

# Generating and Understanding Creative Comparisons

**Keywords:** simile, irony, web-mining

Many of the beliefs that one uses to reason about everyday entities and events are neither strictly true or 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 prototypes comprise the landmarks of our conceptual space against which other, less evocative concepts can be compared and defined. For instance, people readily employ animal concepts like Snake, Bear, 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. But despite this frequent reliance on stereotypes, language users can also be remarkably creative in the comparisons they construct, employing vehicles that are semantically incongruous to make irony-laden statements that are willfully false. Ironically speaking, a bald person may be as hairy as a bowling-ball, a dull vista may be as inspiring as a parking lot, and a painful event may be as pleasurable as a root-canal operation. Understanding how such creative comparisons are employed will not just enable models of creative language, but offer valuable insights about how general mechanisms of human creativity identify and exploit non-obvious similarities between very different concepts.

To model our capacity for creative comparison, we use automated web-harvesting techniques to acquire the largest database of simile-based comparisons of its kind. Using the Google API, we gather and annotate thousands of unique similes of the "*as ADJ as a NOUN*" form, identifying the adjectival properties that they ascribe and the landmark nouns that best evoke those properties. Using human judges, we divide these comparisons into bona-fide and ironic categories, yielding two rich case-bases against which a computer can understand the implications of novel comparisons and with which a computer can generate creative comparisons of its own. Though our approach has mostly concentrated on comparisons in English, we also demonstrate its applicability to other languages such as Chinese, and report cross-cultural differences on the prevalence of creativity (in the guise of novel or ironic comparisons) in both languages. Finally, we describe an implemented, WordNet-based computational system, accessible on-line, that allows web-users to generate and analyze creative comparisons for topics of their own choosing.