JAIST Repository
https://dspace.jaist.ac.jp/

Title	句の言い換えによるテキストの平易化
Author(s)	河原井，翼
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17249
Rights	
Description	Supervisor:白井 清昭，先端科学技術研究科，修士（情報科学）

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修士論文

句の言い換えによるテキストの平易化

河原井 翼

主指導教員 白井 清昭

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和３年３月

Abstract

Text simplification is a technique to paraphrase a sentence or document into a simpler expression while retaining its meaning. It is useful for children, aged people, and non-native speakers of Japanese. For example, if a text that contains many complex words can be paraphrased into a simpler text, it would be helpful for people whose language skill of Japanese is not high to understand the text. The most common method of text simplification is to replace a complex word, which is hard to understand, in a sentence with a simple word, which is easy to understand. The recent studies of text simplification propose various methodology, such as a method to replace a complex word with another word that is easy and the most similar to it, and a method to paraphrase a complex word with a frequently used word in a definition sentence of the complex word in a dictionary. In addition, several studies aim at constructing a database compiling pairs of a complex and simple word as fundamental knowledge for text simplification. However, the paraphrase of words is insufficient for text simplification, although it is relatively easy to implement and the most of the previous studies focus on it. Even when a complex word is paraphrased into a simple word and it is carefully confirmed that they are similar and the latter is easier than the former by word-by-word comparison, the sentence obtained by paraphrase of the words may not be simple or natural. In such a case, paraphrase of the word is not appropriate when a context of the word is considered. Another problem is that variety of simplified texts obtained by paraphrase of words is rather poor. When people simplify a text, they often paraphrase not only words but also phrases, clauses, or a whole document. It is impossible to automatically generate a wide variety of simplified texts as humans write by applying paraphrase of words only.

This study aims at simplifying a text by paraphrase of a phrase. Note that the phrase to be paraphrased is limited to a sequence of words of "noun-particle-verb". Furthermore, when paraphrasing a complex phrase, an appropriate simple phrase is selected from candidates by considering a context of the complex phrase. Comparing with text simplification based on paraphrase of words, simplification by paraphrase of phrases enables us to generate more various and high quality simplified texts as humans do.

In the proposed method, text simplification is performed in five steps. The first step is preprocessing. Morphological analysis of an input text is performed to obtain lexical information of words such as part-of-speech (POS). Then, phrases in the form of "noun-particle-verb" are extracted by referring POSs obtained by morphological analysis. The second step is construction of "simply paraphrased word database". The existing language resources are used to obtain pairs of a complex word and its simply paraphrased word, and they are compiled as the database.

Two language resources are used: one is "Simple PPDB: Japanese" that is a simple paraphrase dictionary, the other is "SNOW D2" that is a paraphrase dictionary. Only pairs of complex and simple words are extracted from them by checking the level of difficulty of words. In addition, pairs of intransitive and transitive verbs are excluded, since their meanings and levels of the difficulty are almost the same. The third step is generation of simple phrases. Candidates of simple phrases to be replaced with a complex phrase are generated using the constructed simply paraphrased word database. Three patterns are used for generation: paraphrasing both a noun and verb, paraphrasing only a noun, and paraphrasing only a verb. The fourth step is selection of the simple phrase. A score is calculated for each candidate of the simple phrase generated in the previous step, considering both the faithfulness and fluency, then the most appropriate simple phrase is selected based on this score. Here the faithfulness evaluates whether the meaning of the sentence is not changed by paraphrase, while the fluency evaluates how natural the paraphrased simple phrase is. The score of the faithfulness is measured by the cosine similarity of sentence embeddings (vector representation of sentences) before and after the paraphrase. Sentence BERT is used to obtain the sentence embedding. The score of the fluency is calculated by the relative frequency of the simple phrase, since a phrase can be recognized as natural when it frequently occurs in a corpus. The frequency of the phrase is obtained by Kyoto University case frame dictionary. The scores of the faithfulness and fluency are combined in two ways, that is, we define two scores for selection of the simple phrase. In these scores, two parameters are introduced: one is used to adjust the scale of the scores of the faithfulness and fluency, the other is the weight parameter for the faithfulness and fluency. In addition, to generate more simple phrases obtained by paraphrase of both the noun and verb, such a phrase is always chosen when its score is not the maximum but greater than a pre-defined threshold. The threshold is determined using development data. The final step is generation of a sentence. The complex phrase of the original sentence is replaced with the chosen simple phrase to generate a simple sentence.

Two experiments were conducted to evaluate our proposed method. In the first experiment, the performance of the selection of the simple phrase was measured by the accuracy, which was the proportion of cases where simple phrases chosen by the system and by the human subject agree to the total number of cases in the test data. The accuracy was changed for the different scores and parameters used in the system as well as the human subjects. It was around 0.45 and 0.51 at maximum. On the other hand, the inter-annotator agreement of two subjects was 0.61. Therefore, it was rather difficult even for humans to select the best simple phrase from candidates. It was found that the idea to rank simple phrase

candidates based on the faithfulness and fluency was effective since the accuracy of the system was close to the inter-annotator agreement. In the second experiment, we evaluated the quality of the simple sentences containing the simple phrases generated by the proposed method. Two subjects compared the simple sentences with the original sentences and gave them a five-point rating with respect to three aspects: simplicity, faithfulness, and fluency. As a result, the average rating of the two subjects was approximately 4 points for each aspect. Since the ratings were high, the effectiveness of the proposed method was confirmed. In addition, the weighted κ coefficient between ratings of two subjects was calculated. They were high, 0.93, 0.89, and 0.95 for the simplicity, faithfulness, and fluency, respectively. It indicated that the evaluation of two subjects was stable. In addition, we performed an error analysis on the results of the first experimental. We investigated the reason why the simple phrases selected by the proposed method and human subjects disagreed, then discussed the current problems of the proposed method and future directions.

The main contribution of this thesis was that it proposed the method of text simplification by paraphrasing phrases considering the faithfulness and fluency, which could generate sufficiently high quality simple sentences for being read by a human.

概要

テキストの平易化とは，文もしくは文章をその意味を保持しながら易しい表現に言い換える技術である．この技術は，子どもや高齢者，あるいは日本語を母語としない話者に有用である．例えば，難解な単語が多く含まれるテキストを平易なテキストに変換できれば，日本語能力がそれほど高くない人によるテキストの理解を助けることができる．平易化の最も一般的な手法は，文中に使用されている理解が難しい単語（難解語）を理解しやすい平易な単語（平易語）へ言い換えることである．テキストの平易化に関する近年の研究では，難解語と最も類似度が高く，かつ易しい単語に言い換える手法や，難解語の国語辞典の語釈文中の単語のうち出現頻度が高いものに言い換える手法が提案されている．また，テキスト平易化のための基礎的な知識として，難解語と平易語を対応させたデータベースを構築する研究も行なわれている．しかしながら，先行研究の多くは単語単位の言い換えにより平易化を実現している．ところが，難解語を平易語に言い換えたとき，単語同士を比べれば平易になっているが，難解語を平易語に置き換えた文は平易になっていなかったり不自然になることがある．これは文脈に適さない言い換えであったためと言える．単語単位の平易化のもうひとつの問題点は，単語しか言い換えないために，平易化されたテキストの多様性が乏しいことである．人がテキストを易しく言い換えるときは，単語を言い換えるだけでなく，句，節，あるいは文全体を易しく言い換えることも多い．単語の言い換えだけでは，人が生成するような様々な平易化テキストを自動生成することはできない．

本研究では，句を単位とした言い換えによって平易な文を生成することを目的とする．ただし，平易化の対象とする句は,「名詞-助詞-動詞」という単語列に限定する．さらに，句を言い換える際，文脈を考慮して複数の平易句の候補から適切なものを選ぶ．文脈を考慮した句単位の平易化を実現することで，単語単位の平易化のみよりも，より多様で人が生成するテキストに近い平易化テキストを生成できる．

提案手法では，大きく分けて 5 つの段階で平易化を実現する．一段階目の処理では，入力したテキストの形態素解析を行ない，それぞれの単語に品詞などの情報を付与する．そして，テキストから品詞の情報から,「名詞-助詞-動詞」から構成される句を抽出する．二段階目の処理では，既存の言語資源を利用して，難解語とその平易な言い換えである平易語の組を大量に取得し，平易言い換え単語対データベースを構築する．利用する言語資源は，平易な言い換え辞書である Simple PPDB: Japanese と言い換え辞書の SNOW D2: 内容語換言辞書であり，これらから単語の難易度をチェックして平易な言い換えのみを収集する．その際に，基本的な意味は共通しかつ難易度に差がないといった考えのもと，自動詞と他動詞の言い換え対は取得しない．三段階目の処理では，抽出した平易言い換え単語対データベースを基に，難解句を 3 つのパターンで言い換え，平易句の候補を生成する．3 つのパターンとは，名詞・動詞両方の言い換え，名詞のみの言い換え，動詞のみの言い換えである．四段階目の処理では，正確性 (言い換え前後で意味が保持され

ているかという観点) と流暢性 (言い換え後の句がどれだけ自然な句であるかという観点) に基づいた評価式で生成した言い換え候補のスコアを算出し，そのスコア付けにより言い換え候補の中から最も適切な候補を選択する．正確性のスコアは言い換え前後の文の分散表現（文の意味を表すベクトル表現）のコサイン類似度で測る．文の分散表現は Sentence BERT を用いて得る．流暢性のスコアは，コーパスによく出現する句ほど自然であるという考えから，句の出現頻度を正規化して測る．句の出現頻度は京都大学格フレームにより求める．2 つのスコアのスケールを合わせるパラメータや両者に対する重み付けのパラメータを導入し，平易句の候補のスコアを 2 つ提案する．さらに，名詞・動詞の両方を言い換える句を多く生成するため，スコアが最大でなくても閾値以上のときにこれを優先して選択する方式を導入する．このときの閾値は開発データを用いて決める．最後の五段階目では，元の文の難解句を平易句に置き換え，さらに活用形の処理をして，平易な文を生成する．

提案手法の評価は 2 種類の実験により行なった．一つ目の実験では，スコア付けにより選択した平易句が人手によって選択された平易句とどの程度一致しているか（正解率）を調べた．その結果，提案手法の正解率はスコア付け，パラメータ，評価者の違いによって変動するものの，0.45 程度であり，最高で 0.51 となった．一方，2 者の判定の一致率は 0.61 と低く，言い換え候補の中から最適な平易句を選択することは人手でも難しい問題といえる．提案手法の正解率は 2 者の判定の一致率と近く，正確性と流暢性の観点に基づいた言い換え候補の選択手法が適切だったと言える．もう一つの実験では，提案手法によって生成された平易句およびそれを含む平易な文の品質評価を行なった．2 名の作業者が，平易な文と元の文を比べ，平易性，正確性，流暢性の 3 つの観点から 5 段階の評点をつけた．その結果，2 者の平均の評点はいずれの観点もおよそ 4 点であった．評点は全般的に高く，提案手法の有効性が確認された．また，2 者の評点の重み付きカッパ係数は 3 つの観点について平易性，正確性，流暢性のそれぞれについて，0.93，0.89，0.95 と高く，評価は安定していることがわかった．さらに，一番目の実験結果に対するエラー分析を行なった．提案手法が選択した平易句と人によって選択された平易句が一致しなかった原因を調査し，提案手法の問題点とその改善案を考察した．

本論文の主たる貢献は，正確性と流暢性を考慮した句の平易化の手法を提案し，人から見ても十分に品質の高い平易な言い換えを実現した点にある．

目 次

第 1 章　はじめに **1**

1.1 背景 1
1.2 目的 2
1.3 本論文の構成 2

第 2 章　関連研究 **3**

2.1 平易化に関する研究 3
2.1.1 平易化のための言語資源 3
2.1.2 単語の言い換えによる平易化 5
2.1.3 統計的機械翻訳に基づく平易化 5
2.2 単語の難易度判定に関する研究 6
2.3 本研究の特色 7

第 3 章　提案手法 **9**

3.1 概要 9
3.2 平易言い換え単語対の収集 11
3.3 平易化する句の検出 15
3.3.1 テキストの形態素解析 15
3.3.2 言い換え対象句の検出 17
3.4 言い換え候補の生成 19
3.5 最適な平易句の選択 19
3.5.1 正確性のスコア 20
3.5.2 流暢性のスコア 21
3.5.3 平易句の候補のスコア 22
3.5.4 名詞・動詞の言い換えを優先した平易句の選択 23
3.6 平易化する文の生成 25

第 4 章　評価実験 **27**

4.1 データセット 27
4.1.1 開発データ 27
4.1.2 評価データ 28
4.2 パラメータの決定 28
4.2.1 スケール調整パラメータ α の決定 30

4.2.2 $Score1$ の閾値 T の決定 30
4.2.3 $Score2$ の閾値 T の決定 31
4.3 平易句選択手法の評価 32
4.4 平易化されたテキストの品質評価 34
4.5 エラー分析 37

第 5 章 おわりに **43**
5.1 まとめ 43
5.2 今後の課題 44

付 録 A 自動詞と他動詞の対応付け **49**

図 目 次

3.1 提案手法の概要 10
3.2 句の検出のフローチャート 18
3.3 京都大学格フレーム Ver 2.0 の抜粋 22

4.1 開発データにおける平易句の正確性・流暢性判定の精度 ($Score1$) . 31
4.2 開発データにおける平易句の適切性判定の精度 ($Score2$) 32

表 目 次

3.1 日本語教育語彙表の抜粋 11
3.2 SNOW D2: 内容語換言辞書の抜粋 12
3.3 Simple PPDB: Japanese の抜粋 13
3.4 平易単語言い換え単語対データベースの概要 15
3.5 収集した平易言い換え単語対の抜粋 16
3.6 「人口に膾炙する」を形態素解析した結果 (IPA 辞書) 17
3.7 「人口に膾炙する」を形態素解析した結果 (mecab-ipadic-NEologd) 17

4.1 開発データ（抜粋） 29
4.2 開発データにおける正確性と流暢性のスコアの統計 30
4.3 平易句の選択手法の評価 (正解率) 33
4.4 平易句の品質評価 35
4.5 平易性の評点の分割表 36
4.6 正確性の評点の分割表 36
4.7 流暢性の評点の分割表 36
4.8 評価者 1 と提案手法によって選択された平易句の一致数 37
4.9 提案手法と評価者の判定が一致した事例 40
4.10 名詞・動詞両方の言い換えスコアが十分に高くないために不一致だった事例 41
4.11 他の名詞・動詞両方の言い換えが選択されて不一致だった事例 . . . 41
4.12 名詞・動詞両方の言い換えを選択して不一致だった事例 42
4.13 名詞のみと動詞のみの言い換えの間で判定不一致だった事例 42

A.1 自・他動詞対判別辞書 49
A.2 自動詞と他動詞の組を判別するルール 50

第1章　はじめに

1.1　背景

テキストの平易化とは，文もしくは文章，あるいはそれらを内包する文書に対して，意味を保持しながら易しい表現に言い換える技術である．近年では，やさしい日本語という共通の概念が広く知られるようになった背景もあり，広く注目されている技術である．テキストの平易化は，子どもや高齢者，あるいは日本語を母語としない話者に有用である．難解な単語が多く含まれるテキストを平易な表現にできれば，日本語能力がそれほど高くない人によるテキストの理解を助けることができる．また，機械翻訳の精度を向上させるために，難解な表現を含むテキストを平易化してから翻訳することも利用例のひとつである．このように，テキストの平易化の技術の活用範囲は多岐にわたる．

平易化の最も一般的な手法は，文中に使用されている理解が難しい単語を理解しやすい平易な単語へ言い換えることである．言い換えた後の単語がどの程度理解しやすいかは，平易化されたテキストを読む人物をどう想定するかに依存する．一般的には，小学生や日本語初学者の外国人を基準として平易化が行なわれることが多い．また，平易化の研究では，理解が難しい単語は難解語，理解しやすい単語は平易語と呼ばれている．基本的な平易化の技術は，難解語を平易語に言い換えることと言える．

