MSFA Encodes Word Meanings in Contexts Multidimensionally

Kow KURODA   Hitoshi ISAHARA

National Institute of Information and Communications Technology (NICT), Japan

In Kuroda and Isahara [3, 4], we defined a framework for fine-grained semantic analysis/annotation called MULTI-LAYERED/DIMENSIONAL SEMANTIC FRAME ANALYSIS (MSFA for short). An evaluation version was released for free, expecting feedback. More work will follow (Besides the "official" result, additional semantic analyses/annotations are being done on a volunteer basis in the form of open development, and made available at http://www.kotonoba.net/~mutiyama/cgi-bin/hiki/hiki.cgi?FrontPage).

In MSFA, real texts are annotated for **semantic roles** differentiated from **semantic types**, an important distinction inspired by Frame Semantics [1]. Roughly, the proposed distinction of semantic roles from semantic types corresponds to the distinction of "relational" categories from "object/entity" categories made by Genter and her colleagues [2]. Also, Gentner's "relational schema" category seems to correspond to "(semantic) frame" in the BFN sense, and "(idealized) situation" in our sense. While semantic types can be equated with natural kindsand encoded in a thesaurus effectively, semantic roles can't: they are **situation-specific concepts** highly dependent on culture, seeming to be more responsible for text understanding than semantic types.

MSFA-based annotation has the following features: (i) it is **not presumptive**, in that there are no presupposed frames before doing annotation/analysis, except when the results of a previous analysis are reused. More specifically, frame identification in MSFA is **necessitated by semantic annotation/analysis itself**; it's rather meant be a "preprocessing" for developing a database of situations against which people understand linguistic forms; (ii) this means that the analysis/annotation needs to be **greedy**, in that as many frames as you need can be identified and added to the analysis, as long as they found necessary for providing deep enough semantic analysis/annotation of a text. Sufficiency is determined by successful specification of semantic co-variations among arguments and adjuncts; (iii) it does not assume (at least currently) established definitions for any frames: they are always open to major modifications, and the annotation task is designed so to make it easy to manage. (iv) it is **exploratory**, in that it aims to "discover" frames in a bottom-up, inductive fashion, through the process of "exhaustive" semantic analysis of a text itself; (v) It is meant to be **exhaustive**, in that every word and multi-word unit are identified as a frame-evoking element: no exception. This means that you are not allowed to ignore certain elements as "uninteresting for our purposes." Rather, you are required to seriously work with capturing the frame-evocation effects by each word within a sentence in a running text. (vi) in this sense, it is **unbiased**, in that identification of frames is not motivated by any specific applications like Machine Translation, QA; (vii) also, it provides semantic analyses/annotations **independent from syntactic parses**, in that it only assumes tokenization, or "shallow parses" in some limited cases only for the sake of simplication. Unlike many other frameworks, it tries to "unground" semantic specifications against syntactic ones. For one, specification of valence-patterns is out of our focus, be-cause it is very likely that other databases will provide such information independently; (viii) this makes annotation **flexible**, in that you are always allowed to add or remove frames. If semantic analysis/annotation depended on certain syntactic parses, it would be a disaster to do such an "editorial" job. (ix) it is meant to be **open**, in that annotation work can be done in the form of open-development. The annotation scheme was so defined that annotators are only required to have a good command of Microsoft Excel, dispensing with house tools.

MSFA was inspired by the Frame Semantics/BFN approach. Its details, however, were specified independently of BFN, for reasons specified above. Their main differences are the following: BFN provides standardized frames in the form of a database, whereas MSFA provides an annotation database using frames discovered in a running text in an unselective way. Frames identified using MSFA are more detailed and specific, having finer granularity than BFN frames.

The relationship among frames is specified in MSFA to reveal how frames are interrelated within a sentence. According to their relative frequencies, "representative" relationships currently assumed in MSFA are presented below, most of which have equivalent, or analogous "frame-to-frame relations" in BFN: (1) A frame F elaborates another frame G; i.e., F inherits information from G (e.g., ⟨Murder⟩ elaborates ⟨Killing⟩); (2) F constitutes G, i.e., F is part of G (e.g., ⟨Paying⟩ constitutes ⟨Buying⟩); (3) G presupposes F (e.g., ⟨Buying⟩ presupposes ⟨Selling⟩); (4) F presumes G. This is the reverse of Presupposition (e.g., ⟨Selling⟩ presumes ⟨Buying⟩); (5) F realizes G (e.g., ⟨Buying⟩ realizes ⟨Obtaining⟩); (6) F targets G. This is specifically introduced in MSFA to describe the "figurative uses" of words including metaphor and/or idioms.

References

[1] C. J. Fillmore and B. T. S. Atkins. Starting where the dictionaries stop: The challenge for computational lexicography. In B. T. S. Atkins and A. Zampoli, editors, *Compuational Approaches to the Lexicon*, pages 349–393. Clarendon Press, Oxford, UK, 1994.

[2] D. Gentner and K. J. Kurtz. Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, and P. W. Wolff, editors, *Categorization Inside and Outside the Laboratory*, pages 151–175. APA, 2005.

[3] K. Kuroda and H. Isahara. Proposing the MULTILAYERED SEMANTIC FRAME ANALYSIS OF TEXT. In *The 3rd International Conference on Generative Approaches to the Lexicon*, pages 124–133, 2005. [Revised version is available as: http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/msfa-gal05-rev1.pdf].

[4] K. Kuroda, M. Utiyama, and H. Isahara. Getting deeper semantics than Berkeley FrameNet with msfa. In *5th International Conference on Language Resources and Evaluation (LREC-06)*, pages P26–EW, 2006. [Available at: http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/msfa-lrec06-submitted.pd%f].