

# 修士論文

## 文章読解支援のための語彙平易化

長岡技術科学大学大学院 工学研究科

電気電子情報工学専攻

学籍番号：11313388

梶原 智之

指導教員 山本 和英 准教授

2015年2月12日



# 文章読解支援のための語彙平易化

梶原 智之

## 概要

Webの発展により、大量かつ多様なテキストデータに多くの人が容易にアクセスできるようになった。しかし、子どもや外国人などを含む全ての人にとってそのテキストの理解までもが容易であるとは言えない。そこで著者は、効率的な情報収集や知識獲得のために言語能力の差を埋める技術が必要であると考え、文章読解支援のための語彙平易化の研究に取り組んで来た。語彙平易化は、文中の難解な語をより平易な同義語に置換する技術である。語彙平易化技術によって、外国人などの言語学習者や子どもをはじめとする幅広い読者の文章読解を支援することができる。

本論文の前半では、語彙平易化の重要な要素技術である語彙的換言の獲得について、2つの手法を提案する。ここでの研究の目的は、換言獲得の精度や網羅性を改善し、語彙平易化への応用のために入力よりも平易な語を得ることである。国語辞典の語釈文の主辞を見出し語の語彙的換言として獲得する1つ目の提案手法は、国語辞典の「語釈文は見出し語を平易に説明している」という特性を利用したアプローチである。この手法は、知識に基づく語彙的換言の先行研究の課題であった網羅性を、品詞を限定しない点、複数の換言知識を併用する点、「語釈文の主辞は文末に位置する」というヒューリスティクスを排除して語釈文の全体から広く換言候補を収集する点、の3点で改善するものである。小学生を対象とする実験の結果、この換言手法が読解支援に有効であることが確認できた。

2つ目の提案手法は、用例に基づく語彙的換言手法である。語の出現頻度を用いる先行研究で課題であった精度を、文脈の種類数という新しい統計量に着目することで改善した。また、文脈の種類数と語の難易度の分布を調査した結果、この手法によって多くの平易語が獲得できる見込みがあることが確認できた。

本論文の後半では、日本語の語彙平易化のための2つの研究資源の構築について述べる。1つ目の研究資源は語彙平易化システムの評価のためのデータセットである。これを公開することにより、日本語の語彙平易化システムの自動評価が初めて可能となる。従来の人手評価には、コストと再現性の課題があったが、この評価セットはこれらの課題を解決し、語彙平易化の研究を加速させるものである。

2つ目の研究資源は、語彙平易化システムである。語彙平易化システムを日本語で初めて公開することにより、読解支援を必要とする読者に技術を届けることができるようになる。また、語彙平易化の典型的な4つの機構を備えた標準的なシステムを構築したので、このシステムをベースラインとして、他の研究者が語彙平易化の研究を効果的に進めていくことができる。

本研究の成果をもとに、日本語の語彙平易化が活発に研究され、子どもや言語学習者をはじめとする幅広い読者の文章読解支援の助けとなる語彙平易化システムが開発されることを期待する。

# Lexical Simplification for Reading Assistance

Tomoyuki Kajiwara

## Abstract

Although extensive and various forms of text data are easily available in the present age, in order for readers to gather information effectively, they need a technology that overcomes any differences in their linguistic competence. Therefore, I came working on the study of lexical simplification for reading assistance. Lexical simplification is a technique that substitutes a complex word or phrase in the sentence with a simpler synonym. Lexical simplification technique supports the reading comprehension of a wide range of readers, including children and language learners.

In the first half of this paper, I propose two lexical paraphrasing techniques. Acquisition of lexical paraphrasing is an important element technique of lexical simplification. The first proposed method is a technique based on the knowledge. I propose a method for acquiring lexical paraphrase using dictionaries. Conventionally the probable last word of the statement in the definition is applied as a paraphrase, for the Japanese is a head-final language. My method collects candidates widely from the whole definition statements. As a result of experiment for primary school students, I was able to confirm that this paraphrasing technique was effective for reading assistance.

The second proposed method is a technique based on the example (corpus). I focus on the new statistics as the different number of the context and calculated the distributional similarity. The effectiveness to lexical simplification of this paraphrasing method was confirmed from the distribution of the different number of the context and the degree of difficulty of the word.

In the second half of this paper, I build and publish two research resources for Japanese lexical simplification. The first resource is a dataset for the evaluation of lexical simplification system. By using this dataset, the automatic evaluation of the Japanese lexical simplification system is made possible for the first time.

The second resource is a lexical simplification system. This system can support the reading comprehension of a wide range of readers, including children and language learners. Additionally, I construct a basic system with four typical mechanisms of lexical simplification. Therefore, researchers can use this system as a baseline.

# 目次

<b>1</b>	<b>序論</b>	<b>11</b>
1.1	語彙平易化と文章読解支援	11
1.2	目的	11
1.3	本論文の構成	12
<b>2</b>	<b>背景</b>	<b>13</b>
2.1	語彙的換言	13
2.1.1	PPDB: Japanese	13
2.1.2	内容語換言辞書 (SNOW D2)	13
2.1.3	日本語 WordNet 同義語データベース	14
2.1.4	動詞含意関係データベース	14
2.1.5	基本的意味関係の事例ベース	14
2.1.6	語彙的換言の獲得	15
2.1.7	その他の語彙的換言	15
2.1.8	換言処理の応用	16
2.2	語の難易度	16
2.2.1	単語親密度	16
2.2.2	学習基本語彙	16
2.2.3	日本語能力試験出題基準	16
2.2.4	日本語教育語彙表	17
2.2.5	統計に基づく基本語彙の推定	17
<b>3</b>	<b>語彙平易化のための知識に基づく語彙的換言</b>	<b>18</b>
3.1	関連研究：知識に基づく語彙的換言	18
3.2	複数の換言知識を併用した語彙的換言の獲得	20
3.2.1	国語辞典からの語彙的換言の獲得	21
3.2.2	シソーラスからの語彙的換言の獲得	21
3.2.3	換言の妥当性の検証	21
3.2.4	換言知識の併用	22

3.3	小学生の読解支援への有効性の検証 . . . . .	23
3.3.1	難解語と平易語の難易度 . . . . .	23
3.3.2	換言前後の文の難易度 . . . . .	24
3.4	語釈文を用いた語彙的換言手法の網羅性の改善 . . . . .	25
3.4.1	換言候補の収集 . . . . .	25
3.4.2	換言候補の選択 . . . . .	27
3.4.3	実験 . . . . .	29
3.4.4	考察 . . . . .	31
3.5	まとめ：知識に基づく語彙的換言 . . . . .	38
<b>4</b>	<b>語彙平易化のための用例に基づく語彙的換言</b>	<b>39</b>
4.1	関連研究：用例に基づく語彙的換言 . . . . .	39
4.2	提案手法 . . . . .	40
4.2.1	分布仮説に基づく換言候補の抽出 . . . . .	40
4.2.2	文脈の種類数に基づく換言候補の選択 . . . . .	41
4.3	単語出現頻度に基づく手法との比較 . . . . .	42
4.3.1	実験設定：単語出現頻度に基づく手法との比較 . . . . .	42
4.3.2	実験結果および考察 . . . . .	43
4.4	語彙平易化への有効性の検証 . . . . .	46
4.4.1	実験設定：語彙平易化への有効性の検証 . . . . .	46
4.4.2	相関係数 . . . . .	47
4.4.3	分布 . . . . .	49
4.5	まとめ：用例に基づく語彙的換言 . . . . .	51
<b>5</b>	<b>日本語の語彙平易化評価セット (SNOW E4)</b>	<b>55</b>
5.1	関連研究：語彙平易化評価セット . . . . .	55
5.1.1	McCarthy の語彙的換言データセット . . . . .	55
5.1.2	Specia の語彙平易化データセット . . . . .	56
5.1.3	De Belder の語彙平易化データセット . . . . .	57
5.2	日本語版データセットの構築 . . . . .	57
5.2.1	語彙的換言データセットの構築 . . . . .	58
5.2.2	語彙平易化データセットへの変換 . . . . .	60
5.2.3	データセットの分析 . . . . .	61

---

5.3	まとめ：語彙平易化評価セット	61
6	日本語の語彙平易化システム (SNOW S3)	63
6.1	関連研究：語彙平易化システム	63
6.2	難解語の検出	64
6.3	語彙的換言の生成	64
6.4	語義曖昧性の解消	67
6.5	難易度に基づく並び替え	68
6.6	システムの性能の検証	68
6.7	まとめ：語彙平易化システム	69
7	結論	70
	参考文献	73
	本論文に関する原著論文	84
	付録 A：3章「知識に基づく語彙的換言」の実験結果の詳細	85
	付録 B：4章「用例に基づく語彙的換言」の実験結果の詳細	112
	付録 C：本論文に関する原著論文	120

# 表目次

3.1	換言の妥当性の評価	22
3.2	難解語と平易語の難易度の評価	23
3.3	換言前後の文の難易度の評価	24
3.4	評価者 2 人ずつの kappa 係数	30
3.5	換言候補の収集	31
3.6	換言候補の選択	31
3.7	辞書ごとの基礎データ	32
3.8	辞書ごとの換言候補の収集	32
3.9	語釈文の文末以外の語を用いた換言例	33
3.10	多義語の換言候補の選択	36
3.11	少なくともある文脈では換言可能な難解語平易語対	36
3.12	6 指標を組み合わせた選択	38
4.1	コーパス	47
4.2	相関係数	48
4.3	低頻度かつ高親密な語	49
5.1	各データセットの品詞の内訳	56
5.2	語彙平易化データセットの文脈依存性	61
6.1	語彙的換言知識の概要	65
6.2	語彙的換言知識の評価	66
6.3	システムの性能	69
6.4	平易化した品詞の内訳	69
7.1	換言の妥当性の評価 ( 小学国語辞典 )	87
7.2	換言の妥当性の評価 ( EDR 日本語単語辞書 )	89
7.3	換言の妥当性の評価 ( 日本語 WordNet )	92
7.4	難解語と平易語の難易度の評価	94
7.5	換言前後の文の難易度の評価 ( 換言前の難解文 )	97



7.6	換言前後の文の難易度の評価（換言後の平易文）	100
7.7	知識に基づく語彙的換言の例	103
7.8	用例に基づく語彙的換言の例	113

## 図目次

3.1	語釈文の末尾で換言できない例	19
3.2	換言知識の併用	23
3.3	従来手法における換言候補の収集	25
3.4	提案手法における換言候補の収集	25
4.1	提案手法による名詞の換言	41
4.2	換言可能と評価された類似度 1 位の名詞数	43
4.3	換言可能と評価された類似度 10 位までの名詞数	44
4.4	類似度の順位と換言可能名詞数の関係（評価者 A）	45
4.5	類似度の順位と換言可能名詞数の関係（評価者 B）	45
4.6	類似度の順位と換言可能名詞数の関係（評価者 C）	45
4.7	新聞記事コーパス：親密度と対数頻度	50
4.8	図 4.7 のヒートマップ	50
4.9	Wikipedia コーパス：親密度と対数頻度	50
4.10	図 4.9 のヒートマップ	50
4.11	論文コーパス：親密度と対数頻度	50
4.12	図 4.11 のヒートマップ	50
4.13	新聞記事コーパス：親密度と対数後文脈種類数	52
4.14	図 4.13 のヒートマップ	52
4.15	Wikipedia コーパス：親密度と対数後文脈種類数	52
4.16	図 4.15 のヒートマップ	52
4.17	論文コーパス：親密度と対数後文脈種類数	52
4.18	図 4.17 のヒートマップ	52
4.19	新聞記事コーパス：JLPT 出題基準と対数頻度	53

---

4.20	新聞記事コーパス：JLPT 出題基準と対数後文脈種類数 . . . . .	53
4.21	Wikipedia コーパス：JLPT 出題基準と対数頻度 . . . . .	53
4.22	Wikipedia コーパス：JLPT 出題基準と対数後文脈種類数 . . . . .	53
4.23	論文コーパス：JLPT 出題基準と対数頻度 . . . . .	53
4.24	論文コーパス：JLPT 出題基準と対数後文脈種類数 . . . . .	53
4.25	新聞記事コーパス：対数頻度と対数後文脈種類数 . . . . .	54
4.26	Wikipedia コーパス：対数頻度と対数後文脈種類数 . . . . .	54
4.27	論文コーパス：対数頻度と対数後文脈種類数 . . . . .	54
6.1	語彙平易化システムの概要 . . . . .	64

# 1 序論

## 1.1 語彙平易化と文章読解支援

Webの発展により、大量かつ多様なテキストデータに多くの人が容易にアクセスできるようになった。しかし、子どもや外国人などを含む全ての人にとってそのテキストの理解までもが容易であるとは言えない。テキストの様々な読者が、効率的に情報収集や知識獲得を行うためには、言語能力の差を埋める技術が必要である<sup>54)</sup>。そこで著者は、読者の言語能力の差を埋める技術として、文章読解支援のための語彙平易化の研究に取り組んで来た。語彙平易化は、文中の難解な語をより平易な同義語に置換する技術である。語彙平易化技術によって、外国人などの言語学習者<sup>61)62)</sup>や子ども<sup>3)103)</sup>、障がい者<sup>7)10)</sup>などの幅広い読者の文章読解を支援することができる。

## 1.2 目的

本研究では語彙平易化によって、日本語学習者や小学生をはじめとする幅広い日本語の読者の文章読解を支援することを目的とする。本論文の前半では、語彙平易化の重要な要素技術である語彙的換言の生成について、語彙平易化への応用の観点から性能を向上させることを目的とする。また本論文の後半では、評価セットやベースラインシステムなどの日本語の語彙平易化のための研究資源を構築して公開し、この分野の研究を加速させることを目的とする。

前半の語彙的換言に関する研究では、語彙平易化の候補を列挙するために高精度に網羅性高く語彙的換言を収集することを目指す。2章でも述べるが、語彙的換言手法は知識に基づく手法と用例に基づく手法の2つに大別される。知識に基づく手法は、人手で整備された辞書から規則的に語彙的換言を抽出する手法が主流であり、比較的高精度に換言が得られる一方で網羅性には課題がある。網羅性の課題とは、ひとつの単語から一種類の換言しか得られない(実際には複数の換言が存在する場合がある)ことや、辞書の整備にコストがかかるため新語に対応できないこと、辞書が語彙全体を網羅できていないことなどで

ある。しかし、例えば国語辞典を用いるアプローチでは、語釈文が見出し語よりも平易に記述されているという特徴を持つため、国語辞典から換言を獲得することで平易化の効果が大きいという語彙平易化への応用のための利点を持つ。そこで知識に基づく手法については、精度を保持しつつ網羅性を向上させることを目指す。

一方で用例に基づく手法は、コーパスから統計的に語彙的換言を抽出する手法が主流であり、比較的網羅性高く換言が得られる一方で精度には課題がある。この精度の課題とは、どんな語とも共起する高頻度語が単語共起頻度の統計を取る際にノイズとなることや、逆に低頻度語が単語共起確率を計算する際にノイズとなることなどである。しかし、低コストで大規模なコーパスが得られるため、用例に基づく手法への期待は大きい。また、語彙平易化への応用を考えたとき、難解語の換言が平易化候補となることから、換言獲得の網羅性は特に重要である。そこで用例に基づく手法については、網羅性を保持しつつ精度を向上することを目指す。

本論文の後半では、語彙平易化のための研究資源として評価セットと語彙平易化システムを構築する。評価セットを構築することで、語彙平易化システムの自動評価が可能となる。これによって評価の際のコストを下げ、研究のサイクルを速く回すことができるようになる。また、共通の評価の枠組みを提供することには、複数のシステムの性能を直接比較できるようになるという利点もある。

最後に語彙平易化システムの構築には、文章読解支援を必要とする読者に語彙平易化の技術を届けるという重要な目的がある。また、標準的なシステムをベースラインとして提供することには、他の研究者との議論を深化させる目的もある。なお、このシステムの性能を上述の評価セットで評価することを通して、今後の語彙平易化システムの研究の方向性を示すことも重要である。

### 1.3 本論文の構成

まず2章では、本研究の重要な背景知識である「語彙的換言」および「語の難易度」について説明する。これらの背景を前提に、3章と4章では、語彙平易化に応用可能な性質の異なる語彙的換言手法を提案する。また、日本語における語彙平易化技術の研究を加速させるため、日本語の語彙平易化システムの評価のためのデータセットと日本語の語彙平易化のベースラインシステムを構築し、公開する。これらについて、5章および6章で説明する。最後に、7章にて本論文の結論を述べる。

## 2 背景

### 2.1 語彙的換言

意味が近似的に等価な言語表現の異形を言い換えと言う<sup>54)</sup>。言い換えを指す用語には他に、言い替え<sup>83)</sup>、書き換え<sup>46)</sup>、書換え<sup>108)</sup>、パラフレーズ<sup>106)</sup>、換言<sup>111)</sup>といったものがあり、本研究では「換言」という用語を使う。換言（特に語彙的換言）は、語彙平易化の重要な要素技術である。本節では、いくつかの日本語の換言知識と換言処理の応用について概説する。

#### 2.1.1 PPDB: Japanese

PPDB: Japanese<sup>30)1</sup> は、日英対訳コーパスから学習された句単位の語彙的換言知識である。このデータは、換言確率の閾値により、S から XL までの 4 種類が提供されている。PPDB では、英語を中間意味表現（ピボット）として用い、同じ英語の句に翻訳される複数の日本語の句を換言として収集している。

#### 2.1.2 内容語換言辞書（SNOW D2）

内容語換言辞書（用言等換言辞書<sup>113)</sup>および普通名詞換言辞書<sup>102)</sup><sup>2</sup> は、それぞれ作業員 1 名の感覚のみで換言を付与している語彙的換言知識である。各辞書の見出し語は、形態素解析器 JUMAN(Ver.7.0)<sup>3</sup> の形態素辞書の見出し語である。作業員はこれらの語について、「日本語初学者からその言葉の意味を問われたときにどのように答えるか」を念頭に置き、語または内容語 3 語までの句で換言を考える。無理な換言を行わないため、見出し語に換言が付与されていない場合もある。

<sup>1</sup><http://isw3.naist.jp/~masahiro-mi/jppdb/>

<sup>2</sup><http://www.jnlp.org/SNOW/D2>

<sup>3</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

### 2.1.3 日本語 WordNet 同義語データベース

日本語 WordNet 同義語データベース<sup>4</sup>は、日本語の概念辞書である「日本語 WordNet<sup>6)</sup>」(Ver.1.1)で同じ synset に登録されている語を組み合わせた語対から、人手で同義関係にあると判定された語対を収集した語単位の語彙的換言知識である。この辞書で扱う同義関係の範囲は、異表記、略記、丁寧語、敬語、方言、専門用語、外来語、旧名称、別称、異称、比喩など、多岐にわたる。なお、以下で紹介する動詞含意関係データベースや基本的意味関係の事例ベースなど、ALAGIN フォーラム<sup>5</sup>にて公開されている言語資源に含まれている語対は、同義関係の判定対象から除外されている。

### 2.1.4 動詞含意関係データベース

動詞含意関係データベース<sup>6</sup>は、含意関係にある動詞対を収集した語単位の語彙的換言知識である。動詞 1 が動詞 2 を含意するとは、動詞 1 の表す事態が成立するなら、同時かそれ以前に、動詞 2 の表す事態も成立しているということの意味する。このデータベースには、まず Web から自動的に含意関係にある動詞対を収集し、さらにその中から人手で含意関係にあると判定された動詞対が登録されている。これらのデータは、文字列上包含関係にあるか否かで、Triv 辞書と Ntriv 辞書に分類される。例えば、動詞“あざ笑う”と動詞“笑う”の対は、文字列上包含関係にあるため Triv 辞書に収録されており、動詞“挑戦する”と動詞“チャレンジする”の対は、文字列上の包含関係はないので Ntriv 辞書に収録されている。

### 2.1.5 基本的意味関係の事例ベース

基本的意味関係の事例ベース<sup>7</sup>は、風間ら<sup>66)</sup>の手法で構築された文脈類似語データベース<sup>8</sup>をもとに、文脈類似度の高い 2 語間の関係を人手で分類したものであり、これらのうち略記対、異形同義語対、異表記対に分類されている語対を語彙的換言知識として用いることができる。これらの辞書では、係り受けのクラスタリング結果を用いて、100 万語の名詞に対して、Web 上での文脈が類似している名詞を最大 500 語収集している。人手での分類に際しては、読みが同じで意味が同じである語対を異表記対、一方の語の文字数または音節数が他方の語より少なく他方の語の短縮形あるいは略称と呼ばれる語対を略記対、読み・形態素数・音節数のいずれかが異なり異表記対にも略記対にも該当しない同一

<sup>4</sup><http://nlpwww.nict.go.jp/wn-ja/jpn/downloads.html>

<sup>5</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-list.html>

<sup>6</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

<sup>7</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

<sup>8</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-1>

の事象/事物を指す語対を異形同義語対としている。

### 2.1.6 語彙的換言の獲得

上述したような語彙的換言知識を自動的に獲得する手法は、知識に基づく手法と用例に基づく手法に大別できる。

知識に基づく手法としては、シソーラスや国語辞典などの辞書から規則的に語彙的換言を獲得する手法が提案されている。上述の日本語 WordNet 同義語データベースや平原ら<sup>41)</sup>の研究では、日本語のシソーラスである日本語 WordNet から概念 ID を基に同義語を獲得している。また、藤田ら<sup>93)90)91)</sup>、鍛冶ら<sup>72)69)17)71)70)18)18)</sup>、土屋ら<sup>45)</sup>、村山ら<sup>55)</sup>、村田ら<sup>75)</sup>、笹野ら<sup>109)</sup>、美野ら<sup>61)62)</sup>の研究では、国語辞典を用いて同義表現を獲得している。国語辞典を用いる手法には、複数の国語辞典の語釈文を見出し語によって対応付けて同義表現を抽出する手法や、語釈文間の類似度に基づき見出し語同士を同義語として抽出する手法、見出し語の同義表現を語釈文から抽出する手法がある。また、これらと類似した手法で、Wikipedia などの Web 上のテキストから定義文を収集して語彙的換言を獲得する研究<sup>110)13)37)36)</sup>もある。

用例に基づく手法としては、対訳コーパスまたは非対訳コーパスから規則的または統計的に語彙的換言を獲得する手法が提案されている。上述の PPDB: Japanese は対訳コーパスから統計的に語彙的換言を獲得しており、動詞含意関係データベースや基本的意味関係の事例ベースは非対訳コーパスから統計的に語彙的換言を獲得している。非対訳コーパスから統計的に語彙的換言を獲得する際には、「似た意味の語は似た文脈で用いられる」という分布仮説<sup>12)</sup>に基づき文脈の類似度を計算する分布類似度と呼ばれる手法がよく用いられている<sup>24)68)88)</sup>。また、近藤ら<sup>51)</sup>、山本ら<sup>112)</sup>、泉ら<sup>98)</sup>は、非対訳コーパスから規則的に語彙的換言を獲得している。

### 2.1.7 その他の語彙的換言

これまで、内容語の語彙的換言について述べてきた。本論文では特に言及しない限り今後も内容語の語彙的換言を扱うが、ここでその他の語彙的換言についても簡単に触れておく。

複合名詞の換言については、佐藤<sup>105)</sup>や鍛冶ら<sup>73)</sup>が複合名詞を名詞句に分解しており、山本ら<sup>112)</sup>が名詞句から複合名詞を構成している。

複合動詞の換言については、内山ら<sup>81)</sup>が複合動詞を動詞句に分解している。

慣用表現の換言については、国語辞典を用いた手法で獲得することができる<sup>92)</sup>。

機能表現の換言については、松吉ら<sup>65)64)</sup>が日本語機能表現辞書「つつじ」<sup>9)</sup>を公開している。

### 2.1.8 換言処理の応用

換言処理は、テキストの評価や生成、理解に関する多くの応用を持つ。テキスト評価の観点では、換言を機械翻訳<sup>19)20)39)56)</sup>や自動要約<sup>38)41)</sup>の自動評価に応用する研究が行われている。テキスト生成の観点でも、機械翻訳<sup>47)48)14)27)</sup>や自動要約<sup>50)1)60)</sup>の分野で換言が応用されている。テキスト理解の観点では、語彙平易化<sup>53)72)16)77)76)</sup>に応用され、言語学習者や子ども、障がい者などの文章読解を助けている。その他、推敲支援<sup>78)</sup>や質問応答<sup>89)</sup>、情報検索<sup>104)</sup>などにも換言が応用されている。

## 2.2 語の難易度

語の難易度の基準は、語彙平易化技術の重要な背景知識のひとつである。本節では、いくつかの日本語の語の難易度の基準について概説する。

### 2.2.1 単語親密度

単語親密度<sup>79)</sup>は、ある単語がどの程度なじみがあると感じられるかを1(なじみがない)から7(なじみがある)までの実数値で表した指標である<sup>80)</sup>。単語親密度が低いほど難解な語、単語親密度が高いほど平易な語として、この指標は先行研究<sup>40)52)97)</sup>で単語の難易度の判定のために利用されている。

### 2.2.2 学習基本語彙

学習基本語彙<sup>99)</sup>は、小学国語教科書等の語彙分析に基づいて選定された、小学生が言語表現に駆使できる語彙である。学習基本語彙に含まれる語を平易な語、学習基本語彙に含まれない語を難解な語として、この指標も単語の難易度の判定のために利用することができる。

### 2.2.3 日本語能力試験出題基準

日本語能力試験(JLPT: Japanese Language Proficiency Test)は、日本語非母語話者のための日本語能力測定試験である。その出題基準<sup>58)</sup>は、単語を1級(難解)から4級(平易)までの4段階に分類している。 JLPT 出題基準の級が高いほど難解な語、 JLPT 出

<sup>9)</sup><http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/>



題基準の級が低いほど平易な語として、この指標も先行研究<sup>67)62)82)</sup>で単語の難易度の判定のために利用されている。

#### 2.2.4 日本語教育語彙表

日本語教育語彙表<sup>101)</sup>は、現代書き言葉均衡コーパス<sup>59)</sup>と日本語教科書コーパス(市販の日本語教科書100冊)の語彙分析に基づき、日本語教師が各単語に6段階の難易度を付与している。初級を平易な語、上級を難解な語として、この指標も先行研究<sup>74)42)</sup>で単語の難易度の判定のために利用されている。

#### 2.2.5 統計に基づく基本語彙の推定

ここまでに語の難易度や基本語彙に関する言語資源を紹介したが、その他に統計に基づく基本語彙の推定手法がいくつか提案されている。美野ら<sup>63)</sup>は、単語の出現頻度や頻度分布に基づく基本語彙の抽出指標を提案している。荒牧ら<sup>44)</sup>は、単語の使用者数に基づく基本語彙の抽出指標を提案している。

言語は変化する<sup>43)</sup>ため、専門家の知見に基づいて人手で設定された基本語彙や語の難易度が、いつまでも適切であるとは言えない。そのため、このような統計的アプローチで語の難易度を動的に決定する手法の洗練は、これからの重要な研究課題のひとつである。

### 3 語彙平易化のための知識に基づく語彙的換言

本章では、文章読解支援に向けた語彙平易化を目的として、国語辞典の語釈文から平易な語彙的換言を獲得する手法を提案する。国語辞典の語釈文は、見出し語を平易な語を用いて説明しており、見出し語から語釈文中の語への換言によって語彙の平易化が見込まれる。従来は主要部終端型である日本語の特徴を利用した語釈文末の語への換言が行われて来た。しかし、見出し語と換言可能な語は必ずしも語釈文の末尾に存在するわけではない。図3.1の例では、入力文「大詰めの大一番」に含まれる名詞の難解語「大詰め」を、語釈文「芝居の最後の場面」中の名詞「芝居」「最後」「場面」にそれぞれ換言している。この例では、語釈文の文末に出現する「場面」には換言することができない。しかし、語釈文の途中に出現する「最後」に換言することで「最後の一番」という文が得られ、換言の前後で意味を保持することができる。そこで本研究では、語釈文全体を用いて換言可能性のある語を広く収集する手法を提案する。また、換言候補から最終的な換言を選択する際には、文脈を考慮するよりもシソーラスに基づく語の類似度を用いた選択の効果が高いことを実験的に示す。

まず3.1節では、知識に基づく語彙的換言手法の関連研究について述べる。続く3.2節では、先行研究の課題であった網羅性を改善するために、国語辞典などの複数の換言知識を併用する提案手法について説明する。そして3.3節では、小学生を対象とした実験によって提案手法による換言の語彙平易化への有効性を検証する。次に3.4節では、国語辞典の語釈文全体から適切な語彙的換言を選択する手法を提案し、網羅性の改善を検証する。最後に3.5節では知識に基づく語彙的換言手法についてのまとめを行う。

#### 3.1 関連研究：知識に基づく語彙的換言

Webテキスト中から換言可能な表現を自動的に獲得する手法がいくつか提案されているが<sup>35)</sup>、これらの品質はまだ不十分である。パラレルコーパスから換言を獲得する研究

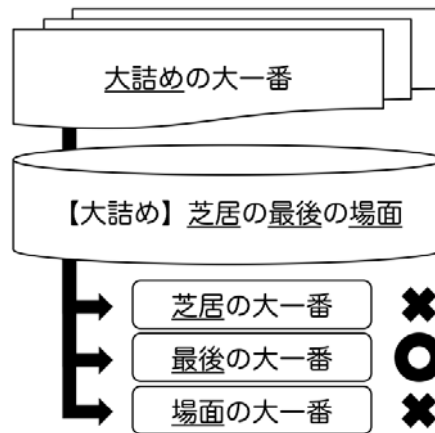


図 3.1 語釈文の末尾で換言できない例

も盛んに行われている。Barzilay ら<sup>2)</sup>は、同じ文書から作られた複数の英訳を用いて換言を獲得している。また、Shinyama ら<sup>33)</sup>は、同じ報道を行っている複数の新聞記事を用いて換言を獲得している。テキスト平易化タスクにおいては、Coster ら<sup>9)</sup>が English Wikipedia と Simple English Wikipedia を対応付けた平行コーパスを作成し、統計的機械翻訳の枠組みで平易化を行っている。これらの平行コーパスを用いる手法では、対応する表現同士のアライメントの精度や利用可能なコーパスの量に課題がある。本研究で扱うのは語彙レベルの換言であり、既存の語彙資源を用いることで、アライメントの精度の問題を度外視することができる。

既存の語彙資源から換言可能な語の対を獲得する手法には、シソーラスを用いるものや国語辞典を用いるものがある。シソーラスを用いる手法は、概念間の階層の距離を計算して意味的な近さを測ることができる利点がある。平原ら<sup>41)</sup>は、語彙的換言を用いてテキスト要約の評価を行っており、日本語 WordNet<sup>6)</sup>などのシソーラスを用いて換言知識を収集している。一方、国語辞典は見出し語を語釈文で説明する語彙資源であり、一般的に語釈文は見出し語よりも平易な語を用いて説明されている。そのため、国語辞典を用いる手法には、平易な語を獲得できる利点がある。本研究では平易な語への換言を目的とするため、国語辞典を用いた手法により平易語を獲得する。

藤田ら<sup>93)</sup>や美野ら<sup>61)</sup>は、名詞の見出し語に関する語釈文間の一致度や類似度を用いて、見出し語同士の換言を行なっている。ただし、美野らの報告にもあるように、これらの手法で得られる換言先の語は、換言元の語より平易なわけではない。本研究では語彙の平易化を目的としているため、見出し語よりも語釈文が平易と期待される国語辞典の特性を活かし、見出し語から語釈文への換言を行う。

鍛治ら<sup>70)</sup>は「見出し語が用言であれば、その語釈文は用言を主辞とする形で記述されており、なおかつ主辞は語釈文の末尾に位置する」と仮定し、国語辞典を使った用言の換言手法を提案した。美野らはまた「語釈文における主要文の最終文節が、見出し語の意味を表している」と考え、国語辞典を使った体言の換言手法も提案している。これらの研究は、どちらも主要部終端型である日本語の特徴を利用して、語釈文の末尾から換言先の語を抽出している。

