Once More, With *Feeling!*
Using Creative Affective Metaphors to Express Information Needs

Tony Veale

Web Science & Technology Division, KAIST / School of Computer Science and Informatics, UCD
Korean Advanced Institute of Science & Technology, South Korea / University College Dublin, Ireland.
Tony.Veale@gmail.com

Abstract

Creative metaphors abound in language because they facilitate communication that is memorable, effective and elastic. Such metaphors allow a speaker to be maximally suggestive while being minimally committed to any single interpretation, so they can both supply and elicit information in a conversation. Yet, though metaphors are often used to articulate affective viewpoints and information needs in everyday language, they are rarely used in information retrieval (IR) queries. IR fails to distinguish between creative and uncreative uses of words, since it typically treats words as *literal mentions* rather than *suggestive allusions*. We show here how a computational model of affective comprehension and generation allows IR users to express their information needs with creative metaphors that concisely allude to a dense body of assertions. The key to this approach is a lexicon of stereotypical concepts and their affective properties. We show how such a lexicon is harvested from the open web and from local web n-grams.

Creative Truths

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 scope to the listener to contribute to the joint construction of meaning, creative language suggests a looser but potentially richer meaning that is amenable to collaborative elaboration by each participant in a conversation.

A metaphor *X is Y* establishes a *conceptual pact* between speaker and listener (Brennan & 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* (Veale

& 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 a speaker and listener to arrive at a mutually acceptable construal of a metaphorical "snake" in the context of X.

Metaphors achieve a balance of suggestiveness and concision through the use of *dense descriptors*, familiar terms like "snake" that evoke a rich variety of stereotypical properties and behaviors (Fishelov, 1992). Though every concept has the potential to be used creatively, 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 richer, more conceptual model of the lexicon is needed to allow any creative uses of stereotypes to be inferred as needed in context. We will show here how a large lexicon of stereotypes is mined from the web, and how stereotypical representations can be used selectively and creatively, to highlight relevant aspects of a given target concept in a specific metaphor.

Because so many familiar stereotypes have polarizing qualities – think of the endearing and not-so-endearing qualities 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 metaphors in response, such as "yes, a true pioneer" or "what an artist!", or even "but he could be such a tyrant!". Proper-names can also be used as evocative stereotypes, as when Steve Jobs is compared to the fictional inventor *Tony Stark*, or Apple is compared to *Scientology*, or Google to *Microsoft*. We use stereotypes effortlessly, and their exploitations are common currency in everyday language.

Information retrieval, however, is a language-driven application where the currency of metaphor has little or no exchange value, not least because IR fails to discriminate literal from non-literal language (Veale 2004, 2011, 2012). Speakers use metaphor to provide and elicit information in

casual conversation, but IR reduces any metaphoric query to literal keywords and key-phrases, which are matched near-identically in texts (Salton, 1968; Van Rijsbergen 1979). Yet everyday language shows that metaphor is an ideal form for expressing our information needs. A query like "Steve Jobs as 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 convey these viewpoints. The IR staple of query expansion (Vernimb, 1977; Vorhees, 1994,1998; Navigli & Velardi, 2003; Xu & Croft, 1996) can be made both *affect-driven* and *metaphor-aware*. In this paper we show how an affective stereotype-based lexicon can both comprehend and generate affective metaphors that capture or shape a user's feelings, and show how this capability can lead to more creative forms of IR.

Related Work and Ideas

Metaphor has been studied within computer science for four decades, yet it remains at the periphery of NLP research. 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 greatest 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., Barden, 2006).

From the earliest computational forays, it has been recognized that metaphor is essentially a problem of knowledge representation. Semantic representations are typically designed for well-behaved mappings of words to meanings – what Hanks (2006) calls *norms* – but metaphor requires a system of soft preferences rather than hard (and brittle) constraints. Wilks (1978) thus proposed his *preference semantics* model, which Fass (1991,1997) extended into a *collative semantics*. In contrast, Way (1990) argues that metaphor requires a dynamic concept hierarchy that can stretch to meet the norm-bending demands of figurative ideation, though her approach lacks computational substance.

