Talking Points in Metaphor: A Concise Usage-based Representation for Figurative Processing

Tony Veale and Yanfen Hao[1]

Abstract. An effective speaker can use metaphor to communicate a wealth of propositions and affective attitudes with a single juxtaposition of ideas [12,8,6,10,7,3,15]. But as such, an effective metaphor requires effective communication, which in turn requires that the speaker has a clear idea of the content to be communicated, and an equally clear understanding of which conceptual vehicles best communicate this content. We present here a concise corpus-derived meaning representation for metaphor processing that captures the most widely-used *talking points* that are evoked in everyday metaphors and similes. We illustrate how these talking points can be acquired by harvesting the web, and further show how comparable but discretely different talking points can be reconciled during metaphor processing. Finally, by replicating the clustering experiments of [1], we show that talking points yield an especially concise representation of concepts in general.

1 INTRODUCTION

Though sometimes fanciful and frequently indeterminate, metaphor is, at heart, a communication device. As such, metaphor can only support effective communication when it employs a vehicle of comparison whose import is well understood by both speaker and audience. For this reason, metaphors make frequent use of a communal inventory of consensus imagery, attitudes and beliefs. Though one can find reflections of this shared knowledge in hand-crafted semantic resources like Cyc [9], WordNet [11] and HowNet [5], the best guide to this communal inventory is provided by the stereotypes that pervade our everyday language. These stereotypes provide the talking *points* that underlie our most effective similes and metaphors: in this inventory, snakes evoke cunning and poisonous charm; lions evoke nobility and bravery; tigers evoke fierceness and feline grace, whales evoke grandeur and massiveness, scientists evoke objectivity and intellectual rigor, and typhoons evoke events of devastating power. These stereotypes comprise both a cultural legacy [8] and a widely accepted linguistic currency [14], so much so that they are often used by speakers who have never actually encountered the physical entities described by them.

We see a talking point as any part of the conventional view of a concept that is primed whenever that concept is employed in discourse. As received wisdom, talking points may reflect an anthropomorphic bias, an idealized world-view [8] or an outdated scientific belief; but what matters most here is that they each embody a belief that is widely held and readily evoked by certain words.

[1] School of Computer Science and Informatics, University College Dublin, Ireland, email: {tony.veale, yanfen.hao}@ucd.ie

Computational models of metaphor processing stand or fall on the knowledge they have available to them [10,7]. But the knowledge required for metaphor processing is not special figurative knowledge that is qualitatively different from that required for other language-processing tasks. Rather, because metaphor is used to communicate the talking points that are most salient for an entity in a given situation, this is the same knowledge needed to categorize those entities and situations and to determine our affective and inferential response to these concepts [14]. It follows that by understanding how metaphor shapes and exploits our shared view of the world, we can design and acquire richer and more flexible models of world knowledge.

In sections 2 and 3 we motivate and describe the construction of a comprehensive knowledge base of the most common talking points in everyday language, by harvesting stereotypical allusions from the texts of the web. In section 4 we then describe how this talking-points database can be used to form a robust computational basis for a scaleable model of metaphor generation and comprehension. In section 5 we show that talking points are not just a very concise means for reasoning within metaphor, but an especially concise representation for capturing the most important elements of conceptual description in general, as judged by the ability to cluster ontologically-related words and ideas.

2 RELATED WORK

Metaphor has long been recognized as a knowledge-driven process, from the hierarchy-traversing approaches of [15,7] to the graph-mapping approaches of [6,13]. However one cuts it, metaphor requires insightful knowledge about the words and ideas it employs. While psychologists, philosophers and cognitive linguists can make a general appeal to the notion of world knowledge when proposing a schematic view of metaphor [8], computationalists must actually furnish this knowledge, in a detailed representational form, if they are to gain enough traction for an implemented approach. Martin [10] attempts a balance between the schematic view of the cognitivists and the representational demands of a real working model; he does so by focusing on conventional metaphors (such as "to *catch* a cold", "to *kill* a process", etc.) and how these can be extended and elaborated in a question-answering/ advice-giving context. Likewise, Barnden and Lee [3] focus on the knowledge needed to comprehend the metaphors of mind that one finds in commonplace utterances such as "to grasp an idea" and "to have doubts at the back of one's mind". Other approaches, such as that of Fass [7], understand metaphor as an aberration relative to literal meaning, and require a rich representation of this literal reality as a diagnostic against which metaphoric statements can be repaired. Veale [13] applies the structure-mapping approach to literal knowledge extracted from the HowNet bilingual ontology [5], but notes that this knowledge is both sparse and unevenly distributed: the lexical concepts for which knowledge can be extracted are not necessarily the ones that are commonly used in metaphors.

