

氏名	かわはら だいすけ 河原 大 輔
学位(専攻分野)	博士(情報学)
学位記番号	情博第171号
学位授与の日付	平成17年7月25日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科知能情報学専攻
学位論文題目	Automatic Construction of Japanese Case Frames for Natural Language Understanding (自然言語理解のための日本語格フレーム自動構築)
論文調査委員	(主査) 教授 松山隆司 教授 河原達也 助教授 佐藤理史

論文内容の要旨

一般に、自然言語で表される意味内容は、ネットワーク状の複雑な関係性を持っていると考えられるが、それを1次元文字列による文章として表現する際に、人間にとって明らかなことは伝達の効率性から省略される。本論文では、このような関係を復元することが、自然言語理解の重要なステップとなると考え、そのために必要となる知識を大規模コーパスから漸進的に求め、単語の格フレームとして表現する手法を提案するとともに、自動構築した格フレームに基づく省略解析システムを作成し、構築された格フレームの有効性を確かめており、9章から構成されている。

第1章は序論であり、文章に含まれる様々な関係と、それらを認識するために必要となる知識について論じている。最も基本的な関係である構文関係は、構文的知識すなわち文法を記述することによって高精度に解析できる。一方、文章中には明示的に表現されない関係が多く含まれており、それらを解析するには語と語の関係性に関する知識が必要となる。このような知識を語の格フレームとして表現し、自然言語理解に利用するには、大規模コーパスから自動的に格フレームを構築する必要があることを述べている。

第2章では、まず日本語の基本的な性質について説明し、次に、二重主語構文、外の関係、格変化という複雑な言語現象について述べている。

第3章では、大規模コーパスに構文解析を適用し、その解析結果から基本的な格フレームを構築する手法について述べている。格フレーム構築における最大の問題は、用言の意味の多義性と、それに伴う用法の多様性である。つまり、同じ表記の用言でも複数の意味を持ち、取りうる格や体言が異なる。この問題に対処するために、用言とその直前の格要素の組を単位として用例を収集し、それらのクラスタリングを行うという方法を提案している。

第4章では、第3章で得られた基本格フレームを高度化する手法を述べている。基本格フレームに基づく格解析を大規模コーパスに適用し、その解析結果を用いて基本格フレームに新たな情報を付加する。この処理によって、二重主語構文、外の関係、格変化に関する情報が得られることを示している。

第5章では、これまでの処理で得られた格フレームの有効性を評価している。評価方法としては、人手によるものと、解析で得られたものの2種類を比較しており、高精度な格フレームが構築できていることを確認するとともに、構文・格解析の精度が格フレームの利用によって改善することを示している。

第6章では、文章に含まれる様々な関係を人手で付与したコーパスの作成について述べている。付与する関係としては、格関係、照応・省略関係、共参照である。このコーパスは、文章中の関係を解析するシステムを作成するための学習データ、また評価データとして用いられる。

第7章では、自動構築した格フレームが文章中の関係の解析に適用可能かどうかを実証的に確かめるために、文章中の関係のうち重要な部分を占める省略関係の解析システムを作成している。本システムは、自動構築した格フレームを知識源として、省略を認識し、その指示対象の同定を行う。第6章で作成したコーパスと解析結果を比較することによって解析シス

テムの評価を行い、格フレームの有効性を示している。

第8章では、省略解析の精度が高くない原因のひとつに、語の多義性の問題があることを指摘し、自動構築した格フレームに基づいて、省略解析と同時に語の多義性解消を行う手法を提案している。2つの異なるドメインについて評価を行い、多義性解消によって省略解析の精度が有意に向上することを示している。

第9章では結論を述べている。

論文審査の結果の要旨

自然言語で表される意味内容は、ネットワーク状の複雑な関係性を持っていると考えられるが、それを1次元文字列による文章として表現する際に、人間にとって明らかなことは伝達の効率性から省略される。本論文は、このような関係を復元することが、自然言語理解の重要なステップとなると考え、そのために必要となる知識を大規模コーパスから漸進的に求め、単語の格フレームとして表現する手法を提案するとともに、構築された格フレームの有効性を確かめたもので、得られた主な成果は以下の通りである。

(1) 自然言語理解を行うために必要となる知識を単語の格フレームとして表し、漸進的に格フレームを構築する枠組みを考案し、言語理解システムの高度化とそれをを用いた解析を交互に繰り返すことによって、知識を徐々に獲得していくことができることを明らかにした。

(2) 大規模コーパスの構文解析結果から、信頼度の高い関係のみを抽出し、それらを基に基本格フレームを構築することによって、構文解析結果をそのまま用いるよりも非常に高い精度で語の間の関係が抽出できることを示した。また、格フレーム構築において問題となる用言の曖昧性が、用言と直前の格要素を組としたクラスタリングを行うことによって低減できることを示した。

(3) 基本格フレームに基づく格解析を大規模コーパスに適用し、その解析結果を集約することによって格フレームを高度化した。その結果、基本的な格関係だけではなく、二重主語構文、外の関係、格変化といった複雑な言語現象にも対応できる格フレームを構築することができた。また、得られた格フレームの評価を行い、大規模かつ高精度なものが自動構築できたことを確認した。

(4) 文章中の単語間の関係について大規模な調査・分析を行い、そのような関係を人手で付与するための仕様を策定し、実際にその仕様に基づくコーパスの作成を行った。本コーパスは、今後の自然言語理解研究において調査・学習・評価を行うための基本的な言語資源となる。

(5) 文章中の関係の中で重要な役割を果たしている省略関係を解析するシステムを作成した。本システムは、格フレームに基づいて省略を認識し、その指示対象を同定する。さらに、省略解析において問題となる語の多義性を格フレームに基づいて解消する手法を考案し、省略解析システムに統合した。正解コーパスと比較することによって評価を行い、格フレームが有効に働いていることを示した。

以上本論文は、自然言語理解に必要な知識を単語の格フレームとして表し、大規模コーパスから格フレームを自動構築する手法を考案し、構築された格フレームの有効性を示したもので、学術上、實際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成17年5月18日実施した論文内容とそれに関連した試問の結果合格と認めた。