More recently, some success has been obtained with statistical approaches that side-step the problems of knowledge representation, by working instead with implied or latent representations that are derived from word distributions. Turney and Littman (2005) show how a statistical model of relational similarity can be constructed from web texts for retrieving the correct answer to proportional analogies, of the kind used in SAT 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 metaphor (such as "to *provoke* excitement" or "*suppress* anger"). Shutova's approach is noteworthy for operating with Lakoff & Johnson's inventory of conceptual metaphors without using an explicit knowledge representation.

Hanks (2006) argues that metaphors exploit distributional norms: to understand a metaphor, one must first recognize the norm that is exploited. 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 found in many figurative expressions, and note that stereotypes and similes enjoy a symbiotic relationship that has some obvious computational advantages. Similes use stereotypes to illustrate the qualities ascribed to a topic, while stereotypes are often promulgated via proverbial similes (Taylor, 1954). Veale and Hao (2007a) show how stereotypical knowledge can be acquired by harvesting "Hearst" patterns of the form "as P as C" (e.g. "*as smooth as silk*") from the web (Hearst, 1992). They show in (2007b) how this body of stereotypes can be used in a web-based model of metaphor generation and comprehension.

Veale (2011) employs stereotypes as the basis of a new *creative information retrieval* paradigm, by introducing a variety of non-literal wildcards in the vein of Mihalcea (2002). In this system, **@Noun** matches any adjective that denotes a stereotypical property of *Noun* (so e.g. @knife matches *sharp*, *cold*, etc.) while **@Adj** matches any noun for which *Adj* is stereotypical (e.g. @sharp matches *sword, laser, razor*, etc.). In addition, **?Adj** matches any property or behavior that co-occurs with, and reinforces, the property denoted by *Adj*; thus, *?hot* matches *humid, sultry* and *spicy*. Likewise, *?Noun* matches any noun that denotes a pragmatic neighbor of *Noun*, where two words are neighbors if they are seen to be clustered in the same ad-hoc set (Hanks, 2005), such as "lawyers and doctors" or "pirates and thieves". The knowledge needed for **@** is obtained by mining text from the open web, while that for **?** is obtained by mining ad-hoc sets from Google n-grams.

There are a number of shortcomings to this approach. For one, Veale (2011) does not adequately model the affective profile of either stereotypes or their properties. For another, the stereotype lexicon is static, and focuses primarily on adjectival properties (like *sharp* and *hot*). It thus lacks knowledge of everyday verbal behaviors like *cutting, crying, swaggering*, etc. 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 behaviors. We determine an affective polarity for each property or behavior and for

each stereotype, and show how polarized +/- viewpoints on a topic can be calculated on the fly. We show how proxy representations for ad-hoc proper-named stereotypes (like *Microsoft*) can be constructed on demand. Finally, we show how metaphors are mined from the Google n-grams, to allow the system to understand novel metaphors (like *Google is another Microsoft* or *Apple is a cult*) as well as to generate plausible metaphors for users' affective information needs (e.g., Steve Jobs was a *great leader*, Google is *too powerful*, etc.).

Once more, with *feeling!*

If a property or behavior P is stereotypical of a concept C, we should expect to frequently observe P in instances of C. In linguistic terms, we can expect to see collocations of "P" and "C" in a resource like the Google n-grams (Brants and Franz, 2006). Consider these 3-grams for "cowboy" (numbers in parentheses are Google database frequencies).

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

N-gram patterns of the above form allow us to find frequent ascriptions of a quality to a noun-concept, but frequently observed qualities are not always noteworthy qualities (e.g., see Almuhareb and Poesio, 2004,2005). However, if we also observe these qualities in similes – such as "*swaggering like a cowboy*" or "*as grizzled as a cowboy*" – this suggests that speakers see these as typical enough to anchor a figurative comparison. So for each hypothesis *P is stereotypical of C* that we derive from the Google n-grams, 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 then dispatch each simile as a phrasal query to Google: a hypothesis is validated if the corresponding simile is found on the web.