Since metaphor pervades language, the knowledge needed for metaphor processing should be evident in everyday language. Almuhareb and Poesio [1,2] describe how conceptual descriptions comprising both attributes (such as *Temperature*) and values (such as *hot*) can be extracted from syntagmatic patterns on the web (such as "the A of X is ..."), with enough insight to permit the concepts so described to be clustered accurately with respect to WordNet's noun hierarchy. Moving closer to metaphor, Veale and Hao [14] describe how more insightful and prototypical feature values for concepts can be derived from simile patterns on the web, such as "as X as a Y". Though the simile pattern is inherently leaky in English, these authors show that with some human annotation, a sizable database of the most stereotypical ascriptions (e.g., pearls are lustrous, wolves are ruthless, prophets are inspiring, etc.) can be acquired and then exploited in the generation of metaphors.

This latter work is limited to feature-centric metaphors only, ones that hinge on the sharing of one of more feature values (e.g., magicians and surgeons are both skilled, supermodels and greyhounds are both skinny, etc.). In this paper we extend the work of Veale and Hao by acquiring detailed talking points that relate, in the style of Almuhareb and Poesio, specific values to salient attributes of a concept; we can thus speak of the eloquent delivery of an orator, the lithe body of a panther, and the powerful aroma or pungent taste of espresso. This in turn supports greater finesse in the generation and comprehension of metaphors, allowing our system to understand that the same (or similar) features relate to two different concepts in the same (or a similar) manner.

3 ACQUIRING TALKING POINTS

We use the simile work of Veale and Hao as the basis for this current foray into conceptual description. To recap here, those authors describe how the query pattern "as ADJ as a *" can be used to harvest web similes in a two-pass bootstrapping process: first, WordNet [11] is used to furnish different values for the adjective field ADJ, while the wildcard * is used to collect associated nouns for those adjectives via the Google API; the query "as * as a NOUN" is then used to collect stereotypical adjectives for those nouns in a second phase of harvesting. By analyzing 200 hits for each query, Veale and Hao acquire 74,704 potential simile instances, yielding 42,618 potentially stereotypical associations between adjectives and nouns. When human judges are used to annotate these associations, many are rejected as noise or explicitly marked as ironic (e.g., "as bullet-proof as a sponge-cake"), but the remaining 12,259 bona-fide pairings yield a comprehensive database of stereotypical descriptions for over 4000 nouns.

But how useful is it to know that drums are taut and pearls are lustrous? When used as the basis of a metaphoric comparison (e.g., to describe a given person), it is equally important to know that these properties stereotypically refer the outer appearance of an entity. In other words, *taut skin* and *lustrous sheen* are conventional talking points of drums and pearls. We now demonstrate how an additional phase of web harvesting can turn these simple property ascriptions into the talking points needed to drive metaphor processing.

For every stereotypical pairing of ADJ and NOUN, we send the query "the ADJ * of a|an|the NOUN" to Google and again scan 200 result snippets for each to identify possible noun values for *. As in Almuhareb and Poesio [1,2], these queries allow us to determine the specific attributes targeted by different property ascriptions. In this way, we find that lions, tigers and cannons all have an angry roar, plums, pearls and ball-bearings all have a smooth surface, and gods, shrines and torahs all have a sacred purpose. These triads of concepts are comparable by virtue of having the same values for the same attributes, or in other words, because they share the same talking points.

Talking points should offer genuine insights into a commonplace idea, but not all adjective:noun pairings will be insightful enough to be considered a talking point. For instance, while *colorful:plumage* and *proud:strut* capture a faithful picture of the stereotypical peacock, the pairings *colorful:environment* and *proud:owner* do not. Notions such as environment and owner can certainly be relevant in some contexts, but they describe qualities that are clearly contingent or extrinsic to the concept of peacock. Likewise, "the skilled son of a surgeon" does not speak directly to the concept of surgeon, so we should take care not to mistake pairings like *skilled:son* as talking points.

We thus make the simplifying assumption that the intrinsic qualities of an entity to which a talking point might pertain can be characterized as one of: the traits possessed by an entity (e.g., grace, strength, skill); the physical properties of an entity (e.g., color, length, weight); the feelings experienced by an entity (e.g., sorrow, courage, soulfulness); the body parts that comprise a physical entity (e.g., hands, eyes, teeth); and the actions that an entity can perform (e.g., roar, bite, cry, gallop). This in turn allows us to use WordNet [11] as a filter for likely talking points, since each of these categories of intrinsic quality corresponds to one or more sub-trees in the WordNet hierarchy of noun senses. Applying this filter to the results of the aforementioned web search, we obtain 22,693 talking points, linking 1360 adjectival values to 1950 different vehicle nouns via 1796 attribute nouns; each stereotypical term has, on average, 9 talking points.

