# *Talking Points* in Linguistic Creativity

*Tony Veale, Yanfen Hao*

## 1. Introduction

Creative linguistic devices like metaphor, simile and analogy serve two important roles in language. The first of these is to make the unfamiliar and the strange seem more familiar and understandable (Indurkhya, 1992). For instance, one might describe a burqa (a full body covering for Muslim women) as a suit of armor, as a shield against prying eyes or, depending on one's communication goal, as a wearable cage. The other role of these linguistic devices is most often associated with the poetic and fanciful use of language, but is no less important: to make the familiar and mundane seem strange and unfamiliar. In this latter guise, metaphors, analogies and similes allow us to view a commonplace idea from a new and revealing category perspective (Camac and Glucksberg, 1984). For instance, one might describe make-up as "the Western burqa", to communicate not just the idea that each involves a covering of the female form, but that each reflects a society-imposed expectation on the public presentation of women.

Each of these roles is a manifestation of the same underlying mechanism for combining concepts, for understanding how they interact (Black, 1962) and for determining how they are connected (Fauconnier and Turner, 1998), even if those connections are tenuous, hidden or not always obvious (Collins and Loftus, 1975). For example, Figure 1 shows the reframing that is needed to appreciate the above metaphors:

| | |
|---|---|
| Burqa ⇒ | Make-up ⇒ |
| *for concealing a Muslim woman* | *typically worn by women* |
| ≈ *for protecting a Muslim woman* | ≈ *expected to be worn by women* |
| ≈ *for protecting a woman* | ≈ *must be worn by women* |
| ≈ *for protecting a person* | ≈ *must be worn by Muslim women* |
| Armor ⇐ | Burqa ⇐ |

*Figure 1.* Linguistically-mediated pathways between concepts Burqa and Armor.

In each case we see how metaphor draws out and highlights, in a modified or exaggerated form, an existing aspect of each target concept. In other words, a creative mapping does not indiscriminately transplant arbitrary aspects of a source concept onto a target, but accommodates a selective graft of the most salient aspects of this source concept onto those aspects of the target that can be highlighted by the juxtaposition (Ortony, 1979). This connection between concepts requires a flexible knowledge representation, one that allows the connections between non-identical source and target aspects to be recognized, reconciled and even compressed (Fauconnier and Turner, 1998). This fluid representation (Hofstadter *et al*., 1995) defines the search space in which the processes of creative linguistic description are cognitively situated (Veale and O'Donoghue, 2000): for generation, fluid connectivity allows a system to search outwards from a given target to find those potential source concepts that offer a new yet appropriate perspective; for understanding purposes, connectivity allows an agent to focus on those key aspects of a source concept that are most apt for a target because they can be linked to that target.

### 1.1. Knowledge Acquisition and Creative Manipulation

Creative uses of language, like those considered above, show that concepts can meaningfully be described at multiple levels of detail and with varying degrees of literal accuracy, to suit the kind of inferences that are required in a particular context. Simple atomic features, for instance, have long held a practical appeal (e.g., see Katz and Fodor, 1963), since sets of such features can easily be used to discriminate between word meanings in a semantic hierarchy (e.g., see Dong and Dong, 2006). Indeed, when enough features are harvested from the web (see Almuhareb and Poesio, 2004), the words that possess them can be clustered into categories that accurately reflect the structure of a semantic hierarchy like WordNet (Fellbaum, 1998). Furthermore, increasing the detail of these features can yield additional rewards: when features *and* their dimensions are harvested from the web for given words (e.g., hot *and* temperature for "coffee", or hot *and* taste for "chilli"), the accuracy of the semantic hierarchy that can be built via clustering also increases (see Almuhareb and Poesio, 2004, 2005).

The concepts described by these features may themselves be composite structures, and so a given feature may apply to some aspects of a

target concept more than others. We say that surgeons are delicate and that poets are sensitive, but it is surely more informative to say that surgeons have delicate hands, or that a poet possesses a sensitive eye. This increased attention to how features are naturally used in everyday linguistic description can pay further dividends when assessing the similarity of two different concepts: for instance, surgeons and artists both have sensitive hands, artists and poets both have sensitive eyes, and poets and orators both have inspiring voices. These descriptions have a naturalistic, almost metaphorical quality that one finds in much of everyday language, in which certain features are communicated by reference to a highly evocative prototype for those features. Thus, a rhinoceros has a thick hide, a lion has a courageous heart, a preacher has an inspiring voice, an eagle has a fierce eye and a statue has a cold visage. These descriptions are "naturalistic" in the sense that they often rely on received linguistic wisdom, stereotypical and even clichéd combinations of ideas that are often literally false. Nonetheless, most English speakers know exactly what is communicated by the description "*the noble soul of a hero*" or "*the cold logic of a computer*".

But, is there truth in metaphors, stereotypes and clichés, or at least enough truth to make this kind of naturalistic, almost fanciful description worth harvesting from the web? If so, then these expressions may allow us to acquire a body of talking points that capture the essence of familiar concepts in a way that a more objective and literal-minded representation can not. We see two ways of testing this hypothesis: if subjective feature sets derived from naturalistic descriptions offer a more productive knowledge representation for creative language processing than their more objective and literal counterparts, they should provide a better and more accurate basis for clustering words into semantic hierarchies; or, we should need fewer such features to achieve the same level of clustering accuracy as their less insightful counterparts. In this current work, we demonstrate that this hypothesis is, in fact, true, and that the naturalistic descriptions we dub *talking points* provide a more insightful basis for acquiring and defining semantic features that are flexible enough to yield to creative manipulation. Our benchmark in this respect is the work of Almuhareb and Poesio (2004, 2005), who demonstrate that large sets of automatically harvested semantic features for given nouns can be used to form a reasonably good semantic hierarchy for those nouns. We argue in this paper that subjective *talking points* yield

comparable semantic clustering ability on the same data sets but with far fewer, and thus more insightful, features.