This mining process gives us over 200,000 validated hypotheses for our stereotype lexicon. We now filter these hypotheses manually, to ensure that the contents of the lexicon are of the highest quality (investing just weeks of labor produces a very reliable resource; see Veale 2012 for more detail). We obtain rich descriptions for commonplace ideas, such as the dense descriptor *Baby*, whose 163 highly salient qualities – a set denoted *typical*(Baby) – includes *crying*, *drooling* and *guileless*. After this manual phase, the stereotype lexicon maps 9,479 stereotypes to a set of 7,898 properties / behaviors, to yield more than 75,000 pairings.

Determining Nuanced Affect
To understand the affective uses of a property or behavior, we employ the intuition that those which reinforce each other in a single description (e.g. "as *lush and green* as a jungle" or "as *hot and humid* as a sauna") are more likely to have the same affect than those which do not. To construct a support graph of mutually reinforcing properties, 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 added to the graph if one or more stereotypes in the lexicon contain both X and Y. If this is not so, we consider whether both descriptors ever reinforce each other in web similes, by posing the web query "*as X and Y as*". If this query has a non-zero hit set, we still add a link between X and Y.

Next, we build a reference set **-R** of typically negative words, and a disjoint set **+R** of typically positive words. Given a few seed members for **-R** (such as *sad*, *evil*, *monster*, etc.) and a few seed members for **+R** (such as *happy*, *wonderful*, *hero*, etc.), we use the **?** operator of Veale (2011) to successively expand this set by suggesting neighboring words of the same affect (e.g., "*sad and pathetic*", "*happy and healthy*"). After three iterations in this fashion, we populate **+R** and **-R** with approx. 2000 words each. If we can anchor enough nodes in the graph with + or − labels, we can interpolate a nuanced positive / negative score for all nodes in the graph. Let N(p) denote the set of neighboring terms to a property or behavior p in the support graph. Now, we define:

(1) $\quad N^{+}(p) \quad = N(p) \cap +\mathbf{R}$

(2) $\quad N^{-}(p) \quad = N(p) \cap -\mathbf{R}$

We assign positive / negative affect scores to p as follows:

(3) $\quad pos(p) \quad = \quad \dfrac{|N^{+}(p)|}{|N^{+}(p) \cup N^{-}(p)|}$

(4) $\quad neg(p) \quad = \quad 1 - pos(p)$

Thus, *pos(p)* estimates the probability that p is used in a positive context, while *neg(p)* estimates the probability that p is used in a negative context. The *X and Y* 3-grams approximate these contexts for us.

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

(5) $\quad pos(S) \quad = \quad \dfrac{\sum_{p \in typical(S)} pos(p)}{|typical(S)|}$

(6) $\quad neg(S) \quad = \quad 1 - pos(S)$

So we simply calculate the mean affect of the properties and behaviors of *s*, as represented in the lexicon via *typical(s)*. Note that (5) and (6) are simply gross defaults.

One can always use (3) and (4) to separate the elements of *typical(s)* into those which are more negative than positive (a negative spin on *s*) and those which are more positive than negative (a positive spin on *s*). Thus, we define:

(7) posTypical(S) = $\{p \in$ typical(S) $\mid$ pos(p) > neg(p)$\}$

(8) negTypical(S) = $\{p \in$ typical(S) $\mid$ neg(p) > pos(p)$\}$

For instance, the positive stereotype of *Baby* contains the qualities such as *smiling*, *adorable* and *cute*, while the negative stereotype contains qualities such as *crying*, *wailing* and *sniveling*. As we'll see next, this ability to affectively "spin" a stereotype is key to automatically generating affective metaphors on demand.

Generating Affective Metaphors, N-gram style

The Google n-grams is also a rich source of copula metaphors of the 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 commonly used to describe T, where commonality is defined as the presence of the corresponding copula metaphor in the Google n-grams. To also find metaphors for proper-named entities like "Bill Gates", we analyse n-grams of the form *stereotype First [Middle] Last*, such as "*tyrant* Adolf Hitler". For example:

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

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

We do not try to discriminate literal from non-literal assertions, nor indeed do we try to define literality at all. Rather, we assume each putative metaphor offers a potentially useful perspective on a topic T.

