氏　　名	松吉　俊（まつよし　すぐる）
学位(専攻分野)	博　士（情報学）
学位記番号	情　博　第　290　号
学位授与の日付	平成20年3月24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	Hierarchically Organized Dictionary of Japanese Functional Expressions: Design, Compilation and Application （階層構造を持つ日本語機能表現辞書の設計、編纂および応用）
論文調査委員	（主査）教授 黒橋禎夫　教授 西田豊明　教授 河原達也

論文内容の要旨

本論文は，自然言語処理のための日本語機能表現辞書の設計，編纂，および応用について論じたものであり，辞書の設計に際して，機能表現を定義し，形態と意味の観点からその分類を行なう問題，さらに，異形を考慮した網羅的な見出し体系を持つ機能表現辞書を編纂する問題，および，作成した機能表現辞書を用いて機能表現を言い換える方法を論じており，全5章から構成されている。

第1章では，日本語の表現を内容的・機能的という対立から2つに分類して，自然言語処理におけるそれぞれの表現に関する辞書の現状について俯瞰し，機能的な表現に関するこれまでの辞書が持つ問題点を指摘している。日本語には，「にたいして」や「なければならない」に代表されるような，複数の形態素から構成されているにもかかわらず，全体として1つの機能語のように働く複合辞が多く存在する。本研究では，機能語とこれら複合辞を合わせて機能表現と呼んでいる。日本語の機能表現が持つ重要な特徴の1つは，個々の機能表現に対して多くの異形が存在することであり，計算機が利用することを想定した辞書はこれらの異形を体系立てて網羅的に扱う必要があるが，先行研究は不十分な状況にあり，更なる研究が必要とされていることを述べ，本研究の位置づけを明確にしている。

第2章では，機能表現辞書の設計にあたり，まず，扱うべき表現の範囲を明確にするために文法機能の観点から機能表現を定義した後，形態と意味のそれぞれの観点に基づき，機能表現を分類する2つの異なる体系を提案している。機能表現の多様な異形を体系立てて網羅的に扱うために，標準形から異形を生成する言語現象を，派生や機能語の交替，表記のゆれなど，7つのカテゴリーに整理している。そして，これら7つのカテゴリーに対応する7つのレベル，および，その上位に見出し語のレベルと，意味を区別した見出し語のレベルの2つのレベルを導入し，9レベルの階層を持つ機能表現の形態階層構造を設計している。意味の観点に基づく分類については，言語学において提案されている機能表現の意味分類は粗く，異形式同意味の判定へ応用することは難しいことから，言い換え可能性の観点に基づいて機能表現を意味分類し，3つの階層を持つ意味階層構造を設計している。

第3章では，異形を網羅的に収録し，かつ，機能表現に関するさまざまな有用な情報を含む機能表現辞書を実際に編纂する方法を提案している。辞書の基本設計において，異形を体系的に整理することができる形態階層構造を見出し体系として採用し，意味階層構造のクラスを各々の見出し語が持つべき属性として定めている。編纂作業においては，機能表現に関する定評のある2冊の文献に存在する機能表現リストを用い，341の見出し語と16,801の出現形を収録する機能表現辞書を作成している。このとき，階層構造が持つ継承を利用して，左接続，文体，難易度などの情報を機能表現に付与している。さらに，一番粗い階層において45クラス，一番細かい階層において199クラスを持つ意味階層構造を構築している。評価実験において，作成した機能表現辞書のエントリーと既存の機能表現リストを比較することにより，各々の見出し語に対してほぼすべての異形が網羅されていることを確かめている。

第4章では，作成した機能表現辞書のひとつの応用として，文体と難易度を制御しつつ，機能表現を言い換える方法を提案している。日本語では，ひとつの文章において，原則として，同じ文体を一貫して使い続けなければならないため，機能表現を言い換える際には文体を制御する必要がある。また，文章読解支援などの応用においては難易度の制御は必須である。予備実験の結果から，人間が文節の機能語部を言い換える時にもっともよく用いる「1つの機能表現を別の機能表現に置換する」という言い換えを行なうシステムを実装している。この言い換えシステムは，入力として文節と文体指定・難易度指定を受け取り，入力文節に対する代替表現の順位付きリストを出力する。文章作成支援への応用を想定し，出力の上位3位以内に適切な代替表現を生成できるかという観点から評価実験を行ない，このシステムは，クローズドテストにおいて入力文節の9割程度，オープンテストにおいて入力文節の8割程度に対してこれを達成できることを示している。

第5章は結論であり，本論文を総括している。

論文審査の結果の要旨

本論文は，日本語機能表現辞書の設計，編纂，および応用に関する研究をまとめたもので，得られた主要な成果は以下のとおりである。

1．日本語機能表現において，標準形から多様な異形を生成する言語現象を整理し，9レベルの階層を持つ機能表現の形態階層構造を設計した。この階層構造を用いることにより，膨大な異形を網羅的に体系化することが可能となる。さらに，既存の言語学における機能表現の意味分類を見直し，言い換え可能性の観点に基づいて意味分類を行ない，3つの階層を持つ意味階層構造を設計した。

2．異形を網羅的に収録し，かつ，それぞれの異形に対して各種の情報を記述した機能表現辞書を編纂する方法を提案した。そして，実際に，341の見出し語と16,801の出現形を収録する日本語機能表現辞書を作成した。この辞書は，現在利用可能な機能表現辞書の中で最大規模の辞書である。辞書の編纂においては，階層構造が持つ継承を利用して，見通し良くかつ簡潔に，少量のコストで左接続，文体，難易度などの情報を機能表現に付与する方法を提案した。作成した機能表現辞書のエントリーと既存の機能表現リストを比較する評価実験において，各々の見出し語に対してほぼすべての異形が網羅されていることを示した。

3．作成した機能表現辞書を用いて，文体と難易度を制御しつつ，機能表現を言い換える方法を提案し，入力として文節と文体指定・難易度指定を受け取り，入力文節に対する代替表現の順位付きリストを出力する言い換えシステムを実装した。このシステムは，上記2の日本語機能表現辞書の形態階層構造，意味階層構造，および各エントリーに記述されている左接続，文体，難易度などの情報を利用し，多様な言い換えを生成することができる。評価実験を行ない，この言い換えシステムは，クローズドテストにおいて入力文節の9割程度，オープンテストにおいて入力文節の8割程度に対して，出力の上位3位以内に適切な代替表現を生成できることを示した。

以上の成果は，学術上，工学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20年2月6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。