

理解しやすい対話を用いた漫才台本の自動生成

青木 哲[†] 義 尚晃^{††} 原口 和貴^{††} 梅谷 智弘^{††} 北村 達也^{††}
灘本 明代^{††}

[†] 甲南大学大学院 自然科学研究科 知能情報学専攻 〒658-8501 兵庫県神戸市東灘区岡本 8 丁目 9 番 1 号

^{††} 甲南大学 知能情報学部 知能情報学科 〒658-8501 兵庫県神戸市東灘区岡本 8 丁目 9 番 1 号

E-mail: †{m1724001,s1417036,s1571098}@s.konan-u.ac.jp, ††{umetani,t-kitamu,nadamoto}@konan-u.ac.jp

あらまし これまで我々は、お題を与えるとそれに関するニュースを用いて漫才台本を自動生成し、それを実演するロボット「漫才ロボット」を提案してきた。しかしながら、提案してきた手法で生成される漫才台本は、ニュース記事を基にしているため、難解な語を用いた台本が多く生成され、わかりやすく面白い漫才であるとは言いがたかく、これらを解消する必要がある。本論文では、この問題のうち、漫才のわかりやすさを対象とし、理解しやすい対話の生成を目的とする。具体的には、漫才台本の基となるニュース記事の中にある難語を抽出し、その難語をわかりやすい類語に言い換え、そのわかりやすい類語の例文を用いた対話文生成手法を提案する。さらに、その生成された対話文を勘違いに基づくボケを生成することにより、よりわかりやすい漫才台本自動生成の手法を提案する。

キーワード 漫才台本, 自動生成, 例文提示

1. はじめに

近年、ロボット技術の発展により、様々なロボットが研究開発されており、コミュニケーションを目的としたロボットも多数開発されている。総務省の情報通信白書 [1] によると、生産性の向上や高齢化、人材不足の懸念から生産ライン用のロボットに加え、NAO^(注1) や Pepper^(注2)、Kibiro^(注3) 等のサービスロボットが注目されている。

そこで我々はこれまで、お題を与えると自動で漫才台本を生成し、漫才を行うロボットである漫才ロボットを提案してきた [2]。我々の提案する漫才台本は「つかみ」「本ネタ」「オチ」の三段構成となっている。「つかみ」は、はじめの挨拶と共に最初の一笑いと同時に、本ネタの話題提示を行っている。「本ネタ」は漫才の軸となる部分で、ニュース記事を基に複数種類の掛け合いを用いることで生成されている。「オチ」は、漫才の締めくくりと最後の一笑いとして謎掛けを行っている。

漫才ロボット全体の流れは、はじめにユーザが漫才台本生成システムに対し、お題を与える。漫才台本生成システムは与えられたお題に関連したニュース記事を Web 上から取得する。そして、取得したニュース記事を基に面白おかしい対話文と、それに合わせた演出を自動生成することにより漫才台本を生成する。最後に、生成した漫才台本を基にロボットが漫才を演じる構成である。

現在の漫才台本生成システムが持つ問題点は種々あるが、その中の一つとして、漫才台本を Web から取得したニュース記事を基に生成しているため、元々ニュース記事上に存在する難

解な語がそのまま台本に用いられ、解りにくい台本が生成される場合が多数ある。そこで、我々は難解となる原因である難解な語 (以下、難語と呼ぶ) を抽出し、抽出された難語を易しい語 (以下、言い換え語と呼ぶ) に言い換え、更にその言い換え語を用いた例文を提示することにより、より理解のし易い対話を生成することが可能になると考えた。そこで本論文では、この手法を提案する。さらに、この手法を漫才台本に適用するために、言い換える語を難語と類似しない易しい語に言い換えることにより、ボケとツッコミの対話文を生成する手法を提案する。