しかし、図3.1に示したように、見出し語と換言可能な語は必ずしも語釈文の末尾に存在するわけではない。図3.1の例では、入力文「大詰めの大一番」に含まれる名詞の難解語「大詰め」を、語釈文「芝居の最後の場面」中の名詞「芝居」「最後」「場面」にそれぞれ換言している。この例では、語釈文の文末に出現する「場面」には換言することができない。しかし、語釈文の途中に出現する「最後」に換言することで「最後の大一番」という文が得られ、換言の前後で意味を保持することができる。そこで本研究では、語釈文全体を用いて換言可能性のある語を広く収集する手法を提案する。

語釈文の全てを扱うことによって、複数の換言先候補を得ることができる。これに伴い、複数の選択肢の中から最適な換言を選択する処理が必要となる。適切な語の選択については、シソーラスから得られる意味の類似度を用いる Ma らの手法<sup>26)</sup>や、分布仮説<sup>12)</sup>に基づいて大規模コーパスから得られる統計的な情報を用いる Lapata ら<sup>23)</sup>や Keller ら<sup>21)</sup>の手法がある。シソーラスは意味のまとまりで単語を階層的に分類した語彙資源であり、シソーラス中での単語間の距離を測ることで、単語間の意味の近さを測ることができる。また、分布仮説によると、意味の似た語は似た文脈で用いられる。この仮説に基づき、Lapata らや Keller らは共起頻度や n-gram を用いて、表現の適格さを判定することができる。本研究では、シソーラスに基づく語の類似度を用いる指標と、共起頻度や n-gram 頻度などの文脈を考慮する指標を比較し、換言の選択手法を検討する。

本章の以降では、まず3.2節で複数の国語辞典やシソーラスを併用して語彙的換言を獲得する手法を提案する。ここでは複数の換言知識の併用によって、語彙的換言の網羅性を改善する。そして、3.3節で小学生を対象とする実験によって、国語辞典を用いる語彙的換言手法の読解支援への有効性を検証する。最後に、3.4節では語釈文全体から見出し語の語彙的換言を選択する手法を提案し、さらに網羅性を改善する。

## 3.2 複数の換言知識を併用した語彙的換言の獲得

本研究では、換言によって語彙を平易化し、小学生のための文章読解支援を行う。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。さらに、小学生を対象とする小学国語辞典<sup>49)</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<sup>95)</sup>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。そこで本研究では、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を獲得する。

### 3.2.1 国語辞典からの語彙的換言の獲得

国語辞典の語釈文から見出し語の平易な換言を獲得する例を以下に示す。日本語は主要部終端型の言語であるので、語釈文の主要部は語釈文の文末に現れやすい。また、語釈文の主要部は見出し語と同じ品詞である場合が多い。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、最も後ろで出現する語を平易語として抽出する。

- 見出し語：控訴（名詞）
- 語釈文：上級裁判所（名詞）に再審（名詞）の申し立て（名詞）をする
- 換言対：控訴（難解語） 申し立て（平易語）

### 3.2.2 シソーラスからの語彙的換言の獲得

シソーラスから獲得できる同義語や上位語の中には、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には、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。また、難解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られる。このようなときに、語釈文から抽出した平易語は説明の一部であるため、難解語の持つ意味を大きく損なってしまう場合がある。しかし、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。

### 3.2.3 換言の妥当性の検証

学習基本語彙<sup>99)</sup>に含まれる単語を平易語、学習基本語彙に含まれない単語を難解語としたとき、100語の難解語を換言した結果を表3.1に示す。この100語の難解語は、2000年の毎日新聞<sup>100)</sup><sup>1</sup>に現れる難解語のうち、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、元の難解語と言い換え可能であると筆者が評価したものである。

<sup>1</sup>[http://www.nichigai.co.jp/cgi-bin/nga\\_search.cgi?KIND=EBCD1&ID=A8131](http://www.nichigai.co.jp/cgi-bin/nga_search.cgi?KIND=EBCD1&ID=A8131)

表 3.1 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
学習基本語彙が得られた語数	83 語	78 語	60 語	96 語
妥当な換言が得られた語数	67 語 (81%)	45 語 (58%)	38 語 (63%)	76 語 (79%)

本研究では、国語辞典として EDR 日本語単語辞書<sup>95)</sup> およびチャレンジ小学国語辞典<sup>49)</sup>、シソーラスとして日本語 WordNet<sup>6)2)</sup> の 3 種類の換言知識を併用して平易語を獲得した。また、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再帰的に換言処理を繰り返した。

表 3.1 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 (81%)、日本語 WordNet (63%)、EDR 日本語単語辞書 (58%) 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 語のうち 96 語を学習基本語彙に換言することができ、そのうち 76 語は実際に換言可能な対を得ることができた。

ところで、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。3 回換言を繰り返す例を以下に示す。

- 競売 競り :【競売】大勢の客を相手に 競り 売りする
- 競り 商い :【競り】商品を携えて売り歩く 商い
- 商い 職業 :【商い】品物を得る 職業

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。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。

### 3.2.4 換言知識の併用

3 種類の換言知識を妥当性の高い順に A、B、C として、換言知識を併用して換言を行う手順を図 3.2 に示す。3.2.3 節の調査の結果に基づき、小学国語辞典 (A)、日本語 WordNet (B)、EDR 日本語単語辞書 (C) の順に優先して換言を行う。

<sup>2</sup><http://nlpwww.nict.go.jp/wn-ja/>

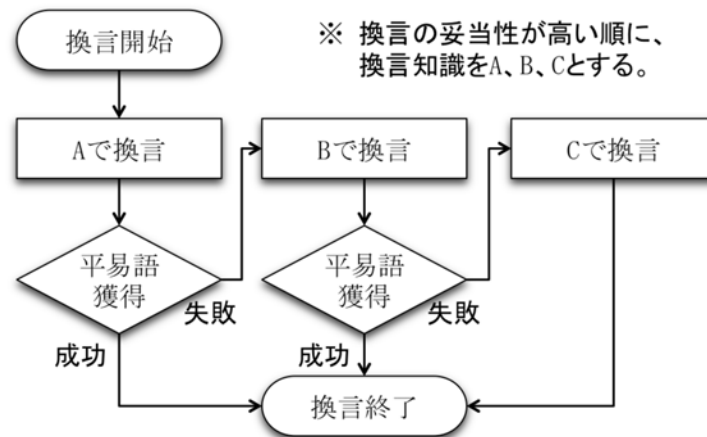


図 3.2 換言知識の併用

### 3.3 小学生の読解支援への有効性の検証

#### 3.3.1 難解語と平易語の難易度

本研究では、語の難易度の尺度として学習基本語彙を用いている。そして学習基本語彙に含まれない難解語を、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。本節では、難解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する。

3.2節で述べた方法で妥当な換言ができる難解語平易語対を、2000年の毎日新聞から頻出する順に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。評価は、被験者に難解語と平易語の対がわからないように、全200語を無作為に提示した。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。この結果を表3.2に示す。

表 3.2 難解語と平易語の難易度の評価

	難解語	平易語
被験者の一人も理解できない語	25語	1語
被験者の過半数が理解できる語	55語	97語
被験者の全員が理解できる語	33語	83語

表3.2より、本研究で平易語に設定した学習基本語彙が、小学生にとって十分理解できる語であることがわかる。被験者の一人も理解できない平易語は「儀式」という語であり、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。また、

平易語よりも難解語の方が理解できる被験者が多かったのは、「国会 議会」「トラブル 災い」という2つの難解語平易語対(難解語 平易語)のみであった。トラブルなどの片仮名の語は、学習基本語彙 5,000 語のうち 128 語しか含まれていない。片仮名語については、小学生が実際に理解できる語も難解語として扱っている可能性がある。

### 3.3.2 換言前後の文の難易度

本節では、2000 年の毎日新聞から、難解語を含む 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する。3.3.1 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 3.3 に示す。

被験者には、難解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 3.3 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。

表 3.3 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 文	50 文	75 文	43 文	78 文	65 文
		93 文	56 文	91 文	71 文	85 文	79 文
	×	0 文	0 文	8 文	7 文	7 文	4 文
		78 文	50 文	67 文	36 文	71 文	60 文
×	×	7 文	44 文	1 文	22 文	8 文	16 文
×		15 文	6 文	24 文	35 文	14 文	19 文
換言後の方が簡単		65 文	59 文	64 文	63 文	68 文	64 文

表 3.3 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。今後は文脈を考慮した換言を行いたい。



### 3.4 語釈文を用いた語彙的換言手法の網羅性の改善

#### 3.4.1 換言候補の収集

##### 3.4.1.1 従来手法

3.1 節でも述べたように、国語辞典の語釈文から見出し語の換言を獲得する従来の手法<sup>70)61)</sup>では、主要部終端型である日本語の特徴を利用して、語釈文の末尾から換言先の語を抽出している。従来手法として例えば3.2.1 節では、次の手順で換言候補の語を抽出する(図3.3)。

- (1) 入力文から難解語を検出
- (2) 難解語を見出し語として国語辞典から語釈文を抽出
- (3) 語釈文を形態素解析
- (4) 難解語と同じ品詞の語のうち最も文末の近くに出現する語を収集
- (5) 収集した語から難解語を取り除き平易語のみを残す

こうして得られた平易語が、従来手法における換言候補となる。図3.3において、「芝居」「最後」の2語は、難解語「大詰め」と同じ品詞ではあるが、最も文末の近くに出現する語ではないので、換言候補にはならない。

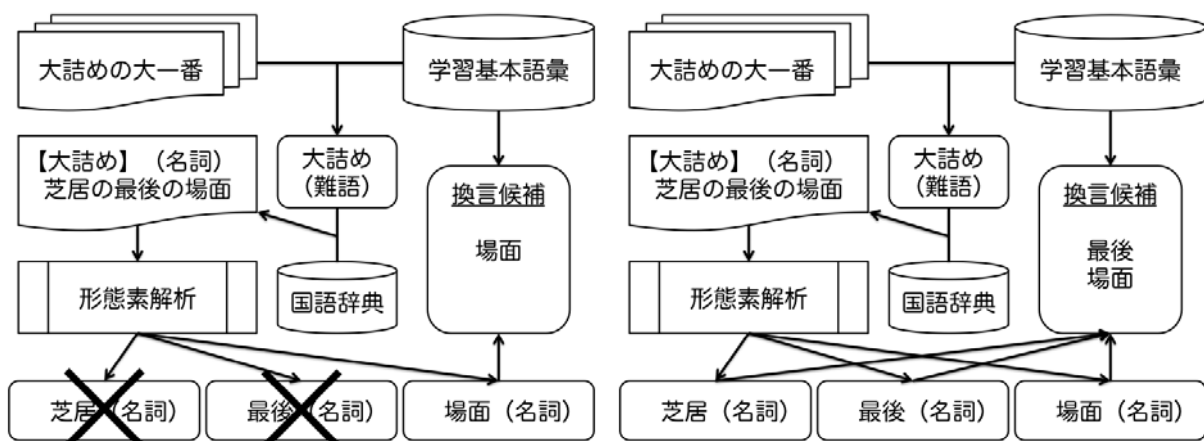


図 3.3 従来手法における換言候補の収集

図 3.4 提案手法における換言候補の収集

### 3.4.1.2 提案手法

図3.1に示したように、見出し語と換言可能な語は必ずしも語釈文の末尾に存在するわけではない。そこで本研究では、語釈文全体を用いて換言可能性のある語を広く収集する手法を提案する。提案手法では、次の手順で換言候補の語を抽出する(図3.4)。

- (1) 入力文から難解語を検出
- (2) 難解語を見出し語として国語辞典から語釈文を抽出
- (3) 語釈文を形態素解析
- (4) 難解語と同じ品詞の語を全て収集
- (5) 収集した語から難解語を取り除き平易語のみを残す

こうして得られた平易語集合が、提案手法における換言候補となる。提案手法の従来手法との違いは、(4)の処理において文末の一語に限定せず、難解語と同じ品詞の語を全て収集することである。これにより、図3.4の例では、文末ではない「最後」という語が換言候補に含まれるようになり、従来手法では扱えなかった図3.1のような換言ができるようになる。また、図3.4において「芝居」は、学習基本語彙に含まれないため換言候補にはならない。このような語は、再び語釈文を参照することによって平易に置き換えられる可能性がある。しかし、3.2.3節でも述べたように、完全に同義な換言対の存在は稀で、二度以上の再帰的な処理によって換言可能な語が得られることは少ない。そのため、本研究では語釈文を一度だけ参照する。

### 3.4.1.3 難解語と同じ品詞の平易語

換言候補の収集に際して、初めに所与の入力文から難解語を検出するが、ここでは難解語を内容語に限定する。内容語とは、名詞、動詞、形容詞および副詞である。また、得られる換言候補語は、難解語と同じ品詞の平易語である。ただし、動詞にはサ変動詞(サ変名詞+する)を含む。その他、形容詞の連用形と副詞、形容詞の連体形と「名詞+の」なども文法上同じ働きをされると考えられるが、本研究ではこれらには対応していない。

また、同じ品詞の語の中には、反義語や否定語「ない」を伴って反対の意味を表す語も含まれる。反義語や否定語が用いられる文脈は、元の語が用いられる文脈と似ていることが実験的に知られている<sup>25)</sup>。ただし、否定語「ない」を含む語釈文は約4%と少ない(EDR日本語単語辞書)。そのため、本研究では反義語や否定語に対して特別な処理は行っていない。

#### 3.4.1.4 複数の語釈文

見出し語に対応する語釈文は語義の数だけ存在するため、ある見出し語が多義である場合、その見出し語には複数の語釈文に対応する。また、一つの語義の中にも複数の語釈文が含まれている場合もある。このように、ある見出し語に複数の語釈文に対応している場合、それら全ての語釈文から換言候補の収集を行う。

### 3.4.2 換言候補の選択

この節では、3.4.1 節で収集した換言候補の中から、適切な換言を選択する方法を説明する。本研究では、シソーラスに基づく語の類似度を用いる指標と、共起頻度などの文脈を考慮する指標を比較し、換言の選択手法を検討する。

#### 3.4.2.1 シソーラスに基づく語の類似度を用いる指標

換言の前後で意味をできるだけ保つために、収集した換言候補の中から、難解語との意味の類似度が最も高い平易語を選択する。類似度の計算には、日本語 WordNet を用いる。WordNet は、同義語の集合が階層的に分類されている言語資源であり、式 3.1 に示すように、2 つの単語が属する同義語集合の深さおよび共通上位概念の深さを用いて意味の類似度を計算することができる。類似度は、単語  $w$  の属する同義語集合の深さを  $d(w)$ 、単語  $w$  と難解語  $n$  の共通の上位概念の深さを  $d_c(w, n)$  として、式 3.1 で定義される。類似度が最も高い平易語が複数存在する場合は、類似度が最も高い平易語の中からランダムに一つを選択することとする。

$$score_{sim}(w, n) = \frac{2d_c(w, n)}{d(w) + d(n)} \quad (3.1)$$

#### 3.4.2.2 語釈文中での出現頻度を用いる指標

より多くの種類の国語辞典から同じ平易語が換言候補として収集されるほど、その平易語が難解語の換言として適切であると考えられる。そこで、単語  $w$  の換言候補中での出現頻度  $freq\_candidate(w)$  を用いた式 3.2 のスコアを定義する。

$$score_{freq}(w) = freq\_candidate(w) \quad (3.2)$$

#### 3.4.2.3 共起頻度を用いる指標

入力文中の内容語とよく共起する平易語は、難解語の換言として適切であると考えられる。そこで、換言候補中の単語  $w$  と入力文中の内容語  $c \in C$  のコーパス中での共起頻度  $co\_freq(w, c)$  を用いた式 3.3 のスコアを定義する。ここで、 $C$  は入力文に含まれる内容語の集合である。

$$score_{co\_freq}(w, C) = \sum_{c \in C} co\_freq(w, c) \quad (3.3)$$

#### 3.4.2.4 自己相互情報量を用いる指標

式 3.3 では共起頻度を用いた指標を定義したが、共起頻度は単語単体での出現頻度から大きく影響を受け、出現頻度の高い単語ほど共起頻度も高くなる場合が多い。これに対して、自己相互情報量は単語単体での出現頻度の影響を差し引いて共起性を測ることができる。そこで、換言候補中の単語  $w$  と入力文中の内容語  $c \in C$  のコーパス中での自己相互情報量を用いた式 3.4 のスコアを定義する。ここで、 $C$  は入力文に含まれる内容語の集合である。また、 $freq(w)$  は単語  $w$  のコーパス中での出現頻度、 $co\_freq(w, c)$  はコーパス中での単語  $w$  と単語  $c$  の共起頻度を表す。

$$score_{pmi}(w, C) = \sum_{c \in C} \log \frac{co\_freq(w, c)}{freq(w)freq(c)} \quad (3.4)$$

#### 3.4.2.5 単語 3-gram 頻度を用いる指標

入力文中の難解語を、換言候補中の単語  $w$  と置き換える。このときの単語  $w$  の周囲の語を用いて、コーパス中での単語 3-gram 出現頻度を求める。この単語 3-gram 出現頻度が高い単語ほど、難解語の換言として適切であると考えることができる。そこで、単語  $w$  で終わる 3-gram、単語  $w$  が中央に来る 3-gram、単語  $w$  から始まる 3-gram の 3 つの単語 3-gram 出現頻度の総和を用いた式 3.5 のスコアを定義する。ここで、入力文中の各語を難解語からの相対位置  $i$  を用いて  $x_i \in X$  で表す。また、 $freq_{3-gram}(x_{-1}, w, x_{+1})$  は  $x_{-1}, w, x_{+1}$  という単語 3-gram のコーパス中での出現頻度を表す。

$$\begin{aligned} score_{3-gram}(w, X) &= freq_{3-gram}(x_{-2}, x_{-1}, w) \\ &+ freq_{3-gram}(x_{-1}, w, x_{+1}) \\ &+ freq_{3-gram}(w, x_{+1}, x_{+2}) \end{aligned} \quad (3.5)$$

#### 3.4.2.6 共起頻度ベクトルの類似度を用いる指標

難解語  $n$  と入力文中の他の内容語  $c \in C$  の共起頻度ベクトルと換言候補中の単語  $w$  と入力文中の難解語を除く内容語の共起頻度ベクトルの類似度を計算する。この類似度が高い単語ほど、難解語と似た文脈で用いられることを表しており、難解語の換言として適切であると考えることができる。ここでは、ベクトルの類似度としてコサイン類似度を用い

た式 3.6 のスコアを定義する。ここで、 $vector(w, C)$  は単語  $w$  と入力文中の内容語集合  $C$  に含まれる各内容語  $c$  のコーパス中での共起頻度ベクトルである。

$$score_{vec}(w, n, C) = \frac{vector(w, C) \cdot vector(n, C)}{|vector(w, C)| |vector(n, C)|} \quad (3.6)$$

### 3.4.3 実験

#### 3.4.3.1 使用したツールおよび語彙資源

3.4.1.2 節でも述べたが、完全に同義な換言対の存在は稀である。そのため、ある文に含まれる複数の難解語を同時に平易化しようとしたとき、個々の平易化が妥当な変換であっても文全体ではニュアンスが異なるという場合も考えられる。そこで本研究では、難解語が 1 語だけ含まれる文を新聞記事から抽出して実験を行った。このような文は、毎日新聞 2000 年度版<sup>100)</sup><sup>3)</sup> に 14,344 文 (全 232,038 文) 含まれる。毎日新聞 2000 年度版には全 26,709 種類の難解語が含まれるが、実験に用いるのは 50 回以上出現する 221 語とする。この 221 語のうち、語釈文中に平易語が含まれない 56 語は換言対象から除いた。さらに全ての換言候補が換言可能であり、選択の必要がない 13 語も換言対象から除き、最終的に 152 種類の難解語を実験で使用した。なお、この 152 種類の難解語は毎日新聞 2000 年度版に延べ 72,153 回出現している。

本実験では 1 種類の難解語につき 1 文ずつ実験対象の文を用意した。これらの文は、毎日新聞 2000 年度版の難解語を 1 語だけ含む 14,344 文から無作為に抽出した。152 語の難解語につき対象とした各 1 文が無作為である限り、難解語 1 語で複数文を対象とする場合の実験結果と同様の傾向が得られると考え、このような実験設定とした。

換言の網羅性を高めるため、換言知識として複数の国語辞典を併用した。本実験では、EDR 日本語単語辞書<sup>95)</sup>、チャレンジ小学国語辞典<sup>49)</sup>、三省堂国語辞典<sup>57)</sup> の 3 種類の国語辞典を用いた。

3.4.1 節で説明した換言候補の収集では、形態素解析に IPADIC(2.7.0)<sup>4)</sup> および MeCab(0.993)<sup>22)</sup><sup>5)</sup> を用いた。

3.4.2 節で説明した換言候補の選択では、単語の出現頻度や共起頻度などを計算するコーパスとして Web 日本語 N グラム<sup>86)</sup><sup>6)</sup> を用いた。Web 日本語 N グラムとは、一般に公開されている日本語の Web ページ約 200 億文から抽出された出現頻度 20 回以上の単語 1-gram から単語 7-gram である。本実験では、最も長い単語 7-gram データから単語出現頻度や

<sup>3)</sup>[http://www.nichigai.co.jp/cgi-bin/nga\\_search.cgi?KIND=EBCD1&ID=A8131](http://www.nichigai.co.jp/cgi-bin/nga_search.cgi?KIND=EBCD1&ID=A8131)

<sup>4)</sup><http://sourceforge.jp/projects/ipadic/releases/24435/>

<sup>5)</sup><https://code.google.com/p/mecab/>

<sup>6)</sup><http://www.gsk.or.jp/catalog/gsk2007-c/>

共起頻度を求めた。出現頻度については 1-gram データをそのまま使うこともできるが、式 3.4 の計算時に分母と分子の母集団を揃えるために 7-gram データを用いて求めた。ただし、単語 3-gram 出現頻度については 3-gram データから求めた。なお、3-gram データの異なり数は 394,482,216、7-gram データの異なり数は 570,204,252 である。

### 3.4.3.2 実験方法

まず、3.4.1 節に示した提案手法および従来手法で換言候補の収集を行った。そして、提案手法で収集した換言候補に対して、3.4.2 節に示した 6 つの指標で換言候補の選択を行った。

評価は、日本語を母語とする工学系大学院生 3 人が行った。3 人中 2 人以上が、選択した平易語が原文において難解語と換言可能であると判断した場合を正解とした。評価者 2 人ずつの kappa 係数を表 3.4 に示す。kappa 係数は全て 0.6 以上であり、評価者間の評価の一致度は十分高い。

表 3.4 評価者 2 人ずつの kappa 係数

評価者	A and B	B and C	C and A
kappa 係数	0.617	0.600	0.662

### 3.4.3.3 実験結果：換言候補の収集

表 3.5 に換言候補の収集の実験結果を示す。「得られた換言候補の数」とは、提案手法または従来手法を用いて 152 語の難解語から収集した換言候補の延べ数である。「換言可能な候補の数」とは、収集された換言候補の中で難解語と換言可能な語の延べ数である。例えば、難解語「サポート」は候補「支える」および候補「助ける」の両方と換言可能である。「換言可能な難解語の数」とは、換言可能な候補によって換言できる難解語の延べ数である。提案手法では従来の文末から換言候補を収集する手法よりも多くの換言候補を収集することができ、従来手法では 64% (難解語 152 語のうち 98 語) のところ、提案手法では 84% (難解語 152 語のうち 128 語) の難解語の平易な換言を収集することができる。

本研究では、国語辞典の見出し語を語釈文中の語に換言することで語彙の平易化ができるという考えのもと、図 3.1 のような従来の語釈文末の語に換言するだけでは扱えない換言について、語釈文中の他の語を用いて換言することを提案した。この実験結果は、語釈文が文末以外にも見出し語と換言可能な語を多く含んでいることを示すものであり、提案手法がより網羅性の高い換言対の収集のために効果的であることを示している。

表 3.5 換言候補の収集

	従来手法	提案手法
得られた換言候補の数	320	1076
換言可能な候補の数	141	239
換言可能な難解語の数	98 (64%)	128 (84%)

#### 3.4.3.4 実験結果：換言候補の選択

表 3.6 に換言候補の選択の実験結果を示す。ここでは、収集した換言候補中に換言可能な平易語が含まれていないものを除き、表 3.5 で示した換言可能な 128 の難解語についてのみ候補の選択を行った。なお、128 の難解語に対して、合計 784 語の換言候補が収集されており、ランダムに選択を行った場合の正解率は 30% である。

表 3.6 より、文脈を考慮する他の指標に比べて、シソーラスに基づく語の類似度を用いる指標で選択した場合の正解率が高いことが分かる。これは、比較的単純な方法で文脈を考慮しても効果は限定的であり、難解語との意味の類似度が高い語を選択することが有効であることを示している。

表 3.6 換言候補の選択

選択の指標	選択の正解数	選択の正解率
WordNet 類似度	86	67%
出現頻度	54	42%
共起頻度	61	48%
自己相互情報量	79	62%
3-gram 頻度	74	58%
共起頻度ベクトル	74	58%

### 3.4.4 考察

#### 3.4.4.1 換言知識の特性

表 3.7 に、難解語 1 語あたりの平均語釈文数（語義数）、語釈文 1 文あたりの平均文長（内容語数）、語釈文 1 文あたりの平均平易語数を、辞書ごとに示す。一般向けの EDR 日本語単語辞書や三省堂国語辞典は見出し語 1 語あたりの語義数が多く、小学生向けのチャレンジ小学国語辞典は 1 語ごとの語義数が少ないことが分かる。一方でチャレンジ小学国語辞典は、語釈文が長く、平易語を多く含んでいる。これは、語釈文でより平易な説明を

行うために、平易な語を積極的に使用して詳細な説明をしているためである。平易な語を用いて詳細に説明するというのは、Simple English Wikipedia<sup>7</sup> の書き方のガイドラインにも記載されている内容であり、平易な文章の書き方の一つの特徴である。

表 3.7 辞書ごとの基礎データ

	チャレンジ	EDR	三省堂
見出し語総数	33,700	270,000	73,000
平均語釈文数（語義数）	1.79	4.74	2.28
平均文長（内容語数）	4.10	3.83	3.12
平均平易語数	3.39	2.55	2.43

次に、表 3.5 の結果を辞書ごとに見る（表 3.8）。EDR 日本語単語辞書と三省堂国語辞典では、従来手法に比べて提案手法で 3.1 倍多くの換言候補を収集することができている。また、換言可能な候補も従来手法より 1.6 倍多く収集することができている。これに対して、チャレンジ小学国語辞典は換言候補を従来手法の 4.2 倍、換言可能な候補を従来手法の 1.9 倍多く収集することができている。これは表 3.7 にも示した通り、小学生向けの国語辞典が語釈文中により多くの平易語を含むためであり、小学生を対象とする本論文の語彙平易化のためには小学国語辞典を用いるのが有効であると言える。ただし、表 3.7 にも示した通り、収録されている見出し語の総数が少ないのは小学生向けの国語辞典の弱みであり、小学国語辞典単体での換言では網羅性は低い。EDR 日本語単語辞書など一般向けの国語辞典は、見出し語数も語義数も多いので、網羅性の高い換言候補の収集のためには、やはり換言知識を組み合わせることは重要である。

表 3.8 辞書ごとの換言候補の収集

	従来手法			提案手法		
	チャレンジ	EDR	三省堂	チャレンジ	EDR	三省堂
得られた換言候補の数	79	232	117	334	712	358
換言可能な候補の数	40	106	55	77	165	90
換言可能な難解語の数	34	76	41	49	102	60

続いて、表 3.5 で換言可能な候補を収集できなかった難解語について、学習基本語彙のいずれの語とも換言ができないのか、換言知識に用いた辞書のために候補が列挙できな

<sup>7</sup>[http://simple.wikipedia.org/wiki/Main\\_Page](http://simple.wikipedia.org/wiki/Main_Page)



かったのかについて調査する。語釈文中に平易な語が含まれないとして表 3.5 の実験に用いなかった 56 語および提案手法で換言可能な候補の収集に失敗した 24 語の計 80 の難解語について調査を行う。3.4.3.2 節の評価方法に従い、学習基本語彙中に換言可能な語が含まれているか否かについて評価を行ったところ、80 語のうち 50 語は換言可能な平易語が存在し、30 語は学習基本語彙のいずれの語を用いても換言できないことが分かった。平易語が存在する 50 語には、難解語「廃棄する」に対しての平易語「捨てる」など、語釈文に含まれていてもおかしくない例があり、換言知識を一層充実させる必要がある。平易語が存在しない 30 語には、難解語「後続」など「後に続く」のように句単位で換言すべき例がある。句単位での換言は今後の課題である。

#### 3.4.4.2 換言候補の収集

表 3.9 に、語釈文の文末以外の語を用いた換言の例を示す。これらの例では、語釈文の末尾に出現する難解語と同じ品詞の語は難解語と換言不可能であり、文中の他の難解語と同じ品詞の語が難解語と換言可能である。「継続」と「冒頭」の例に注目されたい。難解語「継続」はサ変名詞であり、語釈文末は「行う」という動詞であるので、これは換言候補である。しかし、この例では「行う」の手前に出現している「続ける」が難解語の換言として適切である。また、難解語「冒頭」は名詞であり、語釈文末は「部分」という名詞であるので、これは換言候補である。しかし、この例では「部分」の手前に出現している「初め」が難解語の換言として適切である。これらの「行うこと」や「～の部分」という表現は、語釈文の中で意味の中心となるような表現ではなく、換言処理においては不要な表現である。このような不要な表現は、「行うこと」や「～の部分」の他にも「～すること」「～の一種」「～の一つ」など多数存在し、語釈文の末尾からのみ換言を抽出する従来手法では、これらの不要な表現を獲得して換言として適切な他の語を収集できない。

表 3.9 語釈文の文末以外の語を用いた換言例

難解語	語釈文	平易語	換言例
継続	前からやっていることを、続けて行うこと	続ける	話し合いを { 継続し / 続け } ていく
出演	映画・放送・舞台に出て、演じること	出る	日本の映画に { 出演する / 出る }
警戒	よくないことが起きないように注意し、用心すること	注意	{ 警戒 / 注意 } は嚴重
確信	かたく信じて動かない心	信じる	「自分たちが進む道はこれだ」と { 確信し / 信じ } た
講師	大学教師の職の名の一つ	教師	{ 講師 / 教師 } と朗読作品は次の通り
収益	利益を手に入れること	利益	{ 収益 / 利益 } は寄付
悲劇	悲しい出来事や不幸な人生を題材にした劇や映画	不幸	私にとっては大きな { 悲劇 / 不幸 } だ
冒頭	文章や談話の初めの部分	初め	{ 冒頭 / 初め } いきなり次のような場面からはじまる
国籍	その国の国民であるという資格や身分	国	{ 国籍 / 国 } や民族の違いは関係ない

### 3.4.4.3 各選択指標の特徴

換言候補の選択に関する表 3.6 の実験結果を考察する前に、各指標の特徴について考える。語彙的換言を選択するこのタスクでは、単語間の意味の近さを適切に測定することが重要である。3.4.2 節に示した各指標は、いずれもこの単語間の意味の近さを測定するものである。

語釈文中での出現頻度を用いる指標では、難解語の説明に欠かせない語は多くの語釈文の中に出現するであろうというヒューリスティクスに基づいて換言可能性を計算している。しかし、3.4.4.1 節で述べたように、実際には見出し語の平易な換言が語釈文中に含まれる保証はない。語釈文中に換言が含まれていたとしても、その語が平易語だとは限らない。例えば、3.4.4.1 節で挙げた難解語「廃棄する」の例では、換言可能な「捨て去る」という語は語釈文中に出現するが、この語は平易語ではない。また、語釈の方法は一通りではないため、「捨てる」や「捨て去る」という語以外にも「取去る」という語で難解語「廃棄する」の語釈をしている国語辞典もある。このような理由で、見出し語の換言が語釈文中で必ずしも高頻度で出現するという保証はない。

