than 0.5 using (3.3), leading those words to be misclassified as more negative than positive. The misclassified property words are: *evanescent, giggling, licking, devotional* and *fraternal.* Similarly, just twenty-six properties in Ref_{neg} (approx. **1.9**% of the 1,385 properties and behaviors in Ref_{neg}) are assigned a negative affect score of less than 0.5 via (3.4), leading these to be misclassified as more positive than negative. The misclassified words are: *cocky, dense, demanding, urgent, acute, unavoidable, critical, startling, gaudy, decadent, biting,* *controversial, peculiar, disinterested, strict, visceral, feared, opinionated, humbling, subdued, impetuous, shooting, acerbic, heartrending, ineluctable* and *groveling*.

Since Ref_{pos} and Ref_{neg} are populated with words that are chosen for their perceived *pos*/*neg* slants, this result is hardly surprising. Nonetheless, it does verify the intuitions that underpin (3.3) and (3.4) – that the affective polarity of a property or behavior *f* can be reliably estimated as a simple function of the affect of the co-descriptors with which it is commonly used across different contexts. We must still ask whether these polarity scores are consistent with the expected affect of the stereotypical ideas for which these properties and behaviors are typical. The sets Ref_{pos} and Ref_{neg} are populated in §3.1 with adjectives, verbal behaviors and nouns. Ref_{pos} contains 478 positive nouns (such as *saint* and *hero*) while Ref_{neg} contains 677 negative nouns (such as *tyrant* and *monster*). When we use these reference stereotypes to test the effectiveness of (3.5) and (3.6) – and thus, indirectly, of the stereotype lexicon itself – we find that **96.7%** of the positive noun exemplars are correctly assigned a mean positivity of more than 0.5 (so, $pos(S) > neg(S)$) while **96.2%** of the negative noun exemplars are correctly assigned a mean negativity of more than 0.5 (so, $neg(S) > pos(S)$). Though it may seem crude to assess the polarity of a complex stereotype as the mean of the polarity of its features, this does appear to be a reliable measure of the overall polarity of a dense descriptor.

5.2. Placing An Affective Spin on Stereotypes

Nonetheless, stereotypes can be used with varying affect in different contexts. Consider the case of the stereotypical *baby*. We describe loved ones as “baby” to highlight just how much we care for them, and to emphasize features such as lovability and cuteness. But we also use the same word to negatively describe those that “*act like a baby*”, that is, those who are overly dependent on others, or those who are weak, immature and excessively emotional. One can argue that the word “baby” is used in two different dictionary senses here, yet both would ultimately appeal to our mental representation of the same complex stereotype, a human baby. We can conceive of this kind of selective spin as a retrieval task: if *typical*(*S*) specifies the salient features of a stereotypical *S*, then can we retrieve from *typical*(*S*) only the positive features of S (e.g. for “baby” used affectionately), or only the negative features (e.g. of “baby” used as an insult)?

Table 1. *Macro-Average P/R/F1 scores for affective partition, the affective selection of* ***positive*** *and* ***negative*** *properties from 6,230 stereotypes.*

Macro Average **(6,230 stereotypes)**	***Positive*** **properties**	***Negative*** **properties**
Precision	.962	.98
Recall	.975	.958
F-Score	.968	.968

Table 2. *Macro-Average P/R/F1 scores for retrieval of* ***positive*** *and* ***negative*** *stereotypes for 4,536 properties and behaviours.*

Macro Average **(4,536 features)**	***Positive*** **stereotypes**	***Negative*** **stereotypes**
Precision	.986	.965
Recall	.949	.982
F-Score	.967	.973

Suppose we focus on stereotypes with at least one positive feature in Ref_{pos} or at least one negative feature in Ref_{neg} (there are 6,230 in all, with an average of 2.95 features each in Ref_{pos}, and 3.55 features each in Ref_{neg}). We find that the qualities corresponding to each *pos*/*neg* spin can be accurately retrieved. Table 1 reports macro-averages for the selective retrieval of positive only (or negative only) qualities from 6,230 stereotypes. When we focus on features that are associated with at least one stereotype in Ref_{pos} or in Ref_{neg} (there are 4,536 in all), Table 2 reports macro-averages for the retrieval of more-positive-than-negative (or more-negative-than-positive) exemplars for these features. In each case, these results show that a reliable affective partition can be achieved.