4 TALKING POINTS IN METAPHOR

When we speak of the *fluid gait* of a dancer, we employ a specific talking point that ties a given attribute (gait) to a specific value (fluid). Specific talking points thus allow a concept like dancer to serve as a useful vehicle of comparison in metaphors and similes where it can project specific qualities onto a target concept. For instance, one might compare a boxer to a dancer to highlight the former's fluidity of movement and sense of balance. It is not necessary that *fluid:gait* be an established talking point of boxer for this metaphor to be apt, but it is enough that gait is a meaningful attribute of boxers to describe. Indeed, following [12], it is preferable if the target concept does not already possess the talking points in question, since metaphor functions best when it is used to make novel but believable claims about a target.

4.1 Understanding Metaphor with Talking Points

Similarity between concepts, and the ability to measure it, is therefore central to the interpretation of metaphors. Though an elusive idea that can mean different things in different contexts, similarity has been operationalized in the context of WordNet in a variety of simple metrics [4]. Consider then the interpretation of a metaphor in which a vehicle V (such as dancer, or matador) is used to describe a target T (such as boxer or prize-fighter); similarity plays a role in each of the following comprehension strategies:

1. V and T are sufficiently similar (e.g., both are persons, animals, artifacts, events, etc.) that the established talking points of V can be projected directly onto T.

2. V and T are semantically distant, but the talking points of V concern attributes that one can meaningfully ascribe to T. A corpus can be used to ascertain the degree to which these attributes (such as speed, grace, teeth, etc.) are salient of T. For instance, the pattern "the * of a *" identifies *stance* and *grace* as salient attributes of prize-fighter in the Google web IT corpus.

3. Both V and T have their own talking points, but these are similar enough to be reconciled, as in the *skilled hands* of a surgeon and the *creative hands* of an artist. This reconciliation process is described in section 4.3.

4. If the system possesses no talking points for V, a set of potential talking points is established by looking at the semantic neighbors of V (e.g., in WordNet). For instance, if V is gladiator then it can borrow the talking points *solid:strength* and *offensive:capability* from fighter (a generalization of gladiator), and *strong:grip*, *muscular:strength* and *powerful:body* from wrestler (a specialization of fighter and a sibling of gladiator).

These four strategies can apply singly or in combination to produce meaningful interpretations of *T is(like) V*. As noted in strategy 2, a corpus can be used to determine the most salient attributes of T to describe, allowing the elements of the interpretation to be ranked accordingly.

4.2 Generating Metaphor with Talking Points

Metaphor generation is a considerably more open-ended and creative process. We briefly consider here the generation of potentially apt vehicles for a given target T. This first requires that we determine the attributes of T that can be meaningfully described; then we can consider potential values for these attributes, which will yield a set of potential talking points that can be ascribed to T; we should then identify the concepts that best evoke those talking points and which can sensibly be compared to T.

As noted previously, a large corpus (such as the Google web IT corpus of frequent n-grams) can be used to ascertain the different attributes that are relevant to a concept T. For instance, the query pattern "* of a|an|the *" identifies the attributes *work, soul, spirit, understanding, mind, eye, words* and *influence* as the most frequently cited aspects of a philosopher, in that order. Now consider the attribute *understanding*, which underpins talking points that have been acquired for the following entities:

loving:understanding	of mother, father, dog, baby
systematic:understanding	of science
compassionate:understanding	of priest, mother
technical:understanding	of scientist
imaginative:understanding	of poet

To accentuate the imaginative powers of a philosopher, we can thus use *poet* as a vehicle; for technical insight, we should use *scientist*; to suggest compassion, we should use *priest* or *mother*; and so on. WordNet-based similarity metrics will indicate that it requires an unlikely semantic stretch to compare a philosopher to a science, so this potential metaphor is discarded. However, if *technical:understanding* and *systematic:understanding* can be identified as conveying similar ideas, the latter talking point will also be evoked by the vehicle scientist. We now show how related talking points can be reconciled.

4.3 Reconciling Similar Talking Points

Two talking points may have discretely different linguistic forms yet still convey much the same content. This issue arises because talking points are not hand-crafted by a semanticist but harvested automatically from the web. Of course, WordNet can help us to reconcile those talking points whose linguistic elements are synonymous or related by hyponymy, such as the *sleek:beauty* of a greyhound with the *sleek:appearance* of a seal or a yacht. However, as seen in the case of *technical:understanding* and *systematic:understanding*, similarity between talking points can be as much a pragmatic as a semantic issue.

