

## 小学生の読解支援に向けた複数の換言知識を併用した語彙平易化と評価

梶原 智之 山本 和英

長岡技術科学大学 電気系

{kajiwara, yamamoto}@jnlp.org

## 1. はじめに

学習途上である子どもは、大人に比べて理解できる語彙の数が少ない。そこで我々は、大人向けに書かれた文書を子ども向けに平易化することを検討している[1]。本稿では、この語彙の問題を解決することで、子どもの文章読解を支援することを目指す。学習基本語彙[2]への換言によって新聞文書の語彙制限を行い、小学生を対象とした評価実験を通して、語および文の平易化について調査する。

学習基本語彙とは、光村図書出版株式会社が小学校国語教科書等の語彙分析を基に選定した語彙であり、小学生が文章を書く際などさまざまな表現活動に駆使できる5,404語とされている。以下、学習基本語彙に含まれない語を「難語」、難語を換言して得られる学習基本語彙の語を「平易語」とする。この難語が、小学生にとっての文章読解を妨げている一因である可能性があり、平易語への換言には読解支援の効果が期待される。

## 2. 関連研究

我々は小学国語辞典を用いた換言により、大人向け文書の平易化を行なった[1]。国語辞典を用いた換言は、これまでも鍛治ら[3]や美野ら[4,5]により研究が行われてきた。

鍛治らは大規模コーパスから格フレームを学習し、格パターンを持つ用言の換言を実現している。我々は格パターンを持たない体言にも共通に適用できる換言手法を提案した[1]。

美野らは日本語能力試験3,4級の語彙への換言を行い、初歩的な日本語能力を持つ外国人を対象に名詞を平易化した[4]。これには、サ変名詞の換言時に、名詞が動詞に置き換わることによって換言後の文に違和感が残る課題がある。

美野らはさらに、難易度が高い名詞化された用言から、元の用言の平易語への換言を行った[5]。これには、体言が用言へと換言されたことによる換言後の文の整合性の問題が残っている。

本稿では、換言後に活用変換を行うことでこれらの問題を解決し、文の統語構造を崩さずに平易化を実現した。

また既存手法[1]では、一種類の知識のみを用いて換言を行ったことにより、難語の78%しか平易化できないという網羅性の課題が残っていた。そこで本稿では、複数の換言知識を併用することにより、96%まで平易化の網羅性を高めた。

## 3. 提案手法

## 3.1. 平易語の獲得

本研究では、国語辞典としてEDR日本語単語辞書<sup>1)</sup>およびチャレンジ小学国語辞典<sup>2)</sup>、シソーラスとして日本語WordNet<sup>3)</sup>の3種類の換言知識を併用して平易語を獲得する。

一度の換言処理で平易語が獲得できない場合は、平易語が獲得できるまで換言処理を繰り返す。

## 3.1.1. 国語辞典からの平易語獲得

一般に、国語辞典の見出し語を説明したものが語釈文であり、その語釈文は平易な表現になっている場合が多いと考えられる[5]。さらに、小学生を対象とする小学国語辞典<sup>2)</sup>は、語釈文の記述がより平易であると考えられる。しかし、小学国語辞典の収録語数は約3万語であり、一般の国語辞典の約27万語と比べて少ない。そのため、小学国語辞典のみを用いて大人向けの文書を網羅的に換言することは難しい[1]。そこで本稿では、換言の網羅性を高めるために、通常の国語辞典やシソーラスからも平易語の獲

得を行う。

国語辞典の語釈文から、見出し語の平易語を獲得する手順を図1に示す。語釈文には、見出し語と同じ品詞の語が文末に現れやすいという特徴がある。そこで、語釈文中に出現する見出し語と同じ品詞の語のうち、最も後ろで現れる語を平易語候補として抽出することとする。

見出し語：控訴（名詞）
語釈文：上級裁判所 <sub>名</sub> に再審 <sub>名</sub> の申し立て <sub>名</sub> をする
↓
換言対：控訴（難語）→ 申し立て（平易語）