具体的には、ニュース記事中に存在する難語を抽出し、それに対し取り違えた語を抽出する。難語と取り違えた語それぞれに対し難語ではない易しい語とそれを用いた例文を抽出し、それらを用いて掛け合いを生成する。本研究では、これにより生成されるボケを「説明ボケ」と呼ぶ。説明ボケにより難語の説明が漫才中に入り、難語をそのままにしていた時と比べて漫才台本が難解になる問題を緩和することが期待される。

以下、第 2 章では本研究の関連研究に対して述べ、第 3 章では漫才台本生成システムについて、第 4 章では言い換えの手法について、第 5 章では掛け合いの生成手法について、第 6 章では実験とその考察について、第 7 章ではまとめと今後の課題について述べる。

2. 関連研究

ニュース記事に用いられる難しい表現の言い換えや例文抽出は、理解支援を行うに当たって重要な手法と考えられる。藤沢ら [3] は、子供向けのニュース記事と一般向けのニュース記事の文の類似度計算を行い、言い換え表現の抽出を検討している。本研究とは、単語の意味に注目して言い換えを行っている点と、それを用いることで新たな対話文を生成している点で異なる。岡崎ら [4] は、括弧表現を用いた表現に着目し、文中から言い

(注1) : Softbank Robotics <https://www.ald.softbankrobotics.com/ja>

(注2) : SoftBank <https://www.softbank.jp/robot/>

(注3) : FRONTEO <https://www.fronteo-cc.com/kibiro/>

換え表現を抽出した後、それを元に SVM により分類を行っている。本研究とは、Web 上から例文を抽出する点と、対話文の一部にする点で異なる。内田ら [5] は、ブログ記事を用い、オノマトペを含む文の特徴から例文を自動抽出する手法を提案している。本研究とは、抽出する対象が難語であるという点で異なる。

対話文の生成に関する研究も近年行われている。下川ら [6] は、対話文応答のためにチャットにおける対話中の話題語の発見を行う手法を提案している。福岡ら [7] は、未知語の存在する対話に対し、推定を行って応答を行う手法を提案している。塚原ら [8] は、得られた対話にラベリングを行いそのラベルを伝搬させることで、システムが行う対話の前後関係に一貫性をもたせ、対話の内容が二転三転しないように改善している。いずれも、システムが人と対話を行うことを目標としており、ロボット同士の対話を目的とする本研究とは異なる。

漫才に関する研究も近年行われている。竹越ら [9] は、予め用意した文書を基に漫才台本を生成する手法を提案している。竹越らはボケとツッコミを別々の手法を用いることで生成しているため、これらを 1 つの手法を用いて一度に生成する本研究とは異なる。また、用いているロボットもヒト型で動作に着目しているため、腕が無くヒト型ではないロボットを用い台本に着目している本研究とは異なる。

3. 漫才台本生成システム

我々が提案する漫才台本生成システムは、ユーザにより与えられたお題に関連したニュース記事を Web 上から取得し、取得したニュース記事を基に掛け合いと、それに合わせた演出を自動生成することにより漫才台本を生成する。漫才台本はしゃべくり漫才の形式となっており、最初の挨拶となる「つかみ」、漫才の軸となる部分の「本ネタ」、最後の一笑いとなる「オチ」の三段構成となっている。

以下、それぞれについて詳細を述べる。

3.1 つかみ

つかみには、はじめの挨拶を行うと同時に漫才の導入部分としての役割がある。我々は「表情ボケ」と「取り違えボケ」の 2 種類の掛け合いを提案してきた。

表情ボケ

表情ボケでは、ロボットが話す内容に応じて感情を目を用いて表現する。しかし、ボケ役が話している内容に対し、目で表している感情が明らかに不適切である為、これをツッコミ役がツッコミを入れるという掛け合いである。

取り違えボケ

取り違えボケとは、ニュースのカテゴリとそれに対する対立語を用いたボケである。例えば「テニス」に関するニュースに対し、ボケ役が「バトミントン」と取り違えバトミントン選手として「山口茜」を挙げる。しかしそれは間違いであるため、ツッコミ役がテニス選手として「錦織圭」を挙げたうえで訂正するという掛け合いである。