### 1.2. Structure of this paper

In this paper we describe the construction of a fluid knowledge representation for creative language processing, one that is acquired automatically from WordNet (Fellbaum, 1998) and from the texts of the web. In section 2 we summarize related work in the field of metaphor processing as it pertains to flexible knowledge representation. In section 3 we describe two complementary means of acquiring the basic elements of this representation, which we call *talking points*, from an authoritative source like WordNet and from the uncurated texts of the web, before describing how these elements can be placed into a fluid network of connections – what Hofstadter (ibid) calls a *slipnet* – in section 4. We then present in section 5 some empirical evaluation of the acquired talking points on an objective test of term categorization, before concluding in section 6.

## 2. Related Work and Ideas

All discussion of the power of creative language to change how we perceive the world must begin with metaphor. Since metaphor can be viewed as a stretching of linguistic conventions to cover new conceptual ground, the interpretation of metaphor crucially hinges on an agent's ability to recognize these conventions and accommodate the exceptional meaning conveyed by each figurative expression. Indeed, most computational approaches embody a sense of what it means to be literal, and accommodate metaphoric meanings within this conventional scheme through a form of relaxation, mapping or translation. Wilks (1978) advocates that the typically *hard* constraints that define a literal semantics should instead be modeled as *soft* preferences that can accommodate the violations that arise in metaphoric utterances, while Fass (1991) builds on this view to show how these violations can be repaired to thus capture the literal intent behind each metaphor. This repair process in turn relies on the availability of a concept taxonomy through which metaphoric uses can be mapped onto their literal counterparts; a car that "drinks gasoline" would thus be understood as a car

that "consumes gasoline". Way (1991) emphasizes the importance of this taxonomy by positing a central role for a dynamic type hierarchy (DTH) in metaphor, one that can create new and complex taxonyms on the fly. For instance, Way's DTH would understand the "make-up as Western burqa" metaphor via a dynamically created taxonym like *things-women-are-expected-to-wear-in-public*, though Way offers no algorithmic basis for the workings of such a remarkable taxonomy.

Another family of computational approaches combines explicit knowledge about certain metaphors with knowledge about the domains connected by these metaphors. Martin's (1990) *Midas* system encodes schematic knowledge about conventionalized metaphors such as "to kill a process" and "to open a program", and uses this knowledge to fit novel variations of these metaphors into the most appropriate schemas. Barnden and Lee (2002) focus on the role of inference in a metaphorically-structured domain, and describe a system called *ATTMeta* that contains sufficient knowledge about e.g., conventional metaphors of mind to reason about the mental states implied by these metaphors. Each of these approaches sees metaphor interpretation as a process of fitting what is said to what can meaningfully be represented and reasoned about. This fitting process is most explicitly modelled by Hofstadter *et al.* (1995), who focus on the slippage processes that are required to understand analogies in abstract domains that e.g., involve the mapping of letter sequences or the mirroring of actions in a highly stylized tabletop environment. Though simplified and toy-like, these are non-deterministic problem domains that are nonetheless shaped by a wide range of pragmatic pressures. Hofstadter and Mitchell (1994) model these pressures using a *slipnet*, a probabilistic network in which concepts are linked to others into which they can slip or be substituted with. In this view, deeply embedded concepts that are further removed from direct observation are less likely to engage in slippage than more superficial concepts. To take a linguistic example, word choice in natural language generation is more susceptible to slippage (as influenced by synonym availability) than the concepts underlying the meaning of a sentence.

Slippage can be seen as a lossy form of conceptual re-representation: the greater the slippage, the more dramatic the re-representation and the greater the potential for loss of accuracy. For instance, a recent magazine cover proclaims the governor of California, Arnold Schwarzenegger, as "president of 12% of the U.S". This conceptualiza-

tion can be viewed as an intermediate stage in a slippage path from Governor to President, as shown in Figure 2:

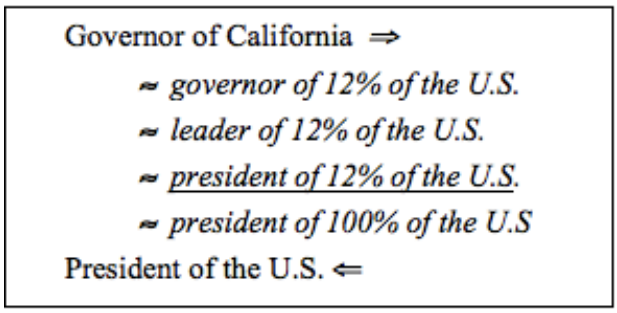


*Figure 2.* Successive re-framing that is needed to view a governor as a president.