テキストの平易化に関する近年の研究では，難解語と最も類似度が高く，かつ易しい単語に言い換える手法や，難解語の国語辞典の語釈文中の単語のうち出現頻度が高いものに言い換える手法が提案されている．また，テキスト平易化のための基礎的な知識として，難解語と平易語を対応させたデータベースを構築する研究も行なわれている．しかしながら，先行研究の多くは単語単位の言い換えにより平易化を実現している．この手法の問題点は，単語単位で平易さを比較したときには明らかに平易になっているにもかかわらず，それを文中の難解語と置き換えてみると，平易とは言い難い文になることがあるという点である．これは文脈に適さない言い換えであるためと言える．単語単位の平易化のもうひとつの問題点は，単語しか言い換えないために，平易化されたテキストの多様性が乏しいことである．人がテキストを易しく言い換えるときは，単語を言い換えるだけでなく，句，節，あるいは文全体を易しく言い換えることも多い．単語の言い換えだけでは，人が生成するような様々な平易化テキストを自動生成することはできない．

1.2 目的

本研究では，句を単位とした言い換えによって，平易な文を生成することを目的とする．ただし，平易化の対象とする句は,「名詞-助詞-動詞」という単語列に限定し，句単位による平易化の可能性を探究する．さらに，難解な句を平易な句に言い換える際に，文脈に応じた適切な言い換えの手法を提案する．

提案手法では,「名詞-助詞-動詞」という句に対し，難解語の名詞を平易語の名詞に言い換える，難解語の動詞を平易語の動詞に言い換える，名詞と動詞の両方を難解語から平易語に言い換える，という 3 つの方法で平易句の候補を生成する．次に，それぞれの平易句の候補について，文として自然であるか，言い換え後の文の意味が元の文の意味と変わっていないか，という 2 つの観点から評価 (スコア付け) し，最良のものを選択する．平易句のスコア付けの際，平易句そのものだけでなく，置き換え対象となる難解句の文脈も考慮する．

句単位の平易化を実現することで，単語単位の平易化のみよりも，より多様で人が生成するテキストに近い平易化テキストを生成できる．このことは，平易化の技術を応用する様々な自然言語処理システムの向上につながる．

1.3 本論文の構成

本論文の構成は，以下の通りである．まず，第 2 章では，本研究の関連研究について述べる．次に，第 3 章では，提案手法について説明する．第 4 章では，提案手法の評価実験の方法や結果について述べる．最後に，第 5 章では，本論文のまとめと今後の課題について述べる．

第2章　関連研究

本章では，本論文の関連研究について述べる．2.1 節では，テキストの平易化に関する過去の研究を紹介する．次に，2.2 節では，単語の難易度を判定する研究について述べる．単語の難易度の判定はテキストの平易化に密接に関係する．最後に，2.3 節では，関連研究と本研究との相違を示し，本研究の特色を論じる．

2.1　平易化に関する研究

平易化に関する研究は大きく分けて 2 つのアプローチがある．まず，2.1.1 項にて，平易化のための言語資源の研究を紹介する．2.1.2 項では，単語単位の言い換えによる平易化の研究を紹介する．2.1.3 項では，統計的機械翻訳の技術を用いて平易化を実現する研究を紹介する．

2.1.1　平易化のための言語資源

テキスト平易化に関する研究では，平易化のための言語資源 (あるいは知識データベース) が構築され，研究者向けに公開されている．ここでは平易化のための言語資源を紹介する．

Paraphrase Database（PPDB）[4] は，英単語の言い換えを収録しているデータベースである．すなわち，ある英語の単語とそれとほぼ同じ意味を持つ別の単語の組が収録されている．PPDB は英語とスペイン語からなる対訳コーパス[1]から自動的に構築されている．すなわち，英単語の言い換えを収集するためにスペイン語のコーパスを利用した．この方法は，機械翻訳分野のピボット翻訳の技術を言い換え対の収集に適用したと言える．具体的には，スペイン語を中間表現とし，英単語に対応するスペイン語の単語を求め，そのスペイン語の単語に対応する別の英単語を収集することで，英単語の言い換え対を収集している．なお，この言語資源は言い換えに関するデータベースであり，平易な言い換えであるかどうかは考慮されていない．

[1]コーパスとは，大量のテキストを収集して，データベースとして取り扱えるようにした言語資源である．対訳コーパスは，2 つの言語の文と文の対，つまり対訳の関係が成立する文の組を大量に収集したコーパスである．一般的に，対訳コーパスは統計的機械翻訳やニューラル機械翻訳の学習データとして利用される．

Pavlick と Callison-Burch は，平易な言い換え辞書である Simple PPDB を構築し，公開した [19]．Simple PPDB は，PPDB に収録されている言い換え対の中から，難解語と平易語の組を選別したものである．したがって，Simple PPDB は平易化の研究に直接利用できる言語資源となっている．

PPDB と Simple PPDB については，日本語版が開発されて公開されている．水上らは PPDB: Japanese を開発した [18]．英語版の PPDB を構築した手法を用いて，英語をピボット言語とし，様々な日本語と英語の対訳コーパスから言い換え対を収集した．梶原と小町は Simple PPDB: Japanese を開発した [9]．英語版の Simple PPDB と同じように，PPDB: Japanese をベースに構築した．平易な言い換えを収集するにあたり，データベース内の各単語に単語の難易度を付与した．難易度は，初級，中級，上級の 3 クラスとした．単語の難易度を識別する分類器を機械学習アルゴリズムのひとつであるサポートベクターマシン（Support Vector Machine, SVM）によって学習した．この際，学習素性として，単語長，文字形態，コーパスにおける単語の頻度，単語分散表現の 4 つを用いた．この分類器の出力（難易度クラス）を単語へ付与した．なお，Simple PPDB: Japanese で採用されている難易度の 3 クラスは日本語教育語彙表 [27] の語彙レベルに由来する．日本語教育語彙表については 2.2 節にて後述する．単語の難易度を推定した後，PPDB: Japanese における言い換え対から，難易度のレベルが高いものと低いものの組を選別し，最終的に約 34 万組からなる平易な言い換え辞書を構築した．また，日本語の平易化のための評価用データセット [11] を利用して，自動構築した言い換え辞書の性能を評価した．その結果，英語の Simple PPDB で行なわれた評価結果と同水準の性能が得られ，高品質な平易な言い換え辞書を作成できたと結論づけている．さらに，平易な言い換え辞書だけでなく，3 つのレベルの難易度が付与された単語辞書もあわせて公開した．

山本と吉倉は SNOW D2: 内容語換言辞書を構築し，研究用に公開した [25]．以降，この言語資源を単に SNOW D2 と呼ぶ．SNOW D2 は，形態素解析[2]ツール Juman[3] の内容語項目を人手で言い換えた辞書である．この辞書の特徴は，作業者 1 名のみの感覚で言い換えていることにある．内容語辞書項目の対象は，普通名詞，サ変名詞，動詞，形容詞，副詞の単語であり，約 3 万件の単語を人手で言い換え，単語の言い換えデータベースを作成した．なお，SNOW D2 は言い換え辞書であり，データベース内の単語対は必ずしも平易な言い換えとは限らない．そのため，これを平易化のために利用するときには注意が必要である．

[2]形態素解析とは，テキストを言語の最小単位である形態素に区切ることである．日本語においては，形態素はおよそ単語に相当する．一般に，文を単語へ分割するときには形態素解析が行なわれる．形態素解析器に内蔵されている辞書により，分割された単語には辞書の情報，例えば品詞の情報が付与される．

2.1.2 単語の言い換えによる平易化

鍛治ら [6]，美野と田中 [17]，梶原と山本 [7] の研究では，国語辞典の語釈文の文末付近から平易語を獲得し，それを難解語と置き換えることで平易化を実現している．国語辞典の語釈文は，ある見出し語に対して，読み手がその見出し語の意味を理解できるように平易な文で説明がされている．したがって，言い換えの研究では国語辞典がよく利用されている．また，日本語における国語辞典の語釈文では，文末に重要な言葉が使われていることが多い．これはほとんどの単語が後ろの単語に係るという日本語の特徴といえる．そのため，これらの先行研究では，語釈文の文末およびその近辺から平易語を獲得していた．梶原と山本は，小学生のための文書読解支援に向けて，同じく国語辞典の語釈文を利用し，より精度の高い平易化を実現した [8]．従来手法とは異なり，語釈文全体から見出し語の言い換え，つまり平易語の獲得を試みた．見出し語の平易語になりうる候補の中から，最も妥当な候補を選択するために，様々な手法を比較検討した．その結果，シソーラスに基づく単語の類似度を用いた選択が最も良いと結論づけた．

Kriz らは，文脈に基づいた難解語判定モデルと言い換えモデルを構築することで，文脈を考慮した平易化を実現した [13]．まず，テキストにおける個々の単語について，それが理解が難しい単語（難解語）であるかを判定する．難解語判定モデルについては 2.2 節にて後述する．テキストから検出された難解語に対して，3 つの言語資源から言い換えの候補となる単語を収集する．一つ目の言語資源は WordNet[16] である．WordNet は，synset と呼ばれる同義語のグループ，それに対応する上位語，下位語など，英語の単語間の関係を記述したデータベースである．上位語とは，ある単語をより広い範囲の概念で捉えた単語のことを指す．下位語は，より狭義な概念を表す単語を指す．つまり，WordNet は，全体ではツリーのような階層構造を持つ言語資源といえる．WordNet を利用し，難解語の同義語，上位語，下位語を検索し，これを言い換え後の平易語の候補とする．なお，以上の説明は名詞に関する説明であり，他の品詞では上位語，下位語以外の単語間の関係も収録されている．ちなみに，日本語版の WordNet[5] も公開されている．2 つ目のデータセットは PPDB である．PPDB は言い換え対の集合であり，これから平易語の候補を得る．3 つ目のデータセットは Simple PPDB である．次に，得られた平易語の候補の中から，語彙的素性と文脈的素性の両方を考慮して，最も適した平易語を選択する．実験では，文脈的素性を利用することで，従来手法よりも高い精度で文脈に応じた好ましい言い換えが実現できることを示した．

2.1.3 統計的機械翻訳に基づく平易化

小藤らは，テキストの平易化を難解なテキストを平易なテキストに翻訳するタスクであるとみなし，統計的機械翻訳の手法を応用してテキストの平易化を実現する手法を提案した [12]．ここで統計的機械翻訳のモデルを学習するためには，難

解なテキストと平易なテキストの組を集めたパラレルコーパスが必要である．この研究では，新聞記事データを採用し，難解なテキストを毎日新聞，平易なテキストを毎日小学生新聞とした．これらの間で文間の対応付け，すなわち文のアライメントを行なう 4 つの手法を試した．最終的には，30,940 対の文の組からなるパラレルコーパスの自動構築に成功した．なお，この論文ではパラレルコーパスの構築を主たる目的としているため，この言語資源を利用した統計的機械翻訳による平易化の結果は示されていない．将来に使用する統計的機械翻訳モデルとしては，統計的機械翻訳ツールキットである cicada[24] やニューラル機械翻訳を提案していた．

統計的機械翻訳では大量の対訳対を必要とするが，難解な文と平易な文の組を大量に用意することは一般には難しい．梶原と小町は，自動構築されたパラレルコーパスを用いてテキストを平易化する手法を提案した [10]．コーパスを用意し，コーパス中の文を擬似的に平易な文に置き換えることで，難解なテキストと平易なテキストの擬似パラレルコーパスを構築した．そして，擬似パラレルコーパスを訓練データとして，フレーズベースの統計的機械翻訳モデルを学習した．そのモデルを用いて英語の難解なテキストを平易化した結果，従来の平易なコーパスを用いる平易化とほぼ同じ性能で平易化が行えることを示した．彼らの擬似パラレルコーパスの構築手法は言語を限定していないため，平易化に関する言語資源が不足している言語に対しても適用できるという点で非常に有用である．

2.2 単語の難易度判定に関する研究

単語の難易度を判定することは，テキストを平易化するための基礎的な技術のひとつである．本節ではこれに関連する研究を紹介する．

2.1.2 項で先述した Kriz らの難解語判定モデル [13] について説明する．難解語の判定に，SVM やランダムフォレストといった機械学習の分類器を用いた．つまり，ある文脈に出現する単語を入力として，その単語が難解語であるかどうかの二値判定を行なった．この研究の主たる貢献は，語彙的素性のみではなく，文脈的素性を利用することで，高い精度で難解語を判定した点にある．語彙的素性として，単語長，音節数， Google Web 1T コーパス [23] を用いた単語頻度，WordNet の synset の数などを用いた．文脈的素性として，文中における単語の平均長，平均音節数，平均単語頻度，WordNet の同義語の synset，文長などを採用した．

単語の難易度を，機械学習の手法で判定するのではなく，言語資源を参照して判定する方法もある．日本語教育語彙表 [27] は，約 18,000 語の日本語教育用の単語を収録している辞書である．辞書に記載されている情報は，標準的な表記，読み，語彙の難易度，品詞，語種である．そのうち，語彙の難易度は，日本語の教育上の語彙レベルが示されている．語彙レベルは，初級前半，初級後半，中級前半，中級後半，上級前半，上級後半の 6 段階と定められている．単語の語彙レベ

ルの推定には，日本語教育コーパス[3]と現代日本語書き言葉均衡コーパス [15] を用いた．前者のコーパスは，初級から上級までの市販の教科書 100 冊からなるコーパスである．後者のコーパスは，日本に流通している現代日本語の書き言葉で書かれたテキストの全体像を把握するために構築されたコーパスであり，さまざまな媒体から無作為に抽出した約一億語のテキストに形態素解析情報，文書構造タグが付与されている．日本語の研究において広く利用されている．

2.3 本研究の特色

2.1.2 項で紹介した関連研究 [6, 17, 7, 8] は，基本的に単語の言い換えで平易化を実現する手法である．これに対し，本研究は，単語の言い換えではなく，句を対象とした平易化を目指している点が異なる．Kriz らの研究 [13] では，文脈に基づいた難易度判定や平易化のための言い換えモデルを実現していた．この研究も単語単位による言い換えである点では本研究とは異なる．ただし，本研究でも文脈に応じて適切な平易化を行なうことを研究の対象としており，この点は共通している．

2.1.3 項で紹介した関連研究 [12, 10] は，平易化を翻訳タスクとみなして言い換えを実現していた．これらの研究は，単語単位による言い換えではなく，文単位による言い換えと言える．つまり，句単位による言い換えも含まれる．しかし，文全体を言い換えるモデルであるため，ユーザの要求に応じて平易化する句としない句を分けるといった調整が難しい．

以上をまとめると，本研究では，単語や文ではなく難解な句を入力として，これを平易な句に置き換える手法を探究する点に特色がある．また，単語単位の平易化に関する多くの先行研究では，平易化の際に難易語が含まれる文脈を考慮しないが，本研究では，難解な句を言い換える際に，それが出現する文脈に応じて適切な平易句を選択する．

また，本研究では，2.1.1 項にて紹介した Simple PPDB: Japanese[9] と SNOW D2[25] を利用して平易化のための辞書を構築する．ただし，SNOW D2 は平易な言い換え辞書ではないため，日本語教育語彙表 [27] を用いて単語の難易度を判定し，難解語と平易語の組を選別して使用する．具体的な方法は 3.2 節にて説明する．

平易化の手法を評価する際には，他の研究と公平な比較ができるため，公開されているデータセットを利用することが望ましい．関連研究 [10] では，Kodaira らによって構築された日本語の平易化のための評価用データセット [11] を用いて評価実験を行なっていた．このデータセットでは，1 つの文に 1 つの難解語が含まれており，難解語にはいくつかの平易語がアノテーションされている．単語に対する言い換えを評価する際には有用なデータセットだが，本研究のような句，すな

[3]非公開資料である．

わち複数の単語を言い換えるような手法の評価に利用することはできない．そのため，本研究では独自に人手で評価用データを作成し，評価実験を行なう．

第3章　提案手法

本章では，本研究の提案手法について説明する．3.1 節では提案手法を概説する．3.2 節では，テキスト平易化のためのデータベースを構築する方法について述べる．3.3 節では，平易化の対象とする句を検出する方法について説明する．本研究では，平易に言い換える句は「名詞-助詞-動詞」という単語列に限定する．3.4 節では，検出した句を平易な句に何通りかに言い換え，言い換え候補を生成する方法について述べる．以降，言い換え後の平易な句を「平易句」と，言い換えの対象とする難解な句を「難解句」と呼ぶ．3.5 節では，生成された複数の候補から最適な平易句を選択する方法を説明する．最後に，3.6 節では，元の文の難解句を平易句に置き換える際に文法的に正しい文になるようにするための処理について述べる．