図1. 語釈文による平易語の獲得手順

### 3.1.2. シソーラスからの平易語獲得

シソーラスから獲得できる同義語や上位語のなかには、対象の語と換言可能なものもある。

語釈文から換言対を獲得する際、見出し語と対応する表現を必ずしも一語で抽出できるとは限らない[1]。難語の持つ概念が難しければ難しいほど、その説明を行う語釈文は長くなると考える。このようにときに、語釈文から抽出した平易語は説明の一部であるため、難語の持つ意味を大きく損なってしまう場合がある。シソーラスによる換言では、同概念集合に属する語も上位概念集合に属する語も必ず一語であるため、これを解決できると考える。

## 3.2. 換言知識の併用

本稿では、換言知識としてシソーラスおよび2種類の国語辞典を併用する。図2に示すように、これらの換言知識を換言の妥当性が高い順に優先して使用する。なお、換言の妥当性については、4.1節で検証を行う。

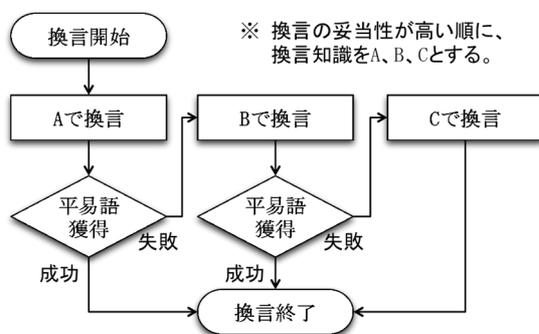


図2. 換言知識の併用

## 3.3. 換言後の活用変換

サ変名詞の換言や動詞の換言において、活用している語については換言後に活用変換を行う。活用変換は、MeCab<sup>4)</sup>による形態素解析によって活用形を取得し、IPADIC<sup>5)</sup>から同じ活用形の語を取得することで実現する。

### 3.3.1. サ変名詞 + “する”の活用

サ変名詞の平易語は、動詞+”こと”の形式で獲得できる場合が多い。換言したい原文がサ変名詞+”する”の場合、”する”の活用形と、平易語から”こと”を除いた動詞の活用形を合わせる活用変換を行う。サ変名詞+”する”の活用変換の例を図3に示す。

原文：～などと供述しているという。【供述：サ変名詞】
換言対：供述 → 述べること 【し：”する”の連用形】
出力：～などと述べているという。【述べ：”述べる”の連用形】

図3. サ変名詞+”する”の活用変換

### 3.3.2. 動詞の活用

動詞の平易語は、動詞で獲得できる場合が多い。換言したい原文で、動詞が活用している場合は、換言前後の動詞の活用形を合わせるように活用変換を行う。動詞の活用変換の例を図4に示す。

原文：ほお寄せ合って～ 【寄せ：”寄せる”の連用形】
換言対：寄せる → 近づける
出力：ほお近づけ合って～ 【近づけ：”近づける”の連用形】

図4. 動詞の活用変換

## 4. 実験

### 4.1. 換言の妥当性

2章で述べた方法で100の難語を換言し、換言の妥当性を評価した。この結果を次項の表1に示す。ここで用いた100語は、2000年の毎日新聞<sup>6)</sup>に現れる難語のうち、頻出する上位100語である。評価は、妥当な換言か否かの2段階評価を行った。妥当な換言とは、学習基本語彙に換言できた平易語のうち、元の難語と言い換え可能であると筆者のうちの1人が評価したものである。

表1より、学習基本語彙に換言できた場合に妥当な換言ができる割合は、小学国語辞典、日本語WordNet、EDR 日本語単語辞書の順に高い。よって、3つの換言知識を併用する際には、この順で優先的に換言を行うこととした。その結果、100語の

表 1. 換言の妥当性の評価

	小学国語辞典	EDR 日本語単語辞書	日本語 WordNet	合計
学習基本語彙に換言できた割合	83%	78%	60%	96%
妥当な換言ができた割合	81% (全体の 67%)	58% (全体の 45%)	63% (全体の 38%)	全体の 76%