This labelling is creative enough to grace a magazine cover because it involves an ambitious level of re-conceptualization, at least from a computational perspective. The pivotal insight is the ad-hoc synonym *California = 12% of the U.S.*, which one is unlikely to find in a dictionary or any general-purpose resource like WordNet (Fellbaum, 1998). Our goal in this current work, while not so ambitious, is to build a slippage network of concepts and their most salient features that combines the principled flexibility of a Hofstadter-style slipnet with the comprehensive scale of a dictionary resource like WordNet.

## 3. Acquiring Conceptual *Talking Points*

We refer to the knowledge elements connected by this slipnet as conceptual *talking points*. We describe in this section the form of these talking points and how they are acquired, before describing in section 4 how slippage operates between different talking points. We discuss two complementary kinds of talking point here: objective talking points, extracted from dictionary definitions (more specifically, the textual glosses offered by WordNet), and informal, stereotypical talking points, harvested from the Web.

It should be noted that automatic knowledge-acquisition from text, in which the knowledge needed to understand language is itself acquired *from* language, is a very well-developed sub-field in computational linguistics. Text-based approaches to knowledge acquisition range from the

ambitiously comprehensive, in which an entire text or resource is fully parsed and analyzed in depth, to the surgically precise, in which highly-specific text patterns are used to eke out correspondingly specific relationships from a large corpus. Endeavors such as that of Harabagiu *et al.* (1999), in which each of the textual glosses in WordNet (Fellbaum, 1998) is linguistically analyzed to yield a sense-tagged logical form, is an example of the former approach. In contrast, foundational efforts such as that of Hearst (1992) typify the latter surgical approach, in which one fishes in a large text for word sequences that strongly suggest a particular semantic relationship, such as hypernymy or, in the case of Charniak and Berland (1999), the part-whole relation. Such efforts offer high precision but low recall, and extract just a tiny (but very useful) subset of the semantic content of a text. The *KnowItAll* system of Etzioni *et al.* (2004) employs the same generic patterns as Hearst (e.g., "NPs such as NP1, NP2, …"), and more besides, to extract a whole range of facts that can be exploited for web-based question-answering. Cimiano and Wenderoth (2007) also use a range of Hearst-like patterns to find text sequences in web-text that are indicative of the lexico-semantic properties of words; in particular, these authors use phrases like "to * a new NOUN" and "the purpose of NOUN is to *" to identify the agentive and telic roles of given nouns, thereby fleshing out the noun's qualia structure as posited by Pustejovsky's (1990) theory of the generative lexicon.

The approach described here is technically similar to these previous approaches. As we shall show, it differs not in how the specific knowledge is acquired, but in the nature of the knowledge that is acquired. Objective literal facts constitute just one part of the talking points knowledge-representation. Equal importance is given to the subjective, folk knowledge that underpins our beliefs about everyday objects and entities, and which pervades everyday language in the form of clichés and stereotypes. We use authoritative linguistic resources that strive for objectivity to acquire objective talking points, and look instead to how language is actually used, in proverbial similes, to acquire a more subjective, non-literal and creative perspective on the world.

### 3.1. Objective Talking Points

Objective talking points are aspects of conceptual description that contribute to the consensus definitional view of a concept. Though Word-

Net does not provide explicit semantic criteria for the definition of each lexical concept, many of these criteria can be gleaned from a shallow parse of the pithy textual gloss it associates with each one. Thus, whenever the head phrase of a concept's gloss has the form "ADJ$^+$ NOUN" where NOUN can denote a hypernym of the concept, we can associate the talking point *is_ADJ:NOUN* with that concept. For example, the WordNet gloss of {Hamas} is "a militant Islamic fundamentalist political movement that …", which yields *is_militant:movement*, *is_islamic:movement, is_fundamentalist: movement* and *is_political:movement* as talking points for Hamas. When a WordNet concept has a hypernym of the form {ADJ_NOUN}, where NOUN can denote a hypernym of this concept, we likewise associate the talking point *is_ADJ:NOUN* with that concept. For example, {Taliban, Taleban} has a hypernym {religious_movement} which yields the talking point *is_religious:movement* for Taliban.

Objective talking points can also be gleaned from the subject-verb-object structure of a WordNet gloss. For instance, the gloss for {conductor, music_director} is "the person who leads a musical group", which yields the talking point *leads:musical_group*. The hypernym of this concept, {musician}, has the gloss "artist who composes or conducts music …", which yields the talking points *composes:music* and *conducts:music* that are then inherited by {conductor, …} and other subtypes of musician in WordNet. A shallow parse will generally not lead to a complete understanding of a concept, but will typically produce some interesting talking points of the *predicate:object* variety that can be used to relate a concept to others that are analogically or metaphorically similar. Using WordNet's noun and verb taxonomies, we can identify the following slippage paths:

*composes:music* → *composes:speech* → *writes:speech* → *writes:oration* → *writes:sermon* → *writes:law* → *writes:philosophy* → *writes:theorem* → *writes:plan* → ...

In all, we extract talking points of the form is_*adj:noun* for over 40,000 WordNet concepts, and talking points of the form *verb:noun* for over 50,000 concepts. However, the real power of these talking points emerges from how they are connected to form a slipnet, which we discuss in section 4.

### 3.2. Talking about Similes and Stereotypes: Subjective Talking Points

If creativity is a reaction against the norm, then to understand and exploit creativity one must first understand and adequately represent this norm. In other words, to produce the extraordinary, as a human or as a machine, one must first understand the ordinary. Linguistic comparisons run the gamut from the ordinary (mundane and commonplace) to the extraordinary (i.e., novel, striking and/or humorous), and provide an excellent vehicle for understanding the interplay between norms and creativity.