以降の説明では,「品詞」という用語は，名詞，動詞などの単語の統語的なカテゴリを指すときもあれば，その品詞を持つ単語を指すこともあることに留意していただきたい．例えば,「名詞を言い換える」という文は，正確には「名詞を品詞とする単語を言い換える」ということを意味する．

3.1　概要

提案手法の概要を図 3.1 に示す．四角の枠で囲まれている項目は処理を表し，矢印の方向に処理が進められる．まず，いくつかの言語資源から,「平易言い換え単語対データベース」を構築する．これは，難解語と，それと置き換え可能な平易語の組を集めたデータベースである．次に，難解な文が与えられたとき，平易化の対象となる難解句を検出する．次に，難解句に対し，それと言い換え可能な平易句の候補を生成する．次に，平易句の候補の中から最適な句を選択する．最後に，選択された平易句を難解句と置き換えて，平易な文もしくは文章を出力する．

図 3.1 における青枠は，上から順に，3.2 節，3.3 節，3.4 節，3.5 節，3.6 節にて，その処理の詳細を説明する．さらに，平易言い換え単語対データベースの構築に利用した言語資源については，次節の 3.2 節にて詳細を述べる．

なお，提案手法はプログラミング言語の Python3.7.3 にて実装する．

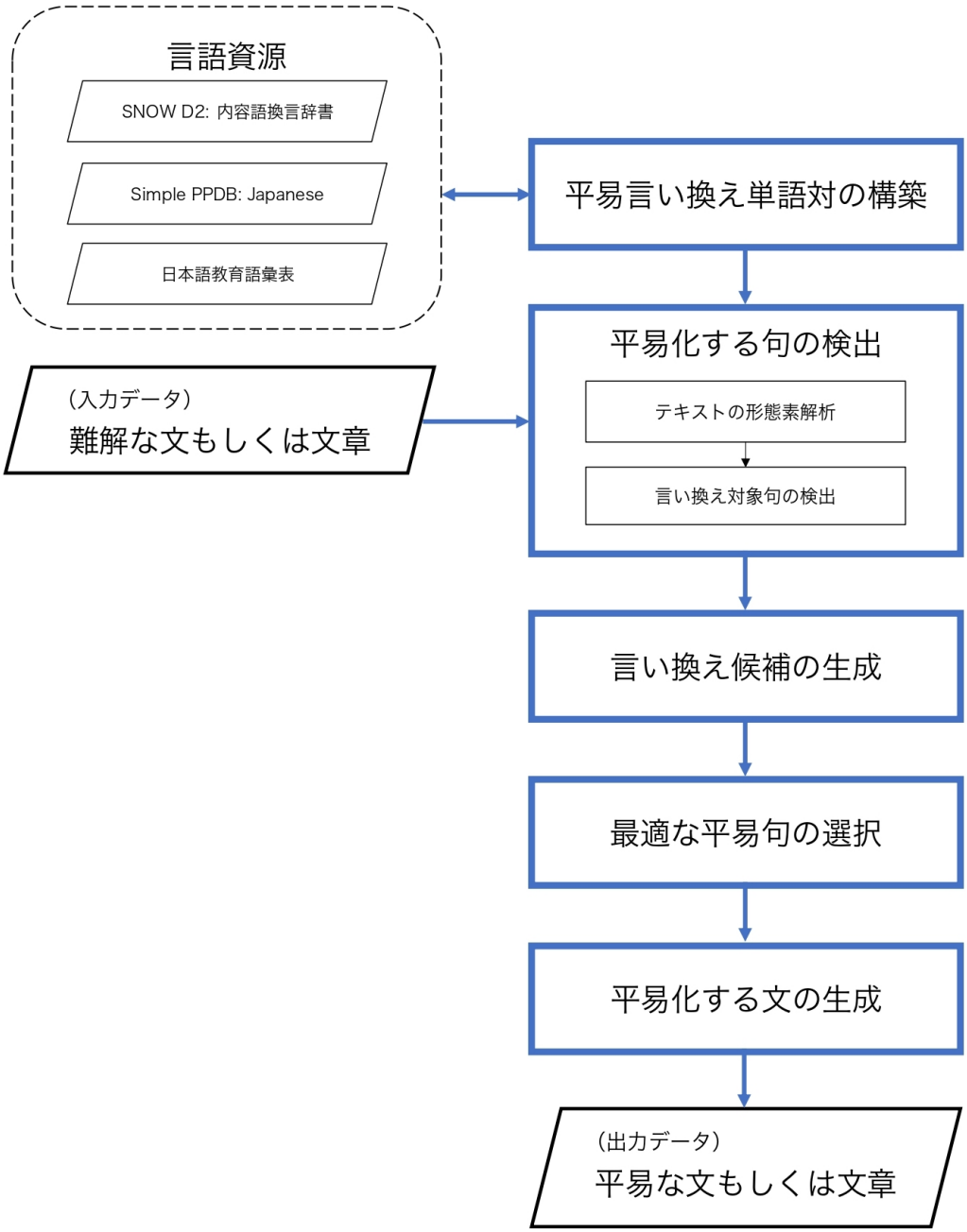
言語資源
SNOW D2: 内容語換言辞書
Simple PPDB: Japanese
日本語教育語彙表
平易言い換え単語対の構築
平易化する句の検出
テキストの形態素解析
言い換え対象句の検出
（入力データ）
難解な文もしくは文章
言い換え候補の生成
最適な平易句の選択
平易化する文の生成
（出力データ）
平易な文もしくは文章

図 3.1: 提案手法の概要

3.2 平易言い換え単語対の収集

難解句を平易句に置き換えるための知識源として，言い換え可能な難解語と平易語の組を収集したデータベースを構築する．以降，このデータベースを平易言い換え単語対と呼ぶ．平易言い換え単語対は，言い換えに関する 2 つの言語資源を用いて構築する．ひとつは SNOW D2，もう一つは Simple PPDB: Japanese である．

2 つの言語資源から平易言い換え単語対を収集する際には，2.2 節で述べた日本語教育語彙表を利用する．そのため，まずこの詳細を説明する．日本語教育語彙表に記載されている情報を表 3.1 に示す．辞書のエントリは，標準的な表記，読み，語彙の難易度，品詞，詳細品詞，語種から構成されている．このうち,「語彙の難易度」は単語の難易度であり，易しいものから順に，初級前半，初級後半，中級前半，中級後半，上級前半，上級後半の 6 つのレベルがある．1〜6 の数字は難易度のレベルを表し，数字が大きいほど難易度が高いことを表す．日本語教育語彙表は，平易言い換え単語対を収集する際，単語の難易度を判定するために利用する．

表 3.1: 日本語教育語彙表の抜粋

標準的な表記	読み	語彙の難易度	品詞 1	品詞 2(詳細)	語種
相方	アイカタ	6. 上級後半	名詞	名詞-普通名詞-一般	和語
青空	アオゾラ	3. 中級前半	名詞	名詞-普通名詞-一般	和語
語り伝える	カタリツタエル	6. 上級後半	動詞 2 類	動詞-一般	和語
返る	カエル	3. 中級前半	動詞 1 類	動詞-一般	和語
返す	カエス	2. 初級後半	動詞 1 類	動詞-一般	和語
示す	シメス	4. 中級後半	動詞 1 類	動詞-一般	和語
退任	タイニン	5. 上級前半	名詞	名詞-普通名詞-サ変可能	漢語
優待	ユウタイ	5. 上級前半	名詞	名詞-普通名詞-サ変可能	漢語
旅行	リョコウ	2. 初級後半	名詞	名詞-普通名詞-サ変可能	漢語
わびる	ワビル	5. 上級前半	動詞 2 類	動詞-一般	和語

以下，SNOW D2 ならびに Simple PPDB: Japanese から平易言い換え単語対を収集する手続きの詳細を述べる．

- SNOW D2: 内容語換言辞書

  2.1.1 項で説明した通り，SNOW D2 は，形態素解析ツール Juman の内容語項目を人手で言い換えた辞書である．SNOW D2 に記載されている情報を表 3.2 に示す．この辞書は，単語，形態素辞書情報，換言語，用例，タグ，備考で構成されている．

  本研究で平易化の対象とする句は「名詞-助詞-動詞」であり，名詞と動詞が言い換えの対象となるため，SNOW D2 から平易言い換え単語対を収集するにあたり，形態素辞書情報が普通名詞，サ変名詞，動詞のいずれかであるエントリを抽出する．エントリのうち,「単語」を難解語,「換言語」を平易語と

して抽出する．日本語教育語彙表から両方の単語の難易度を求め，言い換え前と後の単語を比べて，言い換え後の単語の難易度が低くなっていれば，平易言い換え単語対として収集する．日本語教育語彙表を利用して単語の難易度をチェックしているのは，SNOW D2 が言い換えのための辞書であり，言い換え後の単語が必ずしも言い換え前より平易でないためである．

SNOW D2 には，表 3.2 における「快打」「交わす」のように，換言語がない単語も収録されてた．これは，作業者が換言語を思いつかなかった単語である．当然ながら，このような単語は本研究においては収集対象から外し，必ず換言語がある対のみを収集する．また，表 3.2 の単語「向ける」のように，用例の欄に換言語の使用例を記載している単語が存在する．このような用例は，用例もしくはそれに類似した文を言い換えるときに言い換えの信頼性が高くなるなど，言い換えに関して有用な情報である．しかし，全ての単語に付与されていないこと，後述する Simple PPDB: Japanese にこのような情報がないことから，本研究では利用しない．

表 3.2: SNOW D2: 内容語換言辞書の抜粋

単語	形態素辞書情報	換言語	用例	タグ	備考（メモ）
相対する	動詞...	向かい合う	-	-	-
買い叩く	動詞...	買う	-	-	-
快打	名詞...	-	-	-	-
交わす	動詞...	-	-	-	多義多そう
気苦労	名詞...	精神的疲労	-	-	-
サンプル	名詞...	試供品	-	-	-
達する	動詞...	至る	-	-	-
天王山	名詞...	山	-	場所-その他	-
執る	動詞...	行なう	-	-	相当怪しい
向ける	動詞...	使う	(食費) に-	-	-

- Simple PPDB: Japanese

  2.1.1 項で説明した通り，Simple PPDB: Japanese は，日本語単語の平易な言い換え辞書である．この辞書に記載されている情報を表 3.3 に示す．辞書のエントリは，難解語，平易語，難解語の難易度，平易語の難易度，言い換え確率の組から構成される．このうち，言い換え確率は難解語から平易語への言い換えが妥当である確率を表す．PPDB: Japaense にて付与された値であり，同辞書から Simple PPDB: Japanese を構築する際にそのまま取り入れられた．なお，本研究では，辞書全体を概観し，言い換え確率が 0.1 未満のエントリは言い換えが妥当でない可能性が高いため，使用しないこととした．

  また，前述の SNOW D2 からの平易言い換え単語対の抽出処理に合わせて，難易語と平易語がともに日本語教育語彙表に掲載されていて，かつ言い換え

後の難易度が低い組を収集する．本来，Simple PPDB: Japanese には単語の難易度が付与されているため，日本語教育語彙表を利用して単語の難易度を判定する必要はないと言える．しかしながら，Simple PPDB: Japanese における難易度は，初級，中級，上級の 3 段階であり，機械学習の手法によって自動的に推定されている．すなわち，単語の難易度は必ずしも正確とは限らない．さらに，表 3.3 の難解語「審美的」のように，難解語と平易語の難易度が全く同じ言い換えも存在した．また，前述の通り，日本語教育語彙表は 6 段階の難易度が定められており，Simple PPDB: Japanese における 3 段階の難易度よりも粒度が細かい．以上の理由から，Simple PPDB: Japanese についても日本語教育語彙表による単語の難易度のチェックを行ない，難易度が低くなる単語の組を平易言い換え単語対として収集する．

表 3.3: Simple PPDB: Japanese の抜粋

言い換え確率	難解語	平易語	難解語の難易度	平易語の難易度
0.968	講評	レビュー	3	2
0.732	洋琴	ピアノ	3	1
0.544	息づかい	呼吸	3	2
0.474	川下	下流	3	2
0.420	アド	広告	2	2
0.409	プロポーザル	提案	3	2
0.350	汲み上げ	ポンプ	3	2
0.333	打ち付ける	止める	3	2
0.293	審美的	美学	3	3
0.100	選び出す	選ぶ	3	1

自動詞と他動詞の言い換え

本研究は「名詞-助詞-動詞」といった句を平易に言い換えることを目的とし，その処理には動詞を平易な動詞に言い換える処理を含む．このとき，先に述べた方法で収集した平易言い換え単語対を用いて動詞を平易化する際には注意が必要である．収集された平易言い換え単語対の中には，自動詞と他動詞の組が含まれる．例えば，自動詞「返る」を難解語，他動詞「返す」を平易語とする単語対が存在する．日本語教育語彙表では,「返る」は「返す」よりも難易度が高いと定義されている．しかし，直観的には，自動詞から他動詞への言い換え (あるいは他動詞から自動詞への言い換え) は，平易になっているとは言い難い．自動詞も他動詞も基本的な意味は共通しているからである．さらに，自動詞を他動詞に単純に置き換えると，文法的に不自然になることが多い．例えば,「童心に返る」という句の動

詞「返る」を「返す」に置き換えると,「童心に返す」となり，不自然である．いわゆる助詞の交替により，文法的に正しくない句が生成される．

上記を考慮し，自動詞から他動詞もしくは他動詞から自動詞への言い換えを行なわないようにするために，平易言い換え単語対データベースの中から自動詞と他動詞の関係にあるものを削除する．これを実現するために，難解語の動詞と平易語の動詞が自動詞と他動詞の関係にあるかを 2 つの方法でチェックする．ひとつは自・他判別辞書を利用する方法，もうひとつは自動詞と他動詞の関係にあるかを判定するルールを利用する方法である．

自・他動詞判別辞書は，ウェブ上の公開資料 [1, 2] を用いて，対となる自動詞と他動詞を収集することで構築した．収集された自動詞と他動詞の組の数は 79 と多くはないが，よく使われる基本的な自動詞・他動詞を含んでおり，実用的な場面でも有効に利用できると考える．構築した自・他動詞判別辞書を付録 A の表 A.1 に掲載する．収集した平易言い換え単語対のうち，自・他動詞判別辞書に記載されている組は，データベースから除外する．

辞書のほか，2 つの動詞が自動詞と他動詞の関係にあるかを自動的に判別するルールも併用する．ルールの例を以下に示す．

-eru (自動詞) -asu (他動詞)

この規則は，自動詞をローマ字で表記したときに末尾が -eru であるとき，それを -asu に置き換えると，対応する他動詞が得られることを表す．例えば，自動詞「出る」のローマ字表記は deru であり，他動詞「出す」のローマ字表記は dasu である．上記の規則で deru の末尾を置換すると dasu になるため,「出る」と「出す」は自動詞と他動詞の関係にあることがわかる．このような自動詞と他動詞の関係を判定するルールを 11 個作成した．その一覧を付録 A の表 A.2 に示す．収集した平易言い換え単語対の全てについて，判定ルールを適用し，自動詞と他動詞の関係にあると判定された単語対を除外する．

平易言い換え単語対データベース

上記の手続きで，最終的に 11,769 件のデータを含む平易言い換え単語対のデータベースを構築した．2 つの言語資源のそれぞれから抽出された平易言い換え単語対の数を表 3.4 に示す．両者から同じ単語対が抽出されることがあるため，表 3.4 における 2 行目と 3 行目の数値の和は 11,769 と一致しない．平易言い換え単語対の大部分が Simple PPDB: Japanese から収集されている．

データベースの構成を表 3.5 に示す．各エントリは，難解語，平易語，難解語の難易度，平易語の難易度，出典から構成されている．単語の難易度は日本語教育語彙表における難易度のレベルである．出典は，平易言い換え単語対がどちらの言語資源から収集されたかを表す.「SNOWD2」は SNOW D2 を,「SPPDB」は Simple PPDB: Japanese を表す．両方の言語資源から取得された単語対は，出典の列には

両方を併記する．ただし，以降の処理では，本データベースは難解語を平易語に置き換えるためだけに用い，難易度や出典の情報は利用しない．平易化の際には，難解語と平易語の難易度の差の情報を利用することも考えられるが，本論文では研究の対象外とする．句の平易化における単語の難易度の利用は今後の課題とする．

