

審査の結果の要旨

氏 名 笹野 遼平

本論文は、「**Japanese Anaphora Resolution Based on Automatically Acquired World Knowledge**」（自動獲得した世界知識に基づく日本語照応解析）と題し、日本語テキスト中の照応現象の解析を目的とし、照応解析に必要な知識の自動獲得、および、自動獲得した知識を用いて照応現象を解析する手法を提案し、実験によりその有効性を論じたものであり、7章から構成されている。

第1章は「**Introduction**」（緒言）であり、計算機による言語理解における照応解析の重要性、その実現に向けての高精度な基礎的解析、および、種々の知識の必要性を述べ、自動獲得した世界知識に基づく統合的な照応解析モデルの概要を述べている。

第2章は「**Knowledge Acquisition for Anaphora Resolution**」（照応解析のための知識獲得）と題し、照応解析に必要な様々な知識の獲得について述べている。まず、高度な共参照関係の認識に必要な同義表現辞書の構築を目指し、テキスト中に出現した括弧表現、および、国語辞典から同義表現の獲得を行っている。続いて、用言の取り得る格要素、および、各格要素が取り得る表現の用例を記述した辞書である用言格フレームをWEB文書16億文のコーパスから構築している。さらに、用言格フレームのカバレッジの向上を目指し、コーパスから収集された用例の汎化を、名詞に付与されたカテゴリ情報、および、固有表現解析結果を用いて行う手法を提案している。最後に、名詞とその必須要素、および、その用例を記述した辞書である名詞格フレームをコーパスと辞書を用いて自動構築する手法を提案し、WEB文書16億文のコーパスから構築、さらに、用言格フレームと同様に用例の汎化を行っている。

第3章は「**Named Entity Recognition Using Non-Local Information**」（大域的情報を用いた固有表現認識）と題し、大域的情報を用いた固有表現認識モデルを提案している。固有表現に関する情報は照応解析において重要な手掛りになることから、高精度な固有表現認識システムは、照応解析に有益であると考えられる。従来、固有表現認識は構文解析などの構造的な解析の前に行われ、局所的な情報のみを手掛りとして解析が行われていたが、提案モデルでは構造的な解析を前処理として行うことにより、キャッシュ情報、共参照関係から得られる情報、構文構造から得られる情報、係り先用言の格フレームから得られる情報という4つの大域的情報を新たに用いて固有表現の認識を行っている。CRL固有表現データを用いた実験の結果、大域的情報を用いることで従来手法よりも高いF値を得ている。また、IREX固有表現データ、および、WEBコーパスに固有表現情報を付与したデータを用いた実験においても、大域的情報を用いることにより固有表現認識の精度の向上がみられた。

第4章は「**Coreference Resolution Using Knowledge of Nominal Relations**」（名詞関係辞書を用いた共参照解析）と題し、2章で獲得した同義表現辞書、および、名詞格フレームを用いた共参照解析システムを提案している。日本語における共参照は、普通名詞、または、固有名詞が照応詞となるものが多く、先行詞と照応詞が同一の文字列を含む場合が多い。このため、文字列マッチングに基づく手法で比較的高い精度が得られることが知られている。本研究では、文字列マッチングに基づく手法に、同義表現を用いたマッチング、名詞格フレームを用いた名詞句関係解析に基づく同一性判定を加えることにより、文字列マッチングに基づく共参照解析の改良を行っている。新聞記事、および、WEBコーパスを用いた実験の結果、同義表現辞書、および、名詞格フレームを用いることにより共参照解析の精度が向上することが確認できた。

第5章は「Probabilistic Model for Zero Anaphora and Bridging Reference Resolution」（省略・橋渡し指示解析の確率的モデル）と題し、2章で獲得した用言格フレーム、名詞格フレームを用いた省略・橋渡し指示解析の確率的モデルを提案している。提案モデルでは、まず、テキスト中に出現した談話要素の認識を行い、すべてのテキスト中の表現を認識された談話要素に関連付けている。その上で、格フレームを用いてゼロ代名詞の検出を行い、さらに、格フレームや表層格、出現位置に関する選好を確率的に評価し、談話要素の中から適切な先行詞を決定している。WEB コーパスを用いた実験の結果、従来手法と比べて十分な精度を達成しており、提案モデルが省略・橋渡し指示現象を適切にモデル化できていることが確認できた。

第6章は「The Effect of Corpus Size on Case Frame Construction for Discourse Analysis」（談話解析における格フレーム構築に用いるコーパスサイズの影響）と題し、談話解析における格フレーム構築に用いるコーパスサイズの影響の調査を行っている。この調査のために、160 万文から 16 億文までの 6 種類の異なるサイズのコーパスから格フレームの構築を行い、構築された 6 つの格フレームを構文・格解析、省略解析に適用している。構文・格解析実験、省略解析実験の結果、より大規模なコーパスを用いることにより、これらの解析の精度が向上すること、および、さらに大規模なコーパスを用いることにより、さらなる精度の向上が望まれることを示している。

第7章「Conclusions」（結論）では、本論文の主たる成果をまとめるとともに、今後の方向性について述べている。

以上を要するに、本論文は、計算機による自然言語理解に向けて、照応解析に必要な知識の自動獲得、および、その知識を用いた照応現象の解析システムの構築を行い、実験によりその有効性を示したものであり、電子情報学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。