

形式意味論と計算言語学

峯島 宏次 (Koji Mineshima)
お茶の水女子大学

従来の言語哲学の分野でなされてきた自然言語の分析の多くは、ある種のパズルを生み出す「有名な例文」の周辺で問題を絞り込むことで独自の発展を遂げてきたとも言える。一方、現在では理論言語学の一分野として定着しつつある形式意味論の分野では、言語哲学・論理学の手法を背景にしつつ、分析の対象となる言語現象の細分化・体系化・多様化が進んでいる。この影響を受けて、近年では言語哲学の「言語学化」が進み、言語哲学の一部と形式意味論とは互いに切り離すことのできない学際的領域を形成している。そこで扱われる言語現象は、“the” (Russell) から、“ouch/oops” (Kaplan) まで、多様なものとなっている。

しかし、こうした言語哲学から形式意味論に至る自然言語分析の成果は、そのフォーマルな性格にもかかわらず、統計的機械学習の手法が主流となった現在の計算言語学・自然言語処理の分野ではほとんど使われていない。じっさい、自然言語の実テキストをひとつひとつ分析しようとしてみると、実にさまざまな困難に出会うことがわかる。形式意味論によって言語分析の対象は拡大したものの、じっさいにわれわれが使用している言語との間にはいまだ大きなギャップがある。

この発表では、形式意味論の展開と最近の計算言語学・自然言語処理の分野での意味解析・自然言語推論の研究を概観したのち、コーパスに基づく統計的手法の発展によって従来のフォーマルな意味論研究にどのような変化が起きているのか、また論理的手法は今後、計算機による自然言語の解析にどのように貢献しうるのかについて議論する。特に、日本語・英語の研究を中心として、(1) 組合せ範疇文法 (CCG) に代表される現代的な統語論に基づく構文解析器の出現、(2) 樺ツリーバンクや NPCMJ のような詳細な構文木情報が付与されたコーパス (ツリーバンク) の出現、(3) 含意関係認識タスクと意味論を評価するためのデータセットの発展、(4) 分散表現に基づく意味論の発展をとりあげ、発表者らによる最近の研究[1,2]を交えて議論する。

[1] Koji Mineshima, Pascual Martínez-Goméz, Yusuke Miyao, and Daisuke Bekki. Higher-order logical inference with compositional semantics. Proceedings of EMNLP 2015, 2055–2061.

[2] Koji Mineshima, Ribeka Tanaka, Pascual Martínez-Goméz, Yusuke Miyao, and Daisuke Bekki. Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser. Proceedings of EMNLP2016, 2236–2242.