表 3.4: 平易単語言い換え単語対データベースの概要

言語資源	単語対の数
SNOW D2	1,734
Simple PPDB: Japanese	10,351
両方	11,769

3.3 平易化する句の検出

平易化する句の検出は，大きく分けて 2 つの段階から構成されている．3.3.1 項では，平易化対象の文もしくは文章に対する前処理について説明する．3.3.2 項では，難解語から平易語を取得する方法を説明する．

3.3.1 テキストの形態素解析

提案手法における入力は，難解語を含んでいるとされる文もしくは文章である．文章が入力された場合には，基本的には句点を検出し，分割して文として扱う．以降，文と文章をまとめて指す場合はテキストと呼ぶ．

前処理としてテキストの形態素解析を行なう．形態素解析ツールには，一般的に使われているオープンソースの MeCab[14] を採用する．形態素解析は，内包されている辞書にしたがって，文を言語的に意味を持つ最小単位の形態素に分解する処理である．日本語の場合は単語に分割する処理と言える．MeCab に内包されている標準のシステム辞書は IPA 辞書である．IPA 辞書は，基本的な日本語の単語は網羅しているが，一般的に固有名詞や新語が登録されていないという問題が指摘されている．一方，MeCab では，ユーザは目的によって適切な辞書を選択できる．mecab-ipadic-NEologd[21] は MeCab で利用可能な形態素解析用辞書のひとつである．この辞書は，多数のウェブ上の言語資源からさまざまな単語を収集している．現在も定期的に更新されているため，比較的新しい単語や IPA 辞書では網羅できない固有名詞を多く含むとされるテキストに対する形態素解析に頻繁に利用されている．本研究は句を平易に言い換えることを目的としているが，形態素解析の誤りによって，言い換えるべきではない句や単語を言い換えることは避けるべきである．例を挙げると,「人口に膾炙する」のような慣用句表現は，名詞もしくは動詞の一方を言い換えると，意味が変わってしまう.「人口に膾炙する」

表 3.5: 収集した平易言い換え単語対の抜粋

難解語	平易語	難解語の難易度	平易語の難易度	出典
しのぐ	避ける	6	4	SPPDB
コミック	漫画	3	2	SNOWD2・SPPDB
ステータス	状態	5	3	SPPDB
メタル	金属製	6	2	SPPDB
付ける	見る	3	2	SPPDB
参じる	行く	3	1	SPPDB
和	日本	4	1	SPPDB
図柄	デザイン	5	3	SPPDB
媚びる	持ち上げる	3	2	SPPDB
実る	出来る	5	2	SNOWD2
形状	形	5	3	SNOWD2・SPPDB
戻す	返す	3	2	SPPDB
扱う	使う	4	2	SNOWD2
浴びせる	言う	5	1	SNOWD2・SPPDB
湯水	水	3	2	SPPDB
緯線	彼ら	3	2	SPPDB
蜂	ミツバチ	4	2	SPPDB
負わす	与える	3	2	SPPDB
起き上がる	起きる	4	1	SNOWD2
鍛練	規律	6	5	SPPDB

を IPA 辞書を用いて形態素解析した結果を表 3.6 に，mecab-ipadic-NEologd を用いて形態素解析した結果を表 3.7 に示す．IPA 辞書では慣用句が単語に分割されているのに対し，mecab-ipadic-NEologd では慣用句として 1 つの単語となっている．本研究では「名詞-助詞-動詞」といった単語列を言い換えの対象としているため，一語として解析された慣用句を誤って平易化の対象とすることはない．したがって，本研究では，MeCab の辞書として mecab-ipadic-NEologd を採用する．なお，mecab-ipadic-NEologd は頻繁に更新されているが，本研究では 2020 年 08 月 20 更新の v0.0.7 を用いる．

以上の前処理をまとめると，入力された文に対し，MeCab と mecab-ipadic-NEologd を用いた形態素解析によって文を単語に分割し，それぞれの単語と品詞を同定する．

表 3.6: 「人口に膾炙する」を形態素解析した結果 (IPA 辞書)

単語	品詞	品詞 細分類	品詞 細分類 2	品詞 細分類 3	活用型	活用形	原形	読み	発音
人口	名詞	一般	*	*	*	*	人口	ジンコウ	ジンコー
に	助詞	格助詞	一般	*	*	*	に	ニ	ニ
膾炙	名詞	サ変接続	*	*	*	*	膾炙	カイシャ	カイシャ
する	動詞	自立	*	*	サ変・スル	基本形	する	スル	スル

表 3.7: 「人口に膾炙する」を形態素解析した結果 (mecab-ipadic-NEologd)

単語	品詞	品詞 細分類	品詞 細分類 2	品詞 細分類 3	活用型	活用形	原形	読み	発音
人口に膾炙する	名詞	固有名詞	一般	*	*	*	人口に膾炙する	ジンコウニ カイシャスル	ジンコーニ カイシャスル

3.3.2 言い換え対象句の検出

形態素解析によって得られた品詞情報をもとに,「名詞-助詞-動詞」のように連続する単語を句として抽出する．具体的な処理としては，まず，形態素解析された文に対し，先頭から名詞を探索する．名詞が見つかった場合，後続する単語がそれぞれ助詞，動詞かを判定する．名詞，助詞，動詞を検出できたら，これを言い換え対象の句の候補とする．また，助詞以前に名詞が連続していた場合は，複合名詞としてまとめて扱う．これは，本来ひとつの単語として扱われるべき単語が形態素解析によって複数の単語に分割された場合に対応するためである．特に名詞性接尾辞「-性」「-的」は独立した単語として解析されるため，これより前に出現する名詞とつなげて複合名詞として扱う処理は重要である．複合名詞を検出した後は，単語の名詞を検出したときと同様に，これに続く助詞と動詞を順に検出する．上記の手続きをフローチャートとしてまとめたものを図 3.2 に示す．なお，図 3.2 における単語 n とは，句の探索の起点となる単語を指している．単語 n を文の先頭から末尾まで移動させ，それぞれの起点から「名詞-助詞-動詞」といった句を検出する．

次に，句として検出された名詞と動詞の単語がともに平易言い換え単語対データベースに難解語として登録されているかを判定する．この際，データベースでは動詞は全て基本形で収録されているため，句における動詞を基本形に変換してから探索を行なう．名詞，動詞ともに平易言い換え単語対データベースに登録されているならば，その句を平易化の対象句として検出する．同時に，その難解語に対応する平易語（一般に複数ある）を取得する．また，句の言い換えの際に文脈を考慮するため，これを含む文全体も取得する．

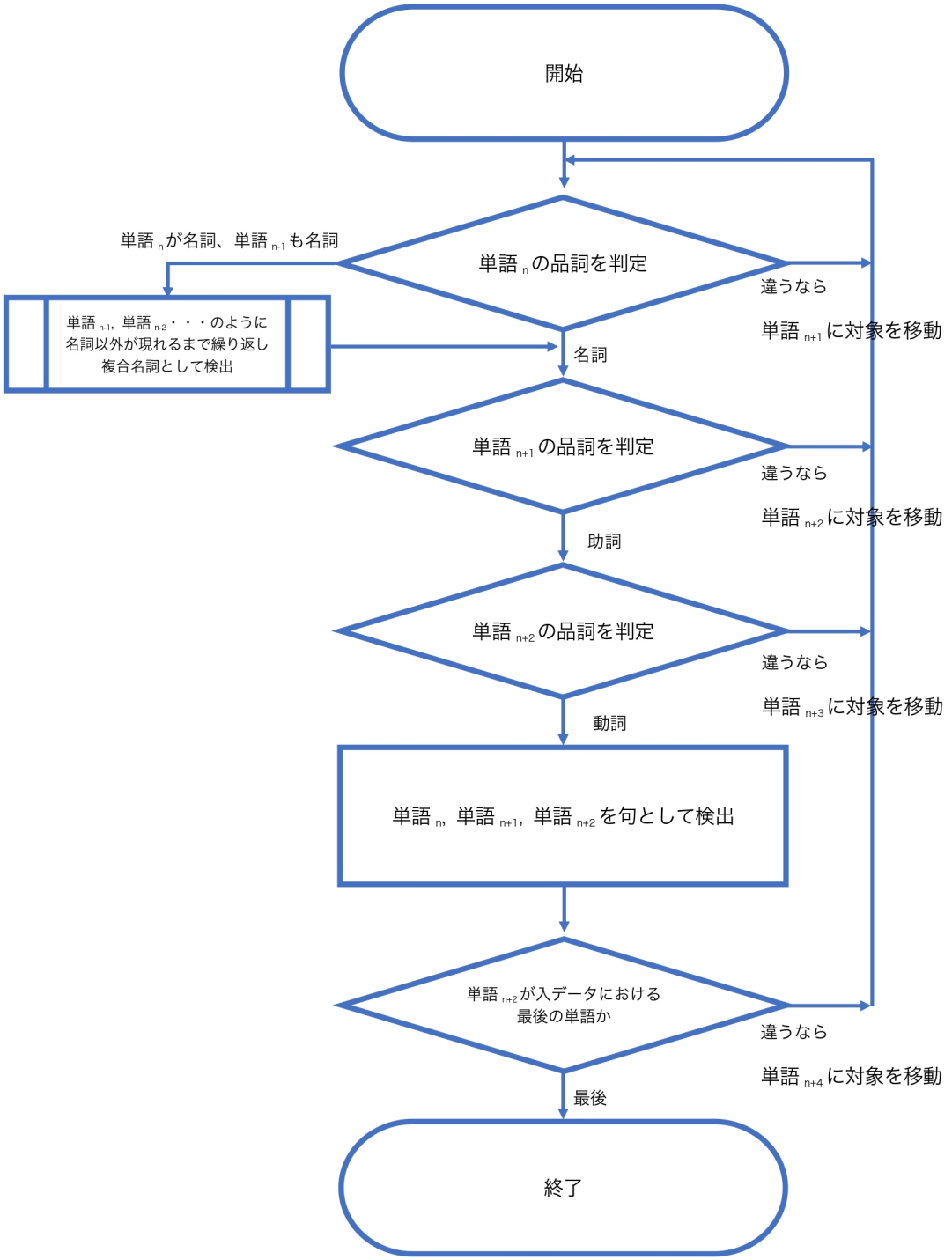

図 3.2: 句の検出のフローチャート

3.4 言い換え候補の生成

検出された難解句「名詞-助詞-動詞」に対し，これを平易に言い換えた平易句の候補を生成する．基本的な方針として，句「名詞-助詞-動詞」の名詞と動詞を両方とも平易に言い換える．しかし，このような言い換えが必ずしも良い平易化とは限らない．そこで，名詞もしくは動詞のいずれかのみを言い換えることも考慮する．したがって，以下に示す 3 種類のパターンによって平易句の候補を生成する．

- 名詞，動詞の言い換え

  名詞と動詞を，平易言い換え単語対に基づき，平易語に言い換える．
  (例) 果敢-に-挑む → 積極的-に-挑戦する

- 名詞のみの言い換え

  名詞のみ，平易言い換え単語対に基づき，平易語に言い換える．
  (例) 果敢-に-挑む → 積極的-に-挑む

- 動詞のみの言い換え

  動詞のみ，平易言い換え単語対に基づき，平易語に言い換える．
  (例) 果敢-に-挑む → 果敢-に-挑戦する

言い換え対象の難解句を検出する際には，名詞，動詞がともに平易言い換え単語対データベースに登録されていることを確認しているため，上記のパターンにより少なくとも 3 つの平易句の候補が必ず生成される．また，平易言い換え単語対データベースでは，1 つの難解語に対して複数の平易語が登録されていることがあるため，ひとつの難解句に対する平易句の候補の数は 3 より大きくなることがある．

3.5 最適な平易句の選択

前節で説明した 3 種類のパターンで生成した言い換え候補のそれぞれについて，その妥当性をスコア付けする．そのスコアに従い，最終的に最も妥当な平易句の候補をひとつ選択する．

スコア付けをするにあたり，平易化の対象句（難解句）を $Ph = N\text{-}P\text{-}V$ と定義する．ここで N は名詞，P は助詞，V は動詞を示す．これに対し，平易言い換え単語対データベースを用いて，名詞と動詞の両方，名詞のみ，動詞のみ，を平易語に言い換えた平易句の候補は，$Ph'_i = N'_i\text{-}P\text{-}V'_i$ と表す．名詞のみを言い換えたときは V と V' は同じ単語であり，動詞のみを言い換えたときは N と N' は同じ単語である．次に，Ph'_i のそれぞれのスコアを計算し，それが最大となる Ph'_i を選択する．

平易句の候補に対するそのスコアは 2 つの観点から算出する．

- **正確性**
  言い換え前後で意味が保持されているかという観点

- **流暢性**
  言い換え後の句がどれだけ自然な句であるかという観点

次の 3.5.1 項にて正確性，3.5.2 項にて流暢性を考慮した平易句の候補のスコア付けを述べる．さらに，3.5.3 項にて，両方の観点を踏まえたスコア付けについて説明する．

3.5.1 正確性のスコア

正確性は，平易句に言い換えた後の文が元の文と同じ意味を持つかという観点で平易句の候補の品質を評価する基準である．S を元の句 Ph を含む文，S'_i を Ph を Ph'_i に言い換えた後の文とする．正確性 (faithfulness) のスコア $FA(Ph'_i)$ は式 (3.1) のように定義する．

$$FA(Ph'_i) = cos(\vec{S}, \vec{S'_i}) \tag{3.1}$$

$\vec{S}, \vec{S'_i}$ は文 S，S'_i の分散表現，cos はコサイン類似度である．すなわち，平易化前の文と後の文の分散表現が似ているほど，Ph'_i の正確性は高いとする．言い換え対象の句だけでなく，それを含む文全体の類似度を考慮することで，言い換え後の平易句が文脈に沿って適しているかを評価する．言い換え後の文 $\vec{S'_i}$ は，単に難解句 Ph を平易句 Ph'_i に置き換えるだけでなく，置き換えた後の文が文法的に正しくなるような後処理を行なう．難解句を平易句に置き換える処理の詳細は 3.6 節で述べる．

文の分散表現は，文の意味を表す抽象表現であり，ベクトルで表現される．文の分散表現を得るための様々な手法が存在するが，本研究では Sentence BERT[20] を用いて文の分散表現を獲得する．Sentence BERT は，事前学習した BERT(Bidirectional Encoder Representationsfrom Transformers) を基に，Siamese Network を用いた fine-tuning によって文の分散表現を学習するモデルである．Sentence BERT は BERT を基にした手法であるが，BERT を利用しても文の分散表現を獲得できる．ただし，文の分散表現を得るための計算コストが高く，処理時間が長くなり，大量の文に対して分散表現を得ることには適していない．Sentence BERT はこの問題を解決するために考案された手法であり，低コストで文の分散表現を得られるほか，自然言語処理におけるいくつかのタスクで BERT による文の分散表現を上回る結果が得られている．本研究では，事前学習された Sentence BERT 日本語モデル [22] を利用する．

3.5.2 流暢性のスコア

流暢性は，平易句がどれだけ自然な表現であるかという観点で平易句の候補の品質を評価する基準である．特に，名詞と動詞の両方を置き換えて平易句の候補を生成する際，名詞と動詞は互いに独立に置き換えられているため，不自然な平易句の候補が生成されることも多い．したがって流暢性の観点からのスコア付けは重要である．言い換え後の句 Ph'_i の流暢性 (fluency) のスコア $FL(Ph')$ は，式 (3.2) のように定義する．

$$FL(Ph'_i) = \frac{count(Ph'_i)}{\sum_j count(Ph_j)} \tag{3.2}$$

式 (3.2) の分子は句 Ph'_i のコーパスにおける出現頻度，分母は全ての句の出現頻度の和である．すなわち，句 Ph'_i がコーパスによく出現するほど，その句は自然な表現の句であるとみなす．

句の出現頻度 $count(Ph_j)$ は京都大学格フレーム Ver 2.0[26] を用いて算出する．京都大学格フレームとは，ウェブから収集した日本語 100 億文から自動構築した格フレーム辞書である．格フレームとは，用言とそれが取り得る格をパターンとしてまとめたものである．日本語では，一般に格は助詞で表される．例えば,「食べる」という動詞は「が」と「を」という 2 つの格を持つ．さらに，格フレームでは，それぞれの格に出現する名詞の情報が記載されることが多い．京都大学格フレームでは，それぞれの格として出現する名詞が列挙されている．これらの名詞はウェブから自動収集されたものであり，その頻度も掲載されている．例えば,「食べる」の「を」格に出現する名詞として「りんご」「みかん」「寿司」などがその頻度とともに掲載されている．この頻度は,「りんご-を-食べる」といった名詞，助詞，動詞の組み合わせが出現する頻度と解釈できる．本研究では $count(Ph')$，すなわち句 N-P-V の頻度は，京都大学格フレームにおける動詞 V の格フレームにおける格 P に出現する名詞 N の頻度とする．また，$\sum_j count(Ph_j)$ は京都大学格フレームにおける全ての格フレームにおける句の出現頻度の総和とする．これで $count(Ph_j)$ を除することにより，流暢性のスコアを 0 から 1 の範囲にスケーリングすることができる．なお，京都大学格フレームでは，$\sum_j count(Ph_j)$ は $1,797,162,224$ である．