Let *srcTypical*(T) denote the aggregation of all properties ascribable to T via metaphors in *src*(T):

(9) $srcTypical\ (T) \quad = \quad \bigcup\limits_{M \in src(T)} typical(M)$

We can also use the *posTypical* and *negTypical* variants of (7) and (8) to focus only on metaphors that place a positive or negatve spin on a topic T. In effect, (9) provides a feature representation for topic T as viewed through the creative lens of metaphor. This is useful when the source *S* in the metaphor *T is S* is not a stereotype in the lexicon, as happens when one describes *Rasputin as Karl Rove*, or *Apple as Scientology*. When the set *typical*(S) is empty, *srcTypical*(S) may not be, so *srcTypical*(S) can act as a proxy representation for S in these cases.

The properties and behaviors that are salient to the interpretation of *T is S* are given by:

(10) *salient* (T,S) = [srcTypical(*T*) $\cup$ *typical*(T)]
$$\cap$$
[srcTypical(S) $\cup$ typical(S)]

In the context of *T is S*, the metaphorical stereotype M $\in$ *src*(S)$\cup$*src*(T)$\cup$\{S\} is an apt vehicle for T if:

(11) *apt*(M, *T,S*) = |*salient*(*T,S*) $\cap$ *typical*(M)| > 0

and the degree to which M is apt for T is given by:

(12) *aptness*(M,*T,S*) = $\dfrac{|salient(T,\ S) \cap typical(M)|}{|typical(M)|}$

We can now construct an interpretation for *T is S* by considering the stereotypes in *src*(T) that are apt for T in the context of *T is S*, and by also considering the stereotypes that are commonly used to describe S that are also potentially apt for T:

(13) *interpretation*(T, S)
        = \{M $\in$ *src*(S)$\cup$*src*(T)$\cup$\{S\} $\mid$ *apt*(M, T, S)\}

In effect, the interpretation of the creative metaphor *T is S* is itself a set of more conventional metaphors that are apt for T and which expand upon S. The elements $\{M_i\}$ of *interpretation*(T, S) can be sorted by *aptness*(M_i *T,S*) to produce a ranked list of interpretations ($M_1 \ldots M_n$). For a given interpretation M, the salient features of M are thus:

(14) *salient*(M, *T,S*) = *typical*(M) $\cap$ *salient* (*T,S*)

So if *T is S* is a creative IR query – to find documents in which T is viewed as S – then *interpretation*(T, S) is an expansion of *T is S* that includes the common metaphors that are consistent with T viewed as S. In turn, for any viewpoint M_i in *interpretation*(T, S), then *salient*(M_i, *T, S*) is an expansion of M_i that includes all of the qualities that T is likely to exhibit when it behaves like M_i.

A Worked Example: Metaphor Generation for IR
Consider the creative query "*Google is Microsoft*", which expresses a user's need to find documents in which Google exhibits qualities typically associated with Microsoft. Now, both *Google* and *Microsoft* are complex concepts, so there are many ways in which they can be considered similar or dissimilar, whether in a good or a bad light. However, 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 metaphors will provide the talking points for an interpretation.

The Google n-grams yield up the following metaphors, 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 ...*}

So the following qualities are aggregated for each:

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

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

Now, the salient qualities highlighted by the metaphor, namely $salient$(Google, Microsoft), are:

{*celebrated, menacing, trusted, challenging, established, threatening, admired, respected, ...*}

Finally, $interpretation$(Google,Microsoft) contains:

{*king, criminal, master, leader, bully, threatening, giant, threat, monopoly, pioneer, dinosaur, ...*}

Let's focus on the expansion "Google is king", since according to (12), $aptness$(*king*, Google, Microsoft) = 0.48 and this is the highest ranked element of the interpretation. Now, $salient$(*king*, Google, Microsoft) contains:

{*celebrated, revered, admired, respected, ruling, arrogant, commanding, overbearing, reigning, ...*}

Note that these properties / behaviours are already implicit in our consensus perception of Google, insofar as they are highly salient aspects of the stereotypical concepts to which Google is frequently compared on the web. These properties / behaviours can now be used to perform query expansion for the query term "Google", to find documents where the system believes Google is acting like Microsoft.