共起頻度や自己相互情報量、単語 3-gram 頻度を用いる指標では、入力文と換言候補の関係を扱っている。これは、難解語と換言候補の関係を間接的に扱っていると考えられることができる。難解語「支援」を含む入力文「政党の支援なんて必要ない」を考える。難解語「支援」と換言可能な候補「助け」は、他の候補「貸し」や「はげまし」と比べて入力文中の語との共起頻度が高く、これらの指標では適切に「助け」が換言として選択される。難解語と換言可能な候補が、入力文との馴染みが良いのは、たしかであろう。一方で、難解語「騒ぎ」を含む入力文「それが騒ぎを大きくした」を考える。これらの指標で選択される候補は「問題」である。この候補で難解語を置換した文「それが問題を大きくした」は、この置換後の文だけを見ると自然な日本語の文である。文中の語と換言候補「問題」の共起頻度も高い。しかし、この候補「問題」は難解語「騒ぎ」とは厳密には換言の関係にはない。難解語「騒ぎ」は候補「問題」を含意するが、「問題」には「騒ぐ」「騒がしい」といった意味は含まれていないからである。また、難解語「騒ぎ」の換言候補には「秩序」も含まれる。換言不可能と評価されたこの候補は、入力文中の語との共起頻度は低い。入力文にそぐわない換言候補は難解語の換言に成り得ないが、入力文との馴染みが良い語が必ずしも難解語と換言可能なわけではない。つまり、共起頻度などで計算される入力文と換言候補の関係は、換言の必要条件には成り得るが十分条件ではない。

共起頻度ベクトルの類似度を用いる指標は、難解語の文脈と換言候補の文脈を比較することで、難解語と換言候補の関係を間接的に扱うことができる。分布仮説によると意味の似た語は似た文脈で用いられるので、文脈の類似度が高い語を選択するこの指標には意味

の近い語を選択することができるという期待がある。しかし、この指標も共起頻度に基づく指標であり、換言候補自身の出現頻度に影響を受ける傾向がある。先の難解語「騒ぎ」の例では、この指標は換言不可能な候補「面倒」を選択する。候補「面倒」の出現頻度は、換言可能な候補「騒動」の出現頻度の約4倍高い。また、難解語「概要」の例を考える。入力文は「インタビューの概要は次の通り」であり、換言可能な候補は「あらまし」である。この指標が選択する換言不可能な候補「物語」の出現頻度は、「あらまし」の出現頻度の約30倍高い。この指標では、出現頻度に大きな差がある場合、出現頻度の高い候補を選択してしまう傾向がある。

シソーラスに基づく語の類似度を用いる指標では、難解語と換言候補の関係を直接的に扱うことができる。これは、シソーラスが意味のまとまりで単語を分類した語彙資源だからである。難解語と換言候補の意味の近さを直接計算することができるのがこの指標の利点であるが、6種類の指標の中で唯一換言候補の出現頻度に影響を受けないという点も特徴である。この指標では、難解語「騒ぎ」の例も難解語「概要」の例も、換言候補の出現頻度に影響されずに、難解語と意味の近い換言可能な候補を選択できる。

#### 3.4.4.4 換言候補の選択

表3.6の実験結果では、文脈を考慮する他の指標に比べて、シソーラスに基づく語の類似度を用いる指標で選択した場合の正解率が高かった。WordNetの同義語(同じsynsetに定義されている語)や上位語を換言として用いることは、平原ら<sup>41)</sup>の研究でも見られ、この指標を用いて一定の精度で適切な換言が選択できるのは妥当な結果である。しかし、WordNetでの単語間の距離が小さくても、必ず換言可能であるというわけではない。例えば、日本語WordNet同義語データベース<sup>8</sup>はWordNetの同じsynsetに属する語対から同義関係にある対を人手で選別しており、単語間の距離が最短であっても全ての対が換言可能なわけではないということを示している。本研究でも、距離が同一の語が複数ある場合には、その中からランダムに一語選択しており、この部分の選択方法の改良により、正解率の更なる向上が期待される。

次に、シソーラスに基づく語の類似度を用いる指標で選択に失敗した例を考える。選択に失敗した42(128-86)語のうち、最も多い67%にあたる28語は難解語が多義語であり、多義性が解消できなかったことによる失敗であった。例えば、難解語「ネット」は多義語であり、「網」「ネットワーク」「インターネット」「正味の量」などの語義がある。自己相互情報量を用いる指標などでは適切な候補「インターネット」を選択することができており、「今のところ、ネットでできるのは申し込みだけ。」という入力文を用いると比較的簡

<sup>8</sup><http://nlpwww.nict.go.jp/wn-ja/jpn/downloads.html>

単に多義性の解消が可能な例である。しかし、WordNet では「網」と「インターネット」の両方が難解語「ネット」と同じ synset に定義されており、多義性を解消せずに選択することは難しい。

しかし、入力文の文脈を考慮する他の指標でも、多義性を解消する効果は限定的であった。表 3.10 で、いずれの国語辞典でも語義が複数存在する語を多義語と定義して、多義語の選択の正解率を比較する。このような多義の難解語は、78 語（128 語中の 61%）存在した。

表 3.10 多義語の換言候補の選択

	選択の正解数	選択の正解率
WordNet 類似度	55	71%
出現頻度	34	44%
共起頻度	37	47%
自己相互情報量	50	64%
3-gram 頻度	47	60%
共起頻度ベクトル	43	55%

表 3.10 より、入力文の文脈を考慮する指標よりも、むしろ文脈を考慮しないシソーラスに基づく語の類似度を用いる指標の方が多義語の換言の正解率が高いことが分かる。文脈を考慮した指標では、共起頻度などの統計量を近似的に語の類似度と見ているが、これらの指標ではシソーラスの距離を用いる指標ほど正確には語の類似度を測定できていないと考える。表 3.6 で選択に失敗した難解語と平易語の対を見ると、シソーラスに基づく指標では 55% が「少なくともある文脈では換言可能」な語であるが、他の指標では「少なくともある文脈では換言可能」な語が 22% から 33% と少ない（表 3.11）。

表 3.11 少なくともある文脈では換言可能な難解語平易語対

	入力文脈で換言不可能	ある文脈では換言可能
WordNet 類似度	42	23 (55%)
出現頻度	74	19 (26%)
共起頻度	67	15 (22%)
自己相互情報量	49	11 (22%)
3-gram 頻度	54	13 (24%)
共起頻度ベクトル	54	18 (33%)

3.4.4.3 節で述べたように、シソーラスに基づく指標は、難解語と換言候補の意味の近さを直接計算することができる点と換言候補の出現頻度に影響を受けない点で、他の指標よりも適切に換言を選択することができる。シソーラスに基づく指標では入力文の文脈によって多義性を解消して換言を選択することはできない。しかし、難解語と換言候補の意味の近さを直接計算できるこの指標では、少なくともある文脈では換言可能な語を選択しやすいため、多義性の解消はできなくても結果としては他の指標よりも適切な換言を選択できる場合が多い。一方、入力文の文脈を考慮する自己相互情報量などに基づく指標では、難解語と換言候補の意味の近さを間接的にしか計算できない。そのため、「少なくともある文脈では」という条件を付けても換言できない候補を選択してしまう場合が、シソーラスに基づく指標に比べて多くなってしまっている。

多義語の換言の選択においては、まず多義性の解消を行い語釈文を選択した上で、語釈文の中から見出し語の換言を選択するという2段階の処理が必要である。選択の実験を行った128語中の61%が多義語であり、最も選択の正解率が高いシソーラスに基づく指標で選択に失敗した42語中の67%も多義性の解消が課題であるため、適切な換言の選択のためには多義性の影響は十分に大きい。今後は多義性を解消した上で高精度に換言を選択できる方法を検討する必要がある。

#### 3.4.4.5 選択指標の組み合わせ

試みに、本研究で扱った6つの指標の投票により128語の難解語の換言候補を選択した(表3.12)。ここで、選択方法1は、表3.6に示した6つの選択指標で一語ずつ換言候補を選択し、最も多くの指標が選択した候補を選択する方法である。一つの候補に絞ることができない場合には、最も多くの指標が選択した候補の中からランダムに一語を選択する。選択方法2は、同様に6つの指標で一語ずつ換言候補を選択するが、一つの候補に絞ることができない場合にシソーラスに基づく指標の選択を優先する方法である。これは、表3.6においてシソーラスに基づく指標の正解率が最も高かったためである。選択方法3も同様に換言候補を選択するが、過半数である4つ以上の指標の選択が一致しなかった場合にシソーラスに基づく指標の選択を優先する方法である。この方法が最も正解率が高く、70%の正解率で適切な語彙的換言を選択することができた。最も正解率の高いシソーラスに基づく指標を優先しつつ、文脈を考慮する指標を組み合わせることで選択の正解率を向上させた。

表 3.12 6 指標を組み合わせた選択

		選択の正解数
1	6 つの指標で投票	85 (66%)
2	票が割れた場合に WordNet 指標を優先	88 (69%)
3	過半数を超えない場合に WordNet 指標を優先	90 (70%)

### 3.5 まとめ：知識に基づく語彙的換言

本章では、小学生のための文章読解支援に向けた語彙平易化を目的として、国語辞典を用いた語彙的換言を行い、新聞記事の語彙を小学生が十分理解可能な学習基本語彙に制限した。本研究の主たる主張は、次の 2 つである。

- (1) 換言候補は、語釈文の末尾に限定せず、語釈文の全体から広く収集すべきである
- (2) 換言候補から適切な換言を選択する際には、比較的簡単な方法で文脈を考慮しても効果は限定的であり、シソーラスに基づく類似度を優先して選択するのが良い

従来は語釈文の末尾の語へ換言が行われて来たが、表 3.5 に示したように、語釈文には末尾以外にも見出し語の換言が多く含まれている。そのため、語釈文全体から文法的に同じ働きをする換言候補を広く収集することで、従来手法よりも 20 ポイント多い 84% の難解語の平易な換言を収集することができた。

換言の選択は、シソーラスに基づく語の類似度を用いる指標の正解率が 67% で最も高く、文脈を考慮する指標の中では入力文中の内容語と平易語の自己相互情報量を用いる指標の 62% の正解率が最も高かった。そして、シソーラスに基づく指標を優先しつつ文脈を考慮する指標を組み合わせることで、選択の正解率は 70% まで向上した。

## 4 語彙平易化のための用例に基づく語彙的換言

本章では、大規模コーパスから得られる文脈の多様性に基づき、入力文の文脈に沿って名詞を換言する。統計情報に基づく手法は、言語に依存しないため、任意の言語のコーパスを用いて、その言語の語彙的換言を得ることができる。提案手法は、単語の出現頻度や共起頻度を用いず、名詞が用いられる文脈の種類数のみを用いて文脈の類似度を計算することが特徴である。これは、換言対象の語とより多くの文脈を共有する換言候補の語は、換言可能性がより高いという考えに基づく。実験の結果、本手法は共起頻度や自己相互情報量に基づく手法よりも適切な換言を多く生成することができた。

### 4.1 関連研究：用例に基づく語彙的換言

コーパスからの換言獲得としては、パラレルコーパスやコンパラブルコーパスを用いる研究が行われてきた。Barzilay ら<sup>2)</sup>は、同じ文書から作られた複数の英訳を用いて換言を獲得している。また、Shinyama ら<sup>33)</sup>は、同じ報道を行っている複数の新聞記事を用いて換言を獲得している。これらのパラレルコーパスやコンパラブルコーパスから換言を獲得する手法では、対応する表現同士のアライメントの精度や利用可能なコーパスの量に課題がある。本研究は、ノンパラレルコーパスを用いるため利用可能なコーパスの量に制限はない点でこれらの研究とは異なる。

ノンパラレルコーパスから得られる文脈の類似性に基づく換言手法は、Yamamoto<sup>35)</sup>、山崎ら<sup>94)</sup>、Marton ら<sup>28)</sup>、Bhagat ら<sup>5)</sup>が提案している。

Yamamoto や山崎らの研究では、コーパスに対する構文解析結果から（内容語、格助詞、内容語）または（係り元文節、対象文節、係り先文節）の三つ組を抽出し、係り受け関係にある表現の共起頻度を計算して換言を獲得している。

Marton らの研究では、機械翻訳システムの改良のために未知語の換言を行っている。コーパスから未知語と同じ文脈で出現する単語を換言候補とし、文脈との共起頻度で特徴

ベクトルを生成する。そして未知語の特徴ベクトルと各換言候補の特徴ベクトルのコサイン類似度を計算し、最も類似度が高い換言候補へ換言を行うことで機械翻訳システムの精度を改善している。

Bhagat らの研究では、250 億語のコーパスから換言を抽出している。コーパス中の単語 5 グラムを句と見なし、句ごとに相互情報量を用いて特徴ベクトルを生成する。そして同じ文脈を持つ語同士の特徴ベクトルのコサイン類似度を計算し、最も類似度が高い語の組を換言として抽出している。

本研究は、単語の出現頻度や共起頻度を計算していない点がこれらの研究と異なる。本研究は文脈の多様性に注目し、語が用いられる文脈の種類 数のみを用いて文脈の類似度を計算し、換言先の語を選択する。これは、換言対象の語とより多くの文脈を共有する換言候補の語は、換言可能性がより高いという考えに基づく。

## 4.2 提案手法

本研究では、大規模コーパスから得られる文脈の多様性に基づき、文中の名詞を他の名詞に換言する。「似た意味の語は似た文脈で用いられる」という分布仮説<sup>12)</sup>に基づき、まず入力文と同じ文脈で用いられる名詞をコーパスから抽出する。そして、抽出した各名詞と入力文中の名詞との文脈の類似度を格フレーム辞書により計算し、類似度の高い名詞へ換言を行う。図 4.1 に提案手法による名詞換言の概要を示す。

### 4.2.1 分布仮説に基づく換言候補の抽出

本手法では、換言対象の名詞の前後 1 文節を文脈と定義し、入力文と同じ文脈で用いられる名詞を換言候補としてコーパスから収集する。

まず、入力文を前文脈と後文脈に分け、各々コーパスを探索する。そして、前文脈の後に出現する名詞と後文脈の前に出現する名詞のうち共通する名詞を換言候補として収集する。例えば、「空港へのアクセスを調べる」という入力文に対して、「アクセス」を換言したい場合、「空港への」という前文脈と「を調べる」という後文脈に分けてコーパスを探索し、に該当する名詞のうち共通する名詞を換言候補として収集する。図 4.1 の例では、前文脈と後文脈で共通して用いられる「乗り換え」「料金」「行き方」の 3 単語が換言候補として収集される。



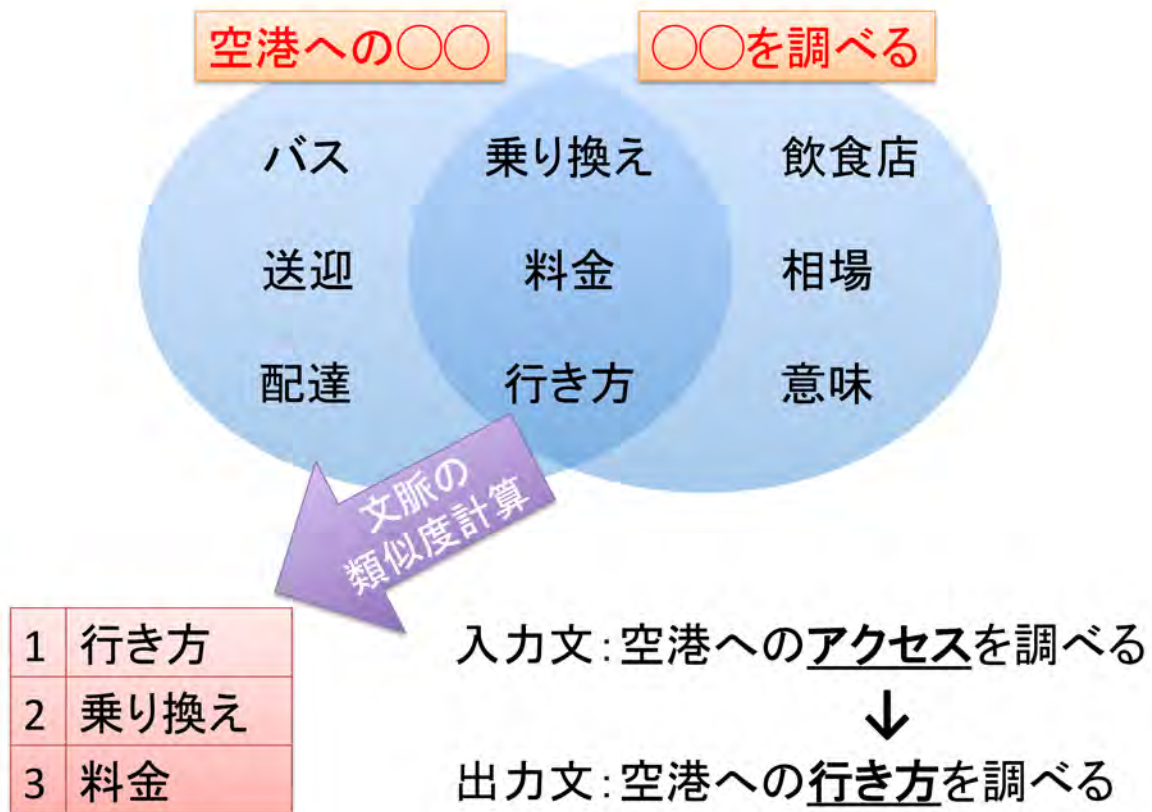


図 4.1 提案手法による名詞の換言

#### 4.2.2 文脈の種類数に基づく換言候補の選択

本研究では、換言対象の語と換言候補の語の間の換言可能性について、次の2つの仮説を立てる。この仮説に基づき、式 4.1 の類似度が最も大きい換言候補の名詞を選択する。

- (1) 換言対象の語と換言候補の語が多く種類の文脈を共有するほど換言可能性は高くなる
- (2) 換言候補の語が多く種類の文脈を持つほど換言可能性は低くなる

$$sim(n_t, n_c) = com(n_t, n_c) * \log\left(\frac{N}{var(n_c)}\right) \quad (4.1)$$

ただし、 $n_t$  は換言対象の名詞、 $n_c$  は換言候補の名詞を表し、 $com(n_t, n_c)$  は  $n_t$  と  $n_c$  が共通して用いられる文脈の種類数、 $N$  は文脈の総数、 $var(n_c)$  は換言候補の名詞  $n_c$  が用

いられる文脈の種類数を表す。前項は共通の文脈の種類が多いほど大きくなり、後項は換言候補の文脈が少ないほど大きくなるため、このスコアが高いほど換言対象の名詞  $n_t$  と換言候補の名詞  $n_c$  の文脈が類似していることを表し、換言可能性が高いことを意味する。

### 4.3 単語出現頻度に基づく手法との比較

#### 4.3.1 実験設定：単語出現頻度に基づく手法との比較

本研究では、Web 日本語 N グラム<sup>86)1</sup> を用いて実験を行った。Web 日本語 N グラムは Web 上の約 200 億文から作成された単語 N グラムで、本実験では最も長い 7 グラムデータを文と見なし、全 570,204,252 文を用いた。これらのうち、先頭が名詞で且つ末尾が動詞の原形である 1,365,705 文を選択し、さらにそのうち頻出する 200 文を抽出して実験対象文とした。この実験対象文のうち、文頭ではない名詞を換言対象の名詞とした。なお、品詞の判別には形態素解析器 MeCab(0.993)<sup>22)2</sup> および形態素解析辞書 IPADIC(2.7.0)<sup>3</sup> を用いた。

抽出した換言対象の名詞と同じ文脈で用いられる名詞群について、用いられる文脈の類似度を京都大学格フレーム<sup>85)4</sup> を用いて計算した。京都大学格フレームは Web 上の約 16 億文から自動構築<sup>84)</sup>された述語とそれが格関係をもつ名詞で、本実験では 34,059 語の述語と 824,639 語の名詞全てを用いた。そして、これらの述語を文脈と仮定し、入力文に含まれる換言対象の名詞を  $n_t$ 、前節で抽出した名詞群に含まれる換言候補の各名詞を  $n_c$  として式 4.1 を用いて文脈の類似度を計算した。

提案手法を評価するために、4.1 節に挙げた文脈の類似性に基づく換言手法との比較を行った。上述した 200 種類の入力文と換言対象の名詞に対して、提案手法および Marton らの手法 (Marton 法)、Bhagat らの手法 (Bhagat 法) を用いて類似度の上位 10 位までに含まれる名詞を集めた。評価は、3 人の評価者が換言対象の名詞と入力文中で換言可能な名詞を 1 語ずつ選んだ。

Marton 法では、名詞と文脈との共起頻度で名詞の特徴ベクトルを作成し、類似度は特徴ベクトル同士のコサイン類似度で求める。また、Bhagat 法では、名詞と文脈との自己相互情報量で名詞の特徴ベクトルを作成し、類似度は特徴ベクトル同士のコサイン類似度で求める。両手法とも、Web 日本語 N グラムを用いて、名詞と係り受け関係にある名詞

<sup>1</sup><http://www.gsk.or.jp/catalog/gsk2007-c/>

<sup>2</sup><https://code.google.com/p/mecab/>

<sup>3</sup><http://sourceforge.jp/projects/ipadic/releases/24435/>

<sup>4</sup><http://www.gsk.or.jp/catalog/gsk2008-b/>

および動詞を文脈と定義し、特徴ベクトルを作成した。式 4.2 に共起頻度、式 4.3 に自己相互情報量、式 4.4 にコサイン類似度をそれぞれ定義する。

$$cooccurrence(w_i, w_j) = \sum_{s_n \in S} freq_n(w_i, w_j) \quad (4.2)$$

$$pmi(w_i, w_j) = \log\left(\frac{cooccurrence(w_i, w_j) \sum_{s_n \in S} \sum_{w_m \in s_n} freq_n(w_m)}{\sum_{s_n \in S} freq_n(w_i) \sum_{s_n \in S} freq_n(w_j)}\right) \quad (4.3)$$

$$cos(u, v) = \frac{vector(u) \cdot vector(v)}{|vector(u)| |vector(v)|} \quad (4.4)$$

ただし、 $s_n \in S$ 、 $w_m \in s_n$ 、 $w_m \in W$  であり、 $S$  は文の集合 (コーパス)、 $W$  は単語の集合、 $freq_n(w_m)$  は文  $n$  における単語  $w_m$  の出現頻度、 $freq_n(w_i, w_j)$  は文  $n$  における単語  $w_i$  と単語  $w_j$  の共起頻度、 $vector(u)$  および  $vector(v)$  は特徴ベクトルを表す。

### 4.3.2 実験結果および考察

4.3.1 節で述べた 200 文に対する換言の評価結果を図 4.2 および図 4.3 に示す。なお、3 人の評価者 A、B、C の kappa 係数は、AB:0.64、BC:0.61、CA:0.59 であり、評価者間の一致度は十分高いと言える。

図 4.2 は、換言可能と評価された類似度 1 位の名詞数であり、提案手法が 2 つの比較手法よりも多くの名詞を換言できることを示している。

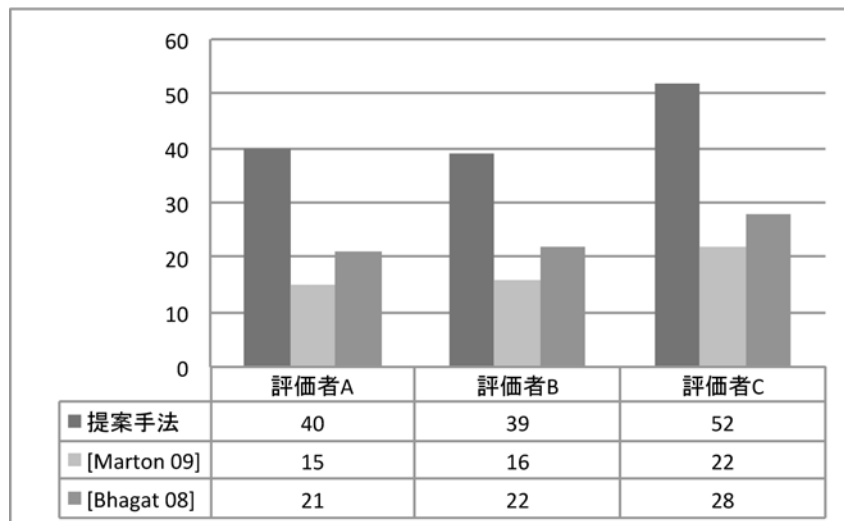


図 4.2 換言可能と評価された類似度 1 位の名詞数

Marton 法では多く共起する文脈を重要な文脈と考え、Bhagat 法では偏って共起する文脈を重要な文脈と考えている。そのため、Marton 法では単体での出現頻度が高い単語が類似度計算に強く反映され、Bhagat 法では単体での出現頻度が低い単語が類似度計算に強く反映されている。例えば、Marton 法では「こと」が 200 組中 100 組で換言候補として現れており、Bhagat 法では「等」「匹」などの接尾辞となる名詞が換言候補として多く現れている。

提案手法は、文脈の出現頻度に依存しないためこれらの影響は少なく、換言対象の名詞と換言可能な名詞のスコアが高くなっている。

図 4.3 は、換言可能と評価された類似度上位 10 位までの名詞数である。上位 10 位まで見ると、Bhagat 法が提案手法の換言可能名詞数に近づいている。

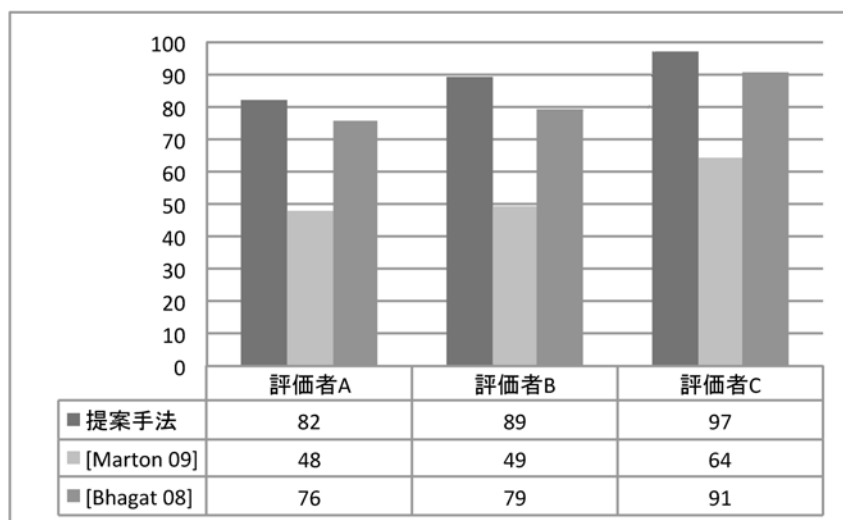


図 4.3 換言可能と評価された類似度 10 位までの名詞数

ここで、図 4.4 から図 4.6 に類似度の順位と換言可能名詞数の関係を示す。提案手法では 1 位の名詞の換言可能数と 2 位の名詞の換言可能数に大きな差があるのに対して、Bhagat 法では 1 位から 3 位までの換言可能数の変化が少ない。これは、提案手法では入力文と同じ文脈で用いられる名詞に換言を行うという制限をかけているためだと考える。Bhagat 法では入力文の文脈を考慮しないため、入力文において換言可能な語のスコアが最大になる保証はない。例えば、「万円以下の【罰金】に処する」という入力文において【罰金】を換言する場合、Marton 法や Bhagat 法では「懲役」のスコアが最も高く、次に「科料」や「過料」といった金銭に関する単語が続いている。一方、提案手法では「万円以下の」という入力文の文脈を考慮し、「懲役」という単語は換言候補中に現れておらず、「罰金刑」のスコアが最高で、次に「過料」が続いている。

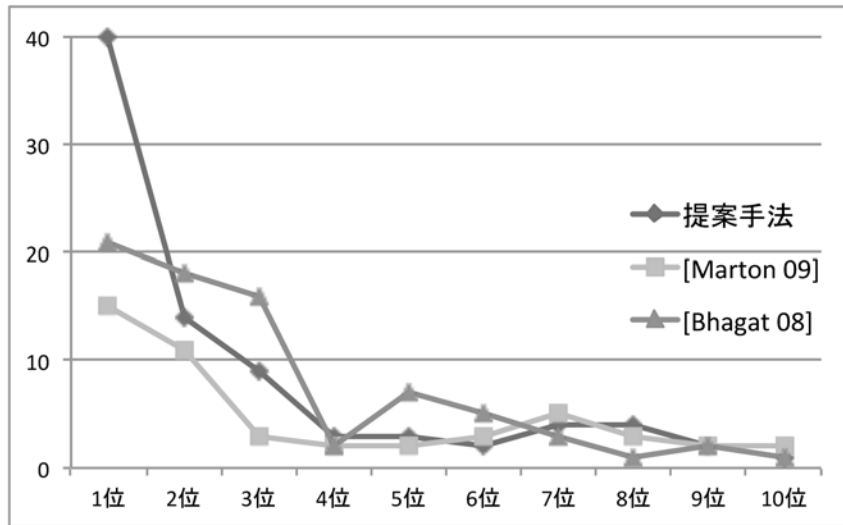


図 4.4 類似度の順位と換言可能名詞数の関係 (評価者 A)

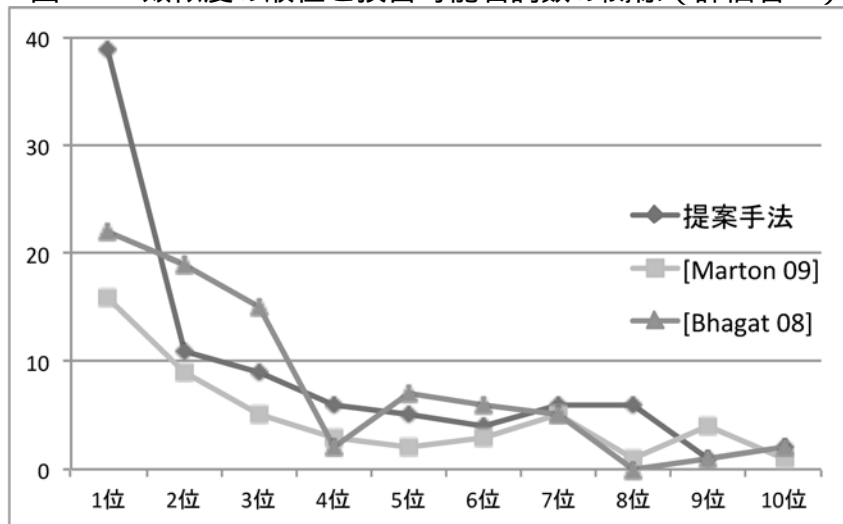


図 4.5 類似度の順位と換言可能名詞数の関係 (評価者 B)

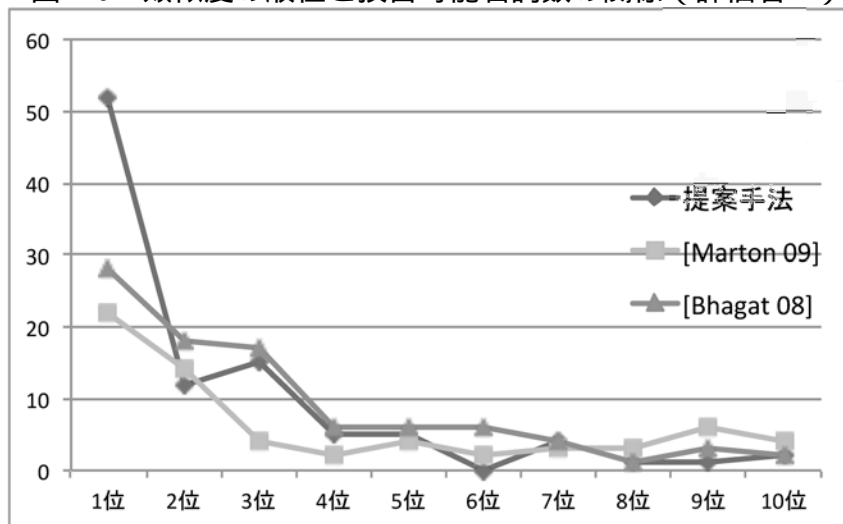


図 4.6 類似度の順位と換言可能名詞数の関係 (評価者 C)