京都大学格フレームにおける格フレームの例を図 3.3 に示す．例えば，平易句の候補が「リクエスト-に-答える」の場合，この句における名詞，助詞，動詞は黄色でマークした箇所に該当する．緑でマークした $3,756$ は「リクエスト-に-答える」という句の出現頻度となる．これを正規化すると 2.09×10^{-6} となる．

平易句の候補の出現頻度を求める際，その句が京都大学格フレームに出現しないこともありうる．このとき，流暢性のスコアを計算することができないため，難解句を平易化の対象から除外する．つまり，生成された平易句の候補のうちひとつでも京都大学格フレームに記載されていない句があれば，その難解句は言い換えができないとし，言い換えの対象から除外する．京都大学格フレームは大規模な格フレーム辞書であり，これに掲載されていない句が平易化の対象となること

```
      ⋮
  <caseframe id="答える/こたえる+する/する:動 2" frequency="21839">
    <argument frequency="21839" case="二格">
      <component frequency="9137">要望/ようぼう</component>
      <component frequency="5099">ニーズ/にーず</component>
      <component frequency="3756">リクエスト/りくえすと</component>
      ⋮
<entry headword="答える/こたえる+する/する+かねる/かねる" predtype="動">
  <caseframe id="答える/こたえる+する/する+かねる/かねる:動 1" frequency="1204">
    <argument frequency="1204" case="ニ格">
      <component frequency="760">質問/しつもん</component>
      <component frequency="348">問い合わせ/といあわせ v</component>
      <component frequency="43">要望/ようぼう</component>
      <component frequency="30">相談/そうだん</component>
      <component frequency="12">リクエスト/りくえすと</component>
      <component frequency="11">苦情/くじょう</component>
    </argument>
      ⋮
<entry headword="答える/こたえる+する/する+こと/こと+出来る/できる" predtype="動">
  <caseframe id="答える/こたえる+する/する+こと/こと+出来る/できる:動 1" frequency="2298">
      ⋮
  <caseframe id="答える/こたえる+する/する+こと/こと+出来る/できる:動 2" frequency="841">
    <argument frequency="841" case="ニ格">
      <component frequency="409">要望/ようぼう</component>
      <component frequency="141">ニーズ/にーず</component>
      <component frequency="100">リクエスト/りくえすと</component>
      <component frequency="64">疑問/ぎもん</component>
      ⋮
```

図 3.3: 京都大学格フレーム Ver 2.0 の抜粋

は多くはない．しかし，京都大学格フレームから句の出現頻度が得られないとき，流暢性のスコアをどのように算出するかは今後の課題である．

3.5.3 平易句の候補のスコア

最終的に平易句の候補のスコアは $FA(Ph')$ と $FL(Ph')$ の対数の重み付き和で算出する．本研究では，$Score1$ と $Score2$ の 2 通りのスコアを提案する．

Score1

$Score1$ の定義を式 (3.3) に示す．

$$Score1(Ph'_i) = \alpha \log FA(Ph'_i) + \log FL(Ph'_i) \tag{3.3}$$

このスコアは，正確性のスコアと流暢性のスコアの対数の重み付き和で平易句の候補の品質を評価することを表す．正確性のスコア $FA(Ph'_i)$ は言い換え前後の文の分散表現同士によるコサイン類似度なので，多くは 1 に近い値を取る．対して，

流暢性のスコア $FL(Ph'_i)$ は句の相対出現頻度であるため，非常に小さな値を取る．パラメータ α は重みであり，$Score1$ では，両者のスケールを合わせるために用いる．これは，正確性と流暢性のスコアを同程度に評価することを意味する．

α は開発データを用いて算出する．開発データとは，ここでは，平易化前の句 Ph を含む文，平易化後の句 Ph' を含む文，それらに対するアノテーションから成るデータセットであるとする．その詳細は 4.1.1 項にて詳述する．α を決定するために，開発データにおける全ての事例について $FA(Ph')$，$FL(Ph')$ を算出する．そして，それぞれの中央値 Me_{fa} と Me_{fl} を求める．最終的に α の値は，式 (3.4) のように算出する．これにより，正確性と流暢性のスコアをスケールを合わせて加算することができる．

$$\alpha = \frac{Me_{fl}}{Me_{fa}} \tag{3.4}$$

Score2

Score2 の定義を式 (3.5) に示す．

$$Score2(Ph'_i) = \beta\, \alpha \log FA(Ph'_i) + (1-\beta) \log FL(Ph'_i) \tag{3.5}$$

$Score2$ におけるパラメータ α は，$Score1$ と同様に，2 つのスコアのスケールを合わせるために用いる．一方，パラメータ β は，正確性と流暢性のスコアに対する重みとして働く．第 4 章で示す評価実験では，$\beta = 0.25, 0.5, 0.75$ と設定する．例えば，$\beta = 0.25$ の場合，正確性よりも流暢性に重みをおいたスコア付けとなる．$\beta = 0.75$ のときは逆に流暢性よりも正確性を重視する．$\beta = 0.5$ のときは，正確性と流暢性のスコアを同程度に評価することになり，このときの計算式は $Score1$ と同じである．

3.5.4 名詞・動詞の言い換えを優先した平易句の選択

前項で説明した平易句の候補のスコアを計算するシステムを実装後，その動作を確認する予備実験を行なった．具体的には，提案した評価式による候補選択と筆者が期待する候補が一致するかをおおまかに確認した．その結果，式 (3.3) または式 (3.5) が最大となる句を選択したところ，名詞と動詞の両方を言い換える句よりも，名詞もしくは動詞のいずれかを言い換えた句が多く選ばれる傾向が見られた．これは，正確性のスコア FA の性質に起因すると考えられる．正確性は言い換え前後の文の類似度で測っているため，名詞と動詞の両方の単語を言い換えた場合よりも 1 つだけ言い換えた場合の方が，高くなることが多い．しかしながら，本研究の目的は句を対象とした言い換えであり，名詞と動詞の両方を言い換えた句をできるだけ生成したい．

以上の背景から，候補の中から平易句をひとつ選択する際，$Score1$ または $Score2$ を用いるだけでなく，名詞と動詞を両方とも言い換えた平易句を優先して選択する手法を提案する．具体的には，以下の手続きにしたがい，言い換え候補の中から句を選択する．

1. 名詞と動詞の両方を言い換えた平易句のスコアの値が閾値 T よりも大きいとき，つまり平易化された句として妥当である可能性が高いとき，その候補を選択する．

2. 上記に該当する句がないとき，全ての候補の中からスコアが最大のものを選択する．

閾値 T は $Score1$ と $Score2$ で個別に設定する．

- $Score1$ のとき

  閾値 T を決定するための開発データを用意する．ここでの開発データとは，難解句 Ph，(自動生成された) 平易句 Ph'，それに対するアノテーションである．ここでのアノテーションとは，言い換え後の平易句 Ph' が正確性を満たすか否か (元の難解句と比べて意味が変わっていないか否か)，流暢性を満たすか否か (平易句が自然な表現であるか否か) のラベルである．開発データの詳細は 4.1.1 項で述べる．

  正確性のスコア $\log FA(Ph')$ が閾値以上のときに平易句 Ph' を生成し，それ以外のときは平易句を生成しないシステムを作成する．このシステムを用いて，開発データの難解句を平易句に言い換える．生成された平易句は正確性を満たすと予測し，その予測が人手による正確性を満たすか満たさないかの判定と一致する割合 (精度) を調べる．閾値を高く設定すると，正確性の高い平易句のみが生成されるので，精度は高くなると予想される．一方，閾値を低く設定すると，より多くの難解句に対して平易句を生成できるが，正確性の低い平易句も生成されるので，精度は低くなると予想される．閾値を変動させ，精度の変化を確認し，精度が十分に大きいときの閾値を T_{fa} とおく．T_{fa} より正確性のスコアが高いと，平易句は真に正確性を満たす可能性が高いといえる．

  同様に，流用性のスコア $\log FL(Ph')$ が閾値以上のときに平易句 Ph' を生成し，それ以外のときは平易句を生成しないシステムを作成する．このシステムを用いたときの流暢な平易句を生成できたという予測が人手による流暢性を満たすか満たさないかの判定と一致する割合 (精度) を調べる．閾値を変動させ，精度の変化を確認し，精度が十分に大きいときの閾値を T_{fl} とおく．T_{fl} より流暢性のスコアが高いと，平易句は真に流暢性を満たす可能性が高いといえる．

最終的に閾値Tは式(3.6)のように決定する．平易句の候補がこの閾値Tを越えたとき，それは正確性，流暢性をともに満たす可能性が高いと考えられる．

$$T = \alpha \times \log T_{fa} + \log T_{fl} \tag{3.6}$$

- $Score2$のとき

  閾値Tを決定するための開発データを用意する．ここでの開発データとは，難解句Ph，(自動生成された)平易句Ph'，それに対するアノテーションである．$Score1$と同じく，アノテーションとは，言い換え後の平易句Ph'が正確性を満たすか否か，流暢性を満たすか否かのラベルである．

  スコア$Score2(Ph')$が閾値以上のときに平易句Ph'を生成し，それ以外のときは平易句を生成しないシステムを作成する．このシステムを用いて，開発データの難解句を平易句に言い換える．生成された平易句は全て妥当と予測し，その予測が人手による平易句の妥当性の判定と一致する割合(精度)を調べる．ここでは，人手によって正確性と流暢性の両方が妥当と判断されたとき，その平易句は妥当な平易句であるとして扱う．先ほどと同様に，この閾値を高く設定すると，平易句を生成できない場合が増えるが，平易句を生成できたときはそれが妥当な平易句である可能性が高くなる．閾値を変動させ，精度の変化を確認し，精度が十分に大きいときの閾値をTとおく．

以上の手続きによって決定された閾値の値については4.2節で報告する．

3.6 平易化する文の生成

本節では，平易句を選択した後，それを元の文の難易句と置き換えて，平易化された文を生成する手続きについて説明する．

基本的には，平易化された文の生成は，難易句を平易句に置き換えることで行なう．ただし，この処理だけでは文法的に正しくない文が生成される可能性がある．平易言い換え単語対データベースに登録されている平易語の動詞は基本形で登録されている．動詞を置き換える際，単純な置換では動詞は必ず基本形に置き換えられるが，元の文で動詞が基本形以外の形で現れていると，文法的に矛盾が生じる．そのため，動詞を言い換える際，元の文の動詞の活用形に合わせて，動詞の平易語も活用させる処理が必要である，先に説明した3種類の平易句のうち，名詞のみを言い換える場合においては単純に難解語を平易語へ言い換えれば問題ないが，名詞・動詞の両方の言い換え，動詞のみの言い換えの場合では，単純な置換では不十分であり，動詞の活用の処理が必要である．

動詞の活用形の処理は以下のように行なう．まず，言い換え前の句における動詞の活用形を判定する．MeCabで付与された品詞情報の中には活用形も含まれて

いるため，これを参照する．活用形の種類としては，未然形，連用形，終止形，連体形，仮定形，命令形がある．このうち，活用形が終止形以外であった場合，元の文に適用する際に活用の処理が必要である．次に，言い換え後の動詞，つまり平易言い換え単語対データベースから取得した平易語の活用型を調べる．活用型には，五段活用，上一段活用，下一段活用，カ行変格活用，サ行変格活用がある．動詞の平易語の活用型は MeCab の辞書によって判定する．このようにして得られた動詞の活用型と活用形の情報をもとに，動詞の平易語の活用形を決定する．最後に，活用させた動詞を元の文の動詞の難解語と置き換えて，平易化された文の生成を終了する．

具体例を用いて，平易化する文を生成する処理の流れを説明する．なお，以下の例文は毎日新聞の 2013 年の記事データから抽出した．

1. 言い換え前の句およびそれを含む文の取得

   「デフレ脱却の名の下に大胆な金融緩和と公共投資の**必要性を唱え**ている.」
   言い換え対象の難解句は「必要性を唱え」である．

2. 動詞の難解語の活用形の判定

   「デフレ脱却の名の下に大胆な金融緩和と公共投資の必要性を**唱え**ている.」

   唱えるの活用形は連用形である．

3. 平易言い換え単語対データベースから平易語を取得

   必要性の平易語: 必要

   唱えるの平易語: 言う

4. 動詞の平易語の活用型を判定

   言うの活用型は五段活用・ワ行促音便である．

5. 動詞の平易語の活用形を決める

   言う→言っ

6. 名詞，活用させた動詞に置き換えて，平易化された文を生成

   「デフレ脱却の名の下に大胆な金融緩和と公共投資の**必要を言っ**ている.」

第4章　評価実験

提案手法の評価実験を行なった．まず，4.1 節では，実験のために構築したデータセットについて説明する．4.2 節では，提案手法のパラメータの決定について述べる．4.3 節では，複数の候補から最適な平易句を選択する処理を評価する．4.4 節では，提案手法によって難解句を平易句に言い換えた結果を評価する．4.5 節では，提案手法についてのエラー分析を行なう．

4.1　データセット

実験にあたり，開発データと評価データの 2 つのデータセットを作成した．まず，4.1.1 項では開発データについて説明する．これは，3.5.3 項で述べた評価式のパラメータ α や 3.5.4 項で述べた閾値 T の最適化に利用する．次に，4.1.2 項では，提案手法の有効性を評価するために作成した評価データについて説明する．

どちらのデータセットも毎日新聞の 2013 年版の記事を利用して作成した．新聞記事は難解語を含んだテキストとして広く利用されている [12]．毎日新聞の記事データには，発刊日，タイトル，本文，キーワードなど様々な情報が記載されている．また，これらの種類を示すタグも付与されている．まず，記事本文を示すタグが付与されているテキスト，すなわち記事本文を抽出した．記事の件数は，106,305 件であった．次に，記事を文に分割した．文の数は 1,080,590 件であった．本実験ではこれらの文に対して平易化を行なう．

4.1.1　開発データ

毎日新聞の 2013 年の記事から取得した約 100 万件の文に対し，3.3 節で述べた手法を用いて難解句を検出し，それを含む文とともに抽出した．そこから無作為に 500 件を選択した．このとき，同じ難解句は一つだけ抽出することにした．例えば,「姿勢を示す」という難解句は，実際には毎日新聞の記事の中で頻出するが，開発データには一度しか含まれないようにした．ただし，提案手法では難解句が出現する文脈も考慮して平易化するため，同じ難解句に対しても，それを含む文が異なれば，異なる平易句を生成する．したがって，重複を考えずに無作為に 500 件をサンプリングすることも妥当である．しかし，多くの種類の難解句に対する平易化の性能を測ることを優先し，ここでは重複を許さずに 500 件を選択した．

500 件の難解句を含む文に対しては，提案手法による言い換え候補が少なくとも 3 件存在する．名詞・動詞を両方とも言い換える候補，名詞のみを言い換える候補，動詞のみを言い換える候補の 3 つである．それらの言い換え候補の中から無作為に一つの候補を決定した．そして，その平易句に対して以下の判定を人手で行なった．

- 平易化した文の正確性

  言い換え前後の文を比較し，正確性を満たしているか否かを判定

  すなわち，言い換え後の文が前の文の意味を保持していれば正確性を満たすと判定する．

- 平易句の流暢性

  言い換え後の平易句が，流暢性を満たしているか否かを判定

  すなわち，言い換え後の文が自然な文であれば流暢性を満たすと判定する．

判定は作業者 1 名が行なった．判定結果のアノテーションの際には，提案手法が算出する正確性や流暢性のスコアを提示すると判定に影響を及ぼすと考え，作業者には提示しなかった．作成された開発データの一部（5 例）を表 4.1 に示す．○は正確性および流暢性を満たすことを，×は満たさないことを表す．