The metaphor "Google is Microsoft" is diffuse and lacks an affective stance. So let's consider instead the metaphor "Google is **-Microsoft**", where **-** is used to impart a negative spin (and where **+** can likewise impart a positive spin). In this case, $negTypical$ is used in place of $typical$ in (9) and (10), so that:

$srcTypical$(**-Microsoft**) =

{*menacing, threatening, twisted, raging, feared, sinister, lurking, domineering, overbearing, ...*}

and

$salient$(Google, **-Microsoft**) =

{*menacing, bullying, roaring, dreaded...*}

Now, $interpretation$(Google, **-Microsoft**) becomes:

{*criminal, giant, threat, bully, evil, victim, devil, ...*}

In contrast, $interpretation$(Google, **+Microsoft**) is:

{*king, master, leader, pioneer, classic, partner, ...*}

More focus is achieved with this query in the form of a simile: "Google is as **-powerful** as Microsoft". For explicit similes, we need to focus on just a sub-set of salient properties, as in this varient of (10):

{$p \in salient$(Google, Microsoft) | $p \in$ N⁻(powerful)}

In this case, the final interpretation becomes:

{*bully, threat, giant, devil, monopoly, dinosaur, ...*}

A few simple concepts can thus yield a wide range of options for the creative IR user who is willing to build queries around affective metaphors and similes.

Empirical Evaluation

The affective stereotype lexicon is the cornerstone of the current approach, and must reliably assign meaningful polarity scores both to properties and to the stereotypes that exemplify them. Our affect model is simple in that it relies principally on +/- affect, but as demonstrated above, users can articulate their own expressive moods to suit their needs: for Stereotypical example, one can express disdain for too much power with the term **-powerful**, or express admiration for guile with **+cunning** and **+devious**.

The Effect of Affect: Stereotypes and Properties
Note that the polarity scores assigned to a property p in (3) and (4) do not rely on any prior classification of p, such as whether p is in **+R** or **-R**. That is, **+R** and **-R** are not used as training data, and (3) and (4) receive no error feedback. Of course, we expect that for $p \in$ **+R** that $pos(p) > neg(p)$, and for $p \in$ **-R** that $neg(p) > pos(p)$, but (3) and (4) do not iterate until this is so. Measuring the extent to which these simple intuitions are validated thus offers a good evaluation of our graph-based affect mechanism.

Just five properties in **+R** (approx. **0.4**% of the 1,314 properties in **+R**) are given a positivity of less than 0.5 using (3), leading those words to be misclassified as more negative than positive. The misclassified property words are: *evanescent, giggling, licking, devotional* and *fraternal*.

Just twenty-six properties in **-R** (approx. **1.9**% of the 1,385 properties in **-R**) are assigned a negativity of less than 0.5 via (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*.

Because **+R** and **-R** have been populated with words that have been chosen for their perceived +/- slants, this result is hardly surprising. Nonetheless, it does validate the key intuitions that underpin (3) and (4) – that the affective polarity of a property *p* can be reliably estimated as a simple function of the affect of the co-descriptors with which it is most commonly used in descriptive contexts.

The sets **+R** and **-R** are populated with adjectives, verbal behaviors and nouns. **+R** contains 478 nouns denoting positive stereotypes (such as *saint* and *hero*) while **-R** contains 677 nouns denoting negative stereotypes (such as *tyrant* and *monster*). When these reference stereotypes are used to test the effectiveness of (5) and (6) – and thus, indirectly, of (3) and (4) and of the stereotype lexicon itself – **96.7%** of the positive stereotype exemplars are correctly assigned a mean positivity of more than 0.5 (so, *pos(S) > neg(S)*) and **96.2%** of the negative 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 affect of a stereotype as the mean of the affect of its properties, this does appear to be a reliable measure of polar affect.

The Representational Adequacy of Metaphors

We have argued that metaphors can provide a collective representation of a concept that has no other representation in a system. But how good a proxy is *src*(S) or *srcTypical*(S) for an *S* like *Karl Rove* or *Microsoft*? Can we reliably estimate the +/- polarity of *S* as a function of *src*(S)? We can estimate these from metaphors as follows:

$$(15) \quad pos(S) \quad = \quad \frac{\sum_{M \in src(S)} pos(M)}{|src(S)|}$$

$$(16) \quad neg(S) \quad = \quad \frac{\sum_{M \in src(S)} neg(M)}{|src(S)|}$$

Testing this estimator on the exemplar stereotypes in **+R** and **-R**, the correct polarity (+ or -) is estimated **87.2%** of the time. Metaphors in the Google n-grams are thus broadly consistent with our perceptions of whether a topic is positively or negatively slanted.

