

Zeitgeist: A Computational Model of Neologism Processing

Tony Veale,

School of Computer Science and Informatics, University College Dublin, Ireland.

Language is a dynamic landscape in which words are not fixed landmarks, but fickle signposts that switch their directions as archaic senses are lost and new, more topical senses, are gained. Frequently, entirely new lexical signposts are added as newly minted word-forms enter the language. Some of these new forms are cut from whole cloth and have their origins in creative writing, movies or games. But many are patchwork creations whose origins can be traced to a blend of existing word forms (e.g., Dent, 2005). This latter form of neologism is of particular interest to the computational lexicographer, since such words possess an obviously compositional structure from which one can begin to infer meaning. In this paper, we demonstrate that, if given enough structural context, an automated system can assign a sufficiently rich semantic structure to these words to allow their broad meanings to be automatically incorporated into an electronic dictionary like WordNet (Miller, 1995). When tied to a system for harvesting new word forms from topical internet resources like Wikipedia, this capability allows for a dynamic computational lexicon that grows itself in response to a changing language and cultural context.

We shall present a computational model of new-word formation, called *Zeitgeist*, that employs a collection of word-formation schemata to harvest previously unseen neologisms from Wikipedia. We further describe how these schemata exploit the semantic context provided by Wikipedia's topology of cross-references to automatically assign meanings to novel portmanteau words. Because this topological context often fails to deterministically capture the precise meaning of a new word, *Zeitgeist* also employs a computational instantiation of the CL notions of *hedging* (e.g., Lakoff, 1987) and *blending* (e.g., Veale and O'Donoghue, 2000). We thus argue that a CL perspective to word formation is necessary even in the context of a practical application such as *Zeitgeist*. Indeed, we report empirical results that suggest *Zeitgeist*'s CL approach to word-formation yields a surprisingly robust model of neologism recognition and interpretation.

Dent, S., *Fanboys and Overdogs: The Language Report III*. Oxford Univ. Press (2003).

G. Lakoff, *Women, Fire and Dangerous Things: How the Mind forms Categories*. University of Chicago Press (1987)

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

Veale, T., O'Donoghue, D., Computation and Blending. *Cognitive Linguistics*, 11(3-4), 2000.