Many of the normative beliefs that one uses to reason about everyday entities and events are neither strictly true nor even logically consistent. Rather, people appear to rely on a large body of folk knowledge in the form of stereotypes, clichés and other prototype-centric structures (e.g., see Lakoff, 1987). These stereotypes comprise the landmarks of our conceptual space against which other, less familiar concepts can be compared and defined. For instance, people readily employ the animal concepts Snake, Bear, Bull, Wolf, Gorilla and Shark in everyday conversation without ever having had first-hand experience of these entities. Nonetheless, our culture equips us with enough folk knowledge of these highly evocative concepts to use them as dense short-hands for all manner of behaviours and property complexes. Snakes, for example, embody the notions of treachery, slipperiness, cunning and charm (as well as a host of other, related properties) in a single, visually-charged package. To compare someone to a snake is to suggest that many of these properties are present in that person, and thus, one would well to treat that person as one would treat a real snake.

In "A Christmas Carol", Dickens (1843/1984) notes that "the wisdom of our ancestors is in the simile; and my unhallowed hands shall not disturb it, or the Country's done for" (chapter 1, page 1). In other words, stereotypical knowledge is passed down through a culture via language, most often in specific linguistic forms. The simile, as noted by Dickens, is one common vehicle for folk wisdom, one that uses explicit syntactic means (unlike metaphor; see Hanks, 2004) to mark out those concepts that are most useful as landmarks for linguistic description. Similes do not always convey truths that are universally true, or indeed, even literally true (e.g., bowling balls are not literally bald). Rather, similes hinge on properties that are possessed by prototypical or stereotypical members of a category (see Ortony, 1979; Norrick, 1986) even

if most members of the category do not also possess them. As a source of knowledge, similes combine received wisdom, prejudice and over-simplifying idealism in equal measure (Taylor, 1954). As such, similes reveal knowledge that is pragmatically useful but of a kind that one is unlikely to ever acquire from a dictionary (or, indeed, from WordNet; see Fellbaum, 1998). Although a simpler rhetorical device than metaphor, we have much to learn about language and its underlying conceptual structure by a comprehensive study of real similes in the wild (e.g., see Roncero *et al.* 2007; Moon, 2008), not least about the recurring vehicle categories that signpost this space (see Veale and Hao, 2007).

Proverbial similes do not offer the kind of authoritative, hand-curated and definitional character we find in hand-crafted resources like WordNet, but they do reflect how people typically talk about (and, perhaps, actually think of) the world. In (Veale and Hao, 2007) we argue that similes present the clearest window into the stereotypical talking points that underpin everyday conversations, and collect from the web instances of the pattern "as ADJ as a *" for thousands of WordNet adjectives. Though the simile frame is somewhat leaky in English, and prone to subversion by irony, (Veale and Hao, 2007) describes the construction of a comprehensive database of more than 12,000 highly stereotypical adjective:noun associations, such as *precise:surgeon*, *straight:arrow*, *balanced:pyramid* and *sharp:knife*. We use this data here, as the basis of an additional web harvesting process to gather stereotypical talking points of the form *has_ADJ:facet*. Simply, for every stereotypical association ADJ:NOUN in their database, we send the query "the ADJ * of a|an|the NOUN" to Google and collect noun values for the wildcard * from the first 200 hits returned for each query.

This pattern allows us to determine the conceptual attributes that are implicit in each stereotypical *adjective:noun* pairing. For instance, "*the delicate hands of a surgeon*" and "*the inspiring voice of a preacher*" reveal that *hand* is a salient attribute of surgeons while *voice* is a salient attribute of preachers. The frequency with which we find these attributes on the web also allows us to build a textured representation for each concept. So while these expanded web patterns also reveal that surgeons have a thorough *eye* and steady *nerves*, "*the hands of a surgeon*" are mentioned far more frequently and are thus far more salient to our understanding of surgeons. To avoid noise, the set of allowable attribute nouns, such as *hands, soul, heart*, *voice*, etc., is limited to the nouns in WordNet that denote a kind of trait, body part, quality,

activity, ability or faculty. This allows us to acquire meaningful talking points such as *has_magical:skill* for Wizard, *has_brave:spirit* for Lion and *has_enduring:beauty* for Diamond, while avoiding misleading talking points like *has_proud:owner* for Peacock that lack any representational value or insight. In effect, this phase of knowledge-acquisition allows us to move from a simple property-ascription representation to a richer, *frame:slot:filler* representation. In such a scheme, the property *sensitive* is a typical filler for the *hands* slot of Surgeon and the *nose* slot of Bloodhound, thereby disallowing any mis-matched comparisons between the two.