4.1.2　評価データ

次に，評価データを作成した．開発データと同じく，毎日新聞の 2013 年版の記事本文から抽出された平易句対象の難解句およびそれを含む文の組の中から，100 件を選択した．このとき，開発データと異なる難解句を選択した．また，開発データと同様に，同じ難解句は重複して選択しなかった．この評価データに対し，作業者 2 名により，以下の 2 つの作業によるアノテーションを行なった．

1. 複数の平易句の候補の中から最適なものを選択する
2. 提案手法で言い換えられた平易句の品質を評価する

これらのアノテーションは，それぞれ 4.3 節，4.4 節の評価実験で使用する．そのため，アノテーションの詳細はそれぞれの節で説明する．

4.2　パラメータの決定

提案手法ではいくつかのパラメータがあるが，本節では開発データを用いた各種パラメータの決定について説明する．一つ目のパラメータは，3.5.3 項で述べた α

表 4.1: 開発データ（抜粋）

	文	正確性	流暢性
言い換え前	「伝える」という**役目を背負い**，この人はマイクに向かっている．	-	-
言い換え後 (名詞・動詞)	「伝える」という**仕事を持ち**，この人はマイクに向かっている．	○	○
言い換え前	だが,「暑くなり過ぎると，逆に**外食に出掛ける**機会が減り，ビール消費に影響しかねない」と懸念の声も上がっている．	-	-
言い換え後 (名詞・動詞)	だが,「暑くなり過ぎると，逆に**レストランに行く**機会が減り，ビール消費に影響しかねない」と懸念の声も上がっている．	×	○
言い換え前	不測の**事態に備えた**ためとも言われる．	-	-
言い換え後 (名詞のみ)	不測の**状態に備えた**ためとも言われる．	○	×
言い換え前	お客様の**要望に応じて**品ぞろえもさらに充実させていく予定だ．	-	-
言い換え後 (動詞のみ)	お客様の**要望に答えて**品ぞろえもさらに充実させていく予定だ．	×	○
言い換え前	地蜂で生計を立て立ち直る，という**物語を思い描いて**，根掘り葉掘り聞く自分が卑しく思えた．	-	-
言い換え後 (動詞のみ)	地蜂で生計を立て立ち直る，という**物語を思って**，根掘り葉掘り聞く自分が卑しく思えた．	○	○

である．これは正確性と流暢性のスコアのスケールを合わせるために利用する．このパラメータは提案手法における平易句の候補に対する 2 つのスコア（$Score1$ と $Score2$）のどちらに対しても同じ値を用いる．α の決定については 4.2.1 項にて説明する．もう一つのパラメータは，3.5.4 項で述べた閾値 T であり，名詞と動詞の両方を言い換えた平易句を優先して選択するために利用される．この値は，$Score1$ と $Score2$ で異なる値を設定するので，4.2.2 項と 4.2.3 項にて，それぞれ説明する．基本的な方針として，閾値 T は，スコアの値がそれを超えたとき，生成した平易句が適切である可能性が高くなる値を設定する．また，$Score2$ では，正確性と流暢性のスコアの重み付けの役割を果たすパラメータ β がある．本実験では，β は $0.25, 0.5, 0.75$ のいずれかとするが，それぞれの場合について閾値 T を個別に設定する．

4.2.1 スケール調整パラメータ α の決定

パラメータ α は式 (3.4) にて決定する．以下に再掲する．

$$\alpha = \frac{Me_{fl}}{Me_{fa}} \tag{4.1}$$

Me_{fa} と Me_{fl} は，それぞれ，開発データにおける正確性，流暢性のスコアの対数の分布の中央値である．まず，開発データにおける全ての平易句の候補 Ph'_i について，その正確性と流暢性のスコアの対数（log）を算出し，その分布を求めた．開発データにおける平易句の候補の総数は 1,669 件であった．このスコアの分布の基本統計量，ここでは平均値，中央値，最小値，最大値，分散を表 4.2 に示す．このうち α の決定に必要な統計量は中央値である．$\log FA(Ph'_i)$ と $\log FL(Ph'_i)$ の中央値は，$Me_{fa} = -0.0111$，$Me_{fa} = -16.9$ となった．これに従い，$\alpha = \frac{Me_{fl}}{Me_{fa}} = \frac{-16.9}{-0.0111} = 1.54 \times 10^3$ と決めた．

表 4.2: 開発データにおける正確性と流暢性のスコアの統計

	平均値	**中央値**	最小値	最大値	分散
$\log FA(Ph'_i)$	-0.0231	**-0.0111**	-0.326	-0.000100	0.00119
$\log FL(Ph'_i)$	-16.5	**-16.9**	-20.6	-8.85	6.06

4.2.2 $Score1$ の閾値 T の決定

$Score1$ における名詞と動詞の両方を言い換えた平易句を優先して選択するための閾値 T の決定について述べる．この決定には，開発データのうち，平易句が名詞と動詞を両方言い換えたものであるデータを利用する．開発データの全 500 件のうち，これに該当するデータは 167 件であった．開発データでは，これらに対して，平易句が正確性の観点から見て妥当か否か，流暢性の観点から見て妥当か否か，の二値ラベルが付与されている．

正確性のスコアが閾値 t のときに平易句が正確性を満たすと判定するシステムを作る．このシステムを上記の開発データのサブセットに適用し，その判定の精度，すなわち正確性と満たすと判定した事例のうち，人によって実際に正確性を満たすとラベル付けされた事例の割合を求める．閾値と精度の関係をグラフとして可視化したものを図 4.1(a) に示す．グラフの X 軸は閾値 t であり，Y 軸が判定精度である．ただし，0.90 未満のデータについては増減の幅が少なかったため描画していない．同様に，流暢性について閾値 t を変化させたときの判定精度の変化も調べた．その結果を図 4.1(b) に示す．同じように，10^{-9} 未満のデータは描画していない．基本的に，これら 2 つのグラフでは，閾値 t を大きくすると精度が高くなる傾向が見られる．

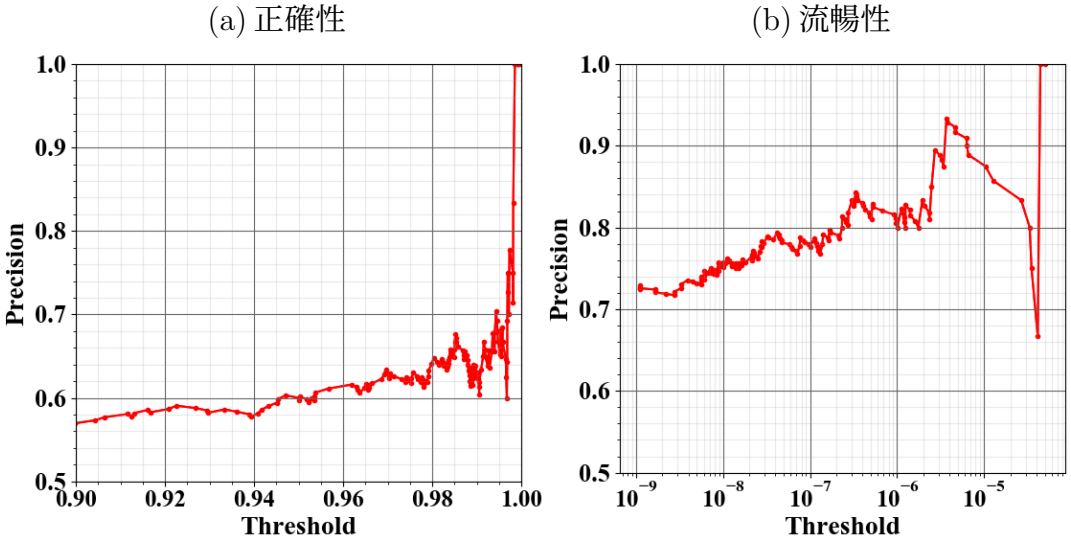

図 4.1: 開発データにおける平易句の正確性・流暢性判定の精度 ($Score1$)

精度が十分に大きいとき，具体的には図 4.1(a) で精度が 0.7 になるときの閾値 T_{fa} は 0.994，図 4.1(b) で精度が 0.8 になるときの閾値 T_{fl} は 2.33×10^{-7} であった．これらを $Score1$ の計算式，すなわち式 (3.6) に当てはめ，$T = \alpha\times\log T_{fa}+\log T_{fl} = -24.5$ と設定した．

4.2.3 $Score2$ の閾値 T の決定

$Score2$ について，名詞と動詞の両方を言い換えた平易句を優先して選択するための閾値 T を決定する．先ほどと同様に，開発データ 500 件のうち，名詞と動詞の両方を言い換えた 167 件のサブセットを用いる．これらの開発データには，平易句が適切であるか否かの二値ラベルが付与されている．実際には，開発データには，正確性を満たすか否か，流暢性を満たすか否か，の 2 つの観点で二値ラベルが付与されているが，正確性と流暢性を両方とも満たす場合にその平易句は適切であり，それ以外は適切でないとして，適切か否かのラベルを決定している．式 (3.5) に示す $Score2$ の値が閾値 t を越えたときに平易句が適切であると判定するシステムを作る．$Score2$ には正確性のスコアと流暢性のスコアの重み付けパラメータ β があり，これを $\beta = 0.25, 0.5, 0.75$ としたときのそれぞれについて，3 通りのシステムを構築した．このシステムの精度，すなわち平易句として適切と判定した事例のうち，人によって実際に適切であるとラベル付けされた事例の割合を求める．閾値と精度の関係を可視化したグラフを図 4.2 に示す．このグラフにおいて，X 軸は閾値 t，Y 軸は判定精度である．グラフでは，閾値 t を増加させると判定精度も上がる傾向が見られた．$\beta = 0.25, 0.5, 0.75$ のそれぞれについて，精度が 0.65 に達したときの -15.5, -16.0, -14.0 を閾値 T と決定した．

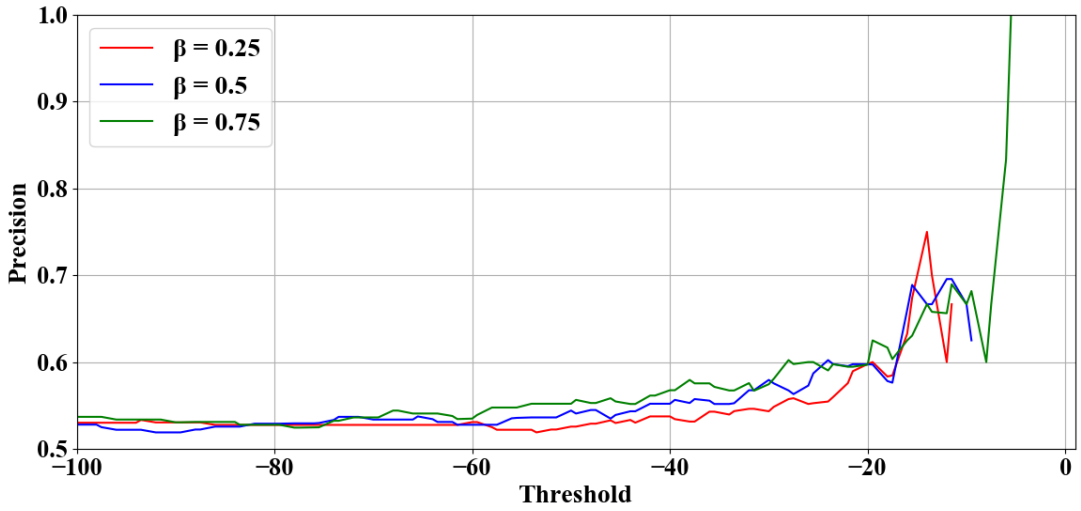

図 4.2: 開発データにおける平易句の適切性判定の精度 ($Score2$)

4.3 平易句選択手法の評価

3.5 節では，複数の平易句の候補の中から最適な平易句を選択する手法について説明した．本節ではこの手法を評価する．複数の候補のうち人手で最適なものを選択し，タグ付けしたデータセットを作成し，提案手法によって選択された平易句が人手によって選択された平易句と一致する割合を正解率とし，これによって提案手法の性能を評価する．

まず，4.1.2 項で述べた評価データに対し，以下のようなタグ付け作業を行なう．評価データにおける難解句に対し，3.4 節で説明した方法を用いて，平易句の候補を生成する．その全ての候補に対して，難解句を含む文 (元の文) と，難解句を平易句に置き換えて平易化した文の両方を提示する．難解句および平易句はマークアップし，容易に識別できるようにする．作業者に，全ての平易句の候補について，難解句を含む文と平易句を含む文の組を提示し，その中から最も適切な平易句を選択してもらう．ここで最も適切な平易句とは，子供や日本語を非母語話者とする人に最もわかりやすい表現とした．

上記のタグ付け作業を 2 名の作業者が独立に行なった．二者のタグ付けの一致率，すなわち 2 名の作業者によって同じ平易句が最適と判定された事例の割合は 0.61 であった．一致率は決して高くないと言える．平易句の候補の中には，適切なものが 1 つしかないわけではなく，複数存在することが多かった．その中から最適なものを選ぶ作業の結果は，人の主観によって大きく異なる．このことが一致率が低い要因と考えられる．

次に，作成した評価データを用いて，提案手法を評価する．ここでは以下のシステムを比較する．

- 正確性のスコアのみ - $FA(Ph'_i)$

平易句の候補の順位付けを正確性のスコアのみで行なう手法.

- 流暢性のスコアのみ - $FL(Ph'_i)$
  平易句の候補の順位付けを流暢性のスコアのみで行なう手法.

- $Score1$
  平易句の候補の順位付けを $Score1$ で行なう手法．正確性と流暢性の両方が考慮される．両者に対する重み付けは行なわない.

- $Score2$
  平易句の候補の順位付けを $Score2$ で行なう手法．正確性と流暢性の両方が考慮される．パラメータ β により両者に対する重み付けを行なう．$\beta = 0.25, 0.5, 0.75$ と設定したときの 3 つのシステムを比較する．なお，$\beta = 0.5$ のときは正確性と流暢性のスコアを同程度に評価するが，閾値 T の決め方が異なるため，$Score1$ と $\beta = 0.5$ と設定した $Score2$ は同じシステムではない.

表 4.3 は，上記のシステムについて，2 名の評価者をそれぞれ正解としたときの正解率を示している．正確性あるいは流暢性のスコアのみを使用したときよりも，両方を考慮したスコア ($Score1$ や $Score2$) の方が正解率が高い傾向が見られた．特に評価者 2 についてその傾向が顕著に現れた．$Score1$ と $Score2$ を比較すると，$Score2$ の正解率は $Score1$ よりも同等もしくはそれ以上であることから，$Score2$ の方が適切な平易句を選択するのに適していると言える．$Score2$ における異なる β の比較については，評価者 1 と評価者 2 で優劣の傾向が異なるため，どれが最適を決めることはできない.

提案手法の中で最高の正解率は 0.51 であり，決して高いとは言えない．しかし，前述したように，本研究で生成する平易句の候補の中には，平易句として妥当なものが複数含まれており，その中から最適なものを選択するのは難しいタスクと言える．二者の判定の一致率は 0.61 であり，提案手法の正解率はその 83%(= 0.51/0.61) に達している.

