

論文内容の要旨

博士論文題目 Semantic Structures of Japanese Verb Phrases
— Acquisition, Representation and Use —
(日本語動詞句の意味構造 — 獲得、表現、利用 —)

氏名 大石 亨

(論文内容の要旨)

大規模な自然言語処理システムを実現するためには、一貫性と構造を備えたな辞書が必要となる。さらに、自然言語理解システムにおいては、局所的推論を可能にすると同時に、文脈処理のために必要なより深い理解を実現するために、辞書はより豊富な意味表現を備えていなければならない。本稿では、辞書概念を、抽象化された句レベルの意味構造を含むように拡張したうえで、

- 必要な知識をどのように獲得するか。
- 獲得された知識をどのように表現するか。
- 表現された知識をどのように利用するか。

という三つの問題に関する手法を提案し、実験によってそれぞれの手法を評価した。

第二章および第三章では、第一の問題、すなわち言語的知識の獲得に関する手法を提案した。第二章では、文の表層に現れる格助詞およびそれと置換されうる語句のパターンに基づいて動詞を細かく分類し、この分類に基づいて、動詞の語彙知識を獲得する手法(格パターン分析法)を提案した。この手法を解析済みコーパスから得られた共起情報に適用して深層格獲得実験を行い、有効性を評価した。

第三章では、共起するアスペクト形式や副詞句に基づいて、動詞の意味を素性によって六種類のカテゴリーに分類する手法を提案した。実験の結果、対象とした動詞のうち、96%のカテゴリーを一意に決定することができた。得られた分類を評価するために、最も基本的なアスペクトの形態である「シテイル」形の意味を自動的に決定する処理を行なった。200文における正解率は71%であった。

第四章では、第二の問題、すなわち獲得された知識の表現について述べた。これまでの章で得られた動詞句の二種類の分類を統合することにより、より詳細に動詞句を分類することが可能となる。それぞれのカテゴリーに対して、意味構造を与えるとともに各カテゴリー間の関係を明らかにした。約800の動詞を網羅している。

第五章では、最後の問題、すなわち表現された知識の利用法を提案した。いくつかの語用論的問題を導入した後、これまでに獲得し表現した知識が、どのようにこれらの問題解決に有効に利用されるかを論じた。意味構造は豊かな内的構造を持っているので、節間の結束関係など、自然言語理解に必要な推論規則を大幅に一般化することが可能となった。

(論文審査結果の要旨)

平成9年12月26日に開催した公聴会の結果を参考に平成10年2月18日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動が続けていくための十分な素養を備えていることを示すものと認める。

大石亨は、本博士論文において、日本語の意味表現のための動詞句レベルの意味構造に関する研究を行った。自然言語解析のために従来から用いられている表層的な統語構造に関する情報を対象にするのではなく、動詞句が持つべき意味構造によってその表層上の可能な表現が予測されるとの立場に立ち、動詞句の意味クラスの分類を中心とする研究を行った。現実には観察できるデータは、語の使用例だけであり、それがどのような意味クラスに属するものかは直接観察することはできない。そこで、分類された意味クラスと可能な統語上の現れの対応を調査し、大規模な言語データにおける実際の動詞の使用例を観察することによって、個々の動詞が属すると考えられる意味クラスの推定を行う方法を提案した。意味構造を中心とし、語の統語的振舞いからその意味構造を推定するというアイデアは極めて独創性の高いものである。

動詞句の意味構造の表現としては、動詞の項構造と事象構造を考慮した意味クラス分類を行い、各クラス間の関係について明らかにした。項構造については、解析済みの日本語データにおける動詞の表層的な格パターンと格助詞の置換の可能性に基づいて、自動的に分類を行う方法を提案し、深層格の予測が80%以上の精度で行えることを確認した。事象構造については、動詞が出現するアスペクト形式や副詞との共起に基づいて動詞を6つのカテゴリに分類する方法を提案するとともに、実際の使用例からカテゴリへの自動分類実験を行い、96%の動詞について一意にカテゴリ分類が可能であることを示した。意味構造の整理により言語データからの語彙知識獲得の自動化の道を開いたことは、今後の語彙知識獲得のあり方への新しい提案であり、極めて意義が深い。

さらに、獲得された語彙知識が言語解析にどのように役立つかを示すため、動詞の「シテイル」形の意味の決定、および、助詞「テ」で接続された節の意味関係の決定に関する検証を行った。両者は、言語理解における極めて困難な問題であり、このような言語現象に対して一般性の高い意味構造からの接近法を示したことは重要な成果である。実験により前者については、7割以上、後者については8割以上の正解が得られることを検証している。

以上のように、本研究は日本語の動詞句の意味構造の表現とその自動獲得に関してまったく新しい方法論を提案しており、かつ、いくつかの言語現象の解析に対してその有用性を示している。これらの成果は、情報科学、言語処理の分野における学術的貢献として高く評価できる。