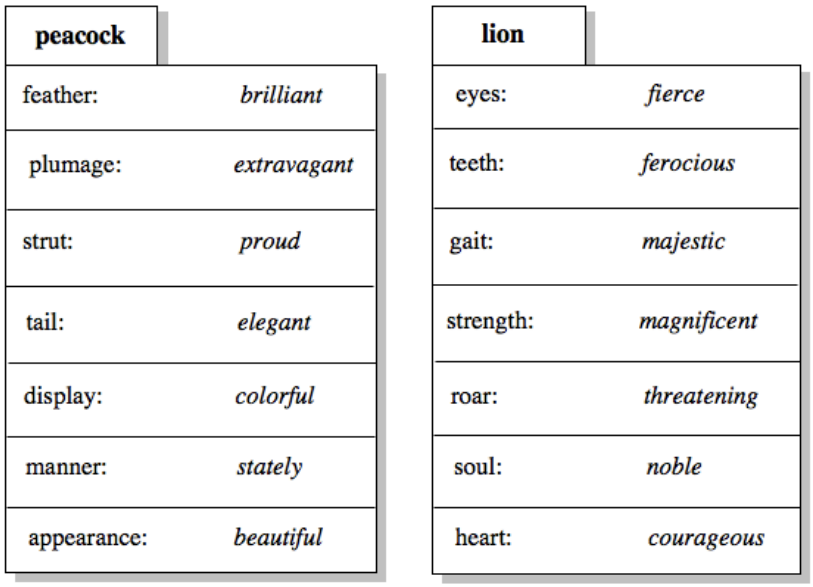


*Figure 3*. Frame:slot:filler stereotype structures for Peacock and Lion.

As can be seen in the examples of Lion and Peacock in Figure 3, the slot:filler pairs that are acquired for each concept do indeed reflect the most relevant cultural associations for these concepts. Moreover, there is a great deal of anthropomorphic rationalization of an almost poetic nature about these representations, of the kind that is instantly recognizable to native speakers of a language but which one would be hard pressed to find in a conventional dictionary (except insofar as some lexical concepts may give rise to additional word senses, such as "peacock" for a proud and flashily dressed person). Naturalistic talking

points capture how people actually conceive of and speak about concepts, and as such, they can be markedly different from the objective descriptions conventionally favored by ontologists and semanticists. Subjective viewpoints such as these often reflect a form of received wisdom that is frequently figurative and often false if judged objectively (e.g., many naturalistic descriptions of animals are based on idealized anthropomorphic models rather than zoological facts). Nonetheless, our results in section 5 will demonstrate that stereotypical talking points can yield a concise and effective means for organizing knowledge, suggesting that metaphors and similes should be taken very seriously indeed (and not simply *spirited away*) when building linguistic representations of the world.

Overall, frame representations of the kind shown in Figure 3 are acquired for 2218 different WordNet noun senses, yielding a combined total of 16,960 slot:filler pairings (or an average of 8 slot:filler pairs per frame). As the examples of Figure 3 demonstrate, these frames provide a level of representational finesse that greatly enriches the basic property descriptions yielded by similes alone, allowing a computer to appreciate e.g., that mimes and ninjas are similar by virtue of each possessing the slot:filler *Has_silent:Art* (that is, both practice the silent arts, but to very different ends).

## 4. Building a Slipnet of Talking Points

To construct a slipnet in the style of Hofstadter and Mitchell (1994), but on the scale of WordNet, we need to connect those talking points that express similar but different meanings, and to quantify the difference between these meanings. Issues of scale mean that we need only connect talking points that are close in meaning, since greater slippage can be achieved by following longer paths through the slipnet. This slippage can be based on semantic *or* pragmatic criteria. For instance, the talking points *has_sacred:authority* (obtained for Pope) and *has_sacred:power* (for God) are semantically similar since the potency sense of "authority" is a specialization of the control sense of "power" in WordNet. Likewise, *composes:speech* and *writes:speech* are similar because "compose" and "write" are synonymous in the context of literary creation, and it is this particular linkage that supports a slippage pathway from *composes:music* to *writes:poetry*. In contrast,

*is_political:movement* (for Hamas) and *is_religious:movement* (for Taliban) are pragmatically similar since movements that are religious often have a political agenda as well. We can use WordNet to construct the semantic links of the slipnet, but pragmatic links like these require world knowledge, of a kind we can find in web texts.

Two talking points *is_ADJ$_1$:OBJ$_1$* and *is_ADJ$_2$:OBJ$_2$* (or, indeed, *has_ADJ$_1$:OBJ$_1$* and *has_ADJ$_2$:OBJ$_2$*) should be connected in the slipnet if: OBJ$_1$ and OBJ$_2$ are semantically close (i.e., synonymous, or semantic siblings in WordNet); and ADJ$_1$ and ADJ$_2$ are synonymous, or ADJ$_1$ frequently implies ADJ$_2$ or ADJ$_2$ frequently implies ADJ$_1$. These implications are recognized and quantified using another web trawling process, in which the query "*as * and * as*" is used to harvest pairs of adjectives that are seen to mutually reinforce each other in web comparisons. For instance, in the web-corpus that we acquire using this query, we find that the pattern "*as religious and superstitious as*" occurs five times. Thus, the corpus reveals not just that "religious" reinforces "superstitious" (5 times), but "moral" (4 times), "political" (3 times), "conservative" (3 times), "intolerant" (2 times) and "irrational" (1 time). These connections support a slippage path from *is_religious:movement* to *is_political:movement* (pragmatic) to *is_political:campaign* (semantic) to *is_military:campaign* (pragmatic), which allows the slipnet to link Taliban (*is_religious:movement*) to Crusade (*is_military:campaign*).