5.3. Representational Adequacy of Metaphors

Even ad-hoc stereotypes such as *Microsoft* or *Donald Trump* – dense decriptors that are not defined in the stereotype lexicon, but which can be given a proxy representation using *srcTypical* in (4.1) – can be given a positive or negative spin in context, since each has their own admirers and detractors. For instance, the n-gram metaphors that populate *src(Trump)* allow these properties to be inferred for *Donald Trump*:

srcTypical (**+Trump**) = {*successful, wealthy, trusted, irrepressible, leading,…}*

srcTypical(**-Trump**) = { *scheming, spoiled, ruthless, overbearing, vain, ...}*

But how good a proxy is *src*(S) or *srcTypical*(S) for an *S* like *Trump* or *Microsoft*? Can we, for instance, reliably estimate the *pos*/*neg* polarity of *S* as a function of *src*(S)? We can estimate *pos*(S) as in (5.1) below (where *neg*(S) follows simply as *1 - pos*(S)):

(4.1)
$$pos(S) = \frac{\sum\limits_{M \in src(S)} pos(M)}{\left\| src(S) \right\|}$$

Testing this estimator on the stereotypes in Ref_{pos} and Ref_{neg}, the correct binary polarity (*pos* or *neg*) is estimated **87.2**% of the time. It follows that the copula metaphors in the Google n-grams – and consequenty the contents of *src*(*S*) – are broadly consistent with our perceptions of whether a topic S has a positive or negative connotation overall.

If we consider all stereotypes *S* for which ||*src*(*S*)|| > 0 (there are 6,904 in our affect lexicon), we find *srcTypical*(S) covers, on average, just **65.7**% of the typical properties of *S* (that is, of *typical*(*S*)). As a proxy representation for *S*, *srcTypical*(*S*) is incomplete. However, this shortfall is often the reason we use novel metaphors in the first place. So consider (5.2), a variant of (4.1) that captures the longer reach of novel metaphors:

(4.2)
$$srcTypical^2(T) = \bigcup_{S \in src(T)} srcTypical(S)$$

Thus, $srcTypical^2(T)$ denotes the set of features that are ascribable to T via the expansive interpretation of all metaphors *T is S* in our Web corpus, since *S* can now project onto T any element of *srcTypical*(*S*). Using macro-averaging over all 6,904 cases where ||*src*(*S*)|| > 0, we find that $srcTypical^2(S)$ covers **99.2**% of *typical*(*S*) on average. Metaphors truly are a descriptive resource, and a well-chosen metaphor can allow us to emphasize almost any feature of a target idea *T* we might wish to highlight.

5.4. Human Judgment

Our Web corpus is a source of potential metaphors that are treated as deliberate for the purposes of novel metaphor generation. Yet a metaphor only truly becomes deliberate when it is framed as such, to evoke in the mind of a reader the distinct spaces of source and target so as to encoruage a comparative analysis of the content of both spaces. If our system's metaphorical outputs are to be judged as deliberate by humans, we shall have to give them an appropriate linguistic framing. The Twitterbot *@MetaphorMagnet* is an autonomous generation system that frames the conceits of (4.5) using a variety of framing strategies and tweets the resulting utterances in 140 characters or less (Veale, 2015). For instance, noting that the feature *pampered* is typical of kings *and* newborns, *@MetaphorMagnet* frames the overlap via this deliberate metaphor: "*To be nurtured by and loved by a mother: This can turn majestic kings into weak newborns.*" When eliciting human judgments on the metaphors produced by (4.5) and related mechanisms, we present judges with the full utterances that are generated by *@MetaphorMagnet*, and not the underlying conceits, as they can seem overly skeletal to non-experts. We ask the Twitterbot to generate 120 figurative utterances that are evenly distributed with regards to their overall affect, and pay raters on the crowd-sourcing platform *CrowdFlower.com* a small sum for each of their judgments. Each metaporical utterance corresponds to a single test unit, and we elicit 20 judgments for each unit from anonymous judges.