表 4.3: 平易句の選択手法の評価 (正解率)

システム	評価者 1	評価者 2
$FA(Ph'_i)$	0.45	0.43
$FL(Ph'_i)$	0.34	0.38
$Score1$	0.41	0.44
$Score2$ ($\beta = 0.25$)	**0.47**	0.44
$Score2$ ($\beta = 0.5$)	0.41	0.48
$Score2$ ($\beta = 0.75$)	0.44	**0.51**

4.4　平易化されたテキストの品質評価

本節では，提案手法によって生成された平易句の品質を評価する．ここで評価の対象とするのは，1 つの難解句に対し，提案手法によって最大のスコアをもつものとして選択された 1 つの平易句である．4.1.2 項で述べた評価データに対して，提案手法によって難解句を平易句に言い換えた．ここで用いたスコア付けの手法は $Score1$ である．正確性と流暢性のスコアに対する重みを設定せずに，基本的な評価式により候補を選択することにした．そして，提案手法によって生成された平易句を「平易性」「正確性」「流暢性」の観点から，1〜5 の評定をつけて評価した．評価基準はおよそ以下のように定めた．

- 平易性

  言い換え前後を比べて平易になっているかを評価する．平易になっているならその度合によって 5 か 4 を，難解になっているならばその度合によって 1 か 2 を，どちらともいい難いならば 3 と評価する．

- 正確性

  言い換え前後の文を比べて，意味が保持されているかを評価する．文の意味が変わっていなければその度合によって 5 または 4 を，文の意味が変わっているときはその度合によって 1 または 2 を，どちらともいい難いならば 3 と評価する．

- 流暢性

  言い換え後の平易句を含む文が自然であるかを評価する．ここでの自然さは，文法的な正しさと，意味を成す文であるかの両方を評価する．十分自然な文であるときはその度合に応じて 5 または 4 を，不自然な文であるときはその度合に応じて 1 または 2 を，どちらともいい難いならば 3 と評価する．

結果を表 4.4 に示す．どの観点についても，おおむね評点は 4 点以上であり，ある程度品質の高い平易句が生成されていることを確認した．また，表 4.3 の正解率が全般的に低いことに対し，表 4.4 の評点が比較的高いのは，既に述べたように，システムが選択した平易句が人手で選ばれた平易句と一致していなくても適切であるとみなせる場合が多かったためである．また，評価者 1 は，平易句の評価に関して，正確性よりも流暢性の方が評点が高い傾向が見られる．ここで評価者 1 の表 4.3 の結果を確認すると，$\beta = 0.25$ のとき，つまり正確性よりも流暢性を重視したとき，正解率が最も高かった．この 2 つの事実は整合性がある．これに対し，評価者 2 は正確性を高く評価する傾向が見られる．表 4.3 における評価者 2 の結果では，$\beta = 0.75$ のとき，つまり正確性を重視するときに正確性が最も高い．4.3 節では，β の最適な値は決められないと結論付けたが，人によって正確性と流暢性のどちらを重視するかに差があることが原因と言える．

表 4.4: 平易句の品質評価

	平易性	正確性	流暢性
評価者 1	4.06	4.02	4.29
評価者 2	3.79	4.38	4.29
平均	3.93	4.20	4.29

評価結果より，評価者 1 と評価者 2 の品質評価には異なる傾向は見られたが，評価値の値自体は大きな差は見られなかった．ここで，2 人の評価結果がどの程度一致していたかをカッパ係数を測ることによって検証する．カッパ係数とは，2 人の評価者が行なった複数のカテゴリに対する判定の一致度を算出する統計量の一つである．カッパ係数は 0 から 1 の範囲で算出され，1 に近いほど評価者間の評価は一致しているとみなせる．また，カテゴリが名義尺度ではなく順序尺度の場合には，重み付きカッパ係数が適しているとされている．この評価実験は 1〜5 の評点をつける順序尺度であるため，重み付きカッパ係数を用いて二者の判定の一致度を測る．

重み付きカッパ係数を算出する際には，2 つの評価者の評点の対応関係を表す分割表を作成する．表 4.5，表 4.6，表 4.7 は，平易性，正確性，流暢性，のそれぞれについて，2 者の評点の分割表を示している．このような分割表に基づき，重み付きカッパ係数は式 (4.2) のように算出される．

$$\sum_{i=1}^{5}\sum_{j=1}^{5}\frac{N\left(\text{評点}_1 : i, \text{評点}_2 : j\right)}{Total} \times w(i,j) \tag{4.2}$$

このとき，$N\left(\text{評点}_1 : i, \text{評点}_2 : j\right)$ は分割表で対応するセルの件数を表す．$Total$ は評価した事例の総数の値であり，この評価実験の場合では 100 である．重みである $w(i,j)$ は $w = 1 - \frac{(i-j)^2}{(k-1)^2}$ で算出され，k はカテゴリの個数を表す．今回の実験では $k = 5$ である．また，$w(i,j)$ におけるべき乗の値により重みを調整できるが，一般的には 2 次が採用されるので，ここでも 2 次の重みを採用する．

各観点のカッパ係数を以下に示す．

- 平易性のカッパ係数: 0.93
- 正確性のカッパ係数: 0.89
- 流暢性のカッパ係数: 0.95

この結果から，カッパ係数による一致度はかなり高いといえるため，評価者間の評点には大きな違いはなかったと考えられる．

表 4.5: 平易性の評点の分割表

		評価者 2					合計
		評点$_2$: 1	評点$_2$: 2	評点$_2$: 3	評点$_2$: 4	評点$_2$: 5	
評価者 1	評点$_1$: 1	0	0	0	0	0	0
	評点$_1$: 2	0	0	0	0	0	0
	評点$_1$: 3	0	1	20	5	9	35
	評点$_1$: 4	0	1	9	2	12	24
	評点$_1$: 5	0	1	8	8	24	41
合計		0	3	37	15	45	100

表 4.6: 正確性の評点の分割表

		評価者 2					合計
		評点$_2$: 1	評点$_2$: 2	評点$_2$: 3	評点$_2$: 4	評点$_2$: 5	
評価者 1	評点$_1$: 1	5	0	0	0	1	6
	評点$_1$: 2	5	0	0	1	0	6
	評点$_1$: 3	4	1	0	1	3	9
	評点$_1$: 4	2	3	0	5	1	11
	評点$_1$: 5	2	4	1	16	45	68
合計		18	8	1	23	50	100

表 4.7: 流暢性の評点の分割表

		評価者 2					合計
		評点$_2$: 1	評点$_2$: 2	評点$_2$: 3	評点$_2$: 4	評点$_2$: 5	
評価者 1	評点$_1$: 1	1	0	0	0	0	1
	評点$_1$: 2	0	3	0	3	0	6
	評点$_1$: 3	2	6	0	5	2	15
	評点$_1$: 4	0	1	0	4	5	10
	評点$_1$: 5	0	2	0	8	58	68
合計		3	12	0	20	65	100

4.5 エラー分析

本節では，平易句のスコア付けの手法についてのエラー分析を行なう．すなわち，複数の平易句の候補の中から最適なものを選択するタスクにおいて，提案手法が選択した平易句と人によって選択された平易句が一致しなかった原因を考察する．ここでは評価者 1 による評価結果を正解として分析を行なう．評価者 1 については $\beta = 0.25$ と設定した $Score2$ を用いる手法が最も高い正解率が得られたため，このシステムの出力をエラー分析の対象とする．確認しておくと，$\beta = 0.25$ のときは正確性より流暢性のスコアが高くなるように重み付けされている．評価者 1 の判定を正解としたときの $Score2$ ($\beta = 0.25$) のシステムの正解率は 0.47 であり，テストデータの全 100 件中 47 件についてはシステムと評価者の判定が一致し，53 件は何かしらの原因で不一致だった．

まず，エラー分析に先立ち，評価者 1 と提案手法によって選択された平易句が一致した 47 件について，その平易句が名詞・動詞の両方の言い換えか，名詞のみか，動詞のみか，の内訳を調べた．また，そのうち提案手法によって選択された平易句と一致する事例の数を調べた．結果を表 4.8 に示す．評価者 1 の判定は，名詞のみの言い換えが最も適切だと選択した例が約半数を占めていたことがわかった．一方，名詞・動詞両方と動詞のみの言い換えについてはほぼ同数であった．これらのうち，提案手法と一致したのは，動詞のみの言い換えが最も多かった．このことから，評価者 1 は名詞のみの言い換えを好むが，提案手法では動詞のみの言い換えを選択する傾向が強いことがわかる．

表 4.8: 評価者 1 と提案手法によって選択された平易句の一致数

	評価者 1 の判定	提案手法との一致数
名詞・動詞の言い換え	24	11
名詞のみの言い換え	47	17
動詞のみの言い換え	29	19
合計	100	47 (0.47%)

次に，提案手法と評価者 1 による判定が一致した例を確認する．一致した言い換え候補の 47 件から抜粋した 3 件の事例を表 4.9 に示す．この表は，言い換え前の文と難解句，名詞・動詞の両方の言い換え候補，名詞のみの言い換え候補，動詞のみの言い換え候補，評価者の判定，提案手法の判定，$Score2$，正確性のスコア $FA(Ph'_i)$，流暢性のスコア $FA(Ph'_i)$ を示す．平易句の前後は言い換え前の文と同じであり，冗長なので省略して表示している．また，説明の都合上，それぞれの言い換えの組に ID を付与した．ID:1 は，名詞・動詞両方の言い換えを優先する閾値 T がうまく作用した例である．名詞・動詞の言い換えはスコア自体は一番低いが，閾値 T を超えているので，優先して選択されている．ID:2 は，正確性のスコアが高く評価されて人による判定と一致した例である．5 つの候補から適切に

平易句を選択できた点も評価できる．ID:3 は，ID:2 と同様に正確性のスコアが高く評価されたが，結果として動詞のみ言い換えを選択して評価者 1 の判定と一致した例である．

以降では，評価者 1 と提案手法による平易句の選択が不一致だった言い換えについて，4 つの大まかなグループに分けて考察する．

一つ目は，評価者 1 は名詞・動詞の両方の言い換えを選択したが，提案手法ではスコアが低く，かつそのスコアが閾値 T を越えずに選択されなかったため，不一致となったグループである．これは，不一致の 53 件中 11 件が該当した．そのうち 2 例を表 4.10 に示す．ID:4 では，名詞・動詞の両方の言い換えの流暢性のスコアが低いことから選択されなかった．さらに，正確性のスコアに関しても，採用された動詞のみの言い換えと比べるとかなり下がってしまっている．ID:5 では，名詞・動詞の両方の言い換えの流暢性のスコアは高かったものの，正確性が低かったために採用されなかった．名詞・動詞の両方の言い換えを選択する方法として，閾値 T を低く設定することが考えられる．ただし，名詞，動詞の一方を言い換えた候補が選択される可能性が低くなるので，単純に閾値 T を下げれば正解率が上がるというわけではない．このトレードオフを解消し，適切な閾値 T を設定する方法を探究することは今後の課題である．

二つ目は，名詞・動詞両方の言い換えのスコアが閾値 T は超えていたものの，他の名詞・動詞の両方の言い換えの方がスコアが高かったため，評価者 1 によって選択された平易句と不一致だったグループである．不一致の事例 53 件中 2 件が該当した．これを表 4.11 に示す．なお，名詞，動詞の一方の言い換え候補については省略する．ID:6 においては，評価者 1 が選択した平易句も提案手法が選択した平易句も違和感のない言い換えといえる．しかし，正確性と流暢性がともにスコアが高い他の言い換えが採用されたために不一致であった．ID:7 では，3 つの言い換えのうち,「何を聞く」という平易句の流暢性が高いため，評価者 1 と異なる平易句が選択された.「何を聞く」を元の文に当てはめると，他の平易句の候補と比べて違和感が強く，これを選択しない評価者 1 の判定は妥当と言える．しかし,「何を聞く」という表現はよく見られるもので，コーパスにおける句の出現頻度により流暢性のスコアを算出すると高く見積もられる．これは，流暢性のスコアを文脈を考慮して算出していないことが問題と言える．

三つ目は，評価者 1 は名詞もしくは動詞の一方の言い換えを最適としていたが，提案手法では名詞・動詞両方の言い換えを選択したために不一致だったグループである．このグループは最も多く，不一致の事例 53 件中 23 件が該当した．そのうち 2 件を表 4.12 に示す．ID:8 は，評価者 1 が選択した言い換えがスコアが最も高く，閾値 T 以上のときに名詞・動詞両方の言い換えを選択するという処理を行なわなければ人手による判定と一致した言い換えである．それに対し，ID:9 は，名詞・動詞を優先的に選択する処理を行なわなくても，スコアが最大の言い換えが評価者が選択した言い換えと一致しない例であった．

最後のグループは，評価者 1 は名詞（動詞）のみの言い換えを選択としている

が，提案手法は動詞（名詞）のみの言い換えを選択しているグループである．不一致の 53 件中 11 件が該当した．そのうち 2 件を表 4.13 に示す．ID:10 では，評価者 1 は名詞のみの言い換えが妥当だと判定したが，提案手法は動詞の言い換えを選択したため，不一致であった．ID:11 はその逆である．

以上のエラー分析の結果を踏まえると，今後の改善策として以下の 3 つが考えられる．

- 閾値の検討

  提案手法ではスコアが閾値以上のときに名詞・動詞の両方の言い換え候補を優先的に採用している．ただし，それにより狙い通りの言い換え候補が選択されない場合もある．基本的には，名詞・動詞両方の言い換えを優先するか，名詞もしくは動詞どちらかの言い換えを優先するかはトレードオフの関係と言えるので，閾値の調整による正解率の向上は難しいが，より良い閾値の設定方法を検討したい．

- 正確性のスコアの改善

  提案手法において，正確性のスコアは言い換え前後の文全体の類似度により算出している．文が長い場合には，言い換える句（3 単語）の長さが文全体に占める割合が低くなるため，いずれの候補も 1 に近い値となり，正確性のスコアに明確な差が現れない可能性がある．文長が与える文間類似度への影響は今後精査する必要がある．

- スコアの評価式の改善

  提案手法では正確性と流暢性を一つのスコアとして算出している．4.3 節の実験結果から，$FA(Ph'_i)$ もしくは $FL(Ph'_i)$ のみで平易句を選択した場合よりも，$Score1$，$Score2$ を用いたときの方が，正解率は同等もしくはそれ以上であった．これにより，正確性と流暢性の両方を考慮することの有効性は示されている．ただし，最終的に一つのスコアとしてまとめて扱うことが適切かは不明である．改善案としては，2 つの観点だけでなくさらに多くの観点を設けて，それぞれの観点の中で平易句の候補を順位付けし，最終的に全ての観点の順位を考慮して一つの候補を選択することが考えられる．ここで重要なのは，正確性，流暢性以外にも，平易句の候補の妥当性を評価する多くの観点を設定することである．

表 4.9: 提案手法と評価者の判定が一致した事例

ID		文	評価式の判定	提案手法	$Score2$ ($\beta = 0.25$)	$FA(Ph')$	$FL(Ph')$
1	言い換え前	尖閣諸島をめぐる日本との対立が背景にあるとみられるが，中国は深刻化する大気汚染などの解決に向けた日本との協力には前向きの**姿勢を示して**おり，代表の派遣見送りは回避した形だ.					
	言い換え候補（名詞・動詞）	～前向きの**態度を見せて**おり～	○	○	-13.8	0.9990	4.340E-08
	言い換え候補（名詞のみ）	～前向きの**態度を示して**おり～			-12.0	0.9994	1.430E-07
	言い換え候補（動詞のみ）	～前向きの**姿勢を見せて**おり～			-11.3	0.9997	6.399E-07
2	言い換え前	内向きだった日本国民の目を国際的な**課題に向け**，より大きな責任を担うことを目指すとした首相の姿勢を私は率直に評価したいと思う.					
	言い換え候補（名詞・動詞）	～国際的な**問題に使い**～			-20.5	0.9721	1.614E-08
	言い換え候補（名詞・動詞）	～国際的な**宿題に使い**～			-51.3	0.8613	7.790E-09
	言い換え候補（名詞のみ）	～国際的な**問題に向け**～	○	○	-14.0	0.9871	5.926E-07
	言い換え候補（名詞のみ）	～国際的な**宿題に向け**～			-47.4	0.8783	2.226E-09
	言い換え候補（動詞のみ）	～国際的な**課題に使い**～			-16.1	0.9895	1.502E-08
3	言い換え前	しかし，測定業者は昨年１１月１７日の測定時，４階足場上での作業の様子を撮影しており，ひさし裏に作業員が**ノズルを向けて**いたと，今年１月に府教委へ報告.					
	言い換え候補（名詞・動詞）	～作業員が**バスを使って**いた～			-32.9	0.9136	1.082E-06
	言い換え候補（名詞のみ）	～作業員が**バスを向けて**いた～			-39.4	0.9042	6.121E-09
	言い換え候補（名詞のみ）	～作業員が**ノズルを使って**いた～	○	○	-13.5	0.9979	3.116E-08

