

Title	機械的に言い換えを実現するシステムの作成
Author(s)	佐藤, 理
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1260
Rights	
Description	Supervisor:佐藤 理史, 情報科学研究科, 修士

機械的に言い換えを実現するシステムの作成

佐藤 理

北陸先端科学技術大学院大学 情報科学研究科

1999 年 2 月 15 日

キーワード： パラフレーズ、言い換え、形態素解析.

近年インターネット環境が大幅に整備されたおかげで、WWW を中心として様々な文書が電子化され、ネットワーク上で容易に入手できるようになってきた。

しかし、それらの文書は必ずしも多くの人々にとってわかりやすい言葉で書かれているわけではない。分野固有の表現（専門用語や特殊な言い回し）が使われている文書や堅苦しく難解な言葉が好んで用いられている文書も多い。

本研究では、このようなわかりにくい文書を、多くの人々にとってわかりやすい文書へと変換する（言い換える）手法について研究し、機械的に言い換えを行うシステムを作成する。

どのような言い換えを行うかは、言い換えを行う対象となる文書の種類や分野に強く依存すると考えられる。そのため、堅苦しく難解な表現が多いと考えられる政府、自治体、公的機関などの報告文書を対象として選び、これらを読みやすいようにするための言い換えを研究対象とした。

WWW 上より調査対象として郵政省「インターネット上の情報流通ルールについて（報告書）」(10764 バイト 5000 語強) と大蔵省「新しい金融行政のあり方について」(7335 バイト 3600 語強) の 2 文書を選び、その文書中の難解な表現をやさしく言い換える作業を行った。

作成された言い換え例は 162 例あったが、それらの言い換え例をまとめた上で、その機械的な実現手法を検討した結果、言い換えシステムの作成方針を次のように立てた。

1. 言い換えの対象となる文は、あらかじめ形態素解析し、その結果得られる形態素列を言い換え処理の対象とする。
2. どのような言い換えを行うかは言い換え規則によって定義できるものとする。1 つの言い換え規則は形態素列の置換規則とする。

3. 活用形の調整などを行う関数部を用意し、言い換え規則中でそれらを参照できるようにする。

作成した言い換えシステムは以下の6つの部分から構成されている。

前処理 入力文を形態素解析し、形態素列に変換する。

言い換えルールセット どのような言い換えを行うかを規則集合として記述したもの。言い換えシステムのプログラムに相当する。

ルールコンパイラ 外部形式で記述された言い換えルールセットを実行形式に変換する。

言い換えエンジン 形態素列に対して実行形式の言い換えルールセットを適用し、ルールに従った部分形態素列の言い換え(置換)を実行する。

関数群 言い換えルールで利用できるサブルーチン群。活用形の調整など行う関数を用意されている。

辞書群 サブルーチンで利用される辞書の集合。

また、入力として一文を取り、言い換え処理を行った後、一文を出力する。言い換え処理は以下の手順で行われる。

1. 入力された一文を形態素解析し、形態素列に変換する。この処理は前処理モジュールによって行われる。
2. 形態素列に対して言い換えルールセットを適用し、言い換えた結果の形態素列を求める。言い換えルールセットはあらかじめルールコンパイラによって内部形式に変換されており、言い換えエンジンは内部形式に変換されたルールセットに従って実際の言い換えを行う。
3. 得られた形態素列を文字列に変換し、出力する。

本システムを用いて言い換えの実験を行った。実験対象として『第百三十六回国会における橋本内閣総理大臣施政方針演説』(14000 語強、28319 バイト)を用いた。実験の結果、総言い換え数 87 件のうち、63 件(72%)は言い換えが正しく行われたが、残り 24 件(28%)は、言い換えの結果文の意味が変わったり文法的に誤った文になったりした。

この実験により、以下のことが明らかになった。

1. 本システムは、単語列から単語列への置換で実現可能な言い換えを機械的に実行することができる。
2. 言い換えに、単語列中の語の係り受けの情報を必要とする場合、本システムでは誤った言い換え結果を出力する場合がある。

3. 言い換え結果が文法的に正しくても意味的に誤っている場合もある。

より多くの種類の言い換えを精度良く実現するためには、構文解析を利用し、構文解析木の一部を別の構文解析木に置き換える機能が必要となる。