Table 3. *Distribution of human judgments for comprehensibility and novelty.*

	Comprehensibility	**Novelty**
Very Low	6.49%	5.26%
Medium Low	17.39%	18.84%
Medium High	20.29%	20.13%
Very High	55.82%	55.77%

Novel metaphorical utterances should be novel *and* comprehensible. Table 3 shows the distribution of mean human judgments for the dimensions comprehensibility and novelty using the same 4-point scale for each. More than **75%** of all judgments deem the machine-generated metaphors to rate satisfactorily high on each dimension. The human raters were also asked to judge the overall affect of each deliberate metaphor by grading each on a scale running from **+2** (very positive) down to **-2** (very negative). We averaged the judgments of the 20 different raters for each metaphor to arrive at a single overall estimate of affect, which shows a **0.85** agreement with the system's own affect score for these metaphors. The Cohen's Kappa coefficient for this agreement is **0.71**.

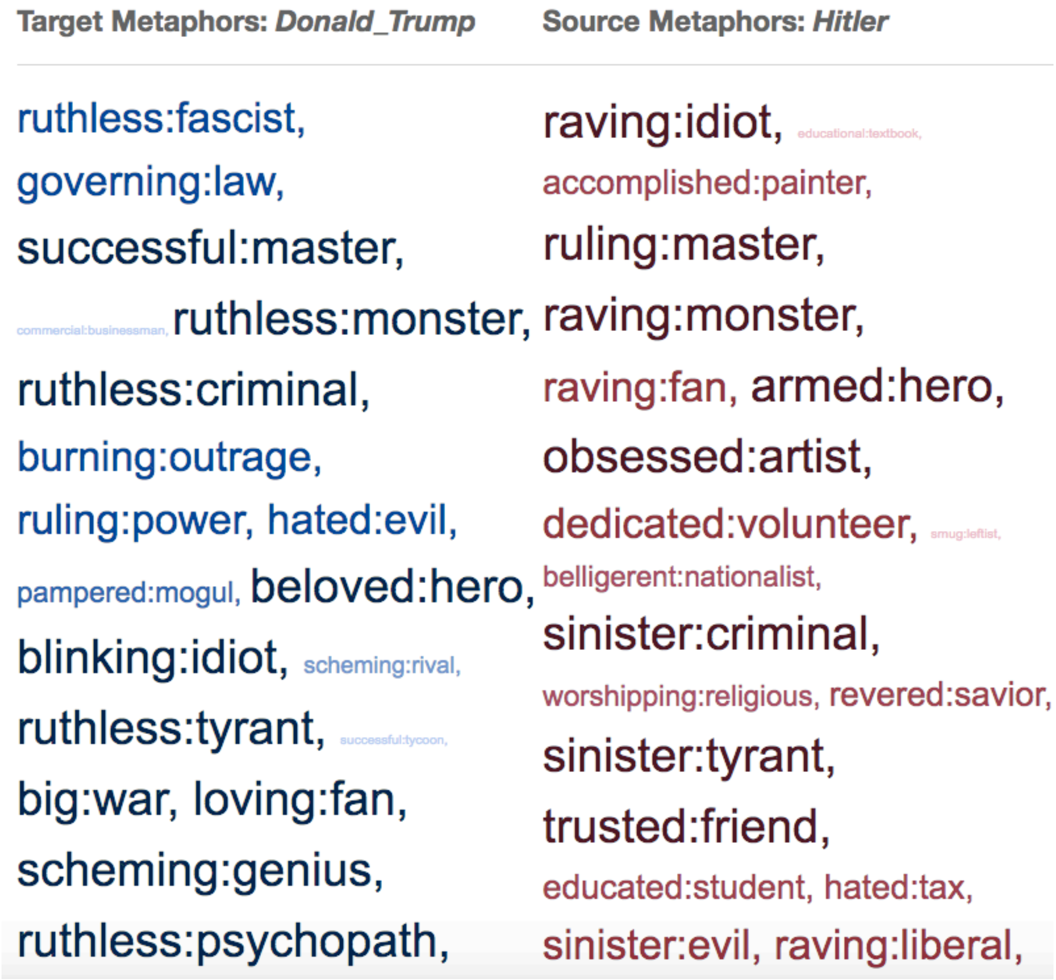
6. Metaphor as a Resource and a Public Web Service

