

研究課題別研究評価

1. 研究課題名：主辞駆動句構造文法を用いた言語獲得モデル

2. 研究者名：鳥澤健太郎

3. 研究のねらい

本プロジェクトでは、言語の学習を大量のテキストを用いて計算機に行わせることを目標とした。この狙いを実現するにあたって、言語の学習が次にあるような二つのステップに分割できるという仮定をおき、そのおのおのについて計算機上でシミュレーションする研究を行うという計画をたてた。

ステップ A. 語彙学習

英語、日本語などの個別言語に関するある程度の文法や品詞レベルでの単語の性質が与えられたという仮定から出発し、単語、特に名詞の意味的な分類をテキストから学習するプロセス

ステップ B. 文法学習

第一のステップで前提となった文法を、より仮定の少ない状況から学習するプロセスである。より具体的には、単語の品詞（例えば、名詞や動詞）が与えられていない状況で、英語や日本語などの様々な言語に共通の性質を表現している「普遍文法」から出発し、与えられた単語列から、単語の品詞を推定しつつ、文法の学習を行う過程である。

また、本研究の特徴として、使用する文法の表現方法として主辞駆動句構造文法（Head-driven Phrase Structure Grammar, HPSG）とよばれる形式を採用し、それを用いて以上のプロセスを行うシステムを構築することを目標とした。

4. 研究結果及び自己評価：

- (1) 研究の狙いで述べたステップ A については、精度の高い確率的な単語分類を生成すると同時に、単語クラス間の意味的な関係をも同時に学習できる方法論を構築することができた。この結果をもとに、従来辞書などがなければ困難であるとされた言語の意味を扱うタスクで、ある程度の精度を出すシステムを構築することができた。さらには、今後ユーザーの意図の認識など、従来、コンピューターが常識を持たないが故に難しいとされてきた課題の解決につながるのではないかと期待している。また、当初設定した研究課題との関係でいえば、予想以上の結果であると考えている。
- (2) 研究の狙いで述べたステップ B については、主辞駆動句構造文法のフレームワークで普遍文法のプロトタイプを記述し、単語の品詞に関して推定を行いつつ、個々の言語（つまり、日本語あるいは英語）に特有なパラメーターを推定するシステムを開発した。しかしながら、記述を行った普遍文法のプロトタイプでは、扱えない言語現象が未だ多数あり、ステップ A で使用したより実用的な日本語文法（人手で開発したもの）との間にはかなりの差が存在する。これについては今後さらなる研究が必要とされる。
- (3) また、研究課題を遂行する上で必要な高速な主辞駆動句構造文法のための構文解析アルゴリズム、ならびに単語の意味分類学習と同様な手法による、WWW 上の表形式の構造解析ならびにオントロジー抽出などの学習に関する研究も行った。これらはサイドワークであるが、

非常に興味深い結果が得られ、今後の研究においても重要な役割を果たすと期待している。

5. 領域総括の見解：

計算機を用いた言語構造の自動獲得は、人間の言語処理の研究における重要な問題であり、また大量の情報の自動処理を必要とするこれからの情報処理社会における重要な応用研究課題であるが、簡単に解決できるテーマではない。鳥澤健太郎は、この困難な問題に挑戦し、主辞駆動句構造文法の利用という極めて正統的方法で、しかも世界的にみて最高速の手法を開発して成果を得たことは評価できる。

6. 主な論文など：

- (1) * Kentaro Torisawa. A nearly unsupervised learning method for automatic paraphrasing of Japanese noun phrases, In Proceeding of Workshop on Automatic Paraphrasing: Theories and Applications, pp.63-72, 2001
- (2) * Kentaro Torisawa,. An unsupervised method for canonicalization of Japanese postpositions, In Proceeding of 6th Natural Language Processing Pacific Rim Symposium, pp. 211-218, 2001.
- (3) Minoru Yoshida, Kentaro Torisawa, Jun'ichi Tsujii. A method to integrate tables of the World Wide Web, In Proceeding of the International Workshop on Web Document Analysis (WDA 2001),2001.
- (4) Hiroshi Kanayama, Kentaro Torisawa, Yutaka Mitsui, Jun'ichi Tsujii. A Hybrid Japanese Parser with Hand-crafted Grammar and Statistics, In Proceeding of the 18th International Conference on Computational Linguistics, pp. 411-417, 2000.
- (5) Kentaro Torisawa, Kenji Nishida, Yusuke Miyao, and Jun-ichi Tsujii . An HPSG Parser with CFG Filtering in Journal of Natural Language Engineering, Cambridge University Press, Vol 6(1), pp. 63--80, 2000.