

Harmonizing WordNet and FrameNet

Christiane D. Fellbaum

Department of Computer Science
Princeton University, Princeton, USA
fellbaum@princeton.edu

Abstract. Lexical semantic resources are a key component of many NLP systems, whose performance continues to be limited by the “lexical bottleneck”. Two large hand-constructed resources, WordNet and FrameNet, differ in their theoretical foundations and their approaches to the representation of word meaning. A core question that both resources address is, how can regularities in the lexicon be discovered and encoded in a way that allows both human annotators and machines to better discriminate and interpret word meanings?

WordNet organizes the bulk of the English lexicon into a network (an acyclic graph) of word form-meaning pairs that are interconnected via directed arcs that express paradigmatic semantic relations. This classification largely disregards syntagmatic properties such as argument selection for verbs. However, a comparison with a syntax-based approach like Levin (1993) reveals some overlap as well as systematic divergences that can be straightforwardly ascribed to the different classification principles. FrameNet’s units are cognitive schemas (Frames), each characterized by a set of lexemes from different parts of speech with Frame-specific meanings (lexial units) and roles (Frame Elements). FrameNet also encodes cross-frame relations that parallel the relations among WordNet’s synsets.

Given the somewhat complementary nature of the two resources, an alignment would have at least the following potential advantages: (1) both sense inventories are checked and corrected where necessary, and (2) FrameNet’s coverage (lexical units per Frame) can be increased by taking advantage of WordNet’s class-based organization. A number of automatic alignments have been attempted, with variations on a few intuitively plausible algorithms. Often, the result is limited, as implicit assumptions concerning the systematicity of WordNet’s encoding or the semantic correspondences across the resources are not fully warranted. Thus, not all members of a synonym set or a subsumption tree are necessarily Frame mates.

We carry out a manual alignment of selected word forms against tokens in the American National Corpus that can serve as a basis for semi-automatic alignment. This work addresses a persistent, unresolved question, namely, to what extent can humans select, and agree on, the context-appropriate meaning of a word with respect to a lexical resource? We discuss representative cases, their challenges and solutions for alignment as well as initial steps for semi-automatic alignment.

(Joint work with Collin Baker and Nancy Ide)