

氏名	むら た まさ き 村 田 真 樹
学位(専攻分野)	博 士 (工 学)
学位記番号	工 博 第 1586 号
学位授与の日付	平 成 9 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	工 学 研 究 科 電 子 通 信 工 学 専 攻
学位論文題目	Anaphora Resolution in Japanese Sentences Using Surface Expressions and Examples (表層表現と用例を利用した日本語文章における照応解析)
論文調査委員	(主 査) 教 授 長 尾 真 教 授 松 山 隆 司 教 授 田 丸 啓 吉

### 論 文 内 容 の 要 旨

日本語のコンピュータによる解析の研究は過去30年あまり行われて来て、形態素解析、構文解析に関してはかなり信頼性の高い技術となっており、意味解析についても多くの成果があがって来ている。しかし文と文との間に存在する文脈関係についてはほとんど研究が行われていなかった。本研究は文脈関係の研究の基本的なものとして文と文との間の照応関係について研究し、主要な照応関係をコンピュータプログラムで自動的に推定することができることを示したもので、7章からなっている。

第1章は本論文の目的と概要を述べている。

第2章は名詞句の指示性と数の推定をする方法を述べている。名詞句は、総称名詞句、定名詞句、不定名詞句に分類される。定名詞句、不定名詞句とは、英語では定冠詞、不定冠詞がつく名詞句に相当し、それぞれ、既知の単語、未知の単語を指すときに用いられるものであり、総称名詞句とは、一般の事物を指し示す場合に用いられるものである。また、名詞句の数には、単数、複数、不可算がある。この名詞句の指示性と数が、その名詞句の文中の言葉(表層表現)により、かなりの程度推定できることを示した。その推定のための規則は確信度付きのエキスパートシステムの書き換え規則に類する形で、文法書などから得られる知識をもとに経験的に作成した。規則を作るのに利用したテキスト(学習サンプル)での正解率は、指示性で85.5%、数で89.0%であった。規則を作るのに利用していないいくつかのテキスト(テストサンプル)での正解率は平均して指示性で68.9%、数で85.6%という結果となった。

第3章は日本語文章における名詞句の指し示す対象(指示対象)が何であるかを把握する方法を述べている。日本語には冠詞がないことから、二つの名詞句が照応関係にあるかどうかを判定することが困難であるが、第2章で明らかにした名詞句の指示性を用いて名詞句が照応するか否かを判定した。例えば、名詞句の指示性が定名詞句ならば既出の名詞句と照応する可能性があるが、不定名詞句ならば既出の名詞句と照応しないと判定できる。さらに、名詞句の修飾語や所有者の情報を用いることによって、より確実に指示対象の推定を行えることを示した。この結果、学習サンプルにおいて適合率82%、再現率85%、テス

トサンプルにおいて適合率79%、再現率77%の精度で名詞句の指示対象の推定をすることができた。

第4章は名詞の間接照応の解析方法を述べている。間接照応とは、文章中に現れていないがすでに言及されたことに関係する事物を間接的に指示する現象のことである。間接照応の解析を行うには、二つの名詞の関係に関する知識が必要となるが、そのような知識辞書はまだ存在していないので、「名詞 X の名詞 Y」といった形の用例を代わりに利用した。この方法で、テストサンプルにおいて再現率63%、適合率68%の精度で解析できた。このことは、名詞間に関する知識辞書が存在しない現在においてもある程度の精度で間接照応の解析ができることを意味している。また、名詞間に関する知識が利用できることを仮定した実験も行ったが、この精度はテストサンプルにおいて再現率71%、適合率82%であった。

第5章は代名詞の指す対象を推定する方法を述べている。従来の研究では、代名詞などの指示対象の推定の際には意味的制約として意味素性が用いられてきたが、本研究では対照実験を通じて用例を意味素性と同様に用いることができることを示した。また、「この」「その」の指示対象の推定に意味的制約として「名詞 X の名詞 Y」の用例を用いるなど、多くの新しい手法を提案した。また、従来から用いられている有力な手法も合わせて用い、総合的に優れた照応解析を実現し、学習サンプルにおいて87%、テストサンプルにおいて78%の正解率で推定することができた。

第6章は省略された動詞の補完方法について述べている。動詞の省略現象は主に文章内にある単語から補完する場合と文章内には現れていない単語を補完する場合の二つがある。文章内にある単語から補完する場合は、ほとんどの場合、その文は典型的な文構造になるので表層表現から推定することができる。また、文章内には現れていない単語を補完する場合は、用例を用いて解析した。その結果、学習サンプルで再現率92%、適合率90%、テストサンプルで再現率84%、適合率82%の精度で解析できた。

### 論文審査の結果の要旨

コンピュータによる日本語文の解析は形態素解析、構文解析、意味解析などが行われて来たが、文脈解析はこれまでほとんど試みられていない。本論文は文脈解析のうちの最も基本的な部分である照応現象について研究したもので、得られた成果は次のようなものである。

1. 文中の名詞が具体的な対象を指す具体名詞として使われているか、一般的・抽象的な名詞として使われているか、またその名詞が単数名詞、複数名詞、あるいは不可算名詞として使われているかを推定する方法を考案し、その有効性を示した。
2. 文中の名詞がそれまでに述べられた対象の何に対応するかを推定する方法を考案し、その有効性を示した。これによって英語に表現する際に定冠詞 the を用いるのが適当かどうか、かなりの程度推定できることを示した。
3. 「名詞の名詞」という形の用例を多数知識として用意し、これとの意味的類似度を調べることによって間接照応の関係を明らかにする方法を考案し、その有効性を示した。
4. 文中の代名詞が指す対象を同定するために用例を参照する方法を作成した。さらに意味素性を用いる方法などを合わせて用いることによって代名詞の照応関係を明らかにする方法を考案し、その有効性

を示した。

5. 前文中の同種文型の動詞を用いたり、用例を参照する方法等によって文中の省略された動詞を復元する方法を考案し、その有効性を示した。

以上のように、本論文は日本語文章における各種の照応関係をコンピュータで解析する方法を示したもので、日本語情報処理に学術上、実際に寄与するところが少なくない。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。また平成9年1月28日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。