様式 C-7-2

自己評価報告書

平成 21 年　4 月 16 日現在

研究種目：特定領域研究
研究期間：2006～2010
課題番号：18061003
研究課題名（和文）　代表性のあるコーパスを利用した日本語意味解析

研究課題名（英文）　Japanese semantic analysis using balanced corpus of contemporary written Japanese
研究代表者
奥村　学（OKUMURA　MANABU）
東京工業大学・精密工学研究所・准教授
研究者番号：60214079

研究分野：総合領域
科研費の分科・細目：情報学・知能情報学
キーワード：語義タグ付コーパス、単語の新語義発見、機械学習、語彙概念構造、クラスタリング

１．研究計画の概要

日本語を対象にした言語処理研究では，形態素解析，構文解析について研究が進み，高精度なツールの開発も行われてきており，それらのツールが日本語学，日本語教育など他の研究分野でも広く利用されるようになってきている．その一方で，意味解析については依然研究が遅れており，一般に利用可能なツールの開発レベルにまで解析精度が到達していない．また，代表性，継時性のあるコーパスを用いた言語処理研究は，これまでそのようなコーパスが存在しなかったため，日本語に関してはまったく行われてこなかったと言って良い．そこで本研究課題では，研究項目 A で構築する代表性，継時性のあるコーパスを用いた実証研究を行う．具体的には，以下の 3 つを柱とした日本語意味解析手法の開発を行う．

(1) 機械学習に基づく多義性解消手法の開発とそれを用いた代表性のある語義タグ付コーパスの半自動構築

タグ付コーパスから学習した多義性解消システムによりタグ付コーパス作成コストの軽減を図るとともに，作成されたコーパスを用いて bootstrap 的に多義性解消システムの性能向上を図る．

(2) 単語の新語義，新用法の自動発見手法の開発

時を経るにしたがって単語の意味は変化し，新しい意味が生まれることが知られている．継時性のあるコーパスで顕著に見られるこの言語現象を自動的に発見する手法を開発する．

(3) 語彙概念構造に基づく動詞の意味構造の自動抽出手法の開発と，それを用いた動詞の述語項構造辞書の自動構築手法の開発

語彙概念構造は動詞の振る舞いに関する分析から動詞の意味をそれが取る名詞同士の意味関係で記述する言語学に基づく意味構造である．文の意味構造は，(1)で特定される単語の語義と(3)で抽出される意味構造の統合により得ることができる．

２．研究の進捗状況

平成 19 年度から語義タグ付コーパスの構築を開始したが，平成 20 年度中に代表性のあ

る語義タグ付コーパスの一部が構築できた．単語の用例をクラスタリングする技術の開発では，従来の教師なし手法よりも高精度な半教師有り手法を平成18年度に提案したが，その有効性を平成19年度に確認し，平成20年度中に異ジャンルコーパスへの対応を実現した．また，用例のクラスタリング技術により，新語義発見の最初のステップとして，コーパスにおける単語の語義を自動識別する手法を開発した．

３．現在までの達成度
②おおむね順調に進展している．
(理由)
研究はほぼ計画通りに進んでおり，研究遂行上深刻な問題は生じていない．

４．今後の研究の推進方策
1.で挙げた 3 つの柱のうち，(2) 単語の新語義，新用法の自動発見手法の開発については，研究分担者を2名加え，新たな研究計画を追加する．新たに追加する研究はコーパス中の特異な用例を検出する手法の開発である．単語の特異な用例は，その単語の使われ方を調査する上で有用である．また特異な用例を検出・排除することで，用例集を精度良く分析することが可能となる．またコーパス内の特異な用例の有無を調べることで，そのコーパスの一般性や特殊性も考察できる．(2)では当初，コーパス中の単語の用例集合をクラスタリングし，同じ意味を持つクラスタを作成した上で新語義を発見する手法を構想していた．しかし，この手法では，一定量同じ意味の用例が出現するまでクラスタが構成できず，したがって，新語義を発見できないという問題点があった．そのため，上述した特異な用例検出手法により，ごく少数の特異な用例しか出現していない時点でも新語義を発見できる手法を開発することで，(2)で当初構想していた手法を補完し，新語義発見手法の完成度を増すことを狙っている．

５． 代表的な研究成果
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）
①白井清昭，コーパスにおける語の意味の自動識別，国文学 解釈と鑑賞．Vol．74，No．1，pp．61-69，2009．（査読無）
②奥村 学，白井清昭，現代日本語書き言葉均衡コーパスを用いた意味解析 -- 語義の自動特定，新語義の発見 --，言語，Vol．37，No．8，pp．66-73，2008．（査読無）

〔学会発表〕（計 15 件）
① Kazunari Sugiyama, Manabu Okumura, Semi-supervised Clustering for Word Instances and Its Effect on Word Sense Disambiguation, The 10th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2009), Mar.5, Mexico City, 2009.
② Koichi Takeuchi, Extraction of Verb Synonyms using Co-clustering Approach, Second International Symposium on Universal Communication (ISUC 2008), Dec. 15-16, Osaka, 2008.
③ Kazunari Sugiyama, Manabu Okumura, Personal Name Disambiguation in Web Search Results Based on a Semi-Supervised Clustering Approach, Proc. of the 10th International Conference on Asian Digital Libraries, Lecture Notes in Computer Science (LNCS), Springer Verlag, Dec.10-13, Hanoi, 2007.

〔その他〕
ホームページ
http://oku-gw.pi.titech.ac.jp/wsd.html