Metaphor is a knowledge multiplier that allows us to expand our knowledge of a target T by using knowledge of other source ideas S as a magnifying lens. We have presented here a robust, stereotype-driven approach that embodies this practical philosophy. Knowledge multiplication is achieved using an IR-like expansionary model, in which an affective query is expanded to include all of the metaphors that are commonly used to convey this affective viewpoint. These viewpoints are expanded in turn to include the salient qualities that are typically implied by each. This expansion approach owes much to IR, and so is, in turn, ideally suited to the creative enrichment of conventional IR.

These ideas have been implemented in the form of a Web service, named *Metaphor Magnet*, which allows users to enter affective metaphors of the form shown here (such as *Google is -Microsoft*, *life is a +game*, and *Steve Jobs is Tony Stark*). Each metaphor is viewed as a conversational gambit, and thus treated as a query to elicit supporting evidence for the affective stance conveyed by the metaphor. The *Metaphor Magnet* service expands each new metaphor-*qua*-query into a set of known metaphors via mappings derived from the Google n-grams; each of these pre-existing metaphors is then expanded into a set of contextually apt properties and behaviors. Ultimately, each of these qualities is re-expressed as a topic-specific IR query that is used to retrieve relevant hits for the topic from Google. In effect, the *Metaphor Magnet* service allows users to interact with a search engine like Google using affective metaphors and other expressive language forms. The service can currently be accessed at this URL:

http://boundinanutshell.com/metaphor-magnet

Metaphor Magnet can exploit the properties and behaviors of its stock of almost 10,000 stereotypes, and can infer salient qualities for many proper-named entities such as *Donald Trump* or *Steve Jobs* using a combination of copula statements from the Google n-grams (e.g., "*Steve Jobs is a visionary*") and category assignments from Wikipedia. When used interactively, the interpretation of the simile "*Donald Trump is as -popular as Hitler*" thus highlights a selection of negative viewpoints on the source concept, *Hitler*, and picks out an apt selection of viewpoints on the target *Donald Trump*. *Metaphor Magnet* displays both selections as side-by-side phrase clouds. The phrase cloud representing *Hitler* in this simile is shown in the screenshot of Figures 2a and 2b (left), while the phrase clouds for *Donald Trump* are shown in Figures 2a and 2b (right).



