

審査の結果の要旨

氏名 鍛治伸裕

本論文は、「Paraphrasing Written Language to Spoken Language」（書き言葉から話し言葉への言い換え）と題し、計算機による言語の理解を目指す研究の一課題として言語の言い換えの問題に着目し、国語辞典と大規模コーパスを使いこなすことにより書き言葉から話し言葉への自動変換を実現する手法について論じたものであり、6章から構成されている。

第1章は「Introduction」（緒言）であり、言語の言い換えという技術の自然言語処理研究における位置付けについて議論し、またその一つの方向性として、書き言葉から話し言葉への言い換えの重要性を指摘するとともに、書き言葉、話し言葉それぞれの特徴を議論し、さらに本論文で提案する書き言葉話し言葉変換の枠組みを示している。

第2章は「Paraphrasing Predicates」（用言の言い換え）と題し、国語辞典を用いて用言の言い換えを自動生成する手法を提案している。国語辞典は、見出し語の意味を言語で説明したものであり、見出し語のより簡易な言い換え表現となっている。しかし、国語辞典に基づく言い換えを実現するためには、語の定義が複数ある場合の扱い、すなわち多義性の解消が必要となり、さらに、用言の格パターンの適切な変換、定義文中からの言い換え表現の選択などを行わなければならない。これらの問題を解決するために、大規模コーパスから自動学習した格フレームを利用し、見出し語の格フレームと定義文主辞の格フレームの対応付けを行うことにより、言い換えパターンを自動学習する方法を考案し、実験によってその有効性を示した。

第3章は「Periphrastic and Overlapping VP」（迂言用言句と重複用言句）と題し、用言の言い換えにおいて問題となる、特殊な用言句の取り扱いを議論している。用言は通常、動作や状態などを表すが、中には副詞や機能語と同等の働きをするものが存在する。また、一般的に句の構成要素には構成性の原理が成立するが、それが成り立たない意味的に重複する表現も存在する。国語辞典を用いることによって、このような特殊な用言句を自動的に判定し、さらにより簡潔な表現に言い換えを行う方法を提案し、実験によってその手法の有効性を確認した。

第4章は「Lexical Choice via Topic Adaptation」（話題適応による単語選択）と題し、ある単語が話し言葉として適格であるか不適格であるかの自動判定を行う手法を提案している。基本的なアイデアは、単語の書き言葉コーパスと話し言葉コーパスにおける出現確率を使うというものである。書き言葉コーパスおよび話し言葉コーパスは、待遇表現と丁寧表現を基にWWWを自動分別したのを利用する。さらに、文章の話題が、単語選択に与える影響についても考慮した手法を提案する。すなわち、WWWから収集した書き言葉コーパスと話し言葉コーパスを話題ごとに分割し、入力文と同じ話題に関して記述されているコーパスを用いて語彙選択を行う。実験によってこのような手法の有効性を示した。

第5章は「Application to Text To Presentation」（テキスト自動プレゼンテーションへの応用）と題し、ここまで提案してきた言い換え技術の応用例について報告している。応用対象は、テキストからプレゼンテーションを自動生成するシステム(Text-To-Presentation System, TTPS)である。ここでは、入力テキストからスライドを自動生成するとともに、書き言葉話し言葉変換によって読み上げ音声

を自動生成する。基礎的な評価実験を行い、書き言葉話し言葉変換がこのシステムにおいて有効に機能していることを確認した。

第6章「Conclusion」（結論）では、本論文の主たる成果をまとめるとともに、今後の課題と展開について述べている。

以上を要するに、本論文は、計算機による言語理解を目指す研究としての言い換え技術の重要性について論じ、書き言葉から話し言葉への自動変換を実現するために、国語辞典と大規模コーパスを利用するいくつかの新しい手法を示すとともに、考案した手法をテキストの自動プレゼンテーションシステムとして統合し、実験によって有効性の確認を行ったものであり、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。