### 4.1. Talking about Similes and Stereotypes: Subjective Talking Points

Slippage is a phenomenon best explained with an example, so consider again the task of creating metaphors for the concept Pope. Even the most objective talking points, when subjected to slippage pressures, can yield highly subjective metaphors. We have already seen that slippage among talking points allows Pope to be linked to the concept God via the path Pope→*has_sacred:authority*→*has_sacred:power*←God. Pope can also be linked to Rabbi by the path Pope→ *has_sacred:words*→ *has_wise:words*←Rabbi and to Judge by further extending this path:

Pope→*has_sacred:words*→*has_wise:words*→*has_solemn:words*←Judge

Black (1962) saw metaphor as an interaction between concepts, in which the interpretation of a particular source concept depends crucially on how it is able to interact with a specific target concept. This con-

cept-sensitive interplay is clearly on display here. The Pope can be metaphorically viewed as a warrior not by considering what it means for a generic person to be a warrior, but by considering how the concept Pope interacts with the concept Warrior, e.g.,

Pope→*has_infallible:voice* →*has_powerful:voice*←Warrior.

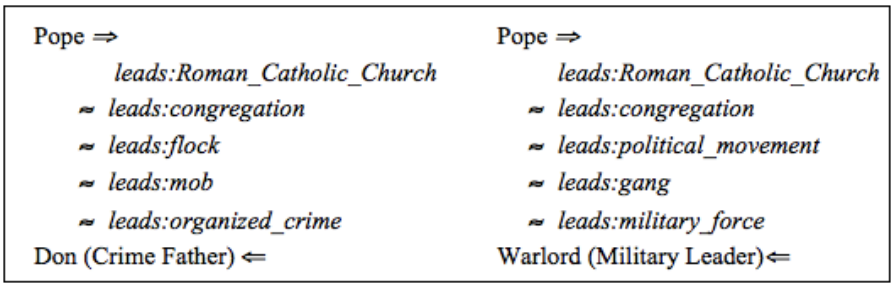


*Figure 4.* Slippage between objective talking points in subjective metaphors.

Figure 4 illustrates the potential for slippage between objective talking points as derived from WordNet. In each case (whether *Pope as Don* or *Pope as Warlord*) an agent can typically terminate a slippage path at any point, to produce different metaphors with varying semantic similarity to the starting concept. Thus, at *leads:flock* one can reach Shepherd, and from *leads:political_movement*, one can reach Civil_rights_leader.

A lexicon alone, like WordNet, is generally insufficient for metaphor processing, but such a resource can still reveal useful lexical resonances that may enrich an interpretation. In the example above, we see a resonance between the Pope, which WordNet also lexicalizes as "holy father", and a mafia Don, which WordNet also lexicalizes as "father". Indeed, since WordNet conceptualizes Roman_Catholic_Church as a specialization of Organized_religion, the metaphor establishes a parallelism between crime and religion as organized activities.

## 5. Empirical Evaluation

Objective talking points, such as those extracted from dictionary definitions, are the very warp and weft of a literal semantics (e.g., see Katz and

Fodor, 1963), of a kind that has long been favored in Artificial Intelligence. Subjective talking points, on the other hand, yield an altogether looser and folk-poetic view of the world. One can ask whether it is sensible to build a conceptual representation of a concept like Lion around the anthropomorphic ideas of a noble soul or a courageous heart, and whether a knowledge-representation that did so would actually be any good at its job. Whatever the superficial merits of a creative representation for creative language processing, there is a practical case to answer here. To understand whether subjective talking points derived from proverbial similes and ancillary linguistic forms really are sufficiently descriptive of the concepts they are acquired for, we need hard empirical evidence. To this end, we replicate here the clustering experiments of Almuhareb and Poesio (2004, 2005), which were designed to measure the effectiveness of web-acquired conceptual descriptions. Almuhareb and Poesio use WordNet as a semantic gold-standard, and attempt to automatically reconstruct the taxonomic structure of WordNet by clustering the objective features they acquire from the web for a variety of different word-concepts. If subjective talking points can be used to achieve a comparable reconstruction for the same word-concepts, we shall know that these talking points are more than a fanciful means of talking about everyday things: we shall know that they mirror, in their own way, the same ontological reality.

Almuhareb and Poesio describe two different clustering experiments. In the first, they choose 214 English nouns from 13 of WordNet's upper-level semantic categories, and proceed to harvest property values for these concepts from the web using the pattern "a|an|the * C is|was" (as in "*the red car is new*"). This pattern yields a combined total of 51,045 values for all 214 nouns; these values are primarily adjectives, such as *hot, black*, etc., but noun-modifiers of C are also allowed, such as *fruit* for cake. They also harvest 8934 attribute nouns, such as *temperature* and *color*, using the query pattern "the * of the C is|was" (as in "*the temperature of the coffee is hot*"). These values and attributes are then used as the basis of a clustering algorithm to partition the 214 nouns back into 13 different categories that hopefully resemble the 13 WordNet categories from whence they were originally drawn. Comparing these clusters with the original WordNet-based groupings, Almuhareb and Poesio report a cluster accuracy of 71.96% using just values like *hot* (all 51,045 values), an accuracy of 64.02% using just attributes like *tem-*

*perature* (all 8,934 attributes), and an accuracy of 85.5% using both together (all 59,979 features combined).

