

やさしい日本語文作成のための複合名詞書き換えシステムの試作

Development of translation system of compound noun in simple Japanese

*足立圭汰朗・**北村達也・***川村よし子

*ADACHI Keitarou・**KITAMURA Tatsuya・***KAWAMURA Yoshiko

甲南大学・東京国際大学

*&***Konan University・***Tokyo International University

〒658-0072 神戸市東灘区岡本8-9-1

E-mail: *t910827@gmail.com

Abstract : In this study, we proposed a system for translating Japanese compound nouns in simple Japanese. Compound nouns are known to make sentence comprehension difficult for non-native speakers. This system first extracts compound nouns that are composed of two nouns and the first noun is a Sa-hen noun by using a morphological analyzer. If the first noun is in the first grade of the previous Japanese-language proficiency test, it is translated with its definition in a web dictionary, Wiktionary. We developed a method to select a suitable tense of the definition depending on the second noun of the compound noun. The system works on web browsers and can be used via the Internet at no charge.

キーワード：書き換え，複合名詞，時制，やさしい日本語

1. はじめに

我々は、非日本語母語話者の文章読解を支援する技術を研究し、開発したシステムを公開している。複数の語句が接続されて1つの語句を構成する複合語は非日本語母語話者にとって難しいものの1つである。従って、これをやさしい日本語で言い換えることができれば、彼らの文章読解を支援できる。そこで、本研究では複合語の1つである複合名詞を自動的に抽出し、前半の語句をその説明文で置き換えるシステムの開発を試みた。

2. 複合名詞

本システムが対象とする複合名詞の条件を以下に挙げる。

1. 前半の語句がサ変接続できる。
 2. 後半の語句が2文字の漢字である。
 3. 後半の語句も2文字の漢字である。
- これらを満たす例として、「変更計画」や「消火活動」などが挙げられる。

3. システムの構成

本システムは、文章の入力を行う「入力部」と、書き換え処理と結果の表示を行う「出力部」で構成されている。文章の入力、結果の表示にはHTML (Hyper-Text Markup Language) を使用しており、Web ブラウザを通じて文章の入力と書き換え結果の情報を取得する。

書き換え処理の形態素解析システムにはMeCab(工藤 2006)を使用した。また、書き換え

時の意味の取得にはフリーの辞書サイト Wiktionary (<http://ja.wiktionary.org/>) を使用している。本来、非日本語母語話者用にやさしい日本語で書かれた辞書を使うべきである。しかし、現時点ではそのような辞書は公開されていないため、便宜的に Wiktionary を採用した。

4. 書き換えの手順

本システムは、まず MeCab を用いて入力された文章の形態素解析を行う。次に、形態素に分解された語句に対して2節に示した条件を満たすものを抽出する。得られた複合名詞の前半の語句を旧日本語能力試験の出題基準データ(1級~4級)と比較し、1級または級外(データに含まれない語句)の場合に書き換える対象とする。書き換え対象語句について、Wiktionary から当該語句の説明文を取得する。最後に、時制の判定を行い、前半語句の説明文を適切な時制に変換した上で置換する。

4.1 複合名詞の抽出

以下の手順で2節に示した条件を満たす複合名詞を抽出する。まず、条件1については、形態素解析によって得られる当該語句の品詞情報により判定する。次に、条件2については、前半の語句が漢字2文字から構成されているか否かを判定する。条件1, 2が満たされる場合、後半の語句も漢字2文字か否かを判定し、これが満たされれば、複合名詞として抽出する。

4.2 単語レベルの判定

旧日本語能力試験の出題基準に基づいて、前半の語句のレベル判定を行う。2級～4級の場合は書き換えを行わず、出力画面にその旨を表示する。一方、前半の語句が1級または級外の場合には書き換えを行う。その際、辞書サイト Wiktionary を利用し、当該語句の説明文を取得する。Wiktionary に当該語句が登録されていない場合には書き換えを行わず、出力画面にその旨を表示する。

4.3 書き換え方法

Wiktionary に登録されている語句でサ変接続できるものの説明文は、文末が「～こと。」で終わる文で記載されている。例えば、「変更」であれば、「変え改めること。」と記載されている。このうち、「こと。」の部分と、さらに後に続く文があればそれを削除し、それに複合名詞の後半の語句を接続することによって書き換えを行う。従って、「変更計画」は「変え改める計画」と書き換えられる。

4.4 時制の判定

複合名詞の書き換えを行う際、時制の判定が必要となる。例として、「変更計画」、「変更結果」という2つの複合名詞を挙げる。これらを書き換える場合、「変え改める計画」、「変え改めた結果」とするのが望ましく、後半の語句によって前半の説明をそれぞれ現在形、過去形に書き換える必要がある。

そこで、本システムでは、抽出された複合名詞について、「」（前半語句）する（後半語句）」、「」（前半語句）した（後半語句）」の2パターンを検索エンジン Google で検索し、ヒット数を比較する。これにより、当該の組合せにおいて、「する」（現在形）と「した」（過去形）のどちらが多く使われているかが分かるので、これに基づいて、時制の決定を行う。

検索の際には、上記のように検索語の前後をダブルクォーテーションマーク（"）で囲っている。これにより、語順も含めて完全に一致する場合のヒット数を調べることができる。例えば、「変更計画」という複合名詞であれば、「変更する計画」, 「変更した計画」の2パターンを Google で検索し、ヒット数の比較を行う。

5. システムの評価

インターネット上の様々な文章を用いて本システムを評価した。実行例を図1に示す。

1. [計算作業]→[計算]は旧日本語能力試験の2級以下なので書き換えません
2. [分割計画]→いくつかに分ける計画
 - ”分割する計画” : 445000件
 - ”分割した計画” : 148000件
3. [分割可能]→いくつかに分ける可能

図1 提案法の実行例

まず、図1の1に示した「計算作業」の例では、前半の「計算」という語句のレベルが2～4級であるため、書き換えが行われなかった。次に、2に示した「分割計画」の例では、前半の「分割」という語句のレベルは級外であり、かつ Google 検索で「分割する計画」の方が多かったため、現在形ので書き換えられた。

3に示した「分割可能」の例は、書き換えの失敗例である。この場合の書き換えは「いくつかに分けることが可能」となるのが正しい。この例のように、後半の語句によっては、Wiktionary に記載されている説明文から単純に「こと。」を削除して接続するだけでは不都合が生じる場合がある。この点については、今後の検討課題としたい。

6. おわりに

本研究では、非日本語母語話者の文章読解を支援することを目的として、複合名詞の書き換えシステムの開発を試みた。本システムは、前半の語句がサ変接続でき、かつ2つの語句が2文字の漢字から成る複合名詞を対象に書き換えを行うことを目指している。しかし、辞書情報取得に用いた Wiktionary に当該語句が登録されていない場合には書き換えを行うことができないことがあった。また、後半の語句によっては前半の語句の説明文と後半の語句の接続がうまくいかないことがあった。

これらの問題に対して、さらに多くの複合名詞の書き換えを可能にすることや、文脈から正しい書き換えを判断し実行することが今後の課題となる。また、本システムの本格的な実用化のため、説明文（書き換え候補）がやさしい日本語で書かれた辞書を開発中である。

謝辞 本研究の一部は平成25年度科学研究費課題番号24320096及び平成25年度私立大学等経常費補助金の支援を得て行われた。

参考文献

工藤拓(2006) MeCab: Yet Another Part-of-Speech and Morphological Analyzer
(<http://mecab.sourceforge.net>)