When we consider all stereotypes *S* for which $|src(S)| > 0$ (there are 6,904 in the lexicon), *srcTypical*(S) covers, on average, just **65.7%** of the typical properties of S (that is, of *typical*(S)). Nonetheless, this shortfall is precisely why we use novel metaphors. Consider this variant of (9) which captures the longer reach of these novel metaphors:

$$(17) \; srcTypical^2(T) \quad = \bigcup_{S \,\in\, src(T)} srcTypical(S)$$

Thus, $srcTypical^2(T)$ denotes the set of qualities that are ascribable to T via the expansive interpretation of all metaphors *T is S* in the Google n-grams, 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. A well-chosen metaphor enables us to emphasize almost any quality of a topic T we might wish to highlight.

Affective Text Retrieval with Creative Metaphors

Suppose we have a database of texts $\{D_1 \dots D_n\}$ in which each document D_i offers a creative perspective on a topic T. We might have texts that view politicians as crooks, popes as kings, or hackers as heroes. So given a query **+T**, can we retrieve only those texts that view T positively, and given **-T** can we retrieve only the negative texts about T?

We first construct a database of artificial figurative texts. For each stereotype *S* in the lexicon, and for each *M* $\in src(S) \cap (+R \cup -R)$, we construct a text D_{SM} in which *S* is viewed as *M*. The title of document D_{SM} is "S is M", while the body of D_{SM} contains all the words in *src*(M). D_{SM} uses the typical language of *M* to talk about S. For each D_{SM}, we know whether D_{SM} conveys a positive or negative viewpoint on *S*, since *M* sits in either in **+R** or **-R**.

The affect lexicon contains 5,704 stereotypes *S* for which $src(S) \cap (+R \cup -R)$ is non-empty. On average, each of these stereotypes is described in terms of 14 other stereotypes (5.8 are negative and 8.2 are positive, according to **+R** and **-R**) and we construct a representative document for each of these viewpoints. We construct a set of 79,856 artificial documents in total, to convey figurative perspectives on 5,704 different stereotypical topics:

Table 1. Macro-Average P/R/F1 scores for affective retrieval of + and - viewpoints for 5,704 topics.

Macro Average (5704 topics)	*Positive* viewpoints	*Negative* viewpoints
Precision	.86	.93
Recall	.95	.78
F-Score	.90	.85

For each document retrieved for T, we estimate its polarity as the mean of the polarity of the words it contains. Table 1 presents the results of this experiment, in which we attempt to retrieve only the positive viewpoints for T with a query **+T**, and only the negative viewpoints for T using **-T**. The results are sufficiently encouraging to support the further development of a creative text retrieval engine that is capable of ranking documents by the affective figurative perspective that they offer on a topic.

Concluding Thoughts: The Creative Web

Metaphor is a creative *knowledge multiplier* that allows us to expand our knowledge of a topic T by using knowledge of other ideas as a magnifying lens. We have presented here a robust, stereotype-driven approach that embodies this practical philosophy. Knowledge multiplication is achieved using an expansionary approach, 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 all the qualities that are typically implied by each. Such an approach is ideally suited to a creative re-imagining of IR.

An implementation of these ideas is available for use on the web. Named *Metaphor Magnet*, the system allows users to enter queries of the form shown here (such as *Google is –Microsoft, Steve Jobs as Tony Stark, Rasputin as Karl Rove*, etc.). Each query is expanded into a set of apt metaphors mined from the Google n-grams, and each metaphor is expanded into a set of contextually apt qualities. In turn, each quality is expanded into an IR query that is used to retrieve relevant hits from Google. In effect, the system – still an early prototype – allows users to interface with a search engine like Google using metaphor and other affective language forms. The system can currently be accessed at this URL:

http://boundinanutshell.com/metaphor-magnet

Metaphor Magnet is just one possible application of the ideas presented here, which constitute not so much a philosophical or linguistic theory of metaphor, but an engineering-oriented toolkit of reusable concepts for imbuing a wide range of text applications with a robust competence in linguistic creativity. Human speakers do not view metaphor as a problem but as a solution. It is time our computational systems took a similarly constructive view of this remarkably creative cognitive tool.