うち 96 語を学習基本語彙に換言することができ、76 語は実際に言い換え可能であると筆者が評価できる換言対を得ることができた。

#### 4.2. 難語と平易語の難易度

本稿では、語の難易度の尺度として学習基本語彙を用いている。我々は学習基本語彙に含まれない難語を、学習基本語彙に含まれる平易語に換言することで、小学生の読解支援を行う。ここでは、難語が小学生にとって本当に難しいのか、平易語は簡単であるのかを検証した。

3章で述べた方法で妥当な換言ができる難語平易語対を 2000 年の毎日新聞<sup>9)</sup>から、頻出する順に 100 組用意した。これらについて、日本語母語話者である小学生高学年の被験者 5 人（5 年生 2 人および 6 年生 3 人）から評価を受けた。この結果を表 2 に示す。

評価は、被験者に難語と平易語の対がわからないように、全 200 語をランダムに提示した。そして、各語の意味が理解できるか否かを 2 段階で評価してもらった。2 段階評価とは、「はっきり理解できる」および「あいまい、もしくは全くわからない」である。

表 2. 難語と平易語の難易度の評価

	難語	平易語
被験者の過半数が理解できる語	55%	97%

#### 4.3. 換言前後の文の難易度

ここでは、2000 年の毎日新聞から、難語を含む 100 文を用いて、平易語への換言によって文の難易

度が下がるかどうかを検証した。4.2 節と同じく、小学生高学年の被験者 5 人に評価を受けた。この結果を表 3 に示す。

被験者には、難語を含む原文と換言後の文を組で提示した。このとき、どちらが原文でどちらが換言後の文であるかは伝えていない。そして、それぞれの文について「はっきり理解できる」または「あいまい、もしくは全くわからない」のどちらかを回答してもらった。表 3 では、前者を○、後者を×で表している。また、どちらの文の方がより簡単に感じるかも回答してもらった。

### 5. 考察

#### 5.1. 妥当な換言ができなかった例

4.1 節の妥当性の評価において、換言を繰り返した場合に、妥当な換言ができない場合が多かった。

2 回換言を行った場合、妥当な換言ができる割合は 36%と低い。さらに、今回実験を行った 100 語については、3 回換言を繰り返した場合は、妥当な換言が全くできなかった。よって、換言回数が増えるほど換言の妥当性は下がると言える。なお、換言回数は 3 回が最大であった。

図 5 に、3 回換言を繰り返す例を示す。

1. 競売 → 競り	【競売】大勢の客を相手に競り売りする
2. 競り → 商い	【競り】商品を携えて売り歩く商い
3. 商い → 職業	【商い】品物を売る職業

図 5. 難語 “競売” を平易語 “職業” に換言する例

3 回とも、それぞれ個別に考えると換言可能であるように見えるが、最終的に得られる“職業”は“競売”と換言可能とは言えない。

表 3. 換言前後の文の難易度の評価

換言前	換言後	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	平均
○	○	78%	50%	75%	43%	78%	65%
○	○	93%	56%	91%	71%	85%	79%
○	×	0%	0%	8%	7%	7%	4%
○	○	78%	50%	67%	36%	71%	60%
×	×	7%	44%	1%	22%	8%	16%
×	○	15%	6%	24%	35%	14%	19%
換言後の方が簡単だと回答した割合		65%	59%	64%	63%	68%	64%

この原因は、換言可能と言える語同士であっても、完全に同義ではないからだと考える。換言を行うたびに語義は少しずつ変わってしまい、換言を繰り返すことで換言不可能な語に変換されてしまう。

語釈文を用いて換言を行う場合の、平易語を獲得するまでに要する換言回数を表4に示す。小学国語辞典の場合、1回で学習基本語彙に換言できる割合が高いことがわかる。これは、小学国語辞典は対象が小学生であるため、一般の国語辞典よりも平易な語で語釈文が記載されていることが原因と考える。このように、換言を繰り返さずとも平易語が得られるような換言知識の作成が今後の課題である。