その他、「腰への【負担】を軽減する」という入力文において【負担】を換言する場合、比較手法では「費用」「経費」「実費」など【負担】の換言先として金銭に関する単語ばかりが換言候補中に出現し、換言可能な語は上位 10 位に存在しない。一方、提案手法では「腰への」という入力文の文脈を考慮し、「負荷」に続いて「ストレス」「ダメージ」「疲労」「緊張」「衝撃」「当たり」「圧迫感」「荷重」「圧迫」と適切な換言候補を挙げることができる。以下に、提案手法で換言できた例を挙げる。

- オーナーの【承認 許可】が必要になる
- 重要な【課題 問題】として取り組んでいる
- 良心的な【料金 価格】を提供する
- 国内農業の【発展 成長】を阻害する
- 教育の【拡充 強化】などがあげられる

## 4.4 語彙平易化への有効性の検証

式 4.1 に基づき、語彙的換言の観点では性能の向上を確認した。本節では、語彙平易化の観点での有効性を検証する。式 4.1 の後項の影響で、本手法では文脈の種類数の多い単語が出力されにくいと考えられる。そこで、文脈の種類数と語の難易度の分布を調査する。また、基本的な統計量である単語の出現頻度と語の難易度や文脈の種類数の関係も同時に調査する。

### 4.4.1 実験設定：語彙平易化への有効性の検証

#### 4.4.1.1 コーパス

頻度や文脈の種類数を求めるために、表 4.1 に示す新聞記事、Wikipedia、論文の 3 種類のコーパスを用いた。新聞記事は、1990 年から 2004 年までの 15 年分の日経新聞<sup>5</sup>の本文を用いた。Wikipedia は、2014 年 3 月 13 日時点での全日本語記事<sup>6</sup>の本文を用いた。論文は、ANLP-20 コーパス<sup>7</sup>として公開されている言語処理学会年次大会予稿集のうち、論文 PDF からの抽出テキストの本文を用いた。

<sup>5</sup><http://www.nikkeibookvideo.com/kijidb/>

<sup>6</sup><http://dumps.wikimedia.org/jawiki/20140313/>

<sup>7</sup><http://nlp20.nii.ac.jp/resources/>

表 4.1 コーパス

	記事数	単語数	異なり単語数
新聞記事	2,479,065	703,761,303	151,514
Wikipedia	853,134	357,761,882	158,971
論文	4,082	14,787,129	37,282

#### 4.4.1.2 文脈の定義

本研究では、語の難易度と文脈の種類数の関係について調査を行うが、ここで文脈を定義する。本研究で扱う文脈は、対象の語が含まれる文節と係り受け関係にある文節中の内容語とする。ここで、内容語とは名詞、動詞、形容詞および副詞を指す。なお、形態素解析、文節区切りおよび係り受け解析には CaboCha 0.68<sup>87)8</sup> を用いた。

#### 4.4.2 相関係数

語の難易度と、頻度および文脈の種類数の相関を表 4.2 に計算した。相関は、式 4.5 に示すピアソンの積率相関係数および式 4.6) に示すスピアマンの順位相関係数を用いた。

$$pearson = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.5)$$

$$spearman = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n^3 - n} \quad (4.6)$$

まず、単語親密度について考える。表 4.2 より、単語親密度と頻度の間には、中程度の相関がある（新聞記事コーパスおよび Wikipedia コーパス、スピアマンの順位相関係数）。また、単語親密度と文脈の種類数の間にも同様の傾向があり、新聞記事コーパスおよび Wikipedia コーパスにおけるスピアマンの順位相関係数を見ると、中程度の相関があると言える。一方、論文コーパスにおけるスピアマンの順位相関係数は、単語親密度と頻度の間でも単語親密度と文脈の種類数の間でも、弱い相関しか見られない。論文コーパスは小規模なコーパスであり、新聞記事コーパスや Wikipedia コーパスと比べて統計量の信頼性に欠けるため、異なる傾向を示している可能性がある。

<sup>8</sup><https://code.google.com/p/cabocha/>

表 4.2 相関係数

		ピアソンの積率相関係数			スピアマンの順位相関係数		
		新聞記事	Wikipedia	論文	新聞記事	Wikipedia	論文
単語親密度	JLPT 出題基準	0.35	0.35	0.38	0.25	0.25	0.27
単語親密度	頻度	0.10	0.09	0.09	0.65	0.57	0.30
単語親密度	前文脈種類数	0.30	0.27	0.14	0.61	0.53	0.23
単語親密度	後文脈種類数	0.32	0.30	0.14	0.64	0.55	0.28
JLPT 出題基準	頻度	0.16	0.13	0.08	-0.02	-0.02	0.03
JLPT 出題基準	前文脈種類数	0.23	0.21	0.14	-0.04	-0.06	0.01
JLPT 出題基準	後文脈種類数	0.25	0.25	0.13	0.00	-0.01	0.03
頻度	前文脈種類数	0.67	0.64	0.80	0.96	0.96	0.84
頻度	後文脈種類数	0.66	0.65	0.76	0.98	0.99	0.91
前文脈種類数	後文脈種類数	0.94	0.90	0.95	0.97	0.96	0.87

続いて、JLPT 出題基準について考える。表 4.2 より、JLPT 出題基準と頻度の間には、コーパスの種類や相関係数の計算方法に関わらず、相関が見られない。同様の傾向ではあるが、JLPT 出題基準と文脈の種類数の間では、新聞記事コーパスおよび Wikipedia コーパスにおけるピアソンの積率相関係数では弱い相関が見られる。

次に、頻度と文脈の種類数の関係に注目する。表 4.2 より、スピアマンの順位相関係数を見ると、コーパスの種類に関わらず、頻度と文脈の種類数の間には強い相関があることがわかる。また、前文脈の種類数と後文脈の種類数の相関も非常に強い。



### 4.4.3 分布

単語親密度と頻度の散布図、JLPT 出題基準と頻度の散布図、単語親密度と文脈種類数の散布図、JLPT 出題基準と文脈種類数の散布図、頻度と文脈種類数の散布図を、図 4.7 から図 4.27 に示す。図 4.7、図 4.9 および図 4.11 の単語親密度と頻度に関しては、寺田ら<sup>96)</sup>の報告にもあるように、基本的に点が左下から右上に位置している。また、寺田らの報告では小さなコーパスでは右下に膨らむ三角形の形状になるとの報告があるが、本研究でも小さな論文コーパスにおいて右下に膨らむ三角形の形状が確認できた。また、新聞記事コーパスや Wikipedia コーパスでは、散布図は三角形とはいえないが、図 4.8 や図 4.10 のヒートマップを確認することで、右下に膨らむ三角形に近い形状が確認できた。これらのヒートマップは、青に近い色ほど疎、赤に近い色ほど密な分布であることを示している。なお、いずれの散布図にも、左上の領域には点がほとんど見られない。これは、高頻度かつ低親密な語がほとんど無いことを表している。一方、右下の領域にも点が見られることは、低頻度かつ高親密な語が存在することを表している。新聞記事コーパスにおける低頻度かつ高親密な語の例を表 4.3 に示す。

表 4.3 低頻度かつ高親密な語

単語	単語親密度	頻度	後文脈種類数
悪運	5.88	11	9
急募	5.78	19	36
×切	6.25	15	26
今頃	5.91	25	30
臆病	5.88	38	55

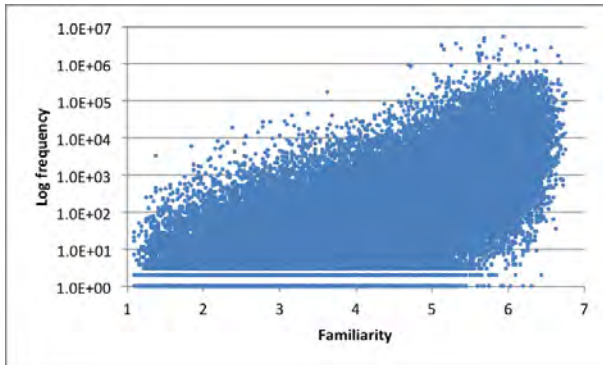


図 4.7 新聞記事コーパス: 親密度と対数頻度

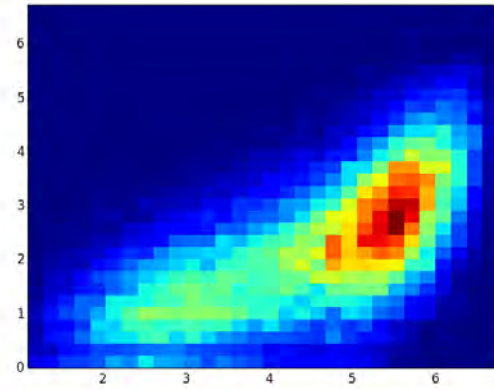


図 4.8 図 4.7 のヒートマップ

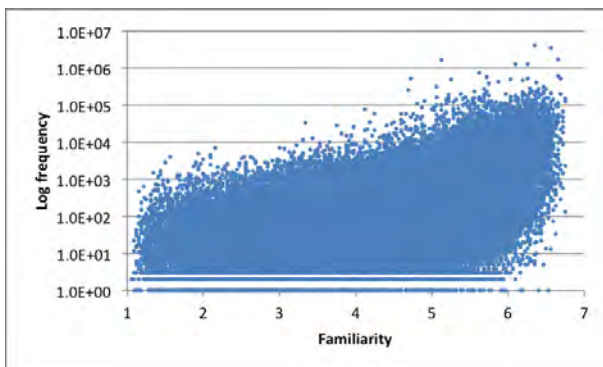


図 4.9 Wikipedia コーパス: 親密度と対数頻度

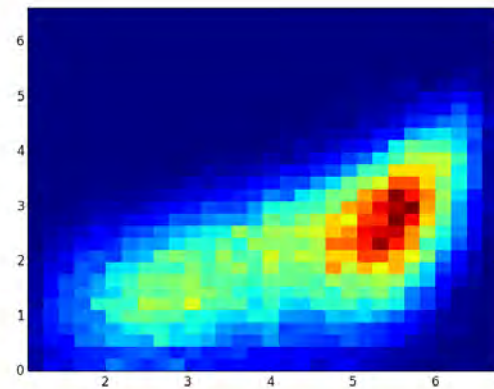


図 4.10 図 4.9 のヒートマップ

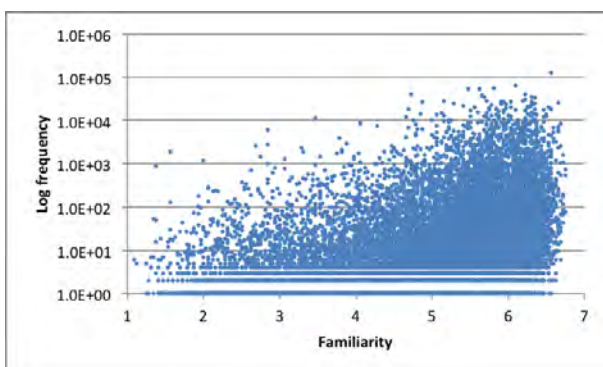


図 4.11 論文コーパス: 親密度と対数頻度

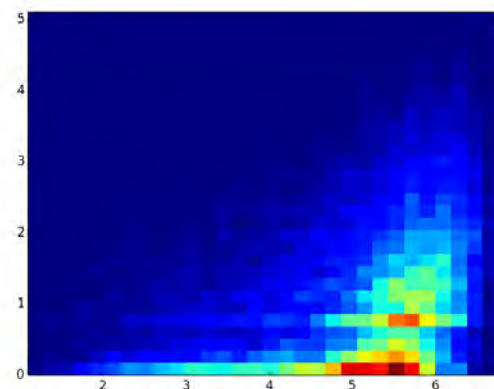


図 4.12 図 4.11 のヒートマップ

続いて、図 4.13 から図 4.18 の単語親密度と文脈種類数に注目する。文脈の種類数についても、頻度と同様の傾向が見られた。つまり、多くの文脈を持ちながら低親密な語はほとんど見られない。一方で、文脈の種類数が少ない、あるいは中程度であっても、高親密な語が多く存在する。文脈種類数に関するこれらのヒートマップでは、特に文脈の種類数が中程度かつ高親密な語が多い様子が確認できる。本研究では後文脈の種類数に関する散布図について調査したが、前文脈の種類数に関する散布図にも同様の傾向が見られた。

次に、図 4.19、図 4.21 および図 4.23 の JLPT 出題基準と頻度の散布図および図 4.20、図 4.22 および図 4.24 の JLPT 出題基準と文脈種類数の散布図について考える。最高頻度の単語は 4 級であり、4 級の最低頻度語は他の級の最低頻度語よりも高頻度である。同様に、最低頻度の単語は 1 級であり、1 級の最高頻度語は他の級の最高頻度よりも低頻度である。しかし、どの級にも低頻度から高頻度まで万遍なく単語が分布している。文脈の種類数に関しても同様の傾向があり、1 級と 4 級を比較するとわずかに差があるが、どの級にも万遍なく単語が分布している。

最後に、図 4.25、図 4.26 および図 4.27 の頻度と文脈種類数を考える。いずれのコーパスでも、単語は左下から右上に向かって分布し、右下の領域にも左下の領域にも点がほとんど見られない。頻度と文脈種類数の間には強い相関があったが、単語の分布からも、高頻度語は多くの文脈を持つということがわかる。

## 4.5 まとめ：用例に基づく語彙的換言

本節では、文脈の種類数に基づく語彙的換言手法を提案した。本手法は、単語の出現頻度に基づく従来の分布類似度の手法よりも高精度に語彙的換言を獲得することができる。また、本手法を用いて語彙の平易化を行うために、単語の頻度や文脈の種類数と難易度の関係について調査した。新聞記事、Wikipedia、論文の 3 種類のコーパスを用いて頻度や文脈の種類数を計算し、単語親密度や JLPT 出題基準に基づく難易度との相関や分布を見た。調査の結果、大規模コーパスには文脈の種類数が中程度かつ高親密な語が多く存在し、比較的小規模なコーパスには文脈の種類数が少なく高親密な語が多いことが分かった。本手法では、文脈の種類数が多い単語は出力されにくいだが、文脈の種類数が中程度以下の領域に多くの平易語が存在するため、語彙平易化への有効性も確認できた。

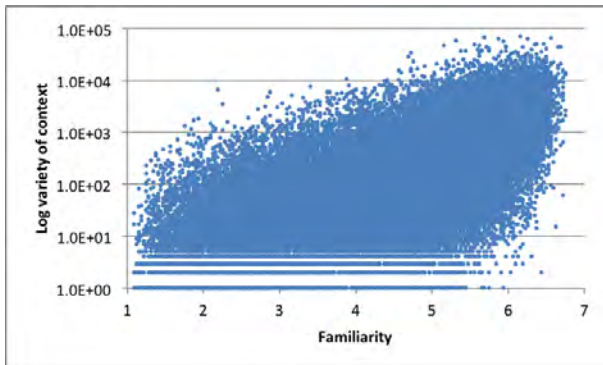


図 4.13 新聞記事コーパス：親密度と対数後文脈種類数

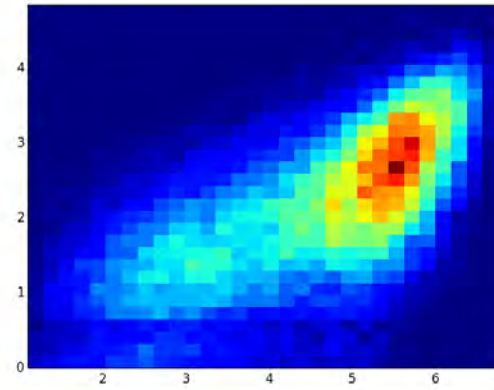


図 4.14 図 4.13 のヒートマップ

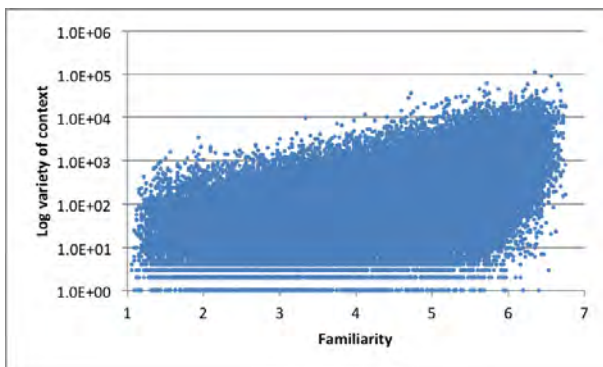


図 4.15 Wikipedia コーパス：親密度と対数後文脈種類数

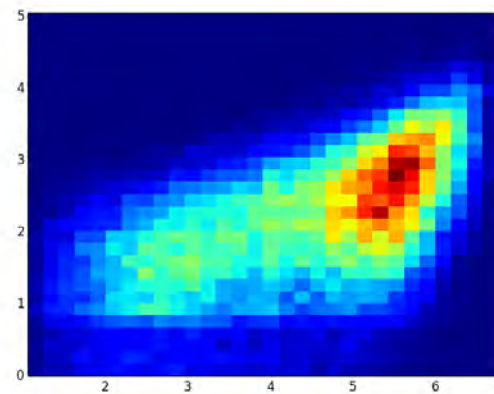


図 4.16 図 4.15 のヒートマップ

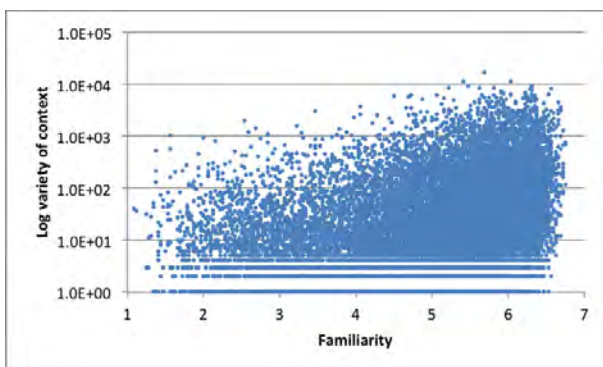


図 4.17 論文コーパス：親密度と対数後文脈種類数

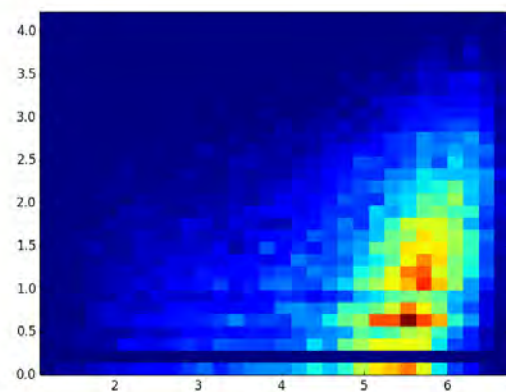


図 4.18 図 4.17 のヒートマップ

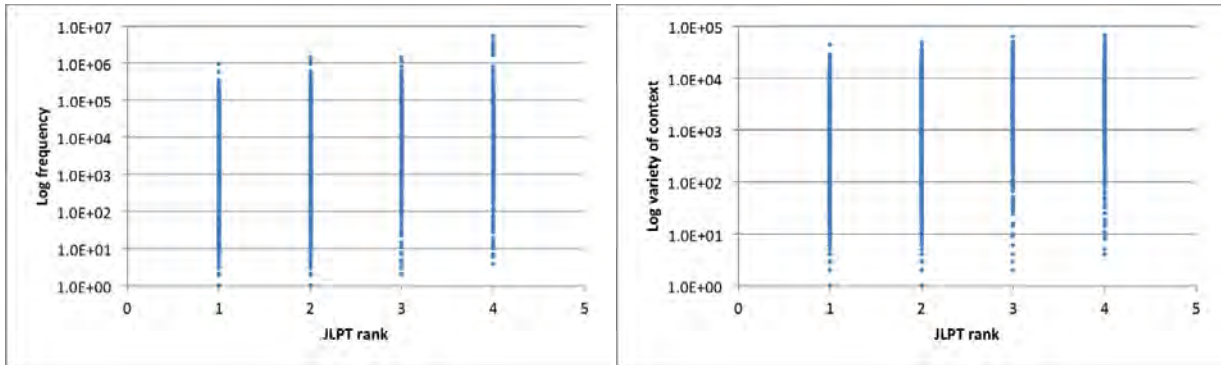


図 4.19 新聞記事コーパス：JLPT 出題基準と対数頻度

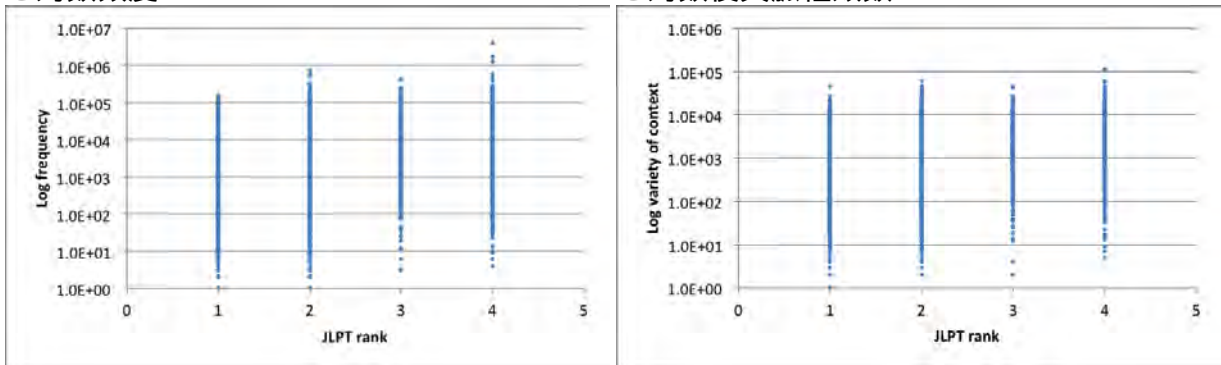


図 4.21 新聞記事コーパス：JLPT 出題基準と対数後文脈種類数

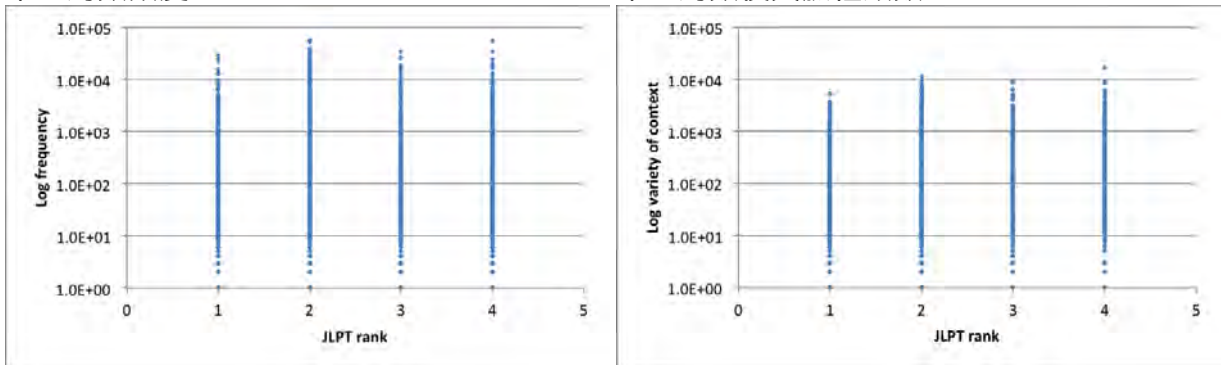


図 4.22 Wikipedia コーパス：JLPT 出題基準と対数頻度



図 4.23 Wikipedia コーパス：JLPT 出題基準と対数後文脈種類数



図 4.24 論文コーパス：JLPT 出題基準と対数頻度



図 4.25 論文コーパス：JLPT 出題基準と対数後文脈種類数

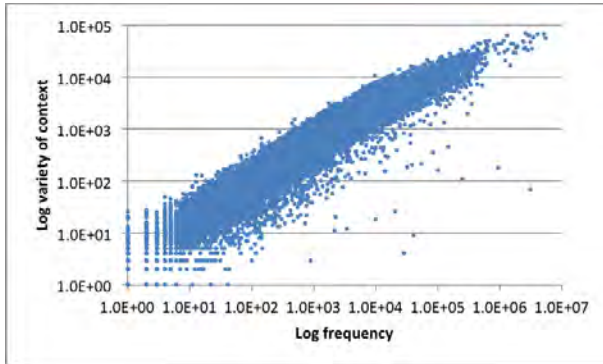


図 4.25 新聞記事コーパス：対数頻度と対数後文脈種類数

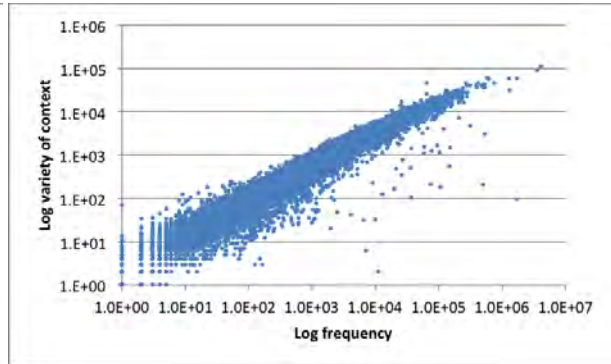


図 4.26 Wikipedia コーパス：対数頻度と対数後文脈種類数

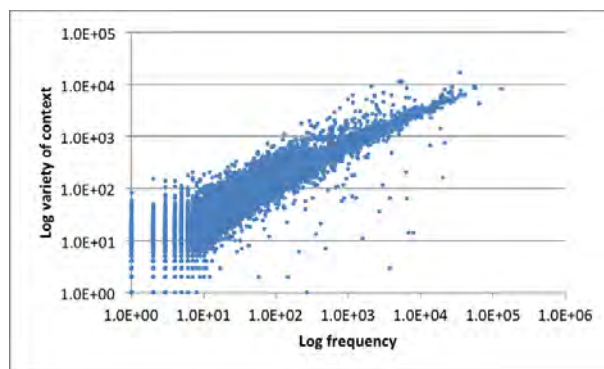


図 4.27 論文コーパス：対数頻度と対数後文脈種類数

## 5 日本語の語彙平易化評価セット (SNOW E4)

本章では、語彙平易化システムの評価のためのデータセットについて説明する。本研究の貢献は、日本語の語彙平易化システムの評価セットを Web で初めて公開したことである。本言語資源の公開によって、語彙的換言および語彙平易化システムの適合率や再現率を自動で計算し、F 値を求めることができるようになる。英語での先行研究を参考に、まず語彙的換言の評価のためのデータセットを構築し、その対象語と複数の換言を「易しさ」の基準で並び替えることにより、語彙平易化の評価のためのデータセットを構築した。本研究ではクラウドソーシングを用いて延べ 500 人の作業者によって、先行研究<sup>29)34)</sup>と同等の規模の日本語のデータセットを構築した。本章で述べる日本語の語彙平易化システムの評価のためのデータセット (SNOW E4) は、次の URL から利用できる。

<http://www.jnlp.org/SNOW/E4>

### 5.1 関連研究：語彙平易化評価セット

語彙平易化システムの評価のために、2 種類の英語のデータセット<sup>34)14)2)</sup>が公開されている。これらはどちらも、SemEval-2007 の English Lexical Substitution Task<sup>29)</sup>のために構築された語彙的換言の評価のためのデータセット<sup>3)</sup>を改良して構築されている。本節では、これらの各データセットについて概説する。

#### 5.1.1 McCarthy の語彙的換言データセット

SemEval-2007 の English Lexical Substitution Task<sup>29)</sup>は、文脈中で対象語と置換可能な語や句を見つけるタスクである。対象語は内容語であり、その内訳は表 5.1 のとおりである。文脈は英語のウェブ均衡コーパスである English Internet Corpus<sup>32)</sup>から選択され

<sup>1</sup><http://www.cs.york.ac.uk/semeval-2012/task1/>

<sup>2</sup><http://people.cs.kuleuven.be/~jan.debelder/lseval.zip>

<sup>3</sup><http://www.dianamccarthy.co.uk/task10index.html>

表 5.1 各データセットの品詞の内訳

データセット	総文数	名詞 (%)	動詞 (%)	形容詞 (%)	副詞 (%)
McCarthy/Specia (Trial)	300	80 (26.7)	80 (26.7)	90 (30.0)	50 (16.7)
McCarthy/Specia (Test)	1,710	500 (29.2)	440 (25.8)	470 (27.5)	300 (17.5)
De Belder	430	100 (23.3)	60 (14.0)	160 (37.2)	110 (25.6)
SNOW E4	2,330	630 (27.0)	720 (30.9)	500 (21.5)	480 (20.6)

ている。このタスクの語彙的換言の評価のためのデータセットは、201 種類の対象語に対して各 10 種類の文脈が付与された 2,010 文で構成されている。これらの各文について、5 人の英語母語話者が文脈中で適切な換言を 3 語まで付与している。作業者は、適切な換言を単語で思いつかない場合、句で回答してもよい。

以下に McCarthy の語彙的換言データセットの例を示す。この文脈中の形容詞 bright の換言として、ある 3 人の作業者は intelligent と回答し、またある 3 人の作業者は clever と回答し、ある 1 人の作業者は smart と回答している。

- <context>During the siegem George Robertson had appointed Shuja-ul-Mulk, who was a <head>bright</head> boy only 12 years old and the youngest surviving son of Aman-ul-Mulk, as the ruler of Chitral.</context>
- Gold: intelligent 3; clever 3; smart 1;

### 5.1.2 Specia の語彙平易化データセット

SemEval-2012 の English Lexical Simplification Task<sup>34)</sup> は、文脈中で対象語の複数の換言を「易しさ」の基準で並び替えるタスクである。「易しい」とは、子どもや英語非母語話者を含む幅広い人々にとって理解しやすいことを指す。このデータセットでは特に、英語が流暢な非母語話者（専攻の異なる大学 1 年生）によって「易しさ」による並び替え作業が行われている。Trial データについては 4 人、Test データについては 5 人の作業者が、5.1.1 節で紹介した McCarthy の語彙的換言データセットの対象語とその換言を「易しさ」の基準で並び替えている。

最終的に、各作業者による並び替えの結果を統合して一つのデータセットを作成する。Specia の語彙平易化データセットでは、各作業者の難易度の順位の平均値を用いて、各単語の難易度の順位を決定している。また、各作業者の順位の平均値が等しい単語同士は、同じ順位としている。



以下に Specia の語彙平易化データセットの例を示す。ある文脈中で、4人の作業員から次のような難易度の順位が得られたとき、clear の順位はそれぞれ、1、2、1、4であり、平均値は2である。同様に、light は3.25、bright は2.5、luminous は4、well-lit は3.25が平均値となる。最終的な統合順位は、これらの単語を平均値の小さい順に並び替えることで得られる。

1: {clear} {light} {bright} {luminous} {well-lit}

2: {well-lit} {clear} {light} {bright} {luminous}

3: {clear} {bright} {light} {luminous} {well-lit}

4: {bright} {well-lit} {luminous} {clear} {light}

Gold: {clear} {bright} {light, well-lit} {luminous}

### 5.1.3 De Belder の語彙平易化データセット

De Belder ら<sup>4)</sup>は、5.1.1 節で紹介した McCarthy の語彙的換言データセットを語彙平易化のためのデータセットに変換するために、まず十分に平易な対象語を除外している。Simple English Wikipedia の Basic English combined word list<sup>4</sup>などに含まれる十分平易な語を除外した結果、201種類の対象語のうち43種類が残っている。これらの対象語とその換言を、それぞれ5人の作業員が「易しさ」の基準で並び替えている。この並び替えでは、難易度が等しいと思った単語同士に同じ順位を付与することが許されている。作業員は、Amazon Mechanical Turk<sup>5</sup>を用いて、アメリカに住んでおり過去の作業承認率が95%以上であるという条件で集められている。

De Belder の語彙平易化データセットでは、雑音のある通信路モデルを用いて、各作業員の並び替え結果と各作業員の信頼度を考慮して、各作業員による並び替えの結果を統合している。

## 5.2 日本語版データセットの構築

本研究では、日本語の語彙平易化システムの評価のためのデータセットを構築する。上述の先行研究を参考に、まず語彙的換言の評価のためのデータセットを構築し、その対象

<sup>4</sup>[http://simple.wikipedia.org/wiki/Wikipedia:Basic\\_English\\_combined\\_wordlist](http://simple.wikipedia.org/wiki/Wikipedia:Basic_English_combined_wordlist)