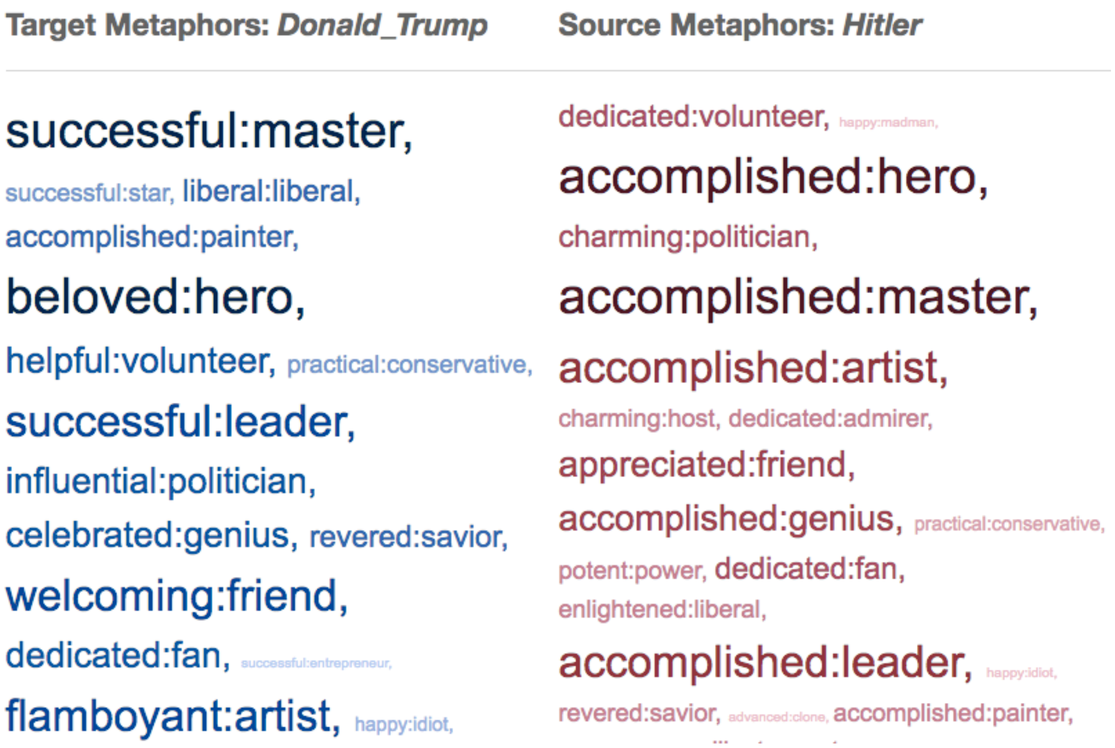
Target Metaphors: *Donald_Trump*

ruthless:fascist,
governing:law,
successful:master,
commercial:businessman, ruthless:monster,
ruthless:criminal,
burning:outrage,
ruling:power, hated:evil,
pampered:mogul, beloved:hero,
blinking:idiot, scheming:rival,
ruthless:tyrant, successful:tycoon,
big:war, loving:fan,
scheming:genius,
ruthless:psychopath,

Source Metaphors: *Hitler*

raving:idiot, educational:textbook,
accomplished:painter,
ruling:master,
raving:monster,
raving:fan, armed:hero,
obsessed:artist,
dedicated:volunteer, smug:leftist,
belligerent:nationalist,
sinister:criminal,
worshipping:religious, revered:savior,
sinister:tyrant,
trusted:friend,
educated:student, hated:tax,
sinister:evil, raving:liberal,

Figure 2a. Godwin's "rule of Hitler analogies" in action. On the left, a screenshot of *Metaphor Magnet*'s phrase cloud for the perspectives cast by the affective metaphor "*Donald Trump is as -popular as Hitler*" on its target, "Donald Trump". On the right, the cloud of negative metaphors typically used for "Hitler" in the Google n-grams.



Target Metaphors: *Donald_Trump*

successful:master, successful:star, liberal:liberal, accomplished:painter, beloved:hero, helpful:volunteer, practical:conservative, successful:leader, influential:politician, celebrated:genius, revered:savior, welcoming:friend, dedicated:fan, successful:entrepreneur, flamboyant:artist, happy:idiot,

Source Metaphors: *Hitler*

dedicated:volunteer, happy:madman, accomplished:hero, charming:politician, accomplished:master, accomplished:artist, charming:host, dedicated:admirer, appreciated:friend, accomplished:genius, practical:conservative, potent:power, dedicated:fan, enlightened:liberal, accomplished:leader, happy:idiot, revered:savior, advanced:clone, accomplished:painter,

Figure 2b. The ugly side of "positive" spin. On the left, *Metaphor Magnet*'s phrase cloud of the perspectives cast by "*Donald Trump is as +popular as Hitler*" on its target. On the right, the cloud of positive metaphors for "Hitler" in the Google n-grams.