In a second, larger experiment, Almuhareb and Poesio select 402 nouns from 21 different semantic classes in WordNet, and proceed to harvest 94,989 property values (again mostly adjectives) and 24,178 attribute nouns from the web using the same retrieval patterns. They then applied the *repeated bisections clustering* algorithm to this larger data set, and report an initial cluster purity measure of 56.7% using only property values like *hot*, 65.7% using only attributes like *temperature*, and 67.7% using both together (cluster purity is measured relative to WordNet; a purity of 100% would indicate a completely faithful reconstruction). Suspecting that noisy features contribute to the perceived drop in performance in the second experiment, Almuhareb and Poesio then apply a variety of noise filters to reduce the value set to just 51,345 values and the attribute set to just 12,345 attributes, for a size reduction of about 50% in each case. This in turn leads to an improved cluster purity measure of 62.7% using property values only and 70.9% using attributes only. Surprisingly, filtering actually appears to reduce the clustering performance of both sets together to 66.4%.

*Table 1.* clustering accuracy for experiment 1 (214 nouns).

| ***Approach*** | ***Values only*** | ***Attr's only*** | ***All (V + A)*** |
|---|---|---|---|
| Almu. + Poesio | 71.96%<br>(51045 vals) | 64.02%<br>(8934 attr) | 85.51%<br>(59979 v+a) |
| Stereotypical talking points | 70.2%<br>(2209 vals) | 78.7%<br>(4974 attr) | 90.2%<br>(7183 v+a) |

*Table 2.* clustering accuracy for experiment 2 (402 nouns).

| ***Approach*** | ***Values only*** | ***Attr's only*** | ***All (V + A)*** |
|---|---|---|---|
| Almu. + Poesio (no filtering) | 56.7%<br>(94989 vals) | 65.7%<br>(24178 attr) | 67.7%<br>(119167 v+a) |
| Almu. + Poesio (with filtering) | 62.7%<br>(51345 vals) | 70.9%<br>(12345 attr) | 66.4%<br>(63690 v+a) |
| Stereotypical talking points | 64.3%<br>(5547 vals) | 54.7%<br>(3952 attr) | 69.85%<br>(9499 v+a) |

We replicate here both of these experiments using the same data-sets of 214 and 402 nouns respectively. For fairness, we collect *raw* descriptions for each of these nouns directly from the web, and use no filtering (manual or otherwise) to remove poor or ill-formed descriptions. We thus use the pattern "as * as a|an|the C" to collect 2209 raw adjectival values for the 214 nouns of experiment 1, and 5547 raw adjectival values for the 402 nouns of experiment 2. We then use the pattern "the ADJ * of a|an|the C" to collect 4974 attributes for the 214 nouns of experiment 1, and 3952 for the 402 nouns of experiment 2; in each case, ADJ is bound to the raw adjectival values that were acquired using "as * as a|an|the C". A comparison of clustering results is given in Tables 1 and 2.

These tables illustrate that clustering is most effective when it is performed on the basis of both values *and* attributes (yielding the highest scores, 90.2% and 69.85%, in each experiment respectively). These results thus support the combination of conceptual attributes with specific adjectival values into integrated talking points which reflect how people actually talk about the concepts concerned.

Overall, the results are quite telling: subjective talking points acquired from proverbial similes out-perform objective feature-sets in reconstructing the hand-crafted category structure of WordNet, a resource that is itself founded on the notion of an objective, literal worldview. As the numbers suggest, subjective talking points achieve greater clustering performance with far greater concision: Table 2 reveals that between 6 and 12 times fewer features are required if these features better capture the essence of what is being described. In these experiments, clearly, an ounce of subjective insight from proverbial similes outperforms a pound of objective information from other sources. As Dickens noted, there is time-tested wisdom in proverbial similes, wisdom enough to build a flexible knowledge-representation for more ambitiously creative ends.

## 6. Conclusions

Metaphor and its creative ilk are knowledge-hungry devices, so any computational treatment can only be as good as the knowledge representation that supports it. Moreover, from a computational perspective, any theory of creative comparison – cognitive, linguistic, or otherwise –

is only as good as the algorithmic and representational insights that it provides, and the scale of the implementation that it ultimately allows us to realize. In this paper we have given computational form to some of the key insights in the metaphor literature, from the interaction theory of Black (1962) to the salience imbalance theory of Ortony (1979) to the theory of conceptual blending of Fauconnier and Turner (1998). We also employ a key insight from the work of Hofstadter and his fluid analogies group (1995), that robust reasoning on a conceptual level requires a degree of slippage that must be supported by the underlying knowledge representation. Our knowledge base of talking points is derived from two complementary information sources: the objective definitions contained in WordNet (Fellbaum, 1998) and the stereotypical comparisons that pervade everyday language and which consequently pepper the texts of the web. These sources yield a knowledge-base that is neither small nor hand-crafted. While the knowledge-base needs to grow by at least an order of magnitude, slippage means that non-identical talking points can be treated as equivalent for purposes of robust processing, which in turn extends the *halo* of talking points that surrounds each concept in the knowledge-base (Hofstadter et al., 1995). The experiments of section 5 also indicate that, in a pinch, new talking points for a previously under-represented concept can be acquired dynamically from the web with reasonable accuracy.

But what does it mean to state, at a knowledge-representation level, that *lions and knights* both have a *brave heart*, that wolves and tyrants both have a *cruel face*, or that eagles and warriors have a *fierce expression*? Stereotypical talking points such as these can be poetic or metaphorical, and may express a viewpoint that is overly simplistic, subjective or even technically inaccurate. Nonetheless, our experiments suggest that the linguistic insights we acquire from non-literal descriptions strongly reflect our ontological intuitions about concepts and are more than mere linguistic decorations. Most significantly, we see from these experiments that stereotypical talking points yield an especially concise representation, since with no filtering of any kind, this approach achieves comparable clustering results with feature sets that are many times smaller than those used in previous work. We anticipate therefore that stereotypical descriptions will be a key growth area for the development of our talking points knowledge-base.

