

論文内容の要旨

博士論文題目 Constructing, Refining and Exploiting Rich Linguistic Resources
(言語資源の構築/精錬/利用方法の研究)

氏名 川井 早苗

(論文内容の要旨)

言語処理分野において、言語資源は、学習データや知識源として必要不可欠なものである。言語資源には、コーパス、辞書、シソーラスなど様々な物が含まれている。例えば、コーパスの場合、文章そのままのものや、形態素解析結果や構文解析結果を付与されたもの、意味情報を付与されたものなどがある。

近年の自然言語処理研究は、字面から語義、形態素解析、構文解析、意味解析へと深化しつつある。それに伴い、深い情報を含む言語資源の重要性が増してきている。

言語資源の構築方法には、人手で構築する方法、新聞記事やWebデータなどから自動的に構築する方法、あるいは、自動的に構築後人手による修正を行なう方法などがある。しかし、一般に、自動的に構築する場合には、早く大量に構築することができる一方、エラーが残るという問題がある。また、意味情報など、複雑で詳細な情報を付与しようとするほど、自動的に構築することは難しい。逆に、人手で構築・修正を行なう場合には、コストと時間はかかるが、意味情報のような深い情報を付与した言語資源の構築も可能である。本論文では、こうした深い情報を含む言語資源に焦点をあて、その構築、拡張、利用方法を提案する。

まず、こうした深い情報を含む言語資源として、主に人手で構築してきたオントロジー(日本語語彙大系)、二言語パターン対辞書、基本語辞書(Lexeed)、コーパス(檜)を紹介する。これらの言語資源は、主に人手で構築されてきたが、核となる部分を作成すれば、後は効率的に拡張することもできる。そこで、第3章では、言語資源のうち、二言語パターン対辞書の効率的拡張方法を提案し、その有効性を翻訳タスクによって評価する。第4章では、より詳細な意味情報として、動詞の交替情報の調査と、その調査結果に基づく二言語パターン対辞書の効率的拡張方法の提案を行なう。更に、意味情報などのリッチな情報と統計手法との融合を試みる。続いて、第5章において、構文解析結果の正解選択への利用を提案し、様々なレベルの意味情報を利用することによってその有効性を示す。また、ここでは人手で付与した意味情報を利用するが、こうした意味情報自体を学習によって獲得する方法について、語義曖昧性解消方法として提案する(第6章)。最後に、こうした実験により、表層情報だけでなく、意味情報を用いた言語処理の有効性を示し、ひいては、こうした情報を持つ言語資源構築方法研究の重要性について再確認する。

氏 名	川井 早苗
-----	-------

(論文審査結果の要旨)

平成 20 年 12 月 22 日に開催した公聴会の結果を参考に平成 21 年 2 月 24 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。

川井 早苗は、本博士論文において、自然言語処理に必要な深い意味情報をもった辞書、オントロジー等の言語資源を構築あるいは拡張する手法を提案し、その評価を行った。本論文の貢献は、次のようにまとめることができる。

1. 既存の動詞格情報をもった辞書を基にして、類似の語は類似の格パターンをもつと仮定して、二言語対訳辞書から格パターンと選択制約に関する情報を半自動的に拡張する方法を提案した。また、その効果を翻訳システムへの実装によって確認した。
2. 動詞の関連し合う自動詞と他動詞のような交替現象を利用して、不足している格パターンを自動的に補う方法を提案した。一方のエントリが選択制約等の意味情報を有していれば、そのような情報が自動的に補完されることを確認した。
3. 構築してきた意味辞書の情報を用いて、HPSG に基づく構文解析システムのあいまい性解消がどのように可能かを調査し、オントロジーを利用した意味情報が正しい構文解析結果の選択に有効に働くことを確認した。
4. 単語の語義あいまい性解消の問題を、オントロジーのような階層的な意味クラス分類を仮定し、上位の意味クラス分類と、より下位の詳細な意味分類に分割し、前者の情報を後者に用いることによって、従来の語義あいまい性解消の手法より高精度での語義分類が可能であることを示した。

以上のように、自然言語処理のための辞書やオントロジーの種々の構築法の提案と、その実際の利用における効果を検証した本研究は、独創的かつ実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。