

氏名(本籍)	兵藤安昭(愛知県)
学位の種類	博士(工学)
学位記番号	甲第52号
学位授与年月日	平成8年3月25日
専攻	電子情報システム工学専攻
学位論文題目	日本語文の骨格構造解析法と構文コーパス及びその応用に関する研究
学位論文審査委員	(主査)教授 池田尚志 (副査)教授 小鹿丈夫 教授 山本和彦 教授 藤田廣志

論文内容の要旨

本論文は、日本語長文に対して安定した高精度な解析を行うことができる構文解析システムについて述べている。さらに、骨格構造解析の応用として、構文情報付きコーパスの構築、構文指定による類似用例検索システムの構築について述べている。本論文は、以下のように5章から構成されている。

第1章では、日本語解析技術、テキストコーパス、テキスト検索技術の研究動向と問題点を示し、本研究の目的を述べている。

第2章では、日本語を解析するための骨格構造解析技術について述べている。日本語の構文解析では、格構造などの意味情報を用いた処理が広く行われている。しかし、大規模なテキストベースに対して正確に解析を行うためには、広範な領域を覆うことができる精密な意味情報が必要となり、現実的には容易ではない。特に、長文の解析は困難である。そこで、本章では、意味情報を用いずに表層的情報のみを用いて日本語長文に対して高精度な解析を行う骨格構造解析法を提案している。ここで述べる骨格構造とは、必ずしも完全な係り受けの木構造をなすものではなく、Nブロック内で複数の係り先が考えられる部分や並列構造などのように意味に立ち入らなければ解析できない部分は曖昧なブロックとしてそのまま残し、文の全体的な構造を把握しようとするものである。

本手法では、まず初めに、日本語文に形態素解析処理を施し、各文節の係り得る文節のタイプを示す文節カテゴリを付与する。次に文頭から順に各文節の係り先を調べる。その際すべての文節について係り可能性を調べることはせず、係り先の範囲をNブロック以内とする仮説および、表層的情報によるその他のいくつかの制約条件に基づいて係り先を決定する。このブロック内には、各文節の係り受けパターンにより、係り先が曖昧な文節が含まれている場合もある。この処理を文末に至るまで繰り返す。

朝日新聞記事300文に対して解析実験を行い、約94%の文に対して正しく構造解析されており、本手法が有効であることを確認している。

第3章では、2章で述べた骨格構造解析法を用いた応用例として、単語、品詞などの形態素情報だけではなく、構文情報を付与したテキストコーパス(構文付きコーパス)の構築について述べている。対象としたテキストコーパスは、講談社和英辞典、オーム社科学技術辞典和英大辞典の対訳付き用例約10万文であり、大量の長文テキストに関して骨格構造解析法の有効性を実証している。

第4章では、構文付きコーパスを対象とする類似用例検索システムについて述べている。本章で述べる類似用例検索システムは、従来のキーワード検索や全文検索などのような単語に基づく検索とは違い、検索パターンに構文的制約(係り受け構造)を指定して検索することができる。このため、単語レベルの検索では一致してしまうような多くの不適切な用例を絞り込むことができる。また分類語彙表を利用した意味分類コード化により類似用例の検索も可能である。このような高度な検索機能を用いれば、用例ベースの翻訳支援システムなどで必要されるような、ある言い回し、ある意味内容を含む用例を検索するといった検索要求に、柔軟に対応することができる。本章では、3章で述べた約8万4千用例からなる構文付きコーパスを対象とした検索実験を示し、本システムの有用性を明らかにしている。

第5章では、本論文で得られた結論をまとめ、今後の展望を述べている。

論文審査の結果の要旨

本論文は、日本語文の解析技術に関して述べたものであり、日本語長文に対して安定した高精度な解析を行うことが可能な骨格構造解析システムについて提案し、さらに、本システムを用いた応用例を示すことで、骨格構造解析の有用性を実証している。本論文により得られた成果は以下のとおりである。

(1) 形態素情報のみを用いて、Nブロック内での係り可能性の組合せにより係り先を決定するという原則と、それに伴うブロック化のアルゴリズムを提案し、これによって日本語長文に対して安定した高精度な骨格構造解析ができることを明らかにしている。大量の新聞記事を用いた解析実験を示すことで、本手法の有効性を実証している。

(2) 骨格構造解析を用いて、大規模で高精度な係り情報付きコーパス(構文付きコーパス:約10万用例文)の構築を行っている。解析結果を評価し、高い精度(約94%)が得られていることを確認している。

(3) 構文付きコーパスを検索対象とした構造指定による類似用例検索システムの構築を行っている。本システムは、係り受け関係を指定することで、単語や品詞などの形態素情報のみを指定した検索では受理されてしまう多くの不要な用例を絞り込み、より精密な検索を可能としている。多数の検索実験を示すことで、本システムの有効性を実証している。

以上、本論文は、日本語長文に対して安定した高精度な解析を行うことが可能な骨格構造解析を提案し、さらに、係り情報付き大規模コーパスの構築、構造指定による類似用例検索システムの構築によって、骨格構造解析の有用性を実証している。このように、本論文は日本文の解析法とその応用に関し、多くの新しい知見を得たものであり、学術上、実際上の価値は極めて高い。よって、本論文は博士(工学)の学術論文として価値あるものと認める。