Metaphor Magnet demonstrates the potential utility of affective metaphors in human-computer linguistic interaction, and provides a Web service via which other natural-language processing applications can acquire a measure of metaphorical competence of their own. When accessed as a service, *Metaphor Magnet* returns both HTML or XML data, and in this way it also serves as the foundation of @*MetaphorMagnet*. Given the resource-intensive nature of metaphor understanding and generation – processes which require lexico-semantic models to formulate hypotheses and vast amounts of corpus data to validate hypotheses – it is good design practice to view these processes as remote services that hide their complexity behind the simplicity of a Web interface.

7. Conclusions: Explaining the World with Deliberate Metaphors

The 2016 conference of the Cognitive Science society was held in Philadelphia just a week after the Democratic National Convention crowned its presidential nominee. The streets around the convention center still buzzed with rhetorical whimsy, and outside a nearby church fiery street preachers waved placards that read "*Ask Me Why You're Going to Hell*.“ The notices of the old church revealed a more temperate character, and advertised a sermon with the eye-catching title "*Jesus Disrupts*." Our metaphors are so pervasive that they often goes unnoticed by speaker and listener alike, but surely this was a metaphor reveling in its status as metaphor. Its author wanted to do more than convey the power of religion to change lives, and so the metaphor self-consciously appropriates the language of disruptive technology to foster the creation of new mental connections between the domains of faith and radical innovation. It was, as Steen (2011, 2015) terms it, a truly "deliberate" metaphor. For the metaphor deliberately subverts the modern view of disruptive pioneers – as best illustrated by Steve Jobs and his *reality distortion field* – as messiahs. If tech pioneers are to be revered as "messiahs" then this new metaphor urges us to have as much faith in the real deal as in the people that design our phones. Steen calls these metaphors *deliberate* because they are designed to be noticed and calculated to make playthings of their source and target domains. Indeed, as metaphors go, this one was not just deliberate but surprisingly self-descriptive too.

Metaphor disrupts. It disrupts our conceptual category systems the way a game of musical chairs disrupts a formal dinner party, licensing guests to ignore the host's place settings in favor of whatever works best when the music stops. Metaphor is the ultimate appropriation device, allowing speakers to appropriate the stereotypical associations and linguistic norms of one domain of experience so as to transplant them wholesale onto another. Wherever metaphor goes, disruption and appropriation are sure to follow, even when we fail to notice, as we so often do, the deep upheaval taking place beneath the beguiling calmness of the metaphor's surface. Though no little skepticism has been expressed regarding the cognitive reality and practical utility of viewing deliberate metaphors as a class apart (e.g. Gibbs, 2015), Steen's hard distinction has significant computational value, not least when it comes to the merits of *potential* metaphors. Suppose the author of "Jesus Disrupts" intended nothing so baroque as the construal above when posting this sermon title, so that "disrupt" means nothing more here than its dictionary sense "to change things dramatically." But the metaphor is no less deliberate for existing only in the mind of a reader as a purposeful evocation of the technology domain, even if its status as a communicative act does change should this be the case.

