

言語能力に合わせた平易な文への言い換えシステム

三戸部 矩倫[†] 横山晶一[‡]
山形大学大学院理工学研究科

1. はじめに

子供や外国人などにとって、難しい言い回しの文章を理解するのは語彙力や日本語能力の関係から困難である。例1に、新聞記事の文例を示す。

例1

残業が続いてうつ病と診断された40歳代の元社員の男性が、症状が出てから5年後に労災を申請し、認定されていたことがわかった。

例1は「診断」や「認定」など、やや難しい単語を使用して事柄を説明しているが、ごく一般的な成人の日本人ならば容易に理解できる。しかし、子供や外国人にとっては、語彙力の低さなどから文の意味が理解できない場合がある。本稿では「利用者の言語能力に合わせた読みやすい平易な文」への簡単な言い換えシステムを作成し、考察を行った結果を報告する。平易な文とは、小学生程度の語彙力でも理解できる文や、やさしい表現を使った文のことである。自動化することにより、未知の単語があった場合に辞書で調べる手間を省くことができ、子供用ニュース文の自動作成や外国人学習者の日本語教育に役立てることができる。

本システムでは比較的理解が困難と思われるサ変名詞と2字の漢字熟語を言い換える対象とし、対象語を類義語の上位概念に置き換えた。その結果、単語レベルで比較的簡単な文に言い換えることができた。

2. 言い換えるの定義と利用法

「言い換え」は意味を保存したまま同一言語内の別の表現に変換する作業である[1]。主に以下のような言い換えるの種類がある。

- ・ 構文的言い換える
言語に関する知識によって実現可能な言い換える。
- ・ 意味的言い換える
言葉が発せられた文脈や談話状況を参照する必要のある言い換える。
- ・ プラグマティック（実用的）な言い換える
上記2つより複雑で、ある状況において同じ効果をもつような文への言い換える。

言い換えるの一般的な利用法として、主に次の5つのような利用法がある。

- ① 利用者の言語能力に合わせた読みやすい平易な文
- ② ニュース原稿から字幕を生成（要約）
- ③ Web文書を携帯端末に上手く表示
- ④ スタイル統一
- ⑤ 機械翻訳の自動修正

本研究では既述のとおり、①の言い換えるを行った。

3. 研究背景

3.1. 「サ変名詞+する」から動詞相当句への言い換える

近藤ら[2]の研究により、「サ変名詞+する+接尾辞」から平易な動詞相当句への言い換えるを機械的に実現する方法が提案された。

例2

入学式に桜が開花する → 入学式に桜が咲く
彼女は英語で会話できる → 彼女は英語で話せる

この言い換えるアルゴリズムを改良してシステムに利用した。「サ変名詞+する」のように動詞として使われているものに加えて、名詞として使われている場合も言い換えるの対象とした。名詞化するために、「動詞相当句+こと」と言い換える処理を追加した。

3.2. 言い換えるのための辞書

言い換える辞書として、類語大辞典[3]を使用した。この辞書は、100のカテゴリに分類されており、その下に約900の小分類が存在する。見出し語には「使う 5400a00」のように、7桁コードが付与されている。

例3

5400 使う	
a 動詞の類	
●使う	●利かせる
00 使う	12 利かせる
01 用いる	13 利かす
02 働かせる	...
...	

例3において、5400は分類番号、アルファベットは品詞の種類（aは動詞）、最後の2桁は5400aの中での番号を表している。さらに5400aの中に、「使う」「利かせる」のような「コード番号のない小分類」が存在する（小分類がない場合もある）。言い換える候補の単語には、この「コード番号のない小分類」を用いた。