表4. 平易語を獲得するまでに要する換言回数

	1回	2回	3回	換言不可能
小学国語辞典	80%	2%	1%	17%
EDR 日本語単語辞書	65%	9%	4%	22%

## 5.2. 換言知識の優先順位

表5に、換言知識の優先順位による妥当な換言ができる割合の変化を示す。

表5より、妥当な換言を最も多く行うことができるのは、小学国語辞典、日本語 WordNet、EDR 日本語単語辞書の順に優先して換言する場合である。これは表1に示した妥当性が高い換言知識の順と一致しており、3.2節で述べた我々の提案手法は有効であると言える。

表5. 換言知識の優先順位の変化と換言の妥当性

換言知識の優先順位	ABC	ACB	BAC	BCA	CAB	CBA
妥当な換言ができる割合	73%	76%	58%	58%	63%	62%

※A: 小学国語辞典、B: EDR 日本語単語辞書、C: 日本語 WordNet

## 5.3. 難語と平易語の難易度

表2より、本稿で平易語と呼んでいる学習基本語彙の語は、小学生にとって十分理解できる語であることがわかる。

また、換言前の難語を被験者の過半数が理解できるにも関わらず、換言後の平易語を理解できないというものは、“儀式”という平易語だけであった。これは、“儀”という漢字を小学校で習わないことが原因であると考えられる。

## 5.4. 換言前後の文の難易度

表3より、全ての被験者が、原文よりも換言後の文の方を理解できると回答した割合が高い。そして、新聞を学習基本語彙に語彙制限することで、小学生が79%の文を理解可能となることがわかった。

また、換言前後の両文の難易度を比較したとき、64%については換言後の文の方が平易であるとの回答が得られた。しかし、被験者からのコメントとして、原文は「言葉が少し難しくても文が自然でわかりやすいときがある」とあった。語同士が換言可能であっても、文単位で考えると換言可能であるとは限らない。今後は文単位での換言が課題である。

## 6. おわりに

本稿では、小学生のための文章読解支援を目的として、換言による学習基本語彙への語彙制限を行い、新聞記事のテキストを平易化した。

シソーラスおよび2種類の国語辞典を用いた換言により、2000年の毎日新聞に頻出する難語100語のうち96%を学習基本語彙に制限することができ、76%は実際に換言可能な換言対を得ることができた。

評価実験の結果、学習基本語彙への換言によって、小学生が理解できる語を得られることを示した。また、学習基本語彙への語彙制限により、小学生が新聞の79%の文を理解できるようになることも示した。

## 使用した言語資源およびツール

- 1) 日本電子化辞書研究所. EDR 電子化辞書仕様説明書. 1995.
- 2) 湊吉正監修, チャレンジ小学国語辞典第五版. 株式会社ベネッセコーポレーション, 2011.
- 3) 日本語 WordNet, Ver.1.1, 独立行政法人情報通信研究機構(NICT), <http://nlpwww.nict.go.jp/wj-ja/>
- 4) 工藤拓. MeCab: Yet Another Part-of-Speech and Morphological Analyzer, Ver.0.993. <http://mecab.sourceforge.net/>
- 5) IPA 品詞体系日本語辞書「IPADIC」, Ver.2.7.0, 奈良先端科学技術大学院大学 松本研究室. <http://sourceforge.jp/projects/ipadic/>
- 6) 毎日新聞社. CD-毎日新聞 2000年度版, 2000.

## 参考文献

- [1] 梶原智之, 山本和英. 小学生の読解支援に向けた語釈文による換言. NLP 若手の会第7回シンポジウム, 2012.
- [2] 甲斐睦朗監修, 松川利広編, 語彙指導の方法: 語彙表編. 光村図書出版株式会社, 2002.
- [3] 鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史. 格フレームの対応付けに基づく用言の言い換え. 自然言語処理, Vol.10, No.4, pp.65-81, 2003.
- [4] 美野秀弥, 田中英輝. 国語辞典を使った放送ニュースの名詞の平易化. 言語処理学会第16回年次大会, pp.760-763, 2010.
- [5] 美野秀弥, 田中英輝. 放送ニュースの動詞連用形名詞の平易化. 言語処理学会第17回年次大会, pp.744-747, 2011.