The real world is festooned with potential metaphors like these, which may or may not have been crafted as deliberate provocations by their makers but which can usefully be appreciated as deliberate by their consumers, to arrive at deeper and more resonant interpretations. Potential metaphors permeate natural-language texts of all kinds, and so procedures for the identification of metaphors in text (e.g. see Steen *et al*. 2010) force annotators to make decisions about vexing constructs whose metaphoricity lies in the eye of the beholder. Not knowing the true intentions of the author, the best one can do is to recognize the potential for these constructs to be interpreted as deliberate metaphors. The texts of the Web are certainly no different in this regard, nor indeed are the free-floating snippets of the Web n-grams that – shorn of their original contexts of use – are even freer to support diverse construals that may go far beyond their authors' intentions. In this paper we have viewed a large corpus of Web n-grams as a large body of potential metaphors, and when the need arises, potential similes too. But what makes an n-gram a potential metaphor? It is the ability of an agent, either cognitive or artificial, to interpret it as a deliberate metaphor that maps a specific source space onto a specific target space. So an agent only sees metaphor potential where it has the knowledge and the dexterity to deliver on that potential, to make it seem resonantly deliberate with the right framing.

Deliberate metaphor presents as much an opportunistic event for consumers as it does a planned and purposeful action for producers. The computational model we have outlined here is especially opportunistic in its use of Web n-grams, which it treats as a source of textual inspiration for the production of novel metaphors and utterances. The world of the model is not the open world of a human, but insofar as the Web holds up a mirror to many diverse aspects of the outside world an opportunistic machine can build vivid pictures of a great many of the domains that feed into human-oriented metaphors. Indeed, we might even think of deliberate metaphor as the means by which a machine can explain the oddities of the outside world that present themselves to it via language. That is, we can view any linguistic stimulus – whether a sermon title on a church or an n-gram on the Web – as an aspect of the world that requires explanation, and view the conversion of potential metaphors into deliberate metaphors as a path to understanding. The danger that our machines may be reading too much into a stimulus is a real one – although it exists for humans too – yet it is not one that should present grave concerns to a creative agent more interested in poetic possibilities than precise facts. In the age of mechanical production our metaphor machines are not mind-readers but world builders.

References

Almuhareb, A. and Poesio, M. (2004). Attribute-Based and Value-Based Clustering: An Evaluation. In *Proc. of EMNLP 2004*. Barcelona.

Almuhareb, A. and Poesio, M. (2005). Concept Learning and Categorization from the Web. In *Proc. of the 27th Annual meeting of the Cognitive Science Society*.

Barnden, J. 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 de Gruyter.

Brants, T. and Franz, A. (2006). *Web 1T 5-gram Ver. 1*. Linguistic Data Consortium.

Brennan, S. E. and Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(6):1482-1493.

Fass, D. (1991). Met*: a method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49-90.

Fass, D. (1997). Processing Metonymy and Metaphor. *Contemporary Studies in Cognitive Science & Technology*. New York: Ablex.

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

Gibbs, R.W. (2015). Does deliberate metaphor theory have a future? *Journal of Pragmatics*, vol. 90, December 2015, pp 73–76.

Glucksberg, S. and Keysar, B. (1990). Understanding Metaphorical Comparisons: Beyond Similarity. Psychological Review, 97(1):3-18.

Hanks, P. (2005). Similes and Sets: The English Preposition ‘like’. In: Blatná, R. and Petkevic, V. (Eds.), *Languages and Linguistics: Festschrift for Fr. Cermak*. Charles University, Prague.

Hanks, P. (2006). Metaphoricity is gradable. In: Anatol Stefanowitsch and Stefan Th. Gries (Eds.), *Corpus-Based Approaches to Metaphor and Metonymy*,. 17-35. Berlin: Mouton de Gruyter.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *In Proc. of the 14th Int. Conf. on Computational Linguistics*, pp 539–545.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York: Cambridge University Press.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. Psychonomic Bulletin Review, 7(2):257-266.

Lakoff, G. (1987). Women, Fire, and Dangerous Things. The University of Chicago Press.

Martin, J. H. (1990). A Computational Model of Metaphor Interpretation. New York: Academic Press.

Mason, Z. J. (2004). CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System, *Computational Linguistics*, 30(1):23-44.

Mihalcea, R. (2002). The Semantic Wildcard. In *Proc. of the LREC Workshop on Creating and Using Semantics for Information Retrieval and Filtering*. Canary Islands, Spain, May 2002.