表 4.10: 名詞・動詞両方の言い換えスコアが十分に高くないために不一致だった事例

ID		文	評価式の判定	提案手法	$Score2$ ($\beta = 0.25$)	$FA(Ph')$	$FL(Ph')$
4	言い換え前	誌名はラテン語で「物事の**真理を探る**」との意味.					
	言い換え候補（名詞・動詞）	～物事の**真実を調べる**～	○		-34.3	0.9221	7.234E-09
	言い換え候補（名詞・動詞）	～物事の**真実を探る**～			-27.2	0.9364	6.087E-07
	言い換え候補（名詞・動詞）	～物事の**真理を調べる**～		○	-15.4	0.9930	1.224E-08
5	言い換え前	しかし西行の知遇を得たことを喜び，**舞を舞って**去ってゆく.					
	言い換え候補（名詞・動詞）	～**踊りを踊って**去ってゆく.	○		-22.7	0.9511	1.231E-06
	言い換え候補（名詞・動詞）	～**ダンスを踊って**去ってゆく.			-25.8	0.9371	2.982E-06
	言い換え候補（名詞・動詞）	～**踊りを舞って**去ってゆく.			-21.8	0.9683	1.057E-08
	言い換え候補（名詞・動詞）	～**ダンスを舞って**去ってゆく.			-26.8	0.9435	7.456E-08
	言い換え候補（名詞・動詞）	～**舞を踊って**去ってゆく.		○	-15.7	0.9840	1.769E-07

表 4.11: 他の名詞・動詞両方の言い換えが選択されて不一致だった事例

ID		文	評価式の判定	提案手法	$Score2$ ($\beta = 0.25$)	$FA(Ph')$	$FL(Ph')$
6	言い換え前	決勝では同五輪覇者のメネセスを破って勝ち上がったフランス選手から小外刈りで**ポイントを奪い**，優勢勝ちした.					
	言い換え候補（名詞・動詞）	～小外刈りで**点数を取り**～	○		-13.8	0.9832	2.905E-06
	言い換え候補（名詞・動詞）	～小外刈りで**点を取り**～		○	-9.20	0.9972	1.190E-05
7	言い換え前	退任直前の**心境を問う**と，流ちょうな日本語で「初めてこの部屋に入ったのは，つい昨日だったという気持ち」と感慨深そうだ.					
	言い換え候補（名詞・動詞）	退任直前の**心を聞く**と～			-14.6	0.9947	2.059E-08
	言い換え候補（名詞・動詞）	退任直前の**気持ちを聞く**と～	○		-11.9	0.9954	6.226E-07
	言い換え候補（名詞・動詞）	退任直前の**何を聞く**と～		○	-10.8	0.9928	5.977E-06

表 4.12: 名詞・動詞両方の言い換えを選択して不一致だった事例

ID		文	評価式の判定	提案手法	$Score2$ ($\beta = 0.25$)	$FA(Ph')$	$FL(Ph')$
8	言い換え前	現場ではダイバーが**水中に沈んだ**とみられる隕石の破片などの捜索作業を行っていた.					
	言い換え候補（名詞・動詞）	～ダイバーが**水に入った**～		○	-13.7	0.9837	2.774E-06
	言い換え候補（名詞のみ）	～ダイバーが**水に沈んだ**～	○		-11.9	0.9928	1.450E-06
	言い換え候補（動詞のみ）	～ダイバーが**水中に入った**～			-14.7	0.9907	7.234E-08
9	言い換え前	新年の**夜空を彩る**天文ショーで，３大流星群の一つ「しぶんぎ座流星群」が，１月３日深夜から４日明け方にかけて出現のピークを迎える.					
	言い換え候補（名詞・動詞）	新年の**空を飾る**天文ショーで～		○	-13.8	0.9919	1.519E-07
	言い換え候補（名詞のみ）	新年の**空を彩る**天文ショーで～	○		-12.0	0.9950	5.759E-07
	言い換え候補（動詞のみ）	新年の**夜空を飾る**天文ショーで～			-11.3	0.9976	6.282E-07

表 4.13: 名詞のみと動詞のみの言い換えの間で判定不一致だった事例

ID		文	評価式の判定	提案手法	$Score2$ ($\beta = 0.25$)	$FA(Ph')$	$FL(Ph')$
10	言い換え前	これらは，日本列島の下にもぐり込むフィリピン海プレートに含まれていた熱水に由来していることが，**湯に含まれる**酸素と水素の同位体分析から分かってきた.					
	言い換え候補（名詞・動詞）	～**水に持たれる**酸素と～			-21.4	0.9754	1.669E-09
	言い換え候補（名詞のみ）	～**水に含まれる**酸素と～	○		-16.9	0.9785	2.459E-07
	言い換え候補（動詞のみ）	～**湯に持たれる**酸素と～		○	-15.6	0.9981	1.669E-09
11	言い換え前	検察の起訴猶予に検審が**異議を唱えて**強制起訴となった事件では，初めての裁判所の判断となった.					
	言い換え候補（名詞・動詞）	～検審が**反対を言って**～			-19.6	0.9771	1.002E-08
	言い換え候補（名詞のみ）	～検審が**反対を唱えて**～		○	-14.9	0.9834	5.848E-07
	言い換え候補（動詞のみ）	～検審が**異議を言って**～	○		-15.2	0.9963	5.564E-09

第5章　おわりに

5.1　まとめ

本研究では，句の言い換えによるテキストの平易化の手法を提案した．言い換え候補の中から正確性と流暢性に基づいて最適な平易句を選択することにより，文脈にあった適切な平易な言い換えの実現を目指した．

提案手法では，大きく 5 つの段階で平易化を実現した．一段階目の処理では，入力したテキストの形態素解析を行ない，それぞれの単語に形態素解析ツール内の辞書を用いて品詞などの情報を付与した．形態素解析には mecab-ipadic-NEologd という辞書を採用することで，固有表現や慣用句表現を言い換えの対象からはずした．そして，形態素解析により得られた品詞の情報から,「名詞-助詞-動詞」から構成される句を抽出した．二段階目の処理では，既存の言語資源を利用して平易言い換え単語対データベースを構築し，難解語とその平易な言い換えである平易語の組を大量に取得した．利用した既存の言語資源は，平易な言い換え辞書である Simple PPDB: Japanese と言い換え辞書の SNOW D2: 内容語換言辞書であり，これらから平易な言い換えのみを収集した．その際には，自動詞と他動詞の言い換え対は構築しないという基準も定め，これを実現した．三段階目の処理では，抽出した平易言い換え単語対データベースを基に，難解句を 3 つのパターンで言い換え，平易句の候補を生成した．3 つのパターンとは，名詞・動詞両方の言い換え，名詞のみの言い換え，動詞のみの言い換えである．四段階目の処理では，正確性と流暢性に基づいた評価式で生成した言い換え候補のスコアを算出し，そのスコア付けにより言い換え候補の中から最も適切な候補を選択した．最後の五段階目では，元の文の難解句を平易句に置き換え，さらに活用形の処理をして，平易な文を生成した．

提案手法の評価は 2 種類の実験により行なった．一つ目の実験は，スコア付けにより選択した平易句が人手によって選択された平易句とどの程度一致しているか（正解率）を調べた．その結果，提案手法の正解率はスコア付け，パラメータ，評価者の違いによって変動するものの，0.45 程度であり，最高で 0.51 となった．言い換え候補の中から最適な平易句を選択することは人手でも難しい問題であり，5 割を超える正解率が得られたことは，正確性と流暢性の観点に基づいた言い換え候補の選択手法が適切だったと言える．もう一つの実験では，提案手法によって生成された平易句およびそれを含む平易な文の品質評価を行なった．2 名の作業者が，平易な文と元の文を比べ，平易性，正確性，流暢性の 3 つの観点から 5 段階

の評点をつけた．その結果，2 人の平均の評点はいずれの観点もおよそ 4 点であった．評点は全般的に高く，提案手法の有効性が確認された．

5.2 今後の課題

本研究の今後の課題を述べる．現段階では，大きく分けて 3 つの課題があると考えている．

一つ目の課題は句の定義である．本論文では句を「名詞-助詞-動詞」といった単語列に限定した．句であることは間違いないが，平易化できる句の種類は限定されていた．今後は，句の定義のパターンを増やすことで，より多様な句を対象に平易化を実現することを目指したい．また，今回平易化の対象としたのは連続した単語列であったため，いわゆる構文解析は必要としなかった．より多様な平易化のためには，連続して出現せず，遠い距離の係り受け関係にある句を平易化の対象とすることも必要であり，そのためには構文解析や文節の係り受け解析は必須の技術である．

二つ目の課題は評価式である．本論文の平易句の候補のスコア付けは，正確性と流暢性という 2 つの観点を用いて構成した．そして，この観点を一つのスコアとして扱い，スコア付けした言い換え候補の中から最も妥当な候補を選択していた．ただし，単純に最もスコアが高くなった言い換え候補を選択するのではなく，閾値を決めて，名詞と動詞の両方を言い換えてかつ信頼できる平易句を優先的に選択するという方策をとった．これは，正確性を表すスコアが言い換え前後の類似性であることから，名詞，動詞の両方の言い換えの正確性のスコアが名詞のみ，動詞のみの言い換えより低くなることを考慮したものであった．この方式の是非はともかく，評価式自体は改善できると考える．例えば，正確性と流暢性の評価式を切り離し，他の観点からの評価式も導入し，一つの候補に関して複数の評価式でスコアを算出した後，それぞれのスコアで順位付けを行ない，その順位の投票制により最適な候補を選択することも考えられる．

三つ目の課題は，評価実験の方法である．特に評価式による平易句の選択手法の評価である．本論文の実験では，複数の言い換え候補の中から最も適切な候補を人手で選び，提案手法で選択された候補と一致したかで正解率を測った．このとき，提示した言い換え候補の数に制限は設けていない．つまり，3 つの言い換え候補が生成された言い換えも，5 つの候補が生成された言い換えも，同じ基準で評価した．言い換え候補が多くなれば人手で選択された平易句と提案手法で選択された平易句が一致しない可能性が高くなり，正解率が上がりにくくなる．言い換え候補の数と正解率の相関を調べるなど，評価方法を洗練させることも必要である．

謝辞

本研究を行なうにあたり，多くの方々に協力していただきました．この場をお借りして感謝の意を表します．

はじめに，終始あたたかいご指導と激励を賜りました北陸先端科学技術大学院大学 白井清昭准教授に心から感謝の意を表します．研究の着想から論文執筆まで，多くのご指導を本当にありがとうございました．また，研究以外の相談も親身に応じていただき，大変お世話になりました．

次に，本論文を審査およびご指導してくださった北陸先端科学技術大学院大学 飯田弘之教授，東条敏教授，池田心准教授に深く感謝いたします．

最後に，これまで私をあたたかく応援してくれた家族，そして心の支えとなってくれた友人らに心から感謝をいたします．

参考文献

[1] 【教材】自動詞他動詞リスト 50.pdf. https://nihon-go.life/jp/jita50-pdf/. (2021 年 1 月閲覧).

[2] 【日本語/非母語話者むけ/文法/自動詞・他動詞. https://ja.wikibooks.org/wiki/日本語/非母語話者むけ/文法/自動詞・他動詞. (2021 年 1 月閲覧).

[3] 日本語形態素解析システム JUMAN. http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN. (2021 年 1 月閲覧).

[4] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. PPDB: The Paraphrase Database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 758–764, Atlanta, Georgia, 2013. Association for Computational Linguistics.

[5] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. Development of the Japanese WordNet. 01 2008.

[6] 鍛治伸裕, 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話の言い換え. 自然言語処理, Vol. 10, No. 4, pp. 65–81, 2003.

[7] 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, pp. 272–275, 2013.

[8] 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, pp. 983–992, mar 2015.

[9] 梶原智之, 小町守. Simple PPDB: Japanese. 言語処理学会 第 23 回年次大会発表論文集, pp. 529–532, 2017.

[10] 梶原智之, 小町守. 平易なコーパスを用いないテキスト平易化. 自然言語処理, Vol. 25, No. 2, pp. 223–249, 2018.

[11] Tomonori Kodaira, Tomoyuki Kajiwara, and Mamoru Komachi. Controlled and Balanced Dataset for Japanese Lexical Simplification. In *Proceedings of*

the ACL 2016 Student Research Workshop, pp. 1–7, Berlin, Germany, August 2016. Association for Computational Linguistics.

[12] 小藤直紀, 難波英嗣, 竹澤寿幸. 新聞記事データを用いたテキスト平易化. 人工知能学会全国大会論文集, Vol. JSAI2019, pp. 2L4J902–2L4J902, 2019.

[13] Reno Kriz, Eleni Miltsakaki, Marianna Apidianaki, and Chris Callison-Burch. Simplification using Paraphrases and Context-based Lexical Substitution. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 207–217, 2018.

[14] Taku Kudo. MeCab: Yet another Part-of-Speech and Morphological Analyzer. http://taku910.github.io/mecab/.

[15] 前川喜久雄. KOTONOHA『現代日本語書き言葉均衡コーパス』の開発 (＜特集＞資料研究の現在). 日本語の研究, Vol. 4, No. 1, pp. 82–95, 2008.

[16] George A. Miller. WordNet: A Lexical Database for English. *Commun. ACM*, 1995.

[17] 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.

[18] 水上雅博, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲. 日本語言い換えデータベースの構築と言語的個人性変換への応用. 言語処理学会 第 20 回年次大会 発表論文集, pp. 773–776, 2014.

[19] Ellie Pavlick and Chris Callison-Burch. Simple PPDB: A Paraphrase Database for Simplification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 143–148, Berlin, Germany, 2016. Association for Computational Linguistics.

[20] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3973–3983, 2019.

[21] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会, pp. 875–878, 2017.

[22] 園部勲. Sentence BERT 日本語モデル. https://qiita.com/sonoisa/items/1df94d0a98cd4f209051. (2021 年 1 月閲覧).

[23] Alex Franz Thorsten Brants. Web 1T 5-gram Version 1. https://catalog.ldc.upenn.edu/LDC2006T13, 2006. (2021 年 1 月閲覧).

[24] 渡辺太郎. cicada. https://github.com/tarowatanabe/cicada. (2021 年 1 月閲覧).

[25] 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, pp. 276–279, 2013.

[26] 京都大学大学院情報学研究科黒橋・河原研究室. GSK2018-B 京都大学格フレーム ver 2.0. https://www.gsk.or.jp/catalog/gsk2018-b. (2021 年 1 月閲覧).

[27] 日本語学習辞書支援グループ. 日本語教育語彙表 ver 1.0. http://jhlee.sakura.ne.jp/JEV.html, 2015. (2021 年 1 月閲覧).

付 録A　自動詞と他動詞の対応付け

3.2 節で述べた自・他動詞判別辞書を表 A.1 に，自動詞と他動詞の組を判別するルールを表 A.2 に示す.

表 A.1: 自・他動詞対判別辞書

自動詞	他動詞	自動詞	他動詞	自動詞	他動詞	自動詞	他動詞
なる	する	見える	見る	乗る	乗せる	付く	付ける
育つ	育てる	見付かる	見付ける	生える	生やす	沸く	沸かす
汚れる	汚す	減る	減らす	切れる	切る	分かれる	分ける
下がる	下げる	現れる	現す	折れる	折る	聞こえる	聞く
加わる	加える	枯れる	枯らす	染まる	染める	並ぶ	並べる
過ぎる	過ごす	降りる	降ろす	増える	増やす	閉まる	閉める
解ける	解く	混ざる	混ぜる	続く	続ける	変わる	変える
回る	回す	残る	残す	直る	直す	片付く	片付ける
壊れる	壊す	始まる	始める	通る	通す	返る	返す
開く	開ける	止まる	止める	点く	点ける	亡くなる	亡くす
外れる	外す	治る	治す	渡る	渡す	無くなる	無くす
掛かる	掛ける	取れる	取る	倒れる	倒す	戻る	戻す
割れる	割る	終わる	終える	当たる	当てる	溶ける	溶かす
乾く	乾かす	集まる	集める	逃げる	逃す	落ちる	落す
喜ぶ	喜ばす	重なる	重ねる	届く	届ける	落ちる	落とす
寄る	寄せる	出る	出す	入る	入れる	離れる	離す
起きる	起こす	助かる	助ける	燃える	燃やす	立つ	立てる
曲がる	曲げる	消える	消す	破れる	破る	冷める	冷ます
決まる	決める	焼ける	焼く	売れる	売る	冷える	冷やす
見つかる	見つける	上がる	上げる	抜ける	抜く	-	-

表 A.2: 自動詞と他動詞の組を判別するルール

自動詞	他動詞	具体例
-eru	-asu	出る - 出す
-eru	-yasu	増える - 増やす
-iru	-osu	落ちる - 落とす
-u	-asu	減る - 減らす
-u	-eru	開く - 開ける
-ru	-seru	乗る - 乗せる
-ru	-su	返る - 返す
-reru	-su	離れる - 離す
-aru	-eru	集まる - 集める
-waru	-eru	変わる - 変える
-eru	-u	焼ける - 焼く