We turn again to usage-based insights from a corpus to resolve this problem. In particular, we turn to similes in which a speaker applies two related adjectives simultaneously to a vehicle, as in "as hot and spicy as a curry". We can expect two adjectives ADJ_1 and ADJ_2 to be related to the extent that we can find instances of the double-adjective construction "as ADJ_1 and ADJ_2 of X". So by collecting the web frequencies of this pattern for all pairings of ADJ_1 and ADJ_2, we construct a confusion matrix of adjectives that indicates the likelihood of one adjective being used to evoke and reinforce the other. The top 10 co-descriptors of *systematic* in this matrix are: *objective, comprehensive, thorough, impartial, rigorous, regular, relentless, unbiased, complete* and *logical*.

Likewise, we can expect two attribute nouns to reinforce and suggest each other to the extent that they are found in double-noun constructions like "with NOUN_1 and NOUN_2" and "the NOUN_1 and NOUN_2 of X". We thus construct a confusion matrix of nouns that indicates the likelihood of one attribute being used to evoke or reinforce another. The top 10 co-descriptors of *understanding* in this matrix are: *knowledge, love, compassion, appreciation, patience, sympathy, wisdom, care, sensitivity* and *insight*.

Taken together, these matrices allow talking points that comprise different adjectives and nouns to be seen to communicate similar ideas if these adjectives and nouns exhibit a sufficient degree of co-description in a corpus. This gives rise to a slippage network of similar talking points that allows a system to more thoroughly explore the space of possible metaphors, similes and blends.

5 EMPIRICAL EVALUATION

Many of the talking points we acquire from the web express viewpoints that are far from objective. In fact, some are strikingly poetic, suggesting that talking points are an ideal basis for capturing the insights needed for metaphor. For instance, we find that lions are believed to have a *kingly:roar*, a *majestic:gait* and a *noble:heart*, while warriors have a *courageous:heart* and a *heroic:path*. But one can ask whether these talking points are merely decorative, or whether they actually reflect the essential qualities of concepts. We aim to demonstrate the latter, by replicating the clustering experiments of Almuhareb and Poesio [1,2], who in turn demonstrated that conceptual features that are web-mined from specific textual patterns can be used to construct WordNet-like concept clusters. These authors used different text patterns for mining adjectival values (like *hot*) and noun attributes (like *temperature*), and their experiments evaluated the relative effectiveness of each as a means of ontological clustering.

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 adjectival values for these concepts from the web using the pattern "a|an|the * C is|was". This pattern yields a combined total of 51,045 adjectival values for all 214 nouns, such as *hot*, *black*, etc. They also harvest 8934 attributes, such as *temperature* and *color*, using the query pattern "the * of the C is|was". These values and attributes are then used as the basis of a clustering algorithm to partition the 214 nouns into 13 categories, in an attempt to re-construct their original semantic groupings. Comparing these clusters with the original WordNet-based groupings, Almuhareb and Poesio report a cluster accuracy of 71.96% using just values (all 51,045), an accuracy of 64.02% using just attributes (all 8934), and an accuracy of 85.5% using both together (59979 features).

In a second, larger experiment, Almuhareb and Poesio select 402 nouns from 21 different semantic classes in WordNet, and proceed to harvest 94,989 adjectival values and 24,178 noun attributes 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 adjectival values only, 65.7% using noun attributes only, and 67.7% using both together. Suspecting that noisy feature sets had contributed to the apparent drop in performance, those authors then proceeded to apply a variety of noise filters to reduce the adjectival value set to just 51,345 values and the attribute set to just 12,345 nouns, for a size reduction of about 50% in each case. This in turn leads to an improved cluster purity measure of 62.7% using adjective values only and 70.9% using noun attributes only. Surprisingly, filtering reduces the clustering performance of both sets together to 66.4%.

We replicate here both of these experiments using the same datasets of 214 and 402 nouns respectively. For fairness, we collect *raw* talking points for each of these nouns directly from the web, and use no filtering (manual or otherwise) to remove poor or ill-formed talking points. 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". In effect then, we harvest not just attributes but talking points. A comparison of clustering results is given in Tables 1 and 2.

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

Approach	*Values only*	*Attr's only*	*All (V + A)*
Almu. + Poesio	71.96% (51045 vals)	64.02% (8934 attr)	85.51% (59979 v+a)
Talking Points	70.2% (2209 vals)	78.7% (4974 attr)	90.2% (7183 v+a)

5.1 Discussion of Results

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 single integrated features, which we have dubbed *talking points* in this paper.