<sup>5</sup><https://www.mturk.com>

語と複数の換言を「易しさ」の基準で並び替えることにより、語彙平易化の評価のためのデータセットを構築する。

本研究ではクラウドソーシングを用いて延べ 500 人の作業者によって、先行研究<sup>34)29)</sup>と同等の規模の日本語のデータセットを構築した。

## 5.2.1 語彙的換言データセットの構築

### 5.2.1.1 換言対象語の選定

本研究では換言の対象語を、IPADIC(2.7.0)<sup>6)</sup>とJUMAN(7.0)<sup>7)</sup>の形態素辞書の共通部分から選択する。ここで対象とするのは内容語(名詞、動詞、形容詞、副詞)であり、サ変名詞は名詞、サ変動詞は動詞、形容動詞は形容詞にそれぞれ含まれる。なお、複合名詞や複合動詞は、その一部の語を置き換えると意味を保持できない場合が多いので、複合語の一部である語は対象語から除外する。

目的が語彙平易化のためのデータセットを構築することなので、すでに十分に平易な語は扱わない。ここでは、小学生のための理解語彙である学習基本語彙<sup>99)</sup>に含まれる語を対象語から除外する。

全ての語に必ずしも語彙的換言が存在するわけではないので、換言が存在しない可能性の高い語は扱わない。ここでは、JUMAN 辞書の内容語に人手で語彙的換言を付与している内容語換言辞書(SNOW D2)<sup>113)102)8)</sup>に含まれない語を対象語から除外する。

先行研究<sup>29)</sup>に倣い、各対象語を 10 種類の文脈中で扱う。ここでは文脈として、1990 年から 2004 年までの 15 年分の日本経済新聞<sup>9)</sup>の本文から、対象語が含まれる文を無作為に抽出する。なお、文脈が 10 種類に満たない低頻度語は、対象語から除外する。

このようにして、名詞および動詞各 75 語、形容詞および副詞各 50 語を対象語として無作為に選定した。

### 5.2.1.2 語彙的換言の列挙

5.2.1.1 節で選定した各対象語について、文脈中で置換可能な語を 5 人ずつの作業者によって列挙する。ここではクラウドソーシングを用いて作業者を募った。それぞれの作業者は、50 語(対象語 5 × 文脈 10)ずつの作業を行う。作業者は、文脈中で対象語と置換可能な単語または句を、思いつく限り列挙する。その際、辞書を参照することは許可しているが、他人に意見を求めることは禁止している。作業者が適切な換言が存在しないと判

<sup>6)</sup><http://sourceforge.jp/projects/ipadic/releases/24435/>

<sup>7)</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

<sup>8)</sup><http://www.jnlp.org/SNOW/D2>

<sup>9)</sup><http://www.nikkeibookvideo.com/kijidb/>

断した場合には、無記入を許している。

本作業ではクラウドソーシングサービスとしてランサーズ<sup>10</sup>を利用し、37人の作業者を得た。これらの作業者から、それぞれの文脈中の対象語について、平均 5.38 語の語彙的換言が得られた。なお、作業者間の一致率を先行研究<sup>29)</sup>と同様に計算したところ、17.8%の一致率であった。

### 5.2.1.3 換言データセットの統合

5.2.1.2 節で 5 人ずつの作業者から得た換言を統合して、ひとつの日本語の語彙的換言の評価のためのデータセットを構築する。ある単語または句が文脈中で対象語の換言であるかどうかは作業者間で基準が一定ではないため、新たに 5 人ずつの作業者を募り、適切な換言であるかどうかの確認を行う。なお、5.2.1.2 節と同様に、作業者はクラウドソーシングで募り、各作業者は 50 語の作業を行う。本研究では、5 人中 3 人以上の作業者が「適切な換言である」と認めた単語または句を採用する。なお、「適切な換言ではない」基準として、次の 2 つを作業者に提示している。

- 文脈中の対象語と置き換えたときに、不自然な文になる場合は、適切な換言ではない。
- 文脈中の対象語と置き換えたときに、文の意味が保持できない場合は、適切な換言ではない。

本作業ではクラウドソーシングサービスとしてランサーズを利用し、83 人の作業者を得た。これらの作業者によって、それぞれの文脈中の対象語について、平均 4.50 語の語彙的換言が適切な換言であると認められた。ただし、全ての換言が不適切であると評価された 17 種類の対象語を含む 170 文は除外した。なお、作業者間の一致率は、66.4%であった。

以下に、構築した日本語の語彙的換言データセットの例を示す。本データセットは、対象語と文脈から成る文脈データと、対象語とその換言から成る換言データの 2 種類で構成されている。下の例の先頭の数値は、2 つのデータを紐付ける ID である。下の例では、この文脈中の名詞「悪気」の換言として、ある 1 人の作業者は「意地悪」と回答し、またある 1 人の作業者は「悪い考え」と回答し、ある 4 人の作業者は「悪意」と回答している。それぞれの換言に続く数値は、本節で「適切な換言である」と認めた作業者の人数ではなく、5.2.1.2 節でその語または句を回答した作業者の人数であることに注意されたい。

<sup>10</sup><http://www.lancers.jp>

- 820, 悪気, 親は悪気で言ったわけではなく、子供をあやすということを本当に知らなかった様子。
- 820, 悪気, 意地悪 1; 悪い考え 1; 悪意 4;

## 5.2.2 語彙平易化データセットへの変換

### 5.2.2.1 複数の換言の並び替え

5.2.1 節で構築した語彙的換言のためのデータセットに含まれる対象語とその換言を、文脈中で「易しさ」の基準で並び替えることによって、語彙平易化のためのデータセットに変換する。ここでは、それぞれ5人の作業者によって並び替えの作業を行う。なお、5.2.1.2 節と同様に、作業者はクラウドソーシングで募り、各作業者は50語の作業を行う。また、先行研究<sup>34)</sup>と同様、各作業者による並び替えの際には、全ての語または句に重複なく難易度の順位を割り当てる。ただし、それらの結果を統合する際には、複数の語または句に同じ難易度の順位が割り当てられる場合がある。

この並び替え作業は、5.2.1.3 節で募った83人の作業者が行った。なお、作業者間の一致率を先行研究<sup>4)</sup>に倣ってスピアマンの順位相関係数で計算したところ、33.2%の一致率であった。

### 5.2.2.2 平易化データセットの統合

5.2.2.1 節で得られた5人ずつの作業者による並び替えの結果を統合して、ひとつの日本語の語彙平易化の評価のためのデータセットを構築する。本研究では先行研究<sup>34)</sup>に倣い、各作業者の難易度の順位の平均値を用いて、各語または句の難易度の順位を決定する。ここで、各作業者の順位の平均値が等しい語または句同士は、同じ順位を割り当てる。本研究では、同じ作業者が換言の評価と並び替えの両方の作業を行っているため、ある作業者がある単語または句を「適切な換言ではない」と評価した場合、その単語または句は難易度の順位を付与されない。これらの順位を付与されなかった単語または句は、最低順位（最も大きな値）として平均値を計算する。

統合して得られたデータセットには、それぞれの文脈中の対象語とその換言について、平均4.94段階の難易度の順位が与えられた。

以下に、構築した日本語の語彙平易化データセットの例を示す。ある文脈中で、5人の作業者から次のような難易度の順位が得られたとき、「意地悪」の順位はそれぞれ、1、2、4、2、2であり、平均値は2.2である。同様に、「悪意」は2.2、「悪気」は2.6、「悪い考え」は3が平均値となる。最終的な統合順位は、これらの単語を平均値の小さい順に並び替えて得られる。

- 1: { 意地悪 } { 悪意 } { 悪気 } { 悪い考え }
- 2: { 悪気 } { 意地悪 } { 悪意 } { 悪い考え }
- 3: { 悪意 } { 悪い考え } { 悪気 } × : 意地悪
- 4: { 悪い考え } { 意地悪 } { 悪気 } { 悪意 }
- 5: { 悪意 } { 意地悪 } { 悪気 } × : 悪い考え
- Gold: { 意地悪, 悪意 } { 悪気 } { 悪い考え }

### 5.2.3 データセットの分析

本節では、構築した日本語の語彙平易化データセットの文脈依存性を調査する。それぞれの対象語について、10種類ずつの文脈で換言を付与し、対象語とその換言を並び替えて語彙平易化のデータセットを構築した。ここには前提として、「文脈によって語彙的換言が異なる」「文脈によって平易な語または句が異なる」という2つの文脈依存性が仮定されている。この仮説を検証するため、表5.2に同じ換言リストから異なる難易度順位が得られた組み合わせの割合などを示す。この表から、語彙的換言では84.8%、語彙平易化では59.5%が文脈に依存して変化することがわかる。

表 5.2 語彙平易化データセットの文脈依存性

	英語 <sup>34)</sup>	日本語
A. 対象語が同じ文脈の組	9,045	10,485
B. 換言リストが等しい組	302 (A. の 3.3%)	1,593 (A. の 15.2%)
C. 難易度順位が異なる組	163 (B. の 54.0%)	948 (B. の 59.5%)
D. 最も平易な語が違う組	57 (C. の 35.0%)	463 (C. の 48.8%)

## 5.3 まとめ：語彙平易化評価セット

本研究では、日本語の語彙的換言および語彙平易化の評価のためのデータセットを構築し、公開した。これらのデータセットを用いることで、語彙的換言および語彙平易化システムの精度や再現率を計算し、F値を求めることができるようになる。このような自動評価の枠組みを設けることで、日本語の語彙的換言および語彙平易化の研究のサイクルが素早く回り、より深い議論を重ねていくことが可能となる。

最後に、本データセットを用いて日本語の語彙的換言および語彙平易化が活発に研究され、子どもや言語学習者をはじめとする幅広い読者の文章読解の助けとなるシステムが開発されることを期待する。

## 6 日本語の語彙平易化システム (SNOW S3)

本章では、語彙平易化システムの構築について説明する。本研究の貢献は、日本語の語彙平易化システムを Web で初めて公開したことである。本システムの公開によって子どもや言語学習者をはじめとする幅広い読者の文章読解を支援すると同時に、日本語の語彙平易化の研究が本システムをベースラインとして活発に研究されることを期待する。本章で述べる日本語の語彙平易化システムは、次の URL から利用できる。

<http://www.jnlp.org/SNOW/S3>

### 6.1 関連研究：語彙平易化システム

英語では、SemEval-2012 の評価型ワークショップにおいて English Lexical Simplification Task<sup>34</sup> が開催されており、語彙平易化システムの評価のための言語資源が整備され、様々な手法を用いた多くのシステムが参加している。また、Wikipedia<sup>1</sup> の平易版である Simple English Wikipedia<sup>2</sup> の存在により、難解な文と平易な文の平行コーパスを用いて統計的に平易化規則を学習するような手法<sup>15</sup> も近年提案されている。このような活発な研究の中で、いくつかの英語の語彙平易化システム<sup>34</sup> が Web で公開されている。

一方で日本語では、このような一般的に利用可能なシステムは存在しない。そのため、読解支援を必要とする読者のためにも、研究を加速させるためにも、日本語の語彙平易化システムの公開が必要である。

本研究で構築する日本語の語彙平易化システムは、典型的な語彙平易化手法の 4 つの機構<sup>31</sup> を備えている。すなわち、まず入力文から難解語を検出する。続いて、その難解語の語彙的換言を列挙する。そして、難解語の換言の中から入力文の文脈で使用可能な語を

---

<sup>1</sup><http://en.wikipedia.org>

<sup>2</sup><http://simple.wikipedia.org>

<sup>3</sup><http://homepages.inf.ed.ac.uk/kwoodsens/demos/simplify.html>

<sup>4</sup><https://rewordify.com>

選択する。最後に、それらの語を難易度で並び替え、最も平易な語を難解語と置き換えることで出力文を生成する。この語彙平易化システムの概要を、例とともに図 6.1 に示す。

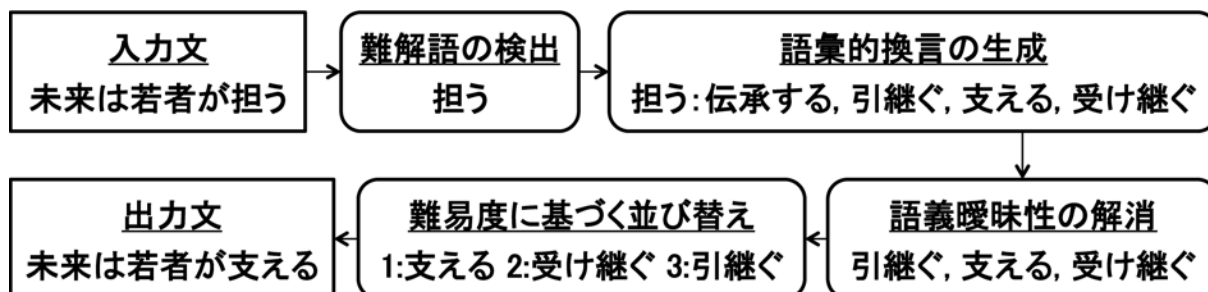


図 6.1 語彙平易化システムの概要

## 6.2 難解語の検出

本システムでは、名詞、動詞、形容詞、副詞などの内容語と呼ばれる単語を平易化の対象とし、機能語、複合語、慣用句などは扱わない。まず、形態素解析によって入力文から内容語を抽出する。ただし、複合語や慣用句などは、そこに含まれる単語の単位で変換を行うと意味を保持できなくなる場合が多いので、複合表現リストを用意し、これらを平易化の対象から除外する。また、すでに十分平易な単語も平易化する必要がないので、十分平易な単語リストも用意し、これらも平易化の対象から除外する。これらの処理を経て残った単語が平易化を試みる難解語である。

なお、形態素解析には MeCab(0.993)<sup>5</sup> および IPADIC(2.7.0)<sup>6</sup> を使用した。また、複合表現リストには、複合名詞として日本語 Wikipedia の見出し語<sup>7</sup>、複合動詞として複合動詞レキシコンの見出し語<sup>8</sup>、慣用句として佐藤らの慣用句リスト<sup>107)</sup> を使用した。十分平易な単語リストとしては、小学生のための理解語彙である学習基本語彙<sup>99)</sup> を使用した。

## 6.3 語彙的換言の生成

続いて、前節で検出した各難解語について、語彙的換言を列挙する。これらの語彙的換言が、平易語の候補となる。本節では 2.1 節で挙げた語彙的換言知識を評価し、品質の高

<sup>5</sup><https://code.google.com/p/mecab/>

<sup>6</sup><http://sourceforge.jp/projects/ipadic/releases/24435/>

<sup>7</sup><http://dumps.wikimedia.org/jawiki/>

<sup>8</sup><http://vvlexicon.ninjal.ac.jp/>



い語彙的換言知識を組合せて換言を得る。

まず、2.1 節で挙げた語彙的換言知識の概要を表 6.1 にまとめる。表 6.2 は、これらの各語彙的換言知識の品質を、それぞれ 3 人の評価者によって評価した結果である。各評価者は、ある文脈中で、換言対象の語または句を換言知識から得られる換言対を用いて置換し、その文脈中での同義性を判定する。文脈と換言対の各組に対して「換言可能」または「換言不可能」の回答を行う。次の例のように、換言対のみを見れば適切な換言だと思われるような { 不熟 / 未成熟 } の例を考えても、周辺の文脈によって換言可能な場合と換言不可能な場合が存在する。

- 換言可能：入選作は、彩調など { 不熟 / 未成熟 } で、…
- 換言不可能：セリフに { 不熟 / 未成熟 } があつたりもするが、…

そこで本研究では、ひとつの換言対を、複数の文脈に照らして評価する。ここでは英語の語彙的換言タスク<sup>29)</sup>の設定に倣い、各換言対を 10 種類の文脈に照らして評価する。ここでは文脈として、1990 年から 2004 年までの 15 年分の日本経済新聞<sup>9)</sup>の本文から、換言対象の語または句が含まれる文を無作為に抽出する。

表 6.1 語彙的換言知識の概要

名称	略称	収録数	品詞	収集方法	形式
PPDB: Japanese (S サイズ)	PPDB	1,522,492	句	対訳コーパス	句 句
内容語換言辞書 (用言等換言辞書)	用言	9,702	用言	人手	語 句
内容語換言辞書 (普通名詞換言辞書)	普通名詞	15,879	名詞		
日本語 WordNet 同義語データベース	WordNet	23,506		言語資源および人手	語 語
動詞含意関係データベース (含意が成り立つ類義 / 上位下位関係)	Ntriv	63,572	動詞	含意スコア計算 (相互情報量等) および人手	
動詞含意関係データベース (文字列上包含関係にあり、 含意が成り立つ類義 / 上位下位関係)	Triv	30,453			
基本的意味関係の事例ベース (略記対)	略記対	3,276	名詞	文脈類似度 および人手	
基本的意味関係の事例ベース (異形同義語対)	同義語対	44,754			
基本的意味関係の事例ベース (異表記対)	異表記対	30,230			

表 6.2 において、評価 1 から評価 3 は、それぞれの評価者が「換言可能」と回答した割合である。また、平均とは 3 人の評価者の「換言可能」と回答した割合の平均値である。次に、多数決とは 3 人の評価者のうち 2 人以上の評価者が「換言可能」と回答した文脈と換言対の組を正解と定義したときの、正解の割合を表す。表 6.2 では各換言対に対して 10 種類の文脈で評価を実施しているが、 は 10 種類全てが正解である換言対の種類数、は 10 種類全てが不正解 (3 人の評価者のうち 2 人以上の評価者が「換言不可能」と回答した

<sup>9)</sup><http://www.nikkeibookvideo.com/kijidb/>

表 6.2 語彙的換言知識の評価

換言知識	評価 1	評価 2	評価 3	平均	多数決	×	一致率	Fleiss	Cohen (1-2, 2-3, 1-3)				
PPDB	0.142	0.132	0.240	0.171	0.156	2	17	31	0.826	0.591	0.687	0.508	0.611
用言	0.652	0.728	0.704	0.695	0.702	24	24	2	0.710	0.544	0.461	0.499	0.671
普通名詞	0.672	0.736	0.802	0.737	0.738	23	24	3	0.728	0.533	0.704	0.446	0.440
WordNet	0.376	0.500	0.530	0.469	0.488	11	29	10	0.662	0.548	0.496	0.604	0.555
Ntriv	0.740	0.634	0.716	0.697	0.700	25	16	9	0.702	0.530	0.472	0.606	0.516
Triv	0.376	0.810	0.738	0.641	0.742	24	19	7	0.414	0.151	0.220	0.376	0.102
略記対	0.976	0.956	0.848	0.927	0.960	43	7	0	0.852	0.274	0.454	0.299	0.218
同義語対	0.830	0.862	0.942	0.878	0.948	41	9	0	0.722	0.135	0.050	0.244	0.174
異表記対	0.856	0.996	0.930	0.927	0.976	47	3	0	0.806	0.040	-0.008	-0.008	0.144

事例)である換言対の種類数、 は正解と不正解が混在する換言対の種類数を、それぞれ表す。最後に、一致率は3人の評価者の評価の単純な一致率(3人とも「換言可能」と回答または3人とも「換言不可能」と回答した割合)であり、Fleiss は3人の評価者間の Fleiss ' kappa<sup>11)</sup>を計算したものであり、Cohen は2人ずつの評価者間の Cohen ' s kappa<sup>8)</sup>を計算したものである。

表6.2の評価結果から、まずPPDB辞書の品質が低いことが分かる。表6.1より、PPDB辞書以外の8つの語彙的換言知識は最終的に人手での判定が行われているが、PPDB辞書のみ人手での判定を経ずに構築されているのが、この最大の原因であると考えられる。

次に、WordNet辞書において、3人の評価者の多数決により正解と判定された割合が、約5割とPPDB辞書に次いで低い。これは、WordNet辞書からは、ALAGINフォーラムにて公開されている言語資源に含まれている語対が除外されていることが大きな原因であると考えられる。基本的意味関係の事例ベースなど、ALAGINフォーラムにて公開されている言語資源は、文脈類似度の高い語同士の対を収集したものである。「似た意味の語は似た文脈で出現する」という分布仮説<sup>12)</sup>によると、これらの語対は換言可能性が高いと考えることができる。よって、これらの語対を対象外としたWordNet辞書は、比較的低品質になっていると考える。

続いて、内容語換言辞書や動詞含意関係データベースは、多数決で正解と判定された割合が7割を超えており、PPDB辞書やWordNet辞書よりも比較的高品質な語彙的換言知識だと言える。Triv辞書以外は、Fleiss ' kappaも0.5を超えており、まずまず(Moderate)の一致が得られている。内容語換言辞書は、人手で換言が付与されており、「無理な換言は行わない」「置換可能な語が思い浮かばない場合は内容語3語までの句に置換して良い」などの工夫により、(文脈によらず常に換言不可能)と判定される換言対が比較的少ない。

最後に、基本的意味関係の事例ベースは、いずれも多数決で正解と判定された割合が9割を超えており、極めて高品質な語彙的換言知識であると言える。Fleiss ' kappaは低いですが、これはFleiss ' kappaが偶然の一致を勘案した指標であるためだと考える。いずれの

評価者もほとんどの回答が「換言可能」である基本的意味関係の事例ベースの評価においては、偶然の一致の重みが大きくなっているため全体のスコアが下がっていると考えられる。3人の評価者の評価の単純な一致率を見ると、比較的高い確率で一致していると言える。

以上の調査に基づいて、本システムでは最終的に人手での確認を行っている比較的高品質な基本的意味関係の事例ベース<sup>10</sup>、内容語換言辞書<sup>11</sup>、動詞含意関係データベース<sup>12</sup>、日本語 WordNet 同義語データベース<sup>13</sup> から語彙的換言を収集した。

## 6.4 語義曖昧性の解消

前節で難解語の語彙的換言を得たが、換言可能と思われる単語対でも、周辺の文脈によって換言可能な場合と換言不可能な場合が存在する。そこで、入力文の文脈を考慮して不適格な換言を除去する手法を提案する。

文の適格性を評価するために、本システムでは文の基本的な意味構造である述語項構造を用いる。述語とその項の関係を見ることで、文の適格性を評価することができる。例えば、図 6.1 の入力文を述語項構造解析することで、述語が「担う」であり、そのガ格の項が「若者」であることが分かる。また、格フレーム辞書には述語の取り得る項の情報が記載されているので、述語「担う」がガ格に「若者」を取ることを格フレーム辞書で調べることにより、この述語項構造が適格であることが分かる。このように、入力文から述語項構造解析によって(項, 格, 述語)の三つ組を抽出し、項または述語を換言した三つ組が格フレーム辞書に存在しなければ、それを不適格な換言として除去する。

図 6.1 の例では、難解語として述語である「担う」を検出している。また「担う」の換言として「伝承する」や「支える」などを得ている。述語項構造解析により、「担う」のガ格の項が「若者」であることが分かっているが、「担う」の換言である「伝承する」の格フレームにはガ格に「若者」が登録されていないため、「伝承する」を除外する。一方、「担う」の別の換言である「支える」の格フレームにはガ格に「若者」が登録されているため、「支える」は平易化候補として残す。

本システムでは、述語項構造解析に SynCha(0.3.0)<sup>14</sup> を、格フレーム辞書に京都大学格フレーム(1.0)<sup>15</sup> を使用した。

<sup>10</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

<sup>11</sup><http://www.jnlp.org/SNOW/D2>

<sup>12</sup><https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

<sup>13</sup><http://nlpwww.nict.go.jp/wn-ja/jpn/downloads.html>

<sup>14</sup><http://www.cl.cs.titech.ac.jp/~ryu-i/syncha/>

<sup>15</sup><http://www.gsk.or.jp/catalog/gsk2008-b/>

## 6.5 難易度に基づく並び替え

前節で、難解語の語彙的換言の中から述語項構造が適格な語が得られた。最後に、これらの単語の中から最も平易な語を入力文中の難解語と置換して出力文を生成する。本システムでは、単語親密度データベース<sup>79)</sup>を用いて各単語に難易度を付与した。親密度の最も高い語が、最も平易な語である。

## 6.6 システムの性能の検証

本章で構築した日本語の語彙平易化システムの性能を、5章で構築した日本語の語彙平易化評価セットを用いて検証した。表 6.3 はシステムの性能である。2,330 文からなる評価セットのうち、対象語よりも平易な語が存在するのは 1,586 文であり、本システムがそれらの対象語を何らかの表現に変換したのは 146 文であった。そのうち、評価セットで対象語よりも平易であると定義されている表現を正しく出力できたのは 130 文であった。なお、全体では本システムは 2,330 文に対して 652 箇所の変換を行っているが、評価セットに含まれる対象語はそのうちの 146 語であった。表 6.4 に平易化できた対象語の品詞の内訳を示す。形容詞が少なく、副詞を全く平易化できないのは、語義曖昧性の解消の機構で述語項構造解析結果を用いたためである。本システムでは述語項構造解析によって得られる主語と述語の組を平易化するが、副詞は主語にも述語にもならないので平易化の対象とならない。また、形容詞は述語になる場合と名詞を修飾して名詞句になる場合があるが、本システムが平易化の対象とするのは述語となる形容詞のみである。

検証の結果、本システムは高い適合率が得られる一方で再現率は極めて低く、F 値も低い。副詞や名詞を修飾する形容詞を中心に、再現率を向上させることが今後の課題である。

以下にシステムの入出力の例を示す。

- 海外からの { 訪問者 お客さん } に { 配る 渡す } ほか、神戸市の海外事務所に送付する。
- 研究開発学校として指定されれば、学習指導要領の { 枠 フレーム } にとらわれない教育も可能。
- 北陸銀の高木 { 頭取 社長 } も「リストラ策と収益拡大策の両面」で統合効果を早期に極大化する考えを示した。

表 6.3 システムの性能

Precision	Recall	F-measure
0.89	0.08	0.15

表 6.4 平易化した品詞の内訳

名詞	動詞	形容詞	副詞
62	65	3	0

## 6.7 まとめ：語彙平易化システム

本章では日本語の語彙平易化システムについて述べた。本システムは、難解語の検出、語彙的換言の生成、語義曖昧性の解消、難易度に基づく並び替えという語彙平易化の典型的な4つの機構を持つ。

特に、語彙的換言の生成においては、公開されている日本語の語彙的換言知識を調査し、品質の高い知識から語彙的換言を収集した。また、語義曖昧性の解消においては、述語項構造解析によって主語と述語の組みを抽出し、意味構造の評価をして語彙を平易化した。5章で構築した日本語の語彙平易化評価セットで本システムの性能を評価したところ、0.89の適合率が得られた一方で、再現率は0.08に留まった。

本システムでは、語彙的換言の生成において、精度の低い句単位の換言知識の多くを除いている。今後は高精度に句単位の換言を収集することで、再現率が改善できるであろう。関連して、格フレーム辞書や単語親密度データベースなどが単語単位の言語資源であるため、6.4節では句の検索時に末尾の単語をその句の主辞と見て主辞で検索を行っている。また6.5節では句の難易度として各単語の難易度の平均値を採用している。これらの基本的な言語資源についても、句単位で整備することにより、語彙平易化をはじめとする各種自然言語処理タスクの性能改善が期待される。

## 7 結論

本論文では、日本語学習者や小学生をはじめとする幅広い日本語の読者の文章読解を支援することを目的として、語彙平易化の研究に取り組んだ。語彙平易化は、難解語の検出、語彙的換言の生成、語義曖昧性の解消、難易度に基づく並び替えの4つの機構で構成されるが、本論文の前半では語彙平易化への応用を意識した語彙的換言の生成に関する手法を提案した。

語彙的換言手法は、知識に基づく手法と用例に基づく手法に大別される。知識に基づく手法では、先行研究の課題であった網羅性について、品詞を限定しない点、複数の換言知識を併用する点、「語釈文の主辞は文末に位置する」というヒューリスティクスを排除して国語辞典の語釈文全体から見出し語の換言候補を広く収集する点の3点で改善した。収集した換言候補から、シソーラスの距離に基づく指標および5種類の統計的な指標を組み合わせて最終的な換言を選択し、70%の適合率および84%の再現率を達成した。また、小学生を対象とした実験によって読解支援への有効性も確認でき、提案手法による換言によって小学生が一般向けの新聞記事の79%を理解できるようになった。

用例に基づく手法では、従来は共起頻度や相互情報量によって計算されていた文脈の類似度を、文脈の種類数という新たな統計に着目して計算する手法を提案した。文脈の種類数に基づく手法で、従来手法の約2倍の精度で語彙的換言を獲得することができるようになった。また、頻度や文脈の種類数と語の難易度の分布を調査した結果、本手法が対象とする領域に多くの平易語が存在することが明らかとなり、本換言手法の語彙平易化への有効性も確認できた。

本論文の後半では、日本語の語彙平易化システムおよびその自動評価のためのデータセットを構築した。本研究の貢献は、日本語で初めてこれらの研究資源を構築し、公開したことである。本研究では、クラウドソーシングを利用して、のべ500人の作業者によって233語×10文の規模で語彙的換言の評価のためのデータセットを構築し、このデータセット中の対象語およびその換言を難易度によって並び替えることによって語彙平易化の評価のためのデータセットを構築した。このデータセットの公開により、日本語の語彙平易化システムの適合率および再現率を自動評価することが可能となった。

---

また、公開されている日本語の語彙的換言知識や語の難易度データベース、テキスト解析器を組み合わせて、標準的な語彙平易化システムを構築した。本システムを先述の評価セットで評価したところ、89%の適合率および8%の再現率が得られた。この標準的な語彙平易化システムの公開により、研究者はこれをベースラインとして研究を効率的に進めることが可能となった。また、日本語の読解支援を必要とする読者に対して、語彙平易化の技術を提供することが可能となった。

本研究の成果をもとに、日本語の語彙平易化が活発に研究され、子どもや言語学習者をはじめとする幅広い読者の文章読解支援の助けとなる語彙平易化システムが開発されることを期待する。

## 謝辞

本研究を遂行するにあたり、多大なる御指導そして御協力いただきました長岡技術科学大学の山本和英准教授に深く感謝致します。山崎克之教授ならびに王龍標産学融合特任准教授には、論文の推敲に際し貴重な御意見をいただきました。深く感謝致します。

最後に、様々な場面でお世話になりました山本研究室の皆様に、心より感謝致します。



## 参考文献

- 1) Regina Barzilay. *Information Fusion for Multi-document Summarization: Paraphrasing and Generation*. Ph.D.thesis, Columbia University, 2003.
- 2) Regina Barzilay and Kathleen R. McKeown. Extracting paraphrases from a parallel corpus. *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, pp. 50–57, 2001.
- 3) Jan De Belder and Marie-Francine Moens. Text simplification for children. *In Proceedings of the SIGIR Workshop on Accessible Search Systems*, pp. 19–26, 2010.
- 4) Jan De Belder and Marie-Francine Moens. A dataset for the evaluation of lexical simplification. *In Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing-2012)*, pp. 426–437, 2012.
- 5) Rahul Bhagat and Deepak Ravichandran. Large scale acquisition of paraphrases for learning surface patterns. *In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pp. 674–682, 2008.
- 6) Francis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese semcor: A sense-tagged corpus of japanese. *In Proceedings of the 6th International Conference of the Global WordNet Association*, pp. 56–63, 2012.
- 7) John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. Practical simplification of english newspaper text to assist aphasic readers. *In Proceedings of the AAAI98 Workshop on Integrating AI and Assistive Technology*, pp. 7–10, 1998.
- 8) Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, 1960.