We conclude by noting that the computational model described in this paper can be accessed on-line, at the Web-site for the *Creative Lan-*

*guage Systems Group* – http://afflatus.ucd.ie – in a variety of online metaphor applications.

## References


Almuhareb, Abdulrahman and Massimo Poesio
2004 Attribute-Based and Value-Based Clustering: An Evaluation. In *Proceedings of EMNLP*. Barcelona, July, 2004.
2005 Concept Learning and Categorization from the Web. In *Proc. of the annual meeting of the Cognitive Sci*ence society, Italy, July

Barnden, John A. and Mark G. Lee
2002 An Artificial Intelligence Approach to Metaphor Understanding. *Theoria et Historia Scientiarum*, 6(1):399-412.

Black, Max
1962 *Models and Metaphor: studies in language and philosophy*. Ithaca, NY: Cornell University Press.

Camac, Mary K. and Sam Glucksberg
1984 Metaphors do not use associations between concepts, they are used to create them. *Journal of Psycholinguistic Research*, **13**(6).

Charniak, Eugene and Mathew Berland
1999 Finding parts in very large corpora. In Proc. of the 37th Annual Meeting of the ACL, pp 57–64.

Cimiano, Philipp and Johanna Wenderoth
2007 Automatic Acquisition of Ranked Qualia Structures from the Web. *In proceedings of ACL 2007, the 45th Annual Meeting of the Association of Computational Linguistics*, pages 888-895. Prague, Czech Republic.

Collins, Allan M. and Elizabeth. F. Loftus.
1975 A Spreading-Activation Theory of Semantic Processing. *Psychological Review* **82**, pp 407-428.

Dickens, Charles
1843/1984 A Christmas Carol.: Puffin Books, Middlesex, UK.

Dong, Zhendong and Qiang Dong
2006 *HowNet and the Computation of Meaning*. World Scientific: Singapore.

Etzioni, Oren, Stanley Kok, Steven Soderland, Michael Cafarella, Ana Popescu., Daniel Weld, Douglas Downey, Tal Shaked and Alexander Yates
2004 Web-scale information extraction in KnowItAll (preliminary results). In *Proceedings of the 13th WWW Conference*, pp 100–109.

Fauconnier, Gilles and Mark Turner
1998 Conceptual Integration Networks. *Cognitive Science*, 22(2):133–187.

Fellbaum, Christianne (ed.).
1998 *WordNet: An electronic lexical database*. The MIT Press. (1985). *A comprehensive grammar of the English*.

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

Hanks, Patrick
2004 The syntagmatics of metaphor. *Int. Journal of Lexicography*, 17(3).

Harabagiu, Sanda, George Miller and Dan Moldovan
1999 WordNet2 - a morphologically and semantically enhanced resource. In *Proceedings of SIGLEX-99*, pp 1-8, University of Maryland.

Hearst, Marti
1992 Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics*, pp 539–545.

Hofstadter, Douglas R. and Melanie Mitchell
1994 The Copycat Project: A model of mental fluidity and analogy-making. *In Holyoak, K.J. & Barnden, J. A. (Eds.) Advances in Connectionist and Neural Computation Theory*, Vol. 2. Norwood, NJ: Ablex.

Hofstadter, Douglas R. and the Fluid Analogy Research Group
1995 *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. NY: Basic Books.

Indurkhya, Bipin
1992 *Metaphor and Cognition: Studies in Cognitive Systems.* Kluwer Academic Publishers, Dordrecht: The Netherlands.

Katz, Jerold J. and Fodor, J. A.
1963 The structure of a semantic theory. *Language* (2): pp. 170–210.

Lakoff, George
1987 *Women, fire and dangerous things*. Chicago University Press.

Martin, James H.
1990 *A Computational Model of Metaphor Interpretation.* NY: Academic Press**.**

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

Norrick, Neal
1986 Stock Similes. *Journal of Literary Semantics* XV(1), 39-52.

Ortony, Andrew
1979 Beyond literal similarity. *Psychological Review, 86,* 161-180.

Pustejovsky, James
1991 The generative lexicon. *Computational Linguistics* 17(4), pp 209–441.

Roncero, Carlos, John M.Kennedy and Ron Smyth
2006 Similes on the internet have explanations. *Psychonomic Bulletin and Review*, 13(1), 74-77.
Taylor, Archer
1954 Proverbial Comparisons and Similes from California. *Folklore Studies* 3. Berkeley: University of California Press.
Veale, Tony and Yanfen Hao
2007 Making Lexical Ontologies Functional and Context-Sensitive. *In proceedings of ACL 2007, the 45th Annual Meeting of the Association of Computational Linguistics*, pages 57–64.
Veale, Tony and Diarmuid O'Donoghue
2000 Computation and Blending. *Cognitive Linguistics*, 11(3-4), special issue on Conceptual Blending.
Way, Eileen Cornell
1991 Knowledge Representation and Metaphor. *Studies in Cognitive systems*. Holland: Kluwer.
Wilks, Yorick
1978 Making Preferences More Active, *Artificial Intelligence* **11.**