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

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

As designed by Almuhareb and Poesio, these experiments are not intended to measure poetic or metaphoric potential but the simple ability to capture those aspects of a concept that are responsible for how the concept is ontologically organized. As such, these experiments suggest that the linguistic insights we acquire from similes and metaphors - even when figurative - strongly reflect the essential qualities of concepts and are more than mere decorations.

Most significantly, we see from these experiments that talking points yield an especially concise representation. With no filtering of any kind, the talking points approach achieves comparable clustering results with feature sets that are many times smaller than those used in [1,2]

6 CONCLUSIONS

In this paper we have presented a contextualized model of metaphor understanding and generation that derives its pragmatic sensibilities from the simple analysis of text corpora. In this usage-based view, widely held and culturally established talking points are harvested from the global texts of the web, while potential talking points are collected from local, context-specific texts. Thus, a comparison between a prize-fighter and a dancer carries greater pragmatic force in contexts (and w.r.t. corpora) where there is a precedent of speaking of the grace of a prize-fighter, or of combatants in general.

Though a simple pairing of adjective and noun, each talking point offers a valuable glimpse into how a word/concept is popularly imagined and construed. Talking points are general enough to be metaphorically transferable between concepts, but specific enough to capture that part of each concept that determines how it is ontologically categorized. Furthermore, this glimpse is frequently insightful in a way that dictionary definitions and ontological specifications are not. For instance, consider the insight we gain from the talking point that pyramids have a *balanced form*. This reflects the physical intuition one has about a stereotypical pyramid, whose broad square base and tapering structure lends it a low center-of-gravity. This intuition is further elaborated in the web simile "as <u>un</u>balanced as an *upturned pyramid*". These insights are clearly grounded in visual imagination and an embodied sense of how objects behave in the real world. Though not linguistic in nature, these insights can nonetheless be acquired from linguistic data sources by observing how speakers employ concepts for the purposes of illuminating comparison and description via simile.

The talking-points model thus offers a bootstrapping approach to figurative processing, in which the knowledge required to understand and generate similes and metaphors can readily be acquired from corpora by observing how others make and appreciate such statements. Furthermore, this is a computational approach to language in which metaphor and simile earn their own keep, by helping us to identify those stereotypical elements of conceptual description that are most useful for inference and categorization. In this view, metaphor shifts from being a vexing and sometimes fanciful problem of language processing to being a convenient window through which a system can acquire insightful knowledge about the world.

REFERENCES

[1] A. Almuhareb and M. Poesio, Attribute-based and Value-based clustering: an evaluation, *In Proc.* of EMNLP, 158–165, Barcelona, 2004.

[2] A. Almuhareb and M. Poesio, Concept Learning from the web, *In Proc. of the 27th Meeting of the Cognitive Science society*, New Jersey: Lawrence Erlbaum, 2005.

[3] J.A. Barnden and M.G. Lee, An Artificial Intelligence Approach to Metaphor Understanding, *Theoria et Historia Scientiarum*, **6(1)**, 399-412, (2002).

[4] A. Budanitsky and G. Hirst, Evaluating WordNet-based Measures of Lexical Semantic Relatedness, *Computational Linguistics*, 32(1):13–47, (2006).

[5] Z. Dong and Q. Dong, *HowNet and the Computation of Meaning*, World Scientific: Singapore, 2006.

[6] B. Falkenhainer, K. Forbus and D. Gentner, Structure-Mapping Engine: Algorithm and Examples, *Artificial Intelligence*, **41**, 1–63, (1989).

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

[8] G. Lakoff, *Women, fire and dangerous things*, Chicago University Press, 1987.

[9] D. Lenat and R.V. Guha, *Building Large Knowledge-based Systems*, Addison Wesley, 1990.

[10] J.H. Martin, *A Computational Model of Metaphor Interpretation*, New York: Academic Press, 1990.

[11] G.A. Miller, WordNet: A Lexical Database for English, *Communications of the ACM*, Vol. 38 No. 11, (1995).

[12] A. Ortony, Beyond literal similarity, *Psychological Review*, **86**, 161–180, (1979).

[13] T. Veale, Analogy Generation with HowNet. *In Proc. of IJCAI05, the 19th joint int. conf on AI*, 1148-53, 2005.

[14] T. Veale and Y. Hao, Making Lexical Ontologies Functional and Context-Sensitive, *In Proc. of ACL, the 45th Ann. Meeting of the Assoc. of Computational Linguistics*, 57–64, 2007.

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