論 文 の 内 容 の 要 旨

論文題目　Extraction, Classification, and Retrieval of Formulaic Expressions in Scientific Papers
（学術論文における定型表現の抽出，分類，検索に関する研究）

氏　名　岩月　憲一

自然言語による表現は，語彙・文法上可能である組合せと比べて，実際には相当に少ないパターンしか出現せず，定型性があることが知られている。定型表現は，連続または非連続の単語列で，都度構成されるのではなく，そのまま記憶され使用されるという特徴を持つ。特に第二言語においては，定型表現の使用がネイティブらしさの観点から重要である。

学術論文においては，‘*in this paper, we propose*’のような，特有の定型表現が多用されている。こうした定型表現には，*showing the aim of the paper*のような伝達機能を具現する働きがあり，文章の論理構造と密接に結びついている。そのため，非英語母語話者も多く執筆し読むことになる学術論文においては，定型表現がスムーズな情報伝達に欠かせないものとなっている。

学術論文における定型表現の活用にあたっては，大量の定型表現の中から目的のものを検索する手法が必要である。これまでの研究では，定型表現の検索手法は，キーワードマッチングによるものが多数であった。しかし，キーワードに依存した検索では，多様な定型表現を検索できず，特定の表現を繰り返し使用することを避けたいといったより洗練された論文執筆というユーザの要求に応えることができないという課題がある。

本論文では，検索意図に添いつつも多様な定型表現を提示するために必要な技術について提案を行う。第1章では，まず本論文の背景及び定型表現の利活用における課題について述べる。更に，キーワードに加え定型表現の伝達機能をクエリとして用いることによって，多様な定型表現を検索するフレームワークを提案する。このフレームワークには，伝達機能ラベル付き定型表現データベースが必要であり，これを構築するためには，伝達機能に基づく文分類技術と，コーパスに対する定型表現抽出技術が必須であることを述べる。第2章では，まず既存の英語論文執筆支援システムを俯瞰し，実質的に定型表現あるいは何らかのフレーズを検索・提示することに集約されることを示す。次

に，学術論文における定型表現及び伝達機能がどのように定義され，また分析されてきたかを述べる。更に，伝達機能に基づく文分類と定型表現抽出に対して，計算機を用いた既存手法を述べる。第3章では，伝達機能に基づく文分類と，定型表現抽出およびそれらの評価に必要なデータセットの構築手法と，そのために用いる論文コーパスについて述べる。第4章では，伝達機能に基づく文分類を教師あり学習を用いて行う手法を提案する。また，訓練データの学術論文の分野と推定データの分野が異なっていても機能することを示す。第5章では，定型表現の抽出手法を提案する。既存の定型表現抽出手法を比較し，提案手法が伝達機能に着目した定型表現を抽出するのに適していることを示す。第6章では，提案手法によって構築した伝達機能ラベル付き定型表現データベースを用い，分野及び伝達機能別の定型表現について分析する。更に，提案した定型表現検索フレームワークによって，実際に多様な定型表現が検索できることを示す。第7章では，多様な定型表現を検索するという観点から，伝達機能の粒度と単位について議論する。第8章では，本論文の貢献をまとめ，今後の課題について述べる。

以上の提案によって，これまで伝達機能に基づく分類に人手を要した故に困難であった大規模な伝達機能ラベル付き定型表現データベースを計算機を用いて自動的に構築することが可能になった。また，伝達機能を用いることで，キーワードマッチングによる検索では不可能だった多様な候補の提示が可能になった。これらの成果は，定型表現および伝達機能という重要な言語現象に対して新たな計算言語学的アプローチをもたらすものであり，また計算機による論文解析や論文執筆支援への応用可能性が示されている点でも有望である。