

子供と高齢者のためのカタカナ語言い換えシステム

竹内 晴彦 稗田 一郎

産業技術総合研究所

{takeuchi, hieda}@nibh.go.jp

1. はじめに

情報革命の進展により、私たちはこれまで経験したことのない大量の情報を容易に入手する事ができるようになってきた。しかし、一方で入手した情報を理解するためには、依然として多大な労力が必要であり、とりわけ高齢者や子供にとっては、この負担が大きい。このような負荷を軽減するためには、情報をわかりやすく表示する技術を開発することが必要である。

このような情報変換をテキストレベルで行うシステムは、日本語言い換えシステムと呼ばれる。このようなシステムの代表的なものとして、失語症患者などの障害者を対象とした医療福祉分野におけるもの^[1]と、文章を自分の好みの文体に変換する娯楽分野を対象としたものがある。本稿では、失語症患者支援を目的とした日本語処理システムをベースとして、高齢者などにとって意味理解が困難なことが多いカタカナ語を、易しく言い換える日本語変換システムを構築する。

2. システム構成

カタカナ語をわかりやすく表示するカタカナ語言い換えシステムの構成を図に示す。通常の日本語変換システムでは、形態素解析部、語彙変換部、及び統語変換部から構成されるが、カタカナ語言い換えシステムでは、統語変換は行っていない。

形態素解析部で、文の分かち書きと統語構造の分析を行う。ここでは、形態素解析プログラムとして「すもも」^[2]を使用する。

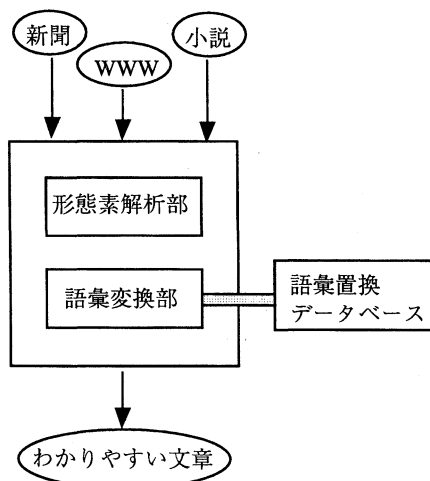


図 カタカナ語言い換えシステムの構成

語彙変換部で、カタカナ語を語彙置換データベースを参照して、わかりやすく言い換える。本研究では、この語彙置換データベースを新規に開発した。

本システムは、WWW上のCGIプログラムとして実現する。ウェブ画面上の指定のウィンドウにソースとなる文章を入力し、幼児向け、子供向け、大人向けのいずれかのスイッチを選択して、変換ボタンを押すと、やさしく言い換えられた文章が表示される。

3. 語彙置換データベースの構築

カタカナ語データベースに登録する語を次の三種類の基準で収集する。まず第一に出現頻度の高い語を採択する。このために

1996年の毎日新聞の家庭面、社会面、文化面の記事の中で使用されたカタカナ語から、高頻度語を抽出した。第二に、医療、福祉、ファッションなど特定の分野を指定し、それぞれの分野でよく使用されると思われる語を収集した。第三に、既存のカタカナ語辞典^[3]から、すぐに意味がわからないと思われる語を適宜選択した。

本システムは、主として高齢者と子供の使用を想定する。そのため、高齢者と子供という2つの異なる世代に対して、それぞれ適切な言い換え語を、データベースの中に記述する。言い換え語については、既存の辞書類^[3]を適宜参照しながらも、すべて独自に記述した。特に、既存の辞書類では、主として語の意味の説明が書かれているが、このカタカナ語データベースでは、言葉の言い換えを目的としているために、文章中でそのまま訳語に置き換えても自然な表現となるように留意した。

言い換え語以外のデータとして、カタカナ語の専門性に関する4段階の評定データと、分類番号を付与した。現在、本データベースには、3112語のカタカナ語が登録されている。

4. 結果と検討

WWWのホームページからカタカナ語を含む文章を本システムに入力した。その結果、難しいカタカナ語が子供向けのやさしい言葉や、大人向けの言い換え語に適切に置き換えられていくことが確かめられた。ただし、語彙置換データベースに登録されていない語については、言い替えはなされないため、今後、より多くの語彙を登録していく必要がある。

また、複数の意味を持つカタカナ語については、語彙置換データベースの中では複数の言い換え語が用意されているが、現在のシステムでは、その第一候補のみを使用して言い換えを行うに留まっている。潜在意味分析法^[4,5] (LSA) などにより、より適

切な言い換え語を選択することを検討しているが、本稿で扱うような広範なカタカナ語空間をカバーするような、適切なコーパスが見あたらないため、まだその実現には至っていない。

4. おわりに

カタカナ語言い換えシステムの概略を述べた。本稿では、語彙変換部で使用するカタカナ語データベースを新たに構築することにより、より適切な語彙変換を行えるようにした。今後の課題として、登録語彙数を増やすことに加え、多義性の解消のために適切なアルゴリズムを検討することが挙げられる。

謝辞

本システムの作成にあたり、北島愛子氏、花田麗奈氏、志村松子氏に、多大な協力を得た。ここに記して感謝いたします。

参考文献

- [1] 竹内晴彦 (2000) 失語症患者支援のための自然言語処理システム, 第16回ファジィシステムシンポジウム.
- [2] 鷲坂光一, 山崎憲一他 (1997) 情報検索のための高速日本語形態素解析システム「すもも」, 情報処理学会第4回全国大会.
- [3] 自由国民社「現代用語の基礎知識」, 小学館「カタカナ語の辞典」, 学研「パーソナルカタカナ語辞典」, 三省堂「カタカナ語辞典」, 講談社「カタカナ語・略語辞典」.
- [4] Landauer, T. K. ,& Dumais, S. T. (1997) A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, vol. 104, pp. 211-240.
- [5] Takeuchi, H., Kitajima, M., Akamatsu, M. (2000) Extracting knowledge from texts: On the effect of context length, *The Fourth Asian Fuzzy Systems Symposium*, 482-486.