

コーパスを用いた自然言語処理のための基礎研究 (A04-45)

Applied Use of Database Corpora in Theoretical Linguistics

仁科弘之 (埼玉大学教養学部 哲学人間システム専攻・教授)

Hiroyuki Nishina (Faculty of Liberal Arts, Dept. of Philosophy & Human Systems)

1. 概要

1.1 研究組織

研究代表者: 仁科弘之 (教養学部教授)

共同研究者: 山中信彦 (教養学部助教授)

1.2 交付決定額

平成16年度: 80万円

2. 研究経過

最近の目覚ましいコンピュータ・ソフトウェア技術、特にコーパス諸技術の発展により大規模言語データの利用が可能になっている。このため、言語理論の検証、一層の精密化、さらには新しい枠組みの利用が可能となりつつある。本プロジェクトにおいては、コーパス・データを利用したリアルな文法記述により、理論の修正、さらには新しい枠組み構築に貢献することが目標である。統語・意味理論と語用・談話理論の両分野を横断する以下の問題に対し、コーパスの豊富な例に基づき実証的な新しい提案を行った。これにより現実の問題解決に有効な言語理論が備えなければならない特性や制約のいくつかが浮き彫りとなってきた。

2.1 (テーマ1) 「コーパスに基づく語彙的多義性の解消の研究」

従来の多義性解消研究は、該当する語を含む文の全体的統語構造と、それに基づく意味構造をもとに解析するのが一般的である。共同研究者の研究においては、該当する文の左文脈をも分析対象の範囲に含め、代名詞の先行詞決定を試みている。該当文に先行する文を文脈に含めるこの手法は、計算論的には従来研究より複雑性を増すが、多義性解消の技術の精度を大幅に高めることができる。

研究代表者(仁科)は、まず多義性の意味表示理論の基礎的研究を行った。研究が依然として乏しい日本語頻度副詞を例に、従来の論理記述ではその意味が十分に記述できない用例を指摘し、その暫定的な新しい表現法(論理式による)を提案した [3]。さらに、日・英の数種のコーパスを用い、談話構造における代名詞化の保持可能な距離とその統語・意味的な条件を研究した [4]。前者はデータの新規性と精密な先行研究において、後者はテーマの新規性と理論的枠組みの将来性において、それぞれ学会発表時に評価された。さらに国際会議(12月:台湾)での発表が受諾され、論文執筆が進行中である。また、文献研究を行った [7, 9]。

共同研究者(山中)は、朝日新聞記事データベース asahi.com perfect (最近20年間)を用い、直接的言及の避けられる表現の迂回法(婉曲表現)の語用論的研究を行った。1年間分の記事を材料として、特定表現の用いられた事件状況と記事類型、その使用例の現れ方(見出し/本文等)、当該表現の統語的役割などを調べ、統計的に分析した。さらに'04年の3か月分についても同様の分析を行った。また、婉曲表現の現れ方を1989年の記事中から98件を抽出し、これらの語の持つ多義性の解消のメカニズムを語用論的に考察した。この手法は語用論ではまれに見る実証的な通時的分析であり、理論的にも興味深い試みである。新規性のある成果

がえられたので学会誌に投稿した[5]。

2.2 (テーマ2) 「手話動作の様相論理に基づく類型的記述」

研究代表者は、基本的な身体動作動詞の意味をそれに対応する身体スケルトン動作の様相論理表現によって記述する提案を行っている[1]。本プロジェクトでは手話辞書から例を採取し、それを動作記述アルゴリズムの検証にもちいた。ユニークなテーマとアプローチの斬新性により発表は認知科学分野の国際学会で注目された[6]。本発表は、言語・範疇・認知関連論文集としてJ・Benjamin社より刊行される予定である。さらに、東北大学21世紀COEプログラム、「言語認知総合科学戦略研究教育拠点」の12月の講演会に招聘され、本研究の関連発表を行った[8]。現在、国内有数の言語科学研究拠点として成果をあげつつあるこのCOEにおいて、強い関心と研究プログラムの方向性について大きな支持をえることができた。今後は動作の論理的公理系を構築し公表する等、次の段階に移行する予定である。

3. 研究成果

3.1 論文

1. Nishina, Hiroyuki: "How Humans Identify Actions? — One Proposal for Extracting Semantic Contents from their Spatiotemporal Information" (*Proceedings of the 17th International Congress of Linguists*, Matfyzpress (CD-ROM), MFF U, Prague, 28pp, 2004年)
2. 仁科弘之: 「行為タイプの動的記述」, 佐藤・堀江・中村編:『対照言語学の新展開——理論研究と——言語記述の接点——』, ひつじ書房, pp. 155-186. 2004年7月.
3. 小澤智明・仁科弘之: 「頻度副詞の多重性とその意味解釈モデル」, 『言語処理学会第11回年次総会発表論文集 CD』, pp.5-4, pp. 1-4, 2005年3月.
4. 春日博・仁科弘之・三島健稔: 「固有名詞の甦りと談話構造モデル」, 『言語処理学会第11回年次総会発表論文集 CD』, pp.1-4, pp. 1-4, 2005年3月.
5. 山中信彦: 「新聞記事コーパスにおける性暴力表現」, 『計量国語学』(投稿中).

3.2 口頭発表・講演

6. Nishina, Hiroyuki: "The Modal-Logical Interpretation of the Causation of Bodily Actions", *International Language and Cognition Conference*, at Novotel, Coffs Harbor, Organized by University of New England, Australia, 2004年9月10日
7. 仁科弘之: 「束縛理論と再構築:これまでの分析法 (A. Barss の論文より)」, 2004年10月23日, 岩崎言語学研究会例会, 東京外国語大学
8. 仁科弘之: 「文法の身体化——動作を様相論理で記述する」, 2004年12月16日, 『21世紀COEプログラム:「言語認知総合科学戦略研究教育拠点」第19回講演会』, 講演東北大学).
9. 仁科弘之: 「束縛理論と再構築:ミニマリスト理論における分析法 (A. Barss の論文より)」, 2005年2月18日, 岩崎言語学研究会例会, 東京外国語大学

4. 外部資金

平成15～18年度科学研究費補助金一般研究(C)

(研究代表:仁科弘之、「様相論理をもちいた行為タイプ抽出のための基礎研究」) 平成16年度:80万円