In this vein, *Metaphor Magnet* continues to evolve as a creative web service. In addition to providing metaphors on demand, the service now also provides a poetic framing facility, whereby the space of possible interpretations for a given metaphor is crystallized into a single poetic form. More generally, poetry can be viewed as a means of reducing information overload, by summarizing a complex metaphor – or the set of texts retrieved using that metaphor via creative IR – whose interpretation entails a rich space of affective possibilities. A poem can thus be seen in functional terms as both an information summarization tool and as a visualization device. *Metaphor Magnet* adopts a simple, meaning-driven approach to poetry generation: given a topic T, a set of candidate metaphors with the desired affective slant is generated. One metaphor is chosen at random, and the elements of its interpretation are sampled to produce different lines of the resulting poem. Each element, and the sentiment it best evokes, is rendered in natural language using one of a variety of poetic tropes.

For example, *Metaphor Magnet* produces the following as a distillation of the space of feelings and associations that arise from the interpretation of *Marriage is a Prison*:

The legalized regime of this marriage

My marriage is a tight prison
The most unitary federation scarcely organizes so much
Intimidate me with the official regulation of your prison
Let your close confines excite me
O Marriage, you disgust me with your undesirable security

Each time we dip into the space of possible interpretations, a new poem is produced. One can use *Metaphor Magnet* to sample the space at will, hopping from one interpretation to the next, or from one poem to another. Here is an alternate rendition of the same metaphor in poetic form:

The official slavery of this marriage

My marriage is a legitimate prison
No collective is more unitary, or organizes so much
Intimidate me with the official regulation of your prison
Let your sexual degradation charm me
O Marriage, you depress me with your dreary consecration

In the context of our earlier worked example, which generated a space of metaphors to negatively describe Microsoft's perceived misuse of power, consider the following, which distills the assertion *Microsoft is a Monopoly* into an aggressive ode:

No Monopoly Is More Ruthless

Intimidate me with your imposing hegemony
No crime family is more badly organized,
    or controls more ruthlessly
Haunt me with your centralized organization
Let your privileged security support me
O Microsoft, you oppress me with your corrupt reign

Poetry generation in *Metaphor Magnet* is a recent addition to the service, and its workings are beyond the scope of the current paper (though they may be observed in practice by visiting the aforementioned URL). For details of a related approach to poetry generation – one that also uses the stereotype-bearing similes described in Veale (2012) – the reader is invited to read Colton, Goodwin & Veale (2012).

Metaphor Magnet forms a key element in our vision of a *Creative Web*, in which web services conveniently provide creativity on tap to any third-party software application that requests it. These services include *ideation* (e.g. via metaphor generation & knowledge discovery), *composition* (e.g. via analogy, bisocation & conceptual blending) and *framing* (via poetry generation, joke & story generation, etc.). Since CC does not distinguish itself through distinct algorithms or representations, but through its unique goals

and philosophy, such a pooling of services will not only help the field achieve a much-needed critical mass, it will facilitate a greater penetration of CC ideas and approaches into the commercial software industry.

Acknowledgements

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):1482-93.

Colton, S., Goodwin, J. and Veale, T. 2012. Full-FACE Poetry Generation.In *Proc. of ICCC 2012, the 3rd International Conference on Computational Creativity*. Dublin, Ireland.

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.

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 Univ., 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.

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*. Spain, May 2002.

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

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.

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*, Vol. 2945/2004, 457-467.

Veale, T. and Hao, Y. 2007a. Making Lexical Ontologies Functional and Context-Sensitive. *In 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 Proc. 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. 2011. Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity. In Proc. of ACL'2011, the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.

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

Vernimb, C. 1977. Automatic Query Adjustment in Document Retrieval. *Information Processing & Mgmt,* 13(6):339-53.

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. MIT Press.

Way, E. C. 1991. Knowledge Representation and Metaphor. *Studies in Cognitive systems*. Holland: Kluwer.

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*.