- 9) William Coster and David Kauchak. Simple wikipedia: A new simplification task. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 665–669, 2011.
- 10) Siobhan Devlin and Gary Unthank. Helping aphasic people process online information. *In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 225–226, 2006.
- 11) Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.
- 12) Zellig Harris. Distributional structure. *Word*, Vol. 10, No. 23, pp. 146–162, 1954.
- 13) Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Junichi Kazama, and Sadao Kurohashi. Extracting paraphrases from definition sentences on the web. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 1087–1097, 2011.
- 14) Wei He, Hua Wu, Haifeng Wang, and Ting Liu. Improve smt quality with automatically extracted paraphrase rules. *In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 979–987, 2012.
- 15) Colby Horn, Cathryn Manduca, and David Kauchak. Learning lexical simplifier using wikipedia. *In Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, pp. 458–463, 2014.
- 16) Kentaro Inui, Atsushi Fujita, Tetsuro Takahashi, Ryu Iida, and Tomoya Iwakura. Text simplification for reading assistance: A project note. *In Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications*, pp. 9–16, 2003.
- 17) Nobuhiro Kaji, Daisuke Kawahara, Sadao Kurohashi, and Satoshi Sato. Verb paraphrase based on case frame alignment. *In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 215–222, 2002.
- 18) Nobuhiro Kaji and Sadao Kurohashi. Recognition and paraphrasing of periphrastic and overlapping verb phrases. *In Proceedings of the 4th International Conference*

- on Language Resources and Evaluation Workshop on Methodologies and Evaluation of Multiword Units in Real-world Application*, pp. 24–30, 2004.
- 19) Hiroshi Kanayama. Paraphrasing rules for automatic evaluation of translation into japanese. *In Proceedings of the Second International Workshop on Paraphrasing*, pp. 88–93, 2003.
  - 20) David Kauchak and Regina Barzilay. Paraphrasing for automatic evaluation. *In Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 455–462, 2006.
  - 21) Frank Keller, Maria Lapata, and Olga Ourioupina. Using the web to overcome data sparseness. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 230–237, 2002.
  - 22) Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2004.
  - 23) Maria Lapata, Frank Keller, and Scott McDonald. Evaluating smoothing algorithms against plausibility judgements. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 346–353, 2001.
  - 24) Dekang Lin and Patrick Pantel. Discovery of inference rules for question answering. *Natural Language Engineering*, Vol. 7, No. 4, pp. 343–360, 2001.
  - 25) Dekang Lin, Shaojun Zhao, Lijuan Qin, and Ming Zhou. Identifying synonyms among distributionally similar words. *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pp. 1492–1493, 2003.
  - 26) Xiaojuan Ma, Christiane Fellbaum, and Perry R. Cook. A multimodal vocabulary for augmentative and alternative communication from sound/image label datasets. *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pp. 62–70, 2010.

- 27) Yuval Marton. Distributional phrasal paraphrase generation for statistical machine translation. *ACM Transactions on Intelligent Systems and Technology*, Vol. 4, No. 3, pp. 1–32, 2013.
- 28) Yuval Marton, Chris Callison-Burch, and Philip Resnik. Improved statistical machine translation using monolingually-derived paraphrases. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 381–390, 2009.
- 29) Diana McCarthy and Roberto Navigli. Semeval-2007 task10: English lexical substitution task. *In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 48–53, 2007.
- 30) Masahiro Mizukami, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. Building a free, general-domain paraphrase database for japanese. *In Proceedings of the 17th Oriental COCOSDA Conference*, pp. 129–133, 2014.
- 31) Matthew Shardlow. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing*, pp. 58–70, 2014.
- 32) Serge Sharoff. Open-source corpora: Using the net to fish for linguistic data. *International Journal of Corpus Linguistics*, 11(4), pp. 435–462, 2006.
- 33) Yusuke Shinyama and Satoshi Sekine. Paraphrase acquisition for information extraction. *Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications*, pp. 65–71, 2003.
- 34) Lucia Specia, Sujay Kumar Jauhar, and Rada Mihalcea. Semeval-2012 task 1: English lexical simplification. *In Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval-2012)*, pp. 347–355, 2012.
- 35) Kazuhide Yamamoto. Acquisition of lexical paraphrases from texts. *Proceedings of 2nd International Workshop on Computational Terminology*, p. no page numbers, 2002.

- 36) Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Junichi Kazama, and Stijn De Saeger. Minimally supervised method for multilingual paraphrase extraction from definition sentences on the web. *In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 63–73, 2013.
- 37) Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Junichi Kazama, and Stijn De Saeger. Multilingual paraphrase extraction from definition sentences on the web. *言語処理学会第 19 回年次大会発表論文集*, pp. 268–271, 2013.
- 38) Liang Zhou, Chin-Yew Lin, Dragos Stefan Munteanu, and Eduard Hovy. Paraeval: Using paraphrases to evaluate summaries automatically. *In Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 447–454, 2006.
- 39) 高地なつめ, 磯崎秀樹. スクランプリングを考慮した和訳の自動評価法の ntcir-9 データによる検証. *情報処理学会研究報告*, NL-219-2, pp. 1–5, 2014.
- 40) 川村よし子, 北村達也. 文章の難易度判定のための単語親密度チェッカーの開発. *日本語教育方法研究会誌*, Vol. 15, No. 2, pp. 24–25, 2008.
- 41) 平原一帆, 難波英嗣, 竹澤寿幸, 奥村学. 言い換えを用いたテキスト要約の自動評価. *情報処理学会論文誌 (TOD)*, Vol. 3, No. 2, pp. 91–101, 2010.
- 42) 荒牧英治, 久保圭, 四方朱子. 老いと <ことば>: ブログ・テキストから測る老化. *情報処理学会研究報告*, IFAT-115-23, pp. 1–6, 2014.
- 43) 荒牧英治, 増川佐知子. 微小時間における日本語の変化とその法則. *言語処理学会第 17 回年次大会発表論文集*, pp. 432–435, 2011.
- 44) 荒牧英治, 増川佐知子, 宮部真衣, 森田瑞樹, 保田祥. 頻出語ではなく使用者が多い語が自然な日本語である. *言語処理学会第 19 回年次大会発表論文集*, pp. 544–547, 2013.
- 45) 土屋雅稔, 黒橋禎夫, 佐藤理史. 辞書定義文の圧縮による定義表現パターンの発見. *人工知能学会論文誌*, Vol. 17, No. 4, pp. 420–430, 2002.

- 46) 吉見毅彦. 英日機械翻訳における自然な和文生成のための英語名詞句の書き換え. 自然言語処理, Vol. 8, No. 4, pp. 55–70, 2001.
- 47) 吉見毅彦, 佐田いち子. 英字新聞記事見出し翻訳の自動前編集による改良. 自然言語処理, Vol. 7, No. 2, pp. 27–43, 2000.
- 48) 吉見毅彦, 佐田いち子, 福持陽士. 頑健な英日機械翻訳システム実現のための原文自動前編集. 自然言語処理, Vol. 7, No. 4, pp. 99–117, 2000.
- 49) 湊吉正. 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.
- 50) 近藤恵子, 奥村学. 言い替えを使用した要約の手法. 情報処理学会研究報告, NL-116-20, pp. 137–142, 1996.
- 51) 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, pp. 4064–4074, 1999.
- 52) 小林健, 安藤一秋. 小学生を対象とした web 新聞読解支援のための説明語抽出の検討. 第 26 会人工知能学会全国大会, 4I1-R-9-7, pp. 1–4, 2012.
- 53) 乾健太郎, 山本聡美, 野上優, 藤田篤, 乾裕子. 聾者向け文章読解支援における構文的言い換えの効果について. 電子情報通信学会技術研究報告, WIT99-2, pp. 9–14, 1999.
- 54) 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151–198, 2004.
- 55) 村山賢洋, 麻岡正洋, 土屋雅稔, 佐藤理史. 語彙の規格とそれに基づく用言の言い換え支援. 言語処理学会第 9 回年次大会発表論文集, pp. 85–88, 2003.
- 56) 赤部晃一, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲. パラフレーズを考慮した機械翻訳の誤り箇所選択. 情報処理学会研究報告, NL-219-19, pp. 1–8, 2014.
- 57) 見坊豪紀, 金田一京助, 金田一春彦, 柴田武, 飛田良文. 三省堂国語辞典第 4 版. 三省堂, 1994.
- 58) 国際交流基金, 財団法人日本国際教育協会. 日本語能力試験出題基準. 凡人社, 1994.
- 59) 国立国語研究所コーパス開発センター. 『現代日本語書き言葉均衡コーパス』利用の手引き. 第 1.0 版, 2011.

- 60) 岩越守孝, 増田英孝, 中川裕志. Web と携帯端末向けの新聞記事の対応コーパスからの文末言い換え抽出. 自然言語処理, Vol. 12, No. 5, pp. 157–184, 2005.
- 61) 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回年次大会発表論文集, pp. 760–763, 2010.
- 62) 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回年次大会発表論文集, pp. 744–747, 2011.
- 63) 美野秀弥, 熊野正, 田中英輝. 放送ニュースの基本語彙の抽出. 言語処理学会第18回年次大会発表論文集, pp. 77–80, 2012.
- 64) 松吉俊, 佐藤理史. 文体と難易度を制御可能な日本語機能表現の言い換え. 自然言語処理, Vol. 15, No. 2, pp. 75–99, 2008.
- 65) 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol. 14, No. 5, pp. 123–146, 2007.
- 66) 風間淳一, Stijn De Saeger, 鳥澤健太郎, 村田真樹. 係り受けの確率的クラスタリングを用いた大規模類似語リストの作成. 言語処理学会第15回年次大会発表論文集, pp. 84–87, 2009.
- 67) 水野淳太, 大山浩美, 小林朋幸, 坂田浩亮, Noah Evans, 谷口雄作, 松本裕治. 日本語読解支援のための語義ごとの用例抽出システムの構築. 言語処理学会第14回年次大会併設ワークショップ, pp. 63–66, 2008.
- 68) 相澤彰子. 大規模テキストコーパスを用いた語の類似度計算に関する考察. 情報処理学会論文誌, Vol. 49, No. 3, pp. 1426–1436, 2008.
- 69) 鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史. 国語辞典とコーパスを用いた用言の言い換え規則の学習. 言語処理学会第8回年次大会発表論文集, pp. 331–334, 2002.
- 70) 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.
- 71) 鍛冶伸裕, 黒橋禎夫. 冗長表現の認識と言い換え. 言語処理学会第9回年次大会発表論文集, pp. 234–237, 2003.

- 72) 鍛冶伸裕, 黒橋禎夫, 佐藤理史. 国語辞典に基づく平易文へのパラフレーズ. 情報処理学会研究報告, NL-144-23, pp. 167-174, 2001.
- 73) 鍛冶伸裕, 喜連川優. 言い換えと逆翻字を用いた片仮名複合名詞の分割. 自然言語処理, Vol. 19, No. 2, pp. 65-88, 2012.
- 74) 宮部真衣, 四方朱子, 久保圭, 荒牧英治. 音声認識による認知症・発達障害スクリーニングは可能か? - 言語能力測定システム“言秤”の提案 -. グループウェアとネットワークサービスワークショップ 2014 論文集, pp. 1-8, 2014.
- 75) 村田真樹, 井佐原均, 金丸敏幸. 複数の辞書の定義文の照合に基づく同義表現の自動獲得. 自然言語処理, Vol. 11, No. 5, pp. 135-149, 2004.
- 76) 空真奈見, 山本和英. 「やさしい日本語」変換システムの試作. 言語処理学会第 19 回年次大会発表論文集, pp. 678-681, 2013.
- 77) 藤沢仁子, 相原健郎, 神門典子. 文化遺産に関する説明文の対象ユーザに合わせた言い換えの提案. 情報処理学会研究報告, NL-174-2, pp. 7-12, 2006.
- 78) 藤沢仁子, 神門典子, 相原健郎, 安達淳. 言い換え箇所と言い換え候補の提示による解説文リライト支援の書き手の評価実験. 情報処理学会研究報告, NL-190-10, pp. 59-66, 2009.
- 79) 天野成昭, 近藤公久. NTT データベースシリーズ日本語の語彙特性 (第 1 期 CD-ROM 版). 三省堂, 1999.
- 80) 天野成昭, 近藤公久. Ntt データベースシリーズ「日本語の語彙特性」について. 音声研究, Vol. 4, No. 2, pp. 44-50, 2000.
- 81) 内山清子, Timothy Baldwin. 機械学習を用いた複合動詞の多義性解消. 言語処理学会第 10 回年次大会発表論文集, pp. 741-744, 2004.
- 82) 中西聖明, 小林伸行, 椎名広光. 辞書データ・web データを利用したクラスタリングによる単語難易度の推定. 言語処理学会第 19 回年次大会発表論文集, pp. 682-685, 2013.
- 83) 関根聡. 複数の新聞を使用した言い替え表現の自動抽出. 言語処理学会第 7 回年次大会ワークショップ論文集, pp. 9-14, 2001.



- 84) 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. 自然言語処理, Vol. 12, No. 2, pp. 109–131, 2005.
- 85) 河原大輔, 黒橋禎夫. 京都大学格フレーム (Ver 1.0). 言語資源協会, 2009.
- 86) 工藤拓, 賀沢秀人. Web 日本語 N グラム第 1 版. 言語資源協会, 2007.
- 87) 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.
- 88) 柴田知秀, 黒橋禎夫. 文脈に依存した述語の同義関係獲得. 情報処理学会研究報告, NL-199-13, pp. 1–6, 2010.
- 89) 高橋哲朗, 乾健太郎, 関根聡, 松本裕治. 質問応答に必要な言い換えの分析. 言語処理学会第 10 回年次大会発表論文集, pp. 309–312, 2004.
- 90) 藤田篤, 乾健太郎. 語釈文を利用した普通名詞の同概念語への言い換え. 言語処理学会第 7 回年次大会発表論文集, pp. 331–334, 2001.
- 91) 藤田篤, 乾健太郎. 語彙的言い換えに必要な知識の部品化. 情報処理学会研究報告, NL-149-5, pp. 31–38, 2002.
- 92) 藤田篤, 乾健太郎. 語彙・構文的言い換えにおける変換誤りの分析. 情報処理学会論文誌, Vol. 44, No. 11, pp. 2826–2838, 2003.
- 93) 藤田篤, 乾健太郎, 乾裕子. 名詞言い換えコーパスの作成環境. 電子情報通信学会技術研究報告, TL2000-32, pp. 53–60, 2000.
- 94) 山崎敦, 沢井康孝, 山本和英. 構文情報を用いた名詞句の換言. 言語処理学会第 12 回年次大会発表論文集, pp. 775–778, 2006.
- 95) 日本電子化辞書研究所. EDR 電子化辞書仕様説明書. 1995.
- 96) 寺田博視, 田中久美子. 単語親密度と単語頻度の関係に関する一考察. 言語処理学会第 14 回年次大会発表論文集, pp. 713–716, 2008.
- 97) 芋野美紗子, 吉村枝里子, 土屋誠司, 渡部広一. 新聞記事中の難解語を平易な表現へ変換する手法の提案. 自然言語処理, Vol. 20, No. 2, pp. 105–132, 2013.

- 98) 泉朋子, 今村賢治, 菊井玄一郎, 藤田篤, 佐藤理史. 正規化を指向した機能動詞表現の述部言い換え. 言語処理学会第15回年次大会発表論文集, pp. 264–267, 2009.
- 99) 甲斐睦朗, 松川利広. 語彙指導の方法: 語彙表編. 光村図書出版株式会社, 2002.
- 100) 毎日新聞社. CD-毎日新聞 2000年度版. 2000.
- 101) 砂川有里子. 学習辞書編集支援データベース作成について-『学習辞書科研』プロジェクトの紹介. 日本語教育連絡会議論文集, Vol. 24, pp. 164–169, 2012.
- 102) 山形祐輝, 山本和英. 普通名詞換言辞書の構築. 言語処理学会第20回年次大会発表論文集, pp. 7–10, 2014.
- 103) 藤沢祐輔, 相原慎太郎, 安藤一秋. Web一般新聞記事を子供向けに言い換える知識の抽出. 言語処理学会第18回年次大会発表論文集, pp. 751–754, 2012.
- 104) 海野裕也, 宮尾祐介, 辻井潤一. 自動獲得された言い換え表現を使った情報検索. 言語処理学会第14回年次大会発表論文集, pp. 123–126, 2008.
- 105) 佐藤理史. 論文表題を言い換える. 情報処理学会論文誌, Vol. 40, No. 7, pp. 2937–2945, 1999.
- 106) 佐藤理史. なぜ言い換え/パラフレーズを研究するのか. 言語処理学会第7回年次大会ワークショップ論文集, pp. 1–2, 2001.
- 107) 佐藤理史. 基本慣用句五種対照表の作成. 情報処理学会研究報告, NL-178-1, pp. 1–6, 2007.
- 108) 林良彦, 菊井玄一郎. 日本文推敲支援システムにおける書換え支援機能の実現方式. 情報処理学会論文誌, Vol. 32, No. 8, pp. 962–970, 1991.
- 109) 笹野遼平, 河原大輔, 黒橋禎夫. 自動獲得した知識に基づく統合的な照応解析. 言語処理学会第12回年次大会発表論文集, pp. 480–483, 2006.
- 110) 橋本力, 鳥澤健太郎, ステイン・デ・サーガ, 風間淳一, 黒橋禎夫. Web上の定義文からの言い換え知識獲得. 言語処理学会第17回年次大会発表論文集, pp. 748–751, 2011.
- 111) 山本和英. 換言処理の現状と課題. 言語処理学会第7回年次大会ワークショップ論文集, pp. 93–96, 2001.

- 
- 112) 山本和英, 大橋一輝. 「サ変動詞+名詞」の複合名詞への換言. 自然言語処理, Vol. 12, No. 3, pp. 19-42, 2005.
- 113) 山本和英, 吉倉孝太郎. 用言等換言辞書を人手で作りました. 言語処理学会第19回年次大会発表論文集, pp. 276-279, 2013.

## 本論文に関する原著論文

- (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, 発表 1, 2012.
- (2) 梶原 智之, 山本 和英. 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. 言語処理学会第 19 回年次大会発表論文集, pp.272-275, 2013.
- (3) 梶原 智之, 山本 和英. 小学生の読解支援に向けた語釈文から語彙的換言を選択する手法. NLP 若手の会第 8 回シンポジウム, 発表 23, 2013.
- (4) Tomoyuki Kajiwara, Hiroshi Matsumoto, Kazuhide Yamamoto. Selecting Proper Lexical Paraphrase for Children. In *Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*, pp.59-73, 2013.
- (5) 梶原 智之, 山本 和英. 文脈の多様性に基づく名詞換言の提案. 言語処理学会第 20 回年次大会発表論文集, pp.769-772, 2014.
- (6) 梶原 智之, 山本 和英. 文脈の多様性に基づく名詞換言の評価. 第 28 回人工知能学会全国大会論文集, 3I4-2, 2014.
- (7) 梶原 智之, 山本 和英. 高頻度語は平易語なのか?. NLP 若手の会第 9 回シンポジウム, 発表 P02, 2014.
- (8) Tomoyuki Kajiwara, Kazuhide Yamamoto. Noun Paraphrasing Based on a Variety of Contexts. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing (PACLIC 2014)*, no page numbers, 2014.
- (9) 梶原 智之, 山本 和英. 日本語の語彙的換言知識の質的評価. 信学技報, vol.114, no.366, NLC2014-37, pp.43-48, 2014.
- (10) 梶原 智之, 山本 和英. 日本語の語彙平易化システムの構築. 第 77 回情報処理学会全国大会講演論文集, no page numbers, 2015.
- (11) 梶原 智之, 山本 和英. 日本語の語彙平易化評価セットの構築. 言語処理学会第 21 回年次大会発表論文集, no page numbers, 2015.
- (12) 梶原 智之, 山本 和英. 語釈文を用いた小学生のための語彙平易化. 情報処理学会論文誌, vol.56, no.3, no page numbers, 2015.

## 付録 A : 3 章「知識に基づく語彙的換言」の実験結果の詳細

- 換言の妥当性の評価（小学国語辞典）
- 換言の妥当性の評価（EDR 日本語単語辞書）
- 換言の妥当性の評価（日本語 WordNet）
- 難解語と平易語の難易度の評価
- 換言前後の文の難易度の評価（換言前の難解文）
- 換言前後の文の難易度の評価（換言後の平易文）
- 知識に基づく語彙的換言の例

表 7.1 から表 7.3 に、各換言知識における換言の妥当性の評価を示す。ここでは、各換言対象語に対して最大 3 回の換言を行っている。学習基本語彙に含まれる語が得られた場合、またはその換言知識の見出し語として登録されていない語である場合、換言対象語あるいは既に出現した換言が再び出現してループとなる場合、そこで換言を終えている。「学習基本語彙」の列は、「○」が最終的に得られた語が学習基本語彙に含まれることを示す。「妥当性」の列は、学習基本語彙が得られた場合について、その語が換言対象語の換言として適切である場合を「○」、不適切である場合を「×」で示す。

表 7.4 に、難解語と、その難解語を換言して得られた学習基本語彙に含まれる平易語の難易度の評価を示す。各行の難解語と平易語が換言の関係にある。評価者はそれぞれ小学校高学年の 5 名である。各語の意味が理解できる場合を「○」、理解できない場合を「×」で示す。

表 7.5 および表 7.6 に、換言前後の文の難易度の評価を示す。表 7.5 は換言前の難解語を含む文、表 7.6 は換言後の平易語を含む文である。評価者は表 7.4 と同じ小学校高学年の 5 名である。文の意味が理解できる場合を「○」、理解できない場合を「×」で示す。

表 7.7 に、知識に基づく語彙的換言の例を示す。各文中で、換言が行われる部分を括弧で示す。括弧内の左側の語が難解語であり、括弧内の右側の語が難解語を換言して得られた平易語である。

表 7.1: 換言の妥当性の評価 (小学国語辞典)

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
いずれ	うち				×
サッカー	競技				
その後	あと				
ホームページ	ページ				
ホテル	ところ				×
意向	考えや気持ち				
遺体	体				×
会見	会うこと				
会合	集まり				
会長	人				
回収	集めること				
開催	開くこと				
開幕	始まること				
獲得	自分				×
確定	決まること				
寄せる	近づける				
規模	大きさ			-	-
起訴	うたえること				
疑い	疑うこと				
供述	述べること				
協議	決めること				
強化	強くすること			-	-
恐れ	こわがること				
掲載	のせること				
景気	動き				×
経営	いくこと				×
懸念	気がかりなこと			-	-
検出	見つけ出すこと			-	-
県	一つ				×
県警				-	-
見直し				-	-
現状	よう				×
言い渡す	伝える				
交渉	話し合うこと				
公表	知らせること				
合意	合うこと				
国会	議会				
国内	中				
今回	今度				
際	とき				
昨年	去年				

次のページに続く

前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
殺害	殺すこと				
市内	中				
支援	助けること				
死去	死ぬこと				
事業	仕事				
自殺	絶つこと				×
自宅	家				
辞任	やめること				
社長	人				
就任	つくこと				×
上回る	なる				×
乗用車	自動車				
申請	願い出ること	申し出ること	言うこと		
人気	評判	有名なこと			×
政権	権力				
盛り込む	とり入れる			-	-
設置	備えつけること			-	-
設立	つくること				
狙い	目的				
狙う	うかがう				
訴訟	うたえること				
喪主				-	-
捜査	たりすること				×
相次ぐ	続く				
葬儀	儀式				
対戦	戦うこと				
逮捕	つかまえること				
大幅	大きいこと				
達す				-	-
地元				-	-
党	なか				×
当地	土地				
同市				-	-
同社				-	-
同署				-	-
同日	日				
同様	よう				×
導入	とり入れること			-	-
日本人				-	-
派遣	せること			-	-
肺炎	病気				
発足	始めること				

次のページに続く



前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
判決	決めること				
判明	わかること				
反発	逆らうこと				
販売	売ること				
被告	人				
表明	表すこと				
負う	受ける				
報じる	知らせる				
本社	会社				
無職	ないこと				×
命じる	命令する				
問い合わせ	確かめること				
野党	政党				
与党	政党				
容疑	疑い	疑うこと			
要請	願うこと				
絡む	関係する				

これで終わり

表 7.2: 換言の妥当性の評価 (EDR 日本語単語辞書)

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
いずれ	どっちみち				×
サッカー	スポーツ				
その後	後				
ホームページ	ページ				
ホテル	施設				
意向	企画や構想	計画や構想			
遺体	人の				×
会見	町				×
会合	集まり				
会長	名誉職	名誉力			×
回収	回収			-	-
開催	始めること				
開幕	催し	催すこと			
獲得	身				×
確定	確か				
寄せる	寄せる			-	-
規模	大きさ			-	-
起訴	起訴			-	-
疑い	思うこと				×
供述	内容				×

次のページに続く

前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
協議	相談				
強化	補強			-	-
恐れ	懸念	心配			
掲載	掲載			-	-
景気	状態				
経営	経営			-	-
懸念	心配				
検出	成分				×
県	直轄領	管轄布	範囲布		×
県警	察本部	測る機関	測る複雑な仕掛け		×
見直し	見ること				
現状	状況				
言い渡す	命じる	命じる		-	-
交渉	的關係				×
公表	公表			-	-
合意	意見				×
国会	国家機関	政府複雑な仕掛け			×
国内	中				
今回	現時点	現時点		-	-
際	時期				
昨年	昨年			-	-
殺害	殺す				
市内	所				×
支援	支援			-	-
死去	動物や人				×
事業	行い	行状	行い	-	-
自殺	自害	自害		-	-
自宅	所有				×
辞任	地位や役目				×
社長	人				
就任	任務や職務	務めや仕事	力や仕事		×
上回る	すぐれる				
乗用車	自動車				
申請	要求				
人気	気風	特性	性質		×
政権	権力				
盛り込む	盛り込む			-	-
設置	物				×
設立	設立			-	-
狙い	事柄				×
狙う	目がける				
訴訟	訴える				

次のページに続く

前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
喪主	当主	主人			
捜査	未知				×
相次ぐ	引き続く	続く			
葬儀	儀式				
対戦	勝負				
逮捕	捕らえること				
大幅	大きいこと				
達す				-	-
地元	地域				
党	政党				
当地	地方				
同市				-	-
同社	神社	施設			
同署				-	-
同日	日				
同様				-	-
導入	段階				×
日本人	人				
派遣	所				×
肺炎	病気				
発足	新た				×
判決	内容				×
判明	はっきりと分かる				
反発	気持ち				×
販売	品物				×
被告	人				
表明	表明			-	-
負う	載せる				
報じる	むくいる				
本社	的存在				×
無職	ないこと				×
命じる	命じる			-	-
問い合わせ	質問				
野党	政党				
与党	政党				
容疑	あること				×
要請	根本命題	根本問題			×
絡む	もつ				×

これで終わり

表 7.3: 換言の妥当性の評価 (日本語 WordNet)

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
いずれ				-	-
サッカー	サッカー			-	-
その後	その後			-	-
ホームページ	ホームページ			-	-
ホテル	ホテル			-	-
意向	考え				
遺体	遺体			-	-
会見	面会				
会合	寄る				
会長	議長				×
回収	回収			-	-
開催	主催				×
開幕	開始				
獲得	得る				
確定	限定				×
寄せる	合計				
規模	規模			-	-
起訴	訴える				
疑い	直感				×
供述				-	-
協議	会談				×
強化	まとめる				×
恐れ	冒険				×
掲載	発表				
景気	景気			-	-
経営	動かす				×
懸念	恐れる				
検出	見つける				
県	県			-	-
県警				-	-
見直し	見直す				
現状	現状			-	-
言い渡す	言い渡す			-	-
交渉	取り引き				
公表	宣言				
合意	約束				×
国会	大会				×
国内				-	-
今回				-	-
際	辺				×
昨年				-	-

次のページに続く

前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
殺害	殺す				
市内				-	-
支援	救助				
死去	亡くなる				
事業	商業				
自殺	自殺			-	-
自宅	住宅				
辞任	退く				
社長	大統領				×
就任	就任			-	-
上回る	超す				
乗用車	車				
申請	需要				×
人気	流行				
政権				-	-
盛り込む	合同				×
設置	築く				×
設立	造る				
狙い	趣旨				
狙う	計画				×
訴訟	裁判				
喪主				-	-
捜査	探検				×
相次ぐ				-	-
葬儀	葬式				
対戦	取り組む				×
逮捕	占領				×
大幅	重大				
達す				-	-
地元	地元			-	-
党	政党				
当地				-	-
同市				-	-
同社				-	-
同署				-	-
同日				-	-
同様	同様			-	-
導入	取り入れる				
日本人	日本人			-	-
派遣	差し出す				
肺炎	肺炎			-	-
発足	開始				

次のページに続く

前のページからの続き

換言対象語	1 回目の換言	2 回目の換言	3 回目の換言	学習基本語彙	妥当性
判決	知恵				×
判明	現す				
反発	反抗				
販売	扱う				
被告	被告			-	-
表明	表現				
負う	負う			-	-
報じる	授業				×
本社	本社			-	-
無職				-	-
命じる	指名				
問い合わせ	質問				
野党	野党			-	-
与党				-	-
容疑	負担				×
要請	要求				
絡む				-	-

これで終わり

表 7.4: 難解語と平易語の難易度の評価

難解語	評価者 1	2	3	4	5	平易語	評価者 1	2	3	4	5
申請	×	×	×	×	×	言うこと					
容疑		×	×	×		疑うこと					
逮捕	×					つかまえること					
自宅						家					
死去						死ぬこと					
葬儀	×	×	×	×	×	儀式	×	×	×	×	×
今回						今度					
本社						会社					
表明		×	×	×		表すこと					
合意	×					合うこと					
開催		×	×		×	開くこと	×				
昨年						去年					
相次ぐ		×	×			続く					
際	×	×	×	×	×	とき					
協議	×	×	×	×	×	決めること					
販売						売ること					
会見		×	×	×	×	会うこと					
起訴	×	×	×	×	×	うたえること					
国会	×		×			議会		×	×	×	×
国内						中					

次のページに続く

前のページからの続き

難解語	評価者 1	2	3	4	5	平易語	評価者 1	2	3	4	5
支援						助けること					
意向	×	×	×	×		考えや気持ち					
設立						つくること					
開幕	×		×			始まること					
公表			×			知らせること					
野党	×	×	×	×	×	政党	×	×	×	×	
負う	×					受ける					
狙う						うかがう					
反発	×		×			逆らうこと	×				
サッカー						競技					
問い合わせ						確かめること					
狙い						目的					
殺害						殺すこと					
命じる	×					命令する					
発足	×	×	×	×	×	始めること					
交渉	×		×			話し合うこと					
乗用車		×				自動車					
辞任	×	×	×	×	×	やめること					
掲載	×	×	×	×		のせること					
その後						あと					
対戦						戦うこと					
政権	×	×	×	×		権力	×				
寄せる						近づける					
ホームページ						ページ					
肺炎						病気					
供述	×	×	×	×	×	述べること					
会合	×	×	×		×	集まり					
確定						決まること					
判明		×		×		わかること					
事業	×		×		×	仕事					
懸念	×	×	×	×	×	恐れる					
派遣	×	×	×	×	×	差し出す					
喪主	×	×	×	×	×	主人					
同社	×		×	×	×	施設		×			
地元						地域					
打ち出す			×		×	表す					
あり方	×			×	×	状態					
法案	×	×	×	×	×	案	×	×			
自治体	×	×	×	×	×	団体					
パソコン						コンピュータ					
運営						進めること					
適用	×	×	×	×		使うこと					

次のページに続く

前のページからの続き

難解語	評価者 1	2	3	4	5	平易語	評価者 1	2	3	4	5
謝罪	×		×			あやまること					
支払い						わたすこと					
関与	×	×	×	×	×	関係すること					
掲げる	×	×	×	×	×	上げる					
上演						見せること					
採択	×	×	×	×	×	とること					
公判	×	×	×	×	×	裁判					
医師						医者					
遺族	×		×			家族や親戚					
維持	×	×	×	×	×	続けること	×				
異例	×	×	×	×		異常	×		×		
トラブル						災い	×		×		
措置	×	×	×	×	×	対策					
総会	×	×	×	×	×	会					
準決勝						試合					
更新	×	×	×			改めること	×		×		
トップ						一番					
一連	×			×		ひと続き					×
後任	×	×	×	×		つぐこと	×				
降す	×		×			倒す	×				
認定			×			認めること					
実態	×	×			×	実際					
出火		×	×			災害					
統合	×	×		×		まとめること					
了承	×			×		承知すること	×		×		
若者	×					若い人					
浮上	×	×	×	×		上がること					
企画	×	×	×		×	計画すること					
出演						出ること					
刊行	×	×	×	×	×	出すこと					
臨む	×	×	×	×	×	出る					
双方	×	×	×	×		両方				×	
帰国						帰ること					
見解	×	×	×	×	×	見方や考え方					
声明	×	×	×	×	×	発表すること					
マンション						高い建物					
現地						現場					
当初		×	×			はじめ					