4. 研究内容

4.1. 言い換えシステム

前節の内容を用いて、図1のアルゴリズムに基づいた言い換えシステムを作成した。

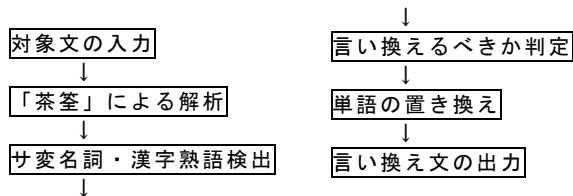


図1 アルゴリズム

まず対象文（またはテキストファイル）を入力し、日本語形態素解析システム「茶筌」[4]で解析をする。次に「サ変名詞・漢字熟語（2字）」を検出し、それらの単語を言い換えるべきかどうか判定を行う。「言い換えるべき」なら、類語大辞典による単語の置き換えを行い、サ変名詞の場合は活用形の修正を行う。

全ての単語を言い換えると、平易な文ではなく回りくどい文になる可能性があるため、以下の基準で判定を行った。

① 上位の小分類が曖昧な場合

「さまざまな～」「いろいろな～」「～の種類」のような場合は意味が変化するため、言い換えるべきではないと判断した。

② 類義（または同義）語があるか

類義（または同義）語が存在しない場合、言い換えることは不可能、あるいは非常に困難であるため、言い換えの対象から外した。

③ 漢字が難しいかどうか

小学校で習う漢字を学年別に分け、指定した学年で習っていない漢字が出現した場合は言い換えの対象とする。

4.2. 実行結果

例4に、システムの実行結果の例を示す。

例4

野球を満喫しよう

→[野球]を[あじわお][楽しむ]う

少年たちとの触れ合いで

→[少年(なし)]たちとの触れ合いで

「枯死して→枯れて」「満喫→あじわう、楽しむ」「少年→少年(なし)」のように言い換えることができた。“（なし）”はコードのない小分類が存在しない、という意味である。

「満喫」のように言い換え候補が複数存在する場合もあるが、今回は全て表示した。

4.3. 考察

言い換えの処理が行われた単語の総数のうち、およそ7割程度がほぼ元の意味を保存したまま言い換えられた（例4）。しかし、元の意味を保存できない例もいくつかあった。

例5

アクション監督で

→アクション[見守ること]で

女性にけがはなかった。

→[女と男]にけがはなかった。

比較的多かった誤りは、本来名詞であるべきサ変名詞が、動詞に言い換えられた場合である。

本処理では名詞化するために「こと」を付与したが、例5の「監督」の場合、言い換えるならば「見守る人」が適切であると考えられる。名詞化のために「動詞相当句+こと、もの、ひと」などの候補を挙げることも可能だが、結合価や意味素性によって解決できる可能性がある。

また、「女性→女と男」のように、上位の小分類に反義対（“男/女”、“買う/売る”等）が充てられている場合、元の文意を保存することができなかった。この場合は言い換え対象の語義を判断し、どちらに言い換えるべきか判断する必要がある。

5. おわりに

本稿では、類義語の概念を用いることで、文章に含まれる比較的難しい単語を簡単な表現に言い換えた。現段階では言い換えの正答率は高いとは言えないものの、今後の土台となるシステムを作成することができた。

問題点として、曖昧な小分類への言い換えや、反義対への言い換えによる語意の変化がある。単純な上位語への言い換えではなく「平易な言い換え語彙集」のようなものを用意したり、反義対の概念を考慮した辞書[5]を用いて言い換えを行う（“大きくない”→“小さい”など）必要がある。

参考文献

- [1] 乾健太郎：言語表現を言い換える技術，言語処理学会 第8回年次大会 チュートリアル資料，pp.1-21(2002)
- [2] 近藤恵子，佐藤理史，奥村学：「サ変名詞+する」から動詞相当句への言い換え，情報処理学会論文誌 Vol.40 No.11, pp4064-4074(1999)
- [3] 柴田武，山田進：類語大辞典，講談社(2002)
- [4] 日本語形態素解析システム茶筌，<http://chasen.naist.jp/hiki/ChaSen/>
- [5] 三戸部矩倫：類義概念を用いた反義対の小分類，山形大学卒業研究論文(2005)