Navigli, R. and Velardi, P. (2003). An Analysis of Ontology-based Query Expansion Strategies. In Proc. of the workshop on Adaptive Text Extraction and Mining (ATEM 2003), at ECML 2003, the 14th European Conf. on Machine Learning, 42–49

Pasca, M. and Van Durme, B. (2007). What You Seek is What You Get: Extraction of Class Attributes from Query Logs. In *the Proc. of the 20th International Joint Conference on Artifical intelligence, 2832-2837.*

Salton, G. (1968). *Automatic Information Organization and Retrieval*. New York: McGraw-Hill.

Shutova, E. (2010). Metaphor Identification Using Verb and Noun Clustering. In *the Proc. of the 23rd International Conference on Computational Linguistics*, 1001-1010.

Steen, G., Dorst, A.G., Herrmann, J.B., Kaal, A., Krennmayr, T. and Pasma, T. (2010). A Method for Linguistic Metaphor Identification: From MIP to MIPVU. *Amsterdam: John Benjamins*.

Steen, G. (2011). The contemporary theory of metaphor -- now new and improved! *Review of Cognitive Linguistics*, vol. 9, 26–64.

Steen, G. (2015). Developing, testing and interpreting deliberate metaphor theory. *Journal of Pragmatics*, vol. 90, December 2015, pp 67–72.

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

Turney, P.D. and Littman, M.L. (2005). Corpus-based learning of analogies and semantic relations. *Machine Learning* 60(1-3):251-278.

Van Rijsbergen, C. J. (1979). *Information Retrieval*. Oxford: Butterworth-Heinemann.

Veale, T. (2004). The Challenge of Creative Information Retrieval. *Computational Linguistics and Intelligent Text Processing: Lecture Notes in Computer Science*, Volume 2945/2004, 457-467.

Veale, T. and Hao, Y. (2007a). Making Lexical Ontologies Functional and Context-Sensitive. *Proc. of the 46th Annual Meeting of the Assoc. of Computational Linguistics.*

Veale, T. and Hao, Y. (2007b). Comprehending and Generating Apt Metaphors: A Web-driven, Case-based Approach to Figurative Language. *In proceedings of AAAI 2007, the 22nd AAAI Conference on Artificial Intelligence*. Vancouver, Canada.

Veale, T. and Hao, Y. (2008). Talking Points in Metaphor: A concise, usage-based representation for figurative processing. *In Proceedings of ECAI'2008, the 18th European Conference on Artificial Intelligence.* Patras, Greece, July 2008.

Veale, T. 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. June 2011.

Veale, T. and Hao, Y. (2012). In The Mood for Affective Search with Web stereotypes. In *Proceedings of the 21st international conference on World Wide Web*, Lyon, France.

Veale, T. (2015). Unnatural Selection: Seeing Human Intelligence in Artificial Creations. *Journal of General Artificial Intelligence*, 6(1):5-20.

Veale, T., Shutova, E. and Beigman Klebanov, B. (2016). *Metaphor: A Computational Perspective.* Morgan Claypool, Synthesis Lectures on Human Language Technologies.

Vernimb, C. (1977). Automatic Query Adjustment in Document Retrieval. *Information Processing & Management*. 13(6):339-353.

Voorhees, E. M. (1994). Query Expansion Using Lexical-Semantic Relations. In *the proc. of SIGIR 94, the 17th International Conference on Research and Development in Information Retrieval*. Berlin: Springer-Verlag, 61-69.

Voorhees, E. M. (1998). Using WordNet for text retrieval. *WordNet, An Electronic Lexical Database*, 285–303. The MIT Press.

Wilks, Y. (1978). Making Preferences More Active, *Artificial Intelligence* 11.

Xu, J. and Croft, B. W. (1996). Query expansion using local and global document analysis. In *Proc. of the 19th annual international ACM SIGIR conference on Research and development in information retrieval.*