これで終わり



表 7.5: 換言前後の文の難易度の評価 (換言前の難解文)

難解文	評価者 1	2	3	4	5
正しいレコード会社のあり方である。	×	×		×	×
息子のサッカーの夏の大会が終わった。					
だがその後の展開は少し違った。					
大竹、トップに並ぶ。				×	
それによると、日本各地で大きなトラブルは報告されていない。		×			
ずらりと並んだ丸テーブルにパソコン。					
とても釣りに出かける余裕などなかったので、釣り関連のホームページをのぞいて、紹介してみることにする。		×			
しかし、地方組織の意向を無視することは現実には難しそうだ。		×	×	×	
異例の首相交代劇。	×	×		×	
今年は遺族も足を運び、そっと手を合わせた。		×			
しかし、家族と医師の介護で、次第に元気を取り戻し、最近は「早く元気になって友達と会いたい」と話している。		×			
これで中山事件をめぐる一連の裁判はすべて終了した。		×		×	
5日午後、民主党の鳩山由紀夫代表は会見で訴えた。				×	
会合は毎月開かれるが、地震活動に応じて臨時に開くこともある。	×			×	
十数人が関与か。	×	×	×	×	×
法律上、起訴と同じ意味がある。	×	×	×	×	
協議は26日も行われる。					
反ナチス記事、懸念。	×	×	×	×	×
後任は決まっていない。	×	×		×	
今後1年間かけ交渉を行う。				×	
初公判は今月下旬に開かれる予定。		×	×	×	
今国会中に成立する見通し。		×	×	×	
日本国内で行われる主な試合は次の通り。					
さあて、今回は何を書こうかな。					
その際、巨人の許可は必要としない。					×
娘は昨年、大学を卒業して神戸で就職した。					
死去による政治的影響はないとみられる。				×	×
主な事業計画は次の通り。			×	×	×
他の自治体にも参考になりそうだ。		×	×	×	×
しかし、人が家の中だけでなく自宅前や公園を清掃することが「変な行動」だろうか。					
辞任の理由は発表されていない。				×	×
実態は数字以上に深刻との指摘もある。		×		×	×
「若者にとってチャンスの多い国」へ、この秋留学する。		×			
出火当時、家族6人がいたが、逃げ出して無事だった。					
中園さん準決勝へ。				×	
男は近くに止めていた黒っぽい乗用車で逃げた。					
そんな政権は日本にはいない。	×	×			

次のページに続く

前のページからの続き

難解文	評価者 1	2	3	4	5
学校側がとった措置については「問題ない」としている。	×	×	×	×	×
それが今、どうなっているのかを探るのが狙いです。				×	
2年、日本で - - 世界に市場拡大狙う。	×	×		×	×
お別れ会の喪主は長男俊樹（としき）さん。	×	×	×	×	
相次ぐ不思議な事件が「失われた1年」に集中している。		×			
15日の総会で報告される。		×	×	×	
21世紀の葬儀はこんなイメージになるのだろうか。		×		×	
中国が他の国や地域と自由貿易圏構想を打ち出すのは初めてだ。	×	×	×	×	×
地元の取り組みを追った。					
代理店は8万店に統合。		×		×	×
これまで同様に健康被害などの連絡はないという。		×			×
派遣選手は次の通り。	×	×	×	×	×
それが伝わったのか母は肺炎を患い、入院することになった。					
1年を超えることがないように結論を出したい」と述べ、近く発足を予定している同会議で1年以内に結論を得たいとの考えを示した。	×	×	×	×	×
政治的発言は極力控えたが、米国の首都でのチベット大集会に対し、中国の反発が予想される。		×		×	×
私は戦争の全責任を負う。			×		
予算関連37法案のうち、36法案は今月中旬までに成立する見通し。		×		×	×
シアトルの本社で一部外国人記者団に語った。		×		×	
しかし、「今後も授業妨害などを起こせば出席停止を命じる可能性はある」という。				×	
正しいのはどっち？」という問い合わせや苦情が増えている。				×	
野党は全員が賛成。		×		×	
2人は容疑を認めている。				×	
中国は、日本政府が「反省」の態度を維持していることに一定の評価をしている模様だ。	×	×		×	×
私が子供のころ、両親は東京都渋谷区幡ヶ谷で「労働クラブ」を運営していた。		×			
全国43会場で開催。					
この日開幕予定だった富山、石川は雨天中止となった。				×	
勝てば「銀」以上確定。					
「泣いてうるさかったので、カッとなって殴ってしまった」などと供述しているという。		×	×		
この投書が掲載されるころには2年を迎えているのだろうか。	×	×	×	×	
結果は1、2週間以内に正式に公表される。				×	
パ・リーグ記録を更新する278万6人の入場者を集めた。		×		×	
日本は必要性を強調し、合意しなかった。		×		×	

次のページに続く

前のページからの続き

難解文	評価者 1	2	3	4	5
このため再検討会議はこの部分を詰め、19日午後にも最終文書を採択するのはほぼ確実にした。		×	×		×
母を殺害した少年は、なぜ北へ向かったのか。					
さらに東ティモールを支援していくことなども確認した。					
1人9万～22万円支払い。					
「深く謝罪します」。				×	
「鐘巻」が大阪で上演されるのは初めて。				×	
28日にも申請する。	×			×	×
来年初めに資本金約3億ウォン(約3億円)で新会社を設立。		×			
キューバ、台湾と対戦。					
近く逮捕する。					
14日から適用する。	×			×	
交通事故は、どちらにどれほどの責任があるのか、事故の後に正しく認定するのはとても難しい。		×		×	
原因が判明するまで同工場での製造を中止した。					×
7日以内の往復に有効で2万9円で販売する。					
近く正式に表明する。					
ご了承ください。		×	×		
よく似た顔が、ほお寄せ合って熱い涙を流す。	×	×			
政府は今年、一つの「旗」を高々と掲げた。		×	×	×	
決勝はヤン・ファンアイデン(ドイツ)がローラン・ガネ(フランス)を2で降し、初優勝した。			×	×	
それぞれ現地の見方をまとめた。		×			
「今の所より安いマンションに移るんです。					
『ザ・ベストミステリーズ2』(講談社・28円)が刊行された。	×	×	×	×	
沖縄訪問も報告会も、メンバーがーから企画したものだ。		×		×	
大統領は東京に約8時間滞在して帰国する。					
あと2、3年かかる」という見解を示した。		×	×		
たまに、日本の映画に出演する。					
日本が中心となって声明をまとめた。	×	×	×	×	
双方の運転手にけがはなかった。	×			×	
作業開始は当初の予定では18日前後だった。				×	
大分は1で大宮を退け、3位に浮上。				×	
「今年は固定メンバーで臨む」と宣言。			×	×	

これで終わり

表 7.6: 換言前後の文の難易度の評価 (換言後の平易文)

平易文	評価者 1	2	3	4	5
正しいレコード会社の状態である。		×		×	×
息子の競技の夏の大会が終わった。					
だがあとの展開は少し違った。					
大竹、一番に並ぶ。				×	
それによると、日本各地で大きな災いは報告されていない。		×			
ずらりと並んだ丸テーブルにコンピュータ					
とても釣りに出かける余裕などなかったので、釣り関連のページをのぞいて、紹介してみることにする。		×			
しかし、地方組織の考えや気持ちを無視することは現実には難しそうだ。		×		×	×
異常の首相交代劇。	×	×		×	×
今年は家族や親戚も足を運び、そっと手を合わせた。					
しかし、家族と医者との介護で、次第に元気を取り戻し、最近は「早く元気になって友達と会いたい」と話している。		×			
これで中山事件をめぐるひと続きの裁判はすべて終了した。		×		×	×
5日午後、民主党の鳩山由紀夫代表は会うことで訴えた。				×	×
集まりは毎月開かれるが、地震活動に応じて臨時に開くこともある。				×	
十数人が関係することか。					
法律上、うったえることと同じ意味がある。		×			
決めることは26日も行われる。					×
反ナチス記事、恐れる。	×	×	×	×	
つくことは決まっていない。		×			
今後1年間かけ話し合うことを行う。				×	×
初裁判は今月下旬に開かれる予定。					
今議会中に成立する見通し。		×		×	
日本中で行われる主な試合は次の通り。					×
さあて、今度は何を書こうかな。					
そのとき、巨人の許可は必要としない。					×
娘は去年、大学を卒業して神戸で就職した。					
死ぬことによる政治的影響はないとみられる。					
主な仕事計画は次の通り。				×	
他の団体にも参考になりそうだ。		×		×	
しかし、人が家の中だけでなく家前や公園を清掃することが「変な行動」だろうか。			×		
やめることの理由は発表されていない。					
実際は数字以上に深刻との指摘もある。		×			×
「若い人にとってチャンスの多い国」へ、この秋留学する。		×			
災害当時、家族6人がいたが、逃げ出して無事だった。					
中園さん試合へ。				×	

次のページに続く

前のページからの続き

平易文	評価者 1	2	3	4	5
男は近くに止めていた黒っぽい自動車で逃げた。					
そんな権力は日本にはいない。	×				
学校側がとった対策については「問題ない」としている。		×			
それが今、どうなっているのかを探るのが目的です。					
2年、日本で - - 世界に市場拡大うかがう。	×	×	×		
お別れ会の主人は長男俊樹(としき)さん。					
続く不思議な事件が「失われた1年」に集中している。		×			
15日の会で報告される。					
21世紀の儀式はこんなイメージになるのだろうか。		×		×	
中国が他の国や地域と自由貿易圏構想を表すのは初めてだ。	×	×		×	
地域の取り組みを追った。				×	
代理店は8万店にまとめる。		×		×	
これまで施設に健康被害などの連絡はないという。		×			
差し出す選手は次の通り。		×			
それが伝わったのか母は病気を患い、入院することになった。				×	
1年を超えることがないように結論を出したい」と述べ、近く始めることを予定している同会議で1年以内に結論を得たいとの考えを示した。		×		×	×
政治的発言は極力控えたが、米国の首都でのチベット大集会に対し、中国の逆らうことが予想される。		×	×	×	×
私は戦争の全責任を受ける。					
予算関連37案のうち、36案は今月中旬までに成立する見通し。		×			×
シアトルの会社で一部外国人記者団に語った。		×		×	
しかし、「今後も授業妨害などを起こせば出席停止を命令する可能性はある」という。					
正しいのはどっち?」という確かめることや苦情が増えている。			×		
政党は全員が賛成。		×			
2人は疑うことを認めている。			×		
中国は、日本政府が「反省」の態度を続けていることに一定の評価をしている模様だ。		×			×
私が子供のころ、両親は東京都渋谷区幡ヶ谷で「労働クラブ」を進めていた。		×	×	×	
全国43会場で開く。					
この日始まる予定だった富山、石川は雨天中止となった。					
勝てば「銀」以上決まる。					
「泣いてうるさかったので、カッとなって殴ってしまった」などと述べているという。		×			
この投書がのせられるころには2年を迎えているのだろうか。	×	×		×	
結果は1、2週間以内に正式に知らせられる。					
パ・リーグ記録を改める278万6人の入場者を集めた。		×			

次のページに続く

前のページからの続き

平易文	評価者 1	2	3	4	5
日本は必要性を強調し、合わなかった。		×	×		
このため再検討会議はこの部分を詰め、19日午後にも最終文書をとるのはほぼ確実にした。		×		×	×
母を殺した少年は、なぜ北へ向かったのか。					
さらに東ティモールを助けていくことなども確認した。					
1人9万~22万円わたす。					
「深くあやまります」。					
「鐘巻」が大阪で見せられるのは初めて。				×	
28日にも言う。					
来年初めに資本金約3億ウオン(約3億円)で新会社をつくる。		×			
キューバ、台湾と戦う。					
近くつかまえる。					
14日から使う。					
交通事故は、どちらにどれほどの責任があるのか、事故の後に正しく認めるのはとても難しい。		×			
原因がわかるまで同工場での製造を中止した。					
7日以内の往復に有効で2万9円です。					
近く正式に表す。					
ご承知ください。		×			
よく似た顔が、ほお近づけ合って熱い涙を流す。	×	×			
政府は今年、一つの「旗」を高々と上げた。		×			
決勝はヤン・ファンアイデン(ドイツ)がローラン・ガネ(フランス)を2で倒し、初優勝した。					
それぞれ現場の見方をまとめた。		×		×	
「今の所より安い高い建物に移るんです。			×	×	
『ザ・ベストミステリーズ2』(講談社・28円)が出された。		×			
沖縄訪問も報告会も、メンバーが一から計画したものだ。		×		×	
大統領は東京に約8時間滞在して帰る。				×	
あと2、3年かかる」という見方や考え方を示した。		×			
たまに、日本の映画に出る。					
日本が中心となって発表することをまとめた。		×			
両方の運転手にけがはなかった。					
作業開始ははじめの予定では18日前後だった。					
大分は1で大宮を退け、3位に上がる。					
「今年は固定メンバーで出る」と宣言。					

これで終わり

表 7.7: 知識に基づく語彙的換言の例

難解語	平易語	難解文 平易文
表明	あらわす	近く正式に{表明する あらわす}。
表明	表す	近く正式に表{明する す}。
表明	しめす	近く正式に{表明する しめす}。
合意	一致	日本は必要性を強調し、{合意 一致}しなかった。
開催	始める	毎年春に{開催 始める}。
開催	ひらく	毎年春に{開催 ひらく}。
開催	開く	毎年春に開{催 く}。
設置	組織	同病院は近く調査委員会を{設置 組織}し調べる。
設置	持ち	同病院は近く調査委員会を{設置し 持ち}調べる。
設置	置き	同病院は近く調査委員会を{設置し 置き}調べる。
設置	おき	同病院は近く調査委員会を{設置し おき}調べる。
設置	作り	同病院は近く調査委員会を{設置し 作り}調べる。
設置	つくり	同病院は近く調査委員会を{設置し つくり}調べる。
設置	定め	同病院は近く調査委員会を{設置し 定め}調べる。
設置	用意	同病院は近く調査委員会を{設置 用意}し調べる。
調べ	調査	調{べ 査}では、首の下から腹にかけての部分で、顔はなかった。
導入	入れ	インターネット{導入 入れ} 教育はどう変わるか。
会見	話	きのうの{会見 話}の前、首相は何時間考えたのだろう。
会見	面会	きのうの{会見 面会}の前、首相は何時間考えたのだろう。
捜査	調べ	「{捜査 調べ}は続ける」としている。
捜査	探り	「{捜査 探り}は続ける」としている。
判決	判断	この事件での判{決 断}は初めて。
判決	決定	この事件での{判決 決定}は初めて。
盛り込む	入れる	農家以外の地域住民の意見を聞く制度も{盛り込む 入れる}方針。
支援	助け	政党の{支援 助け}なんて必要ない。
多く	たくさん	こんなに{多く たくさん}の方々に出席していただき、大変うれしく思います。
取引	売買	消費者が{取引 売買}する企業を選ぶための判断材料にしてもらう。
負う	背負う	私は戦争の全責任を{背}負う。
負う	引き受ける	私は戦争の全責任を{負う 引き受ける}。
負う	ひきうける	私は戦争の全責任を{負う ひきうける}。
負う	せおう	私は戦争の全責任を{負 せお}う。
負う	受ける	私は戦争の全責任を{負う 受ける}。
獲得	入れ	優秀な学生を{獲得し 入れ}たい。
狙う	目指す	「ポスト森」{狙う 目指す}動きにイライラ!?
狙う	向かう	「ポスト森」{狙 向か}う動きにイライラ!?
狙う	めざす	「ポスト森」{狙う めざす}動きにイライラ!?

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
反発	逆らい	{反発 逆らい}を招くことが一番だ。
反発	反抗	反{発 抗}を招くことが一番だ。
サッカー	競技	{サッカー 競技}くじは果たして売れるのか。
サッカー	スポーツ	{サッカー スポーツ}くじは果たして売れるのか。
狙い	目的	{狙い 目的}は何なのか。
狙い	目当て	{狙い 目当て}は何なのか。
狙い	目標	{狙い 目標}は何なのか。
交渉	話し合い	今後、製作者と{交渉 話し合い}に入る。
交渉	話し	今後、製作者と{交渉 話し}に入る。
子ども	子	いまの子{ども }たちは幸せだ。
子ども	子供	いまの子{ども 供}たちは幸せだ。
辞任	退き	また「最後まで責任を果たす」と{辞任 退き}の考えはないことを強調した。
辞任	やめ	また「最後まで責任を果たす」と{辞任 やめ}の考えはないことを強調した。
地元	地域	この不安は地{元 域}では大きく、解決はこれからです。
景気	商売	いったい日本の{景気 商売}はいいのか、悪いのか。
景気	経済	いったい日本の{景気 経済}はいいのか、悪いのか。
事業	仕事	頭の中はこの町でスタートを切る新しい{事業 仕事}のことでいっぱいだったのだ。
あり方	状態	正しいレコード会社の{あり方 状態}である。
異例	特別	少年事件の調査は{異例 特別}という。
開設	始め	相談電話を{開設し 始め}たり、学校の卒業式・入学式での強制に反対するなどの活動に取り組む。
謝罪	わび	深く{謝罪し わび}ます。
謝罪	あやまり	深く{謝罪し あやまり}ます。
謝罪	詫び	深く{謝罪し 詫び}ます。
継続	持続	話し合いを{継 持}続していくということだ。
継続	続け	話し合いを{継続し 続け}ていくということだ。
主婦	女性	私は{主婦 女性}です。
主婦	女	私は{主婦 女}です。
主婦	つま	私は{主婦 つま}です。
企画	づくり	紙面の{企画 づくり} 取材、写真撮影など。
企画	計画	紙面の{企 計}画、取材、写真撮影など。
所属	入る	{所属す 入}るチームは未定。
所属	加わる	{所属す 加わ}るチームは未定。
出演	出る	たまに、日本の映画に出{演す }る。
臨む	行く	「今年は固定メンバーで{臨む 行く}」と宣言。
臨む	あたる	「今年は固定メンバーで{臨む あたる}」と宣言。
臨む	出る	「今年は固定メンバーで{臨む 出る}」と宣言。
現地	現場	まず現{地 場}の状況を的確に知ること。

次のページに続く



前のページからの続き

難解語	平易語	難解文 平易文
棄却	しりぞけ	国への請求は{棄却し しりぞけ}た。
棄却	捨て	国への請求は{棄却し 捨て}た。
棄却	すて	国への請求は{棄却し すて}た。
証言	証明	「適切でなかった」と証{言 明}した。
ミス	失敗	作業に{ミス 失敗}はなかったと説明している。
ファン	者	数はともかく、これが魅力で、と通う{ファン 者}が多い。
現行	現在	補助水準は現{行 在}と本来の家賃の中間となる。
アピール	うったえる	透明性を{アピール うったえる}。
株式	取り引き	普通の{株式 取り引き}とどう違うのですか。
出来る	できる	私のこの体で、この年齢で世話が{出来 でき}るだろうか。
促す	急ぐ	特に夫の理解を{促す 急ぐ}ことが、患者の負担を和らげるとい う。
促す	すすめる	特に夫の理解を{促す すずめる}ことが、患者の負担を和らげる という。
逃走	にげだし	男は車で{逃走 にげだ}した。
逃走	逃げ出し	男は車で逃{走 げ出}した。
逃走	逃げ	男は車で逃{走し げ}た。
規制	制限	海外の{規制 制限}や日本企業の取り組みも。
サミット	会議	きょうからはじまる{サミット 会議}。
サミット	会談	きょうからはじまる{サミット 会談}。
バス	自動車	電車や{バス 自動車}でもそうだった。
バス	乗り物	電車や{バス 乗り物}でもそうだった。
通報	通知	健康被害の通{報 知}はないという。
通報	知らせ	健康被害の{通報 知らせ}はないという。
機能	作用	豊富な{機能 作用}つき。
データ	資料	{データ 資料}にはっきりあるわけではないが、実感だ。
データ	情報	{データ 情報}にはっきりあるわけではないが、実感だ。
データ	材料	{データ 材料}にはっきりあるわけではないが、実感だ。
データ	事実	{データ 事実}にはっきりあるわけではないが、実感だ。
プロ	専門	こういうときは{プロ 専門}に聞くに限る。
ボランティア	奉仕	十数年前に初めて、{ボランティア 奉仕}をしました。
移植	内蔵	{移植 内蔵}を待つ患者にその可能性が広がる。
移植	治療	{移植 治療}を待つ患者にその可能性が広がる。
警戒	備え	{警戒 備え}は厳重、ピリピリしている。
警戒	注意	{警戒 注意}は厳重、ピリピリしている。
警戒	用心	{警戒 用心}は厳重、ピリピリしている。
検証	調べ	時代も大きく変わる中、三つの疑問を{検証し 調べ}た。
解消	けし	そういう悩みを{解消 け}してくれるのがこの本だ。
解消	なくし	そういう悩みを{解消 なく}してくれるのがこの本だ。
同行	行き	担当記者も{同行し 行き}ました。
デザイン	設計	表紙の{デザイン 設計}がとて美しい。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
デザイン	つくり	表紙の{デザイン つくり}がとてもいい。
スーパー	店	駅の近くの{スーパー 店}でのことです。
受け付ける	受け入れる	はがきとインターネットで受け{付け 入れ}る。
受け付ける	受ける	はがきとインターネットで受け{付 }ける。
レース	競争	運も多分に作用しそうな{レース 競争}だ。
レース	競技	運も多分に作用しそうな{レース 競技}だ。
逆転	ひっくりかえる	最後の最後に{逆転 ひっくりかえる}。
知人	知り合い	海外に住む家族や友人・知{人 り合い}を訪ねる人も少なくない。
経緯	事情	講演から、国との闘いの{経緯 事情}をまとめた。
革命	改革	{革命 改革}ではない。
革命	変化	{革命 変化}ではない。
コメント	述べる	結果が出るまで{コメントす 述べる}べきではない。
コメント	説明	結果が出るまで{コメント 説明}すべきではない。
コメント	のべる	結果が出るまで{コメントす のべる}べきではない。
コメント	意見	結果が出るまで{コメント 意見}すべきではない。
コメント	解説	結果が出るまで{コメント 解説}すべきではない。
コメント	評論	結果が出るまで{コメント 評論}すべきではない。
賠償	損害	{賠償 損害}の請求額は明らかにしなかった。
賠償	つぐない	{賠償 つぐない}の請求額は明らかにしなかった。
ドラマ	演劇	今度はどんな{ドラマ 演劇}が生まれるのだろうか。
ドラマ	作品	今度はどんな{ドラマ 作品}が生まれるのだろうか。
ドラマ	劇	今度はどんな{ドラマ 劇}が生まれるのだろうか。
ドラマ	番組	今度はどんな{ドラマ 番組}が生まれるのだろうか。
講師	先生	{講師 先生}と朗読作品は次の通り。
講師	教師	{講 教}師と朗読作品は次の通り。
痛み	苦痛	腕を振り下ろすと左の背中に{痛み 苦痛}が走る。
決断	決定	とはいえ、勇気ある決{断 定}だ。
小説	文学	読者が想像する部分大きいほどいい{小説 文学}だ。
小説	作品	読者が想像する部分大きいほどいい{小説 作品}だ。
小説	創作	読者が想像する部分大きいほどいい{小説 創作}だ。
小説	物語	読者が想像する部分大きいほどいい{小説 物語}だ。
ビル	建物	一方で「{ビル 建物}では寂しい」という声も。
入札	売買	来週にも{入札 売買}を実施する。
ポイント	点	主な{ポイント 点}をまとめた。
ポイント	箇所	主な{ポイント 箇所}をまとめた。
ポイント	場所	主な{ポイント 場所}をまとめた。
ポイント	部分	主な{ポイント 部分}をまとめた。
ポイント	要点	主な{ポイント 要点}をまとめた。
首脳	人物	{首脳 人物}と会談する。
男児	少年	{男児 少年}、母親とも状態は安定しているという。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
男児	男性	男{児 性} 母親とも状態は安定しているという。
男児	男	男{児 } 母親とも状態は安定しているという。
ホール	施設	温かい木の{ホール 施設}に合う、実に良いプログラムである。
ホール	へや	温かい木の{ホール へや}に合う、実に良いプログラムである。
ホール	場所	温かい木の{ホール 場所}に合う、実に良いプログラムである。
ホール	建物	温かい木の{ホール 建物}に合う、実に良いプログラムである。
移行	うつり	春から夏、夏から秋、秋から冬の{移行 うつり}はおとなしい。
金利	額	なぜ{金利 額}を上げるかの回答は難しい。
金利	割合	なぜ{金利 割合}を上げるかの回答は難しい。
チェック	比べ	よく{チェックし 比べ}ましょう。
チェック	たしかめ	よく{チェックし たしかめ}ましょう。
チェック	確認	よく{チェック 確認}しましょう。
チェック	検査	よく{チェック 検査}しましょう。
プレゼント	おくる	ほかの子にも{プレゼントす おく}るから。
プレゼント	贈る	ほかの子にも{プレゼントす 贈}るから。
在り方	状態	中央と地方の関係の{在り方 状態}。
ネット	インターネット	今のところ、{ インター } ネットでできるのは申し込みだけ。
早急	急	{早 } 急に総選挙。
審理	判断	{審理 判断}に影響なし。
審理	調べ	{審理 調べ}に影響なし。
促進	はかどる	元からの傾向を{促進 はかどる}。
促進	高める	元からの傾向を{促進 高める}。
促進	推進	元からの傾向を{促 推}進。
虐待	いじめ	どんな大人が{虐待し いじめ}ているのか。
相撲	競技	{相撲 競技}の世界は勝負の世界。
相撲	勝負	{相撲 勝負}の世界は勝負の世界。
伸び	あくび	緑のなかで、木が{伸 あく}びをしている。
概要	あらまし	インタビューの{概要 あらまし}は次の通り。
思惑	考え	当てるには、あなたの見方が正しいかどうかでなく、他人がどの株を買うか売るか{思惑 考え}による。
思惑	見込み	当てるには、あなたの見方が正しいかどうかでなく、他人がどの株を買うか売るか{思惑 見込み}による。
思惑	期待	当てるには、あなたの見方が正しいかどうかでなく、他人がどの株を買うか売るか{思惑 期待}による。
思惑	予想	当てるには、あなたの見方が正しいかどうかでなく、他人がどの株を買うか売るか{思惑 予想}による。
釈明	説明	「重要書類の管理に関する認識が甘かった」と{釈 説}明した。
釈明	言い訳	「重要書類の管理に関する認識が甘かった」と{釈明 言い訳}した。
勤務	つとめる	それも父親の{勤務す つとめ}る学校だ。
勤務	勤める	それも父親の勤{務す め}る学校だ。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
勤務	仕事	それも父親の{勤務 仕事}する学校だ。
傷つける	こわす	暴力は人の心を{傷つける こわす}。
火災	災害	{火災 災害}の危険性指摘、早くから。
火災	火事	火{災 事}の危険性指摘、早くから。
プレー	競技	正々堂々と{プレー 競技}してほしい。
プレー	試合	正々堂々と{プレー 試合}してほしい。
財政	金	渋滞、{財政 金}に不安。
財政	経済	渋滞、{財政 経済}に不安。
監視	つけ	{監視し つけ}ていると告げる。
監視	見	{監視し 見}ていると告げる。
素材	もと	{素材 もと}にとことんこだわる。
素材	材料	{素材 材料}にとことんこだわる。
収益	利益	{収 利}益は寄付。
ヒト	人物	ただ、私は形から入る{ヒト 人物}。
ヒト	性格	ただ、私は形から入る{ヒト 性格}。
ヒト	人間	ただ、私は形から入る{ヒト 人間}。
ヒト	人類	ただ、私は形から入る{ヒト 人類}。
スタッフ	人々	監督、選手、{スタッフ 人々}などはこれまで通り。
スタッフ	係	監督、選手、{スタッフ 係}などはこれまで通り。
スタッフ	担当	監督、選手、{スタッフ 担当}などはこれまで通り。
ネットワーク	つながり	テーマは「伝統」・「{ネットワーク つながり}」。
ネットワーク	通信	テーマは「伝統」・「{ネットワーク 通信}」。
兆し	しるし	かつてない変化の{兆 しる}し。
ブーム	流行	「おいしい水」が{ブーム 流行}である。
評判	うわさ	「まじめで成績優秀」と、近所でも{評判 うわさ}のよい少年だったという。
評判	評価	「まじめで成績優秀」と、近所でも評{判 価}のよい少年だったという。
メス	女	{メス 女}は卵を採るため。
ビジネス	商売	でも、{ビジネス 商売}はやはり人と人とのつながりの中から生まれる。
ビジネス	仕事	でも、{ビジネス 仕事}はやはり人と人とのつながりの中から生まれる。
後退	下がる	白、少しずつ{後退 下がる}。
後退	退く	白、少しずつ{後退 退く}。
後退	しりぞく	白、少しずつ{後退 しりぞく}。
前期	前	前{期}と同じカード。
日取り	日	お別れ会を行うが、日{取り}は未定。
幼稚園	施設	さかのぼると、{幼稚園 施設}の遠足になる。
機関	団体	こうした{機関 団体}は全国でも珍しいという。
機関	はたらき	こうした{機関 はたらき}は全国でも珍しいという。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
機関	組織	こうした{機関 組織}は全国でも珍しいという。
対処	対応	世界はそれにどう対{処 応}するのか。
対処	行動	世界はそれにどう{対処 行動}するのか。
対処	とりくむ	世界はそれにどう{対処する とりくむ}のか。
対処	処理	世界はそれにどう{対処 処理}するのか。
構図	位置	特に感心したのは{構図 位置}。
構図	構成	特に感心したのは構{図 成}。
一言	言葉	私はつい、夢のない{一言 言葉}。
騒ぎ	騒動	それが騒{ぎ 動}を大きくした。
騒ぎ	面倒	それが{騒ぎ 面倒}を大きくした。
騒ぎ	問題	それが{騒ぎ 問題}を大きくした。
差別	区別	僕は{差 区}別を感じたことがない。
ペース	速度	計画の倍の{ペース 速度}。
ペース	勢い	計画の倍の{ペース 勢い}。
損失	損害	歴史の限らない損{失 害}である。
清算	処分	過去は既に{清算 処分}したのだろう。
トイレ	便所	すでに使用中の{トイレ 便所}に後から加工する方法もある。
数値	数	社会貢献を数{値 }で評価する。
数値	数字	社会貢献を数{値 字}で評価する。
大詰め	最後	{大詰め 最後}の一番。
一転	変化	{一転 変化}して神経戦。
側面	方面	スポーツの世界は、いつもこんな{側 方}面がある。
側面	表面	スポーツの世界は、いつもこんな{側 表}面がある。
側面	性質	スポーツの世界は、いつもこんな{側面 性質}がある。
レベル	能力	{レベル 能力}はあまり高くない。
レベル	水準	{レベル 水準}はあまり高くない。
レベル	評価	{レベル 評価}はあまり高くない。
イラスト	図	{イラスト 図}で未来世界を表現。
イラスト	挿絵	{イラスト 挿絵}で未来世界を表現。
イラスト	絵	{イラスト 絵}で未来世界を表現。
イラスト	図解	{イラスト 図解}で未来世界を表現。
悲劇	出来事	私にとっては大きな{悲劇 出来事}だ。
悲劇	不幸	私にとっては大きな{悲劇 不幸}だ。
メニュー	表	食事と飲み物の{メニュー 表}は次の通り。
メニュー	献立	食事と飲み物の{メニュー 献立}は次の通り。
メニュー	項目	食事と飲み物の{メニュー 項目}は次の通り。
パーティー	仲間	先輩議員の{パーティー 仲間}にも拡大したい考えだ。
パーティー	会	先輩議員の{パーティー 会}にも拡大したい考えだ。
パーティー	一行	先輩議員の{パーティー 一行}にも拡大したい考えだ。
パーティー	集まり	先輩議員の{パーティー 集まり}にも拡大したい考えだ。
パーティー	集団	先輩議員の{パーティー 集団}にも拡大したい考えだ。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
議席	席	より少数政党が{議}席を得にくい仕組みだ。
柔軟	しなやか	その{柔軟 しなやか}な頭なら、「その日」はまだまだ遠いでしょうけど。
執行	行い	早ければ今月中にも{執行 行い}を決める。
執行	実行	早ければ今月中にも{執 実}行を決める。
執行	おこない	早ければ今月中にも{執行 おこない}を決める。
後押し	おし	家族の{後押 お}し武器に。
援助	助ける	なぜそんな国に{援助す 助け}る必要があるのか。
援助	救う	なぜそんな国に{援助する 救う}必要があるのか。
援助	たすける	なぜそんな国に{援助す たすけ}る必要があるのか。
資質	性質	また首相の{資 性}質の問題だ。
資質	才能	また首相の{資質 才能}の問題だ。
波乱	変化	その{波乱 変化}の生涯を描く。
功績	利益	自動車関連で今年最大の{功績 利益}。
功績	はたらき	自動車関連で今年最大の{功績 はたらき}。
功績	仕事	自動車関連で今年最大の{功績 仕事}。
冒頭	初め	{冒頭 初め}いきなり次のような場面からはじまる。
軌跡	跡	少年の{軌}跡を追う。
軌跡	あと	少年の{軌跡 あと}を追う。
驚き	びっくり	喜びよりも{驚き びっくり}の方が大きい。
レポート	報告	全国の大学の落語研究会が参加した大会の様態を{レポート 報告}する。
レポート	放送	全国の大学の落語研究会が参加した大会の様態を{レポート 放送}する。
レポート	レポート	全国の大学の落語研究会が参加した大会の様態を{リ レ}ポートする。
顔ぶれ	人々	すごい{顔ぶれ 人々} ありがたい。
顔ぶれ	ひと	すごい{顔ぶれ ひと} ありがたい。
顔ぶれ	メンバー	すごい{顔ぶれ メンバー} ありがたい。
顔ぶれ	集まり	すごい{顔ぶれ 集まり} ありがたい。
後続	続き	{後続 続き}に影響はなかった。
ママ	お母さん	笑顔の{ママ お母さん}が一番好き！
ママ	おかあさん	笑顔の{ママ おかあさん}が一番好き！
ママ	母親	笑顔の{ママ 母親}が一番好き！
コスト	費用	どう{コスト 費用}を下げるかの視点が必要だ。
コスト	値段	どう{コスト 値段}を下げるかの視点が必要だ。
実情	事情	現場から{実 事}情を報告する。
実情	状況	現場から{実情 状況}を報告する。
作戦	計画	「懐かしさ」で大人の気もひく{作戦 計画}だ。
作戦	方法	「懐かしさ」で大人の気もひく{作戦 方法}だ。
フリー	自由	「組織」から{フリー 自由}に。

次のページに続く

前のページからの続き

難解語	平易語	難解文 平易文
骨子	中心	宣言の{骨子 中心}。
骨子	要点	宣言の{骨子 要点}。
国籍	国	国{籍 }や民族の違いは関係ない。
リスク	危険	つまり{リスク 危険}が大きい。
神様	神	神{様 }がこうつぶやく。
温泉	ゆ	寒くなると、{温泉 ゆ}が恋しくなる。
温泉	湯	寒くなると、{温泉 湯}が恋しくなる。
強風	風	{強 }風のいたずらだ。
受け付け	承認	インターネットまたは署名用紙で{受け付け 承認}。
サポート	支える	期間中、全国の子供の生活を{サポートす 支え}る教育分野などで活動する。
サポート	助ける	期間中、全国の子供の生活を{サポートす 助け}る教育分野などで活動する。
遺書	手紙	{遺書 手紙}はなかった。
遺書	文書	{遺 文}書はなかった。
遺書	書類	{遺書 書類}はなかった。
確信	信じ	「自分たちが進む道はこれだ」と{確信し 信じ}た。
破損	こわれる	ガス管や水道管が{破損す こわれ}ることもある。
半面	半分	ほっとした半{面 分} どこか寂しい。
シンボル	象徴	科学の勝利の{シンボル 象徴}のはずだった。
シンボル	しるし	科学の勝利の{シンボル しるし}のはずだった。

これで終わり

## 付録B：4章「用例に基づく語彙的換言」の実験結果の詳細

- 用例に基づく語彙的換言の例

表7.8に、用例に基づく語彙的換言の例を示す。各入力文中で、換言対象語を1語ずつ括弧で囲んで示す。これらの対象語について、提案手法で文脈の類似度を計算し、類似度の高い上位10語を換言候補として掲載する。また、3人の評価者がこの10語の換言候補の中から適切な換言であると選択した語を1語ずつ掲載する。「-」は、その10語の候補中に適切な換言がないと評価者が判断したことを示す。



表 7.8: 用例に基づく語彙的換言の例

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
すべての (機能) をご利用いただく	サービス	サービス	サービス
換言候補: (01) システム (02) サービス	(03) 環境	(04) 情報	(05) 方法
換言候補: (06) 製品 (07) 内容	(08) 商品	(09) 施設	(10) 事業
特定商取引に関する (法律) に基づく	法	法	法
換言候補: (01) 法 (02) 情報	(03) 内容	(04) 法令	(05) 設定
換言候補: (06) 表示 (07) 規約	(08) 事項	(09) 通信	(10) 法規
法等の (法令) に違反する	法律	基準	法律
換言候補: (01) 法律 (02) 法	(03) 基準	(04) 規定	(05) 措置
換言候補: (06) 内容 (07) 規制	(08) ルール	(09) 等	(10) 事項
オーナーの (承認) が必要になる	許可	許可	許可
換言候補: (01) 許可 (02) 登録	(03) 確認	(04) 更新	(05) 使用
換言候補: (06) 認証 (07) 情報	(08) サービス	(09) 設定	(10) 支援
以上の (評価) を受けている	-	-	-
換言候補: (01) 活動 (02) 教育	(03) 事業	(04) 情報	(05) サービス
換言候補: (06) 調査 (07) 制度	(08) 管理	(09) 処理	(10) 企業
記入された (キーワード) を含む	項目	情報	情報
換言候補: (01) 情報 (02) 項目	(03) 物	(04) 文字	(05) データ
換言候補: (06) 内容 (07) 写真	(08) メール	(09) 名前	(10) 部分
浴衣にも (洋服) にも合う	-	普段着	ドレス
換言候補: (01) ドレス (02) 着物	(03) ジーンズ	(04) 水着	(05) 女性
換言候補: (06) 系 (07) 和	(08) 普段着	(09) カジュアル	(10) デニム
複数の (ジャンル) に属している	分野	分野	分野
換言候補: (01) 分野 (02) 中	(03) 日本	(04) 地域	(05) サイト
換言候補: (06) 世界 (07) 会社	(08) 企業	(09) タイプ	(10) 国
旅館・ホテルの (トップ) に戻る	ホームページ	-	サイト
換言候補: (01) 中 (02) サイト	(03) 方	(04) 部屋	(05) ホームページ
換言候補: (06) 情報 (07) スタッフ	(08) イベント	(09) 空	(10) 紹介
重要な (課題) として取り組んでいる	問題	問題	問題
換言候補: (01) 問題 (02) 活動	(03) 研究	(04) 事業	(05) 対策
換言候補: (06) 分野 (07) テーマ	(08) プロジェクト	(09) 学習	(10) 目標
生活環境の (改善) にもつながる	向上	回復	向上
換言候補: (01) 向上 (02) 回復	(03) 確保	(04) 実現	(05) 事
換言候補: (06) 問題 (07) 評価	(08) 維持	(09) 理解	(10) 創出
一つの (コンポーネント) となっている	システム	プログラム	データ
換言候補: (01) プログラム (02) システム	(03) データベース	(04) ファイル	(05) データ
換言候補: (06) ソフト (07) ツール	(08) コード	(09) プロセス	(10) 文書
普及している (ケース) もある	-	物	物
換言候補: (01) 物 (02) 中	(03) カード	(04) 商品	(05) パソコン
換言候補: (06) 製品 (07) 間	(08) 建物	(09) システム	(10) 電話
合致した (方法) で実践する	-	-	-
換言候補: (01) システム (02) 形	(03) デザイン	(04) 企業	(05) 本
換言候補: (06) 投資 (07) 個人	(08) 家庭	(09) 上	(10) 銘柄
十分な (確認) をしていただく	検査	検査	検査
換言候補: (01) 処理 (02) 管理	(03) 検査	(04) 対応	(05) サービス
換言候補: (06) 調査 (07) 対策	(08) 作業	(09) 調整	(10) 評価
法等の (一部) を改正する	-	-	-
換言候補: (01) 方法 (02) 制度	(03) 基準	(04) ルール	(05) 法律
換言候補: (06) 対象 (07) 部	(08) 規定	(09) 規格	(10) 等

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
万社以上の (実績) がある	登録	顧客	-
換言候補: (01) 企業 (02) 方 (03) ホームページ (04) 会員 (05) 広告			
換言候補: (06) 利益 (07) 登録 (08) 顧客 (09) 採用 (10) 不動産			
出品した (場合) に通知する	0:00:00	後	後
換言候補: (01) 後 (02) 0:00:00 (03) ため (04) ところ (05) 人			
換言候補: (06) 者 (07) 本人 (08) 様 (09) 所有者 (10) 加盟店			
良心的な (料金) を提供する	価格	価格	価格
換言候補: (01) 価格 (02) 費用 (03) サービス (04) 商品 (05) 内容			
換言候補: (06) システム (07) サイト (08) 製品 (09) 機能 (10) コース			
地方消費税の (対象) となる	-	-	-
換言候補: (01) 市町村 (02) 計算 (03) 金額 (04) 導入 (05) 負担			
換言候補: (06) 額 (07) 収入 (08) 課税 (09) 充実 (10) 合計			
サイトの (コンテンツ) をご利用いただく	情報	システム	情報
換言候補: (01) 情報 (02) サイト (03) システム (04) データ (05) サービス			
換言候補: (06) 商品 (07) プログラム (08) 製品 (09) 内容 (10) ホームページ			
重要な (位置) を占めている	部分	場所	場所
換言候補: (01) 場所 (02) 中 (03) 部分 (04) 方 (05) 上			
換言候補: (06) 部屋 (07) ポイント (08) 地域 (09) 人 (10) 中心			
著作権上の (問題) がある	課題	問題点	課題
換言候補: (01) 課題 (02) 問題点 (03) トラブル (04) 質問 (05) 保護			
換言候補: (06) 理由 (07) 観点 (08) 都合 (09) 制限 (10) 侵害			
ダウンロードした (データ) を再生する	ファイル	ファイル	情報
換言候補: (01) 情報 (02) ファイル (03) 画像 (04) 商品 (05) 写真			
換言候補: (06) 会社 (07) 関係 (08) 物 (09) コンテンツ (10) 記事			
ホットな (話題) も教えてくれる	情報	情報	情報
換言候補: (01) 内容 (02) 情報 (03) 人 (04) サイト (05) 方			
換言候補: (06) 技術 (07) 動き (08) 場所 (09) ところ (10) バス			
的な (意味) を持っている	-	-	内容
換言候補: (01) 0:00:00 (02) 結果 (03) 場合 (04) 間 (05) 度			
換言候補: (06) 自分 (07) 内容 (08) 物 (09) 事 (10) 関係			
他の (物件) を見てみる	-	商品	住宅
換言候補: (01) 商品 (02) 製品 (03) 施設 (04) 住宅 (05) サイト			
換言候補: (06) 車 (07) 会社 (08) サービス (09) 店 (10) 情報			
関連しない (記事) が含まれる	情報	情報	情報
換言候補: (01) サイト (02) 情報 (03) 内容 (04) 話 (05) コメント			
換言候補: (06) 人 (07) 国 (08) 処理 (09) 点 (10) 行為			
マニアの (地図) を使っている	-	-	-
換言候補: (01) 写真 (02) 情報 (03) サイト (04) 本 (05) ホームページ			
換言候補: (06) 商品 (07) 作品 (08) 製品 (09) リスト (10) ポイント			
的能力の (向上) にも役立つ	育成	育成	育成
換言候補: (01) 開発 (02) 育成 (03) 強化 (04) 研究 (05) 活性化			
換言候補: (06) 形成 (07) 評価 (08) 分析 (09) 維持 (10) 解明			
持続的な (施設) を整備する	設備	設備	環境
換言候補: (01) 設備 (02) 環境 (03) 制度 (04) 機能 (05) 方法			
換言候補: (06) 産業 (07) リンク (08) 改善 (09) 成長 (10) 森林			
農業の (成長) が阻害される	発展	発展	発展
換言候補: (01) 発展 (02) 生活 (03) 環境 (04) 研究 (05) 効果			
換言候補: (06) 経営 (07) 形成 (08) 生産 (09) 発達 (10) 進展			

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
施設・設備の (整備) がある	設置	管理	管理
換言候補: (01) 管理 (02) 計画	(03) 改善	(04) 活用	(05) 利用
換言候補: (06) 提供 (07) 内容	(08) 設置	(09) 状況	(10) 中
国外からの (輸入) が行われる	購入	導入	供給
換言候補: (01) 投資 (02) 購入	(03) 供給	(04) アクセス	(05) 電話
換言候補: (06) 登録 (07) 申請	(08) 導入	(09) 修理	(10) 攻撃
ような (影響) を与えている	効果	変化	環境
換言候補: (01) 問題 (02) 形	(03) 環境	(04) 効果	(05) 仕事
換言候補: (06) 変化 (07) 方法	(08) 感じ	(09) 場所	(10) 被害
評価する (原因) になっている	理由	-	理由
換言候補: (01) 為 (02) 結果	(03) 理由	(04) 問題	(05) 中
換言候補: (06) 0:00:00 (07) 場合	(08) 情報	(09) システム	(10) 物
不利な (状況) を引き起こしている	状態	状態	状態
換言候補: (01) 状態 (02) 事	(03) 環境	(04) 事態	(05) 活動
換言候補: (06) 結果 (07) 変化	(08) 何	(09) 影響	(10) 競争
楽天ブログの (ブログ) を表示する	-	日記	サイト
換言候補: (01) サイト (02) 中	(03) 方	(04) 掲示板	(05) 日記
換言候補: (06) テーマ (07) 機能	(08) トップ	(09) 更新	(10) 書き込み
手紙と (写真) が入っている	-	-	-
換言候補: (01) 絵 (02) 作品	(03) 言葉	(04) 自分	(05) 花
換言候補: (06) 電話 (07) カメラ	(08) 広告	(09) 資料	(10) 子供
普段の (生活) で目にする	-	状態	中
換言候補: (01) 中 (02) 日本	(03) 会社	(04) 状態	(05) 商品
換言候補: (06) 街 (07) 旅行	(08) 町	(09) シーン	(10) ネット
上記の (画像) に似ている	写真	絵	写真
換言候補: (01) 写真 (02) 映像	(03) 情報	(04) 絵	(05) 画面
換言候補: (06) 商品 (07) 物	(08) 作品	(09) 部分	(10) 内容
関係の (仕事) をしている	作業	作業	作業
換言候補: (01) 作業 (02) 活動	(03) 事	(04) 中	(05) 生活
換言候補: (06) 関係 (07) 会社	(08) 人	(09) 問題	(10) 作り
以外の (目的) で使用される	-	-	-
換言候補: (01) 中 (02) 形	(03) 方法	(04) 地域	(05) 会社
換言候補: (06) サイト (07) 日本	(08) 施設	(09) 学校	(10) 状態
絶大な (支持) を得ている	評価	支援	支援
換言候補: (01) 支援 (02) 評価	(03) 協力	(04) 応援	(05) 技術
換言候補: (06) 情報 (07) シェア	(08) 力	(09) 機能	(10) アクセス
検討する (動き) が出ている	声	声	姿勢
換言候補: (01) 事 (02) 状況	(03) 声	(04) 内容	(05) 関係
換言候補: (06) 方 (07) 形	(08) 人	(09) 情報	(10) 姿勢
的な (雰囲気) を醸し出している	感じ	感じ	感じ
換言候補: (01) 感じ (02) 感覚	(03) デザイン	(04) 表情	(05) 音
換言候補: (06) スタイル (07) イメージ	(08) 世界	(09) 物	(10) ムード
接続できる (環境) になっている	状況	状況	システム
換言候補: (01) システム (02) 技術	(03) 機能	(04) サービス	(05) 事
換言候補: (06) 状況 (07) 状態	(08) 施設	(09) 人	(10) 方法
ような (状態) になっている	状況	状況	状況
換言候補: (01) 中 (02) 事	(03) 状況	(04) 形	(05) 感じ
換言候補: (06) 関係 (07) 環境	(08) 場所	(09) 方法	(10) 問題

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
僕の (人生) は輝いている	-	-	-
換言候補: (01) 世界 (02) 自分	(03) 人	(04) 彼	(05) あなた
換言候補: (06) 国 (07) 作品	(08) 人間	(09) 女性	(10) 彼女
保護者との (関係) に悩む	関わり	関わり	対応
換言候補: (01) 教育 (02) 情報	(03) 対応	(04) 電話	(05) 学習
換言候補: (06) コミュニケーション (07) 風	(08) トラブル	(09) 関わり	(10) 交渉
サイトの (リンク) が切れている	-	関係	-
換言候補: (01) 情報 (02) 画像	(03) 電話	(04) ネット	(05) 部分
換言候補: (06) 関係 (07) 間	(08) 保険	(09) ライン	(10) リング
公開されない (求人) を含む	-	-	案件
換言候補: (01) 物件 (02) 情報	(03) 商品	(04) 記事	(05) システム
換言候補: (06) 作品 (07) 発表	(08) 映画	(09) 案件	(10) ケース
ような (活動) をしている	事業	取り組み	研究
換言候補: (01) 研究 (02) 事業	(03) 教育	(04) 運動	(05) 作り
換言候補: (06) 開発 (07) 取り組み	(08) 仕事	(09) 学習	(10) 生活
編集したい (部分) などがある	ところ	ところ	ところ
換言候補: (01) 方 (02) ところ	(03) 情報	(04) 人	(05) 写真
換言候補: (06) 画像 (07) メール	(08) ファイル	(09) 記事	(10) 曲
発言者の (名前) をクリックする	項目	-	情報
換言候補: (01) 中 (02) 言葉	(03) 情報	(04) 方	(05) 問題
換言候補: (06) 声 (07) 手	(08) アドレス	(09) 項目	(10) 下
ような (意見) を持っている	感想	感想	感想
換言候補: (01) 感想 (02) 事	(03) 内容	(04) 情報	(05) 話
換言候補: (06) 物 (07) 問題	(08) 方	(09) 考え方	(10) 言葉
士の (資格) を持っている	-	技術	技術
換言候補: (01) サービス (02) 試験	(03) 情報	(04) システム	(05) 商品
換言候補: (06) 関係 (07) カード	(08) 事業	(09) 制度	(10) 技術
温泉街の (中心) に位置する	中央	中央	中
換言候補: (01) 中 (02) 前	(03) 先	(04) 中央	(05) 周辺
換言候補: (06) 通り (07) 真ん中	(08) 奥	(09) 街	(10) 近く
投稿された (質問) を見る	-	-	問題
換言候補: (01) 言葉 (02) 内容	(03) 問題	(04) 話	(05) 事
換言候補: (06) メール (07) 中	(08) 物	(09) 方	(10) 発言
系の (作品) を見ていただく	-	-	映画
換言候補: (01) 映画 (02) 本	(03) 商品	(04) 写真	(05) 物
換言候補: (06) 中 (07) サイト	(08) 製品	(09) 情報	(10) 内容
着目した (抜け毛) を改善する	-	-	肌
換言候補: (01) 髪 (02) 肌	(03) 疾患	(04) 商品	(05) カード
換言候補: (06) 活動 (07) 効果	(08) 方	(09) 生活	(10) 評価
発行された (マガジン) を見る	雑誌	情報	雑誌
換言候補: (01) サイト (02) メール	(03) 雑誌	(04) 新聞	(05) 情報
換言候補: (06) 本 (07) カード	(08) 記事	(09) 中	(10) 会社
教育の (拡充) などがあげられる	強化	強化	強化
換言候補: (01) 強化 (02) 拡大	(03) 改善	(04) 整備	(05) 向上
換言候補: (06) 充実 (07) 支援	(08) 開発	(09) 導入	(10) 削減
重要な (地位) を占めている	役割	役割	役割
換言候補: (01) 市場 (02) 心	(03) 地域	(04) 力	(05) 分野
換言候補: (06) 空間 (07) 人	(08) 役割	(09) 収入	(10) 資産

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
スノーボードの (インストラクター) をしている	ガイド	-	-
換言候補: (01) プロ (02) ガイド (03) 方 (04) 中 (05) ショップ			
換言候補: (06) スクール (07) サイト (08) サークル (09) 販売 (10) 活動			
駅からの (距離) を指定する	-	場所	コース
換言候補: (01) 場所 (02) 方 (03) コース (04) 写真 (05) 道			
換言候補: (06) 道路 (07) 金額 (08) アクセス (09) 事務所 (10) 駅			
的な (意味合い) を持っている	意味	意味	意味
換言候補: (01) 意味 (02) 感じ (03) ニュアンス (04) 事 (05) 形			
換言候補: (06) 気持ち (07) 関係 (08) 名前 (09) 内容 (10) イメージ			
事業による (取り組み) が始まっている	仕事	事業	-
換言候補: (01) 研究 (02) 事業 (03) 開発 (04) 整備 (05) サービス			
換言候補: (06) 支援 (07) 調査 (08) 仕事 (09) 治療 (10) 利用			
経済活動の (手段) を強化する	-	-	-
換言候補: (01) 関係 (02) 管理 (03) 調査 (04) 市場 (05) 支援			
換言候補: (06) 分析 (07) 現場 (08) 連携 (09) 基盤 (10) 確保			
時の (扱い) から計算される	-	-	-
換言候補: (01) 結果 (02) ポイント (03) 量 (04) 収入 (05) 表			
換言候補: (06) 体重 (07) 性 (08) ここ (09) 数 (10) アクセス数			
食料の (確保) が困難になる	調達	調達	調達
換言候補: (01) 管理 (02) 利用 (03) 提供 (04) 販売 (05) 調達			
換言候補: (06) 保存 (07) 供給 (08) 継続 (09) 購入 (10) 計算			
話の (ネタ) にしたくなる	話題	話題	話題
換言候補: (01) 物 (02) 事 (03) 言葉 (04) 歌 (05) 店			
換言候補: (06) 記事 (07) 話題 (08) 絵 (09) カード (10) 顔			
阻害する (要因) となっている	-	要素	要素
換言候補: (01) 要素 (02) 問題 (03) 環境 (04) 効果 (05) 作用			
換言候補: (06) 事 (07) 行動 (08) 情報 (09) 物質 (10) 障害			
記事や (他人) を誹謗中傷する	個人	-	個人
換言候補: (01) 私 (02) 会社 (03) 個人 (04) 社長 (05) サイト			
換言候補: (06) 会員 (07) 人物 (08) 関係者 (09) 大学 (10) 作品			
者自身の (行為) に起因する	行動	行動	行動
換言候補: (01) 行動 (02) 事 (03) 問題 (04) 活動 (05) 関係			
換言候補: (06) 事業 (07) 運動 (08) 事故 (09) 利用 (10) 状態			
エコな (暮らし) が付いてくる	-	-	-
換言候補: (01) 人 (02) 車 (03) 商品 (04) 情報 (05) 言葉			
換言候補: (06) 本 (07) 広告 (08) 名前 (09) アイテム (10) プレゼント			
上の (アイコン) をクリックする	画像	画像	-
換言候補: (01) 画像 (02) リンク (03) ボタン (04) メニュー (05) 画面			
換言候補: (06) 物 (07) 写真 (08) バー (09) 地図 (10) 文字			
医師の (治療) を受けている	療法	管理	療法
換言候補: (01) 活動 (02) 手術 (03) 療法 (04) 教育 (05) 事業			
換言候補: (06) 仕事 (07) サービス (08) 管理 (09) 指導 (10) 検査			
日本人の (活躍) が期待される	-	活動	活動
換言候補: (01) プレー (02) 事 (03) 活動 (04) 姿 (05) 動き			
換言候補: (06) 戦い (07) 試合 (08) 作品 (09) 発表 (10) 演技			
的な (研究) を結合させる	-	-	-
換言候補: (01) 人 (02) 組織 (03) 機能 (04) 物 (05) 基			
換言候補: (06) ファイル (07) 物質 (08) 細胞 (09) サウンド (10) 成分			

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
管理人の (判断) で削除する	基準	基準	基準
換言候補: (01) 中 (02) 設定	(03) 方	(04) 人	(05) 操作
換言候補: (06) 自分 (07) 基準	(08) 責任	(09) 感じ	(10) 思考
圧倒的な (人気) を誇っている	評価	評価	-
換言候補: (01) 技術 (02) サービス	(03) 内容	(04) レベル	(05) 能力
換言候補: (06) 環境 (07) 効果	(08) 仕事	(09) 施設	(10) 評価
情報通信の (技術) を利用する	-	設備	理論
換言候補: (01) 研究 (02) 会社	(03) 世界	(04) 設備	(05) 学習
換言候補: (06) 本 (07) 理論	(08) 領域	(09) 活用	(10) 人材
あなたの (悩み) に応えてくれる	疑問	疑問	思い
換言候補: (01) 思い (02) 情報	(03) 希望	(04) 気持ち	(05) あなた
換言候補: (06) サービス (07) 疑問	(08) 質問	(09) 私	(10) 声
迅速な (対応) を可能にする	対策	対策	対策
換言候補: (01) 対策 (02) 管理	(03) 活動	(04) システム	(05) 処理
換言候補: (06) サービス (07) 開発	(08) 技術	(09) 作業	(10) 仕事
お客様の (ニーズ) にお応えする	需要	需要	需要
換言候補: (01) 問題 (02) 状況	(03) 需要	(04) 内容	(05) 課題
換言候補: (06) 変化 (07) スタイル	(08) 条件	(09) 中	(10) 処理
支払い方法の (変更) をお願いする	-	設定	設定
換言候補: (01) 登録 (02) 情報	(03) 設定	(04) 手続き	(05) 中
換言候補: (06) 確認 (07) 選択	(08) 入力	(09) チェック	(10) 改善
韓国の (ホームページ) をのぞいてみる	サイト	サイト	サイト
換言候補: (01) サイト (02) 掲示板	(03) 中	(04) 店	(05) 方
換言候補: (06) 記事 (07) 活動	(08) 本	(09) ショップ	(10) 世界
重要な (動き) をしている	行動	行動	機能
換言候補: (01) 事 (02) 作用	(03) 力	(04) 機能	(05) 中
換言候補: (06) 運動 (07) 体	(08) 行動	(09) 活動	(10) 生活
万円以上の (収入) がある	利益	所得	収益
換言候補: (01) 利益 (02) 収益	(03) 売り上げ	(04) 所得	(05) 売上
換言候補: (06) 価格 (07) 金額	(08) 事業	(09) 費用	(10) 仕事
以上の (利益) をあげている	収入	-	収益
換言候補: (01) 収入 (02) 収益	(03) 売上	(04) 売り上げ	(05) 効果
換言候補: (06) 情報 (07) 機能	(08) 能力	(09) 活動	(10) 問題
ような (気分) になってしまう	気持ち	気持ち	気持ち
換言候補: (01) 気持ち (02) 思い	(03) 状態	(04) 中	(05) 感じ
換言候補: (06) 笑顔 (07) 形	(08) 事	(09) 感覚	(10) 車
社会的な (業務) を委託する	事業	仕事	事業
換言候補: (01) 事業 (02) 仕事	(03) 教育	(04) 情報	(05) 支援
換言候補: (06) 方法 (07) 評価	(08) 対応	(09) 生産	(10) 行為
腰への (負担) を軽減する	負荷	負荷	負荷
換言候補: (01) 負荷 (02) ストレス	(03) ダメージ	(04) 疲労	(05) 緊張
換言候補: (06) 衝撃 (07) 当たり	(08) 圧迫感	(09) 荷重	(10) 圧迫
英語の (勉強) をしている	学習	学習	学習
換言候補: (01) 仕事 (02) 練習	(03) 活動	(04) 学習	(05) 事
換言候補: (06) 生活 (07) 研究	(08) 教育	(09) 授業	(10) ゲーム
一つ以上の (場所) を指定する	地域	地域	-
換言候補: (01) 地域 (02) 情報	(03) 商品	(04) 分野	(05) プログラム
換言候補: (06) 装置 (07) 記事	(08) ファイル	(09) 文字	(10) 項目

次のページに続く

前のページからの続き

入力文 (換言対象語)	評価者 A	評価者 B	評価者 C
的な (存在) となっている	-	人	事
換言候補: (01) 事 (02) 言葉 (03) 物 (04) 姿 (05) 方			
換言候補: (06) 行動 (07) 人 (08) 問題 (09) 関係 (10) 声			
夕食の (準備) をしている	-	-	-
換言候補: (01) 事 (02) 対応 (03) 中 (04) アップ (05) 設定			
換言候補: (06) サービス (07) 企画 (08) 内容 (09) 間 (10) 部分			
田舎の (価値) を再発見する	-	-	-
換言候補: (01) 文化 (02) 生活 (03) 地域 (04) 力 (05) 日本			
換言候補: (06) 部分 (07) 自分 (08) 人間 (09) 方 (10) 歴史			
業務支援システムの (開発) を行う	構築	構築	構築
換言候補: (01) 構築 (02) 導入 (03) 販売 (04) 設計 (05) 運用			
換言候補: (06) 企画 (07) 提供 (08) テスト (09) 決定 (10) ノウハウ			
製造、(販売) を行っている	-	-	-
換言候補: (01) サービス (02) 事業 (03) 開発 (04) システム (05) 管理			
換言候補: (06) 研究 (07) 業務 (08) 工事 (09) 技術 (10) デザイン			
全体の (方向) が間違っている	-	-	-
換言候補: (01) 場所 (02) 部分 (03) 事 (04) 地域 (05) 世界			
換言候補: (06) 位置 (07) 人 (08) 関係 (09) 自分 (10) システム			

これで終わり

## 付録 C : 本論文に関する原著論文

- (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, 発表 1, 2012.
- (2) 梶原 智之, 山本 和英. 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. 言語処理学会第 19 回年次大会発表論文集, pp.272-275, 2013.
- (3) 梶原 智之, 山本 和英. 小学生の読解支援に向けた語釈文から語彙的換言を選択する手法. NLP 若手の会第 8 回シンポジウム, 発表 23, 2013.
- (4) Tomoyuki Kajiwara, Hiroshi Matsumoto, Kazuhide Yamamoto. Selecting Proper Lexical Paraphrase for Children. In *Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*, pp.59-73, 2013.
- (5) 梶原 智之, 山本 和英. 文脈の多様性に基づく名詞換言の提案. 言語処理学会第 20 回年次大会発表論文集, pp.769-772, 2014.
- (6) 梶原 智之, 山本 和英. 文脈の多様性に基づく名詞換言の評価. 第 28 回人工知能学会全国大会論文集, 3I4-2, 2014.
- (7) 梶原 智之, 山本 和英. 高頻度語は平易語なのか?. NLP 若手の会第 9 回シンポジウム, 発表 P02, 2014.
- (8) Tomoyuki Kajiwara, Kazuhide Yamamoto. Noun Paraphrasing Based on a Variety of Contexts. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing (PACLIC 2014)*, no page numbers, 2014.
- (9) 梶原 智之, 山本 和英. 日本語の語彙的換言知識の質的評価. 信学技報, vol.114, no.366, NLC2014-37, pp.43-48, 2014.
- (10) 梶原 智之, 山本 和英. 日本語の語彙平易化システムの構築. 第 77 回情報処理学会全国大会講演論文集, no page numbers, 2015.
- (11) 梶原 智之, 山本 和英. 日本語の語彙平易化評価セットの構築. 言語処理学会第 21 回年次大会発表論文集, no page numbers, 2015.
- (12) 梶原 智之, 山本 和英. 語釈文を用いた小学生のための語彙平易化. 情報処理学会論文誌, vol.56, no.3, no page numbers, 2015.