3.2 本ネタ

本ネタは漫才の軸となる部分で、取得された Web ニュース

を説明するとともに、様々な掛け合いを行うことで笑いを取っていく。ユーザが記事の内容を把握できるように、句点や読点で分割された 1 文に対し最大 1 ボケを挿入する。ただし、記事のはじめの 1 文に対しては、ニュース記事の内容を顕著に表わしているため、テキストを変更するボケの挿入は行わない。ボケが挿入されない行に対しても、ツッコミ役は相槌を打つことにより参加する。「一つの漫才の中には、いろんな種類のボケが入っていた方がいい」[10] とのことから、本研究でも複数の掛け合いを用いており、その目的で複数の掛け合いを提案している。掛け合いのパターンとしては、「言葉遊びボケ」「ノリツッコミ」「対立ボケ」「過剰ボケ」がある。また、この本ネタはニュース記事を基に生成されているため、漫才台本が難解となる部分でもある。本研究で提案する「説明ボケ」をこの本ネタで用いることで漫才台本が難解となる問題を緩和することが期待される。

言葉遊びボケ

言葉遊びボケとは、ボケ役がニュース記事中の単語を別の単語と読み違えるボケである。例えば、「投票(とうひょう)」という語に対し「投了(とうりょう)」や「凍傷(とうしょう)」というように読み違えるものである。ツッコミ役はこれに対し訂正する形でツッコミを入れることで掛け合いが成立する。

ノリツッコミ

ノリツッコミは、通常のツッコミと異なり、ツッコミ役が一度ボケ役のボケに対し同調を行い(のっかり)、話題を展開した上で改めてツッコミを入れるというツッコミである。また、このツッコミを行う際に、ボケに用いた単語と同じタイトルを持つ Wikipedia の記事を用いることでボケた単語の補足を行う。

対立ボケ

対立ボケとは、ニュース記事内にある単語に対し、別の語と勘違いする掛け合いである。掛け合いの流れとしては、まずツッコミ役はボケ役が行っているニュースの読み上げを中断し、最後に読み上げた文中にある単語に対し、知っているかと問う。ボケ役はこの問いに対し、問われた単語と対立関係にある別の語と勘違いし、その勘違いした語に関する説明を行う。これに対しツッコミ役は訂正を入れ、正しい説明を行う、というものである。

過剰ボケ

過剰ボケとは、ボケ役がニュース記事中に存在する数値を過剰に大きく読んだり、過小に小さく読んだりするボケである。例えば、「7冠達成」という文に対し「7万冠達成」と読んだり、「観客が1万人」という文に対し「観客が1人」というように読み間違えるものである。

3.3 オチ

オチは漫才を締めくくる部分である。本研究では、本ネタの内容であるニュース記事の主題を用いた謎掛けを自動生成している。謎掛けとは、「A とかけて B ととく、その心はどちらも $C_1(C_2)$ 」といった形式で行われる言葉遊びの一種である。ここで、 C_1 と C_2 は同音異義語である。例えば「たこ焼きとかけて音楽と解く、その心はどちらも店舗(テンポ)がつきものです」という具合である。

4. 難語、言い換え、例文の抽出

説明ボケには難語と言い換え語、例文を用いている。そこで、以下に示すように、まずはその基本となる Web ニュース上の難語の抽出、及びその難語の言い換えとその言い換えた語を用いたわかりやすい例文の抽出手法を提案する。

- (1) 対象となる文から難語を抽出する。
- (2) 抽出した難語から言い換える語を決定する。
- (3) 言い換えた語を用いた例文を抽出する。

以下に手法の詳細を示す。

4.1 難語の抽出

本研究では、以下の条件を全て満たす語を難語とする。

条件 1-1 仮名文字と小学 3 年生までに習った漢字だけでは構成されていない

条件 1-2 単語の親密度が低い

条件 1-3 子供用の国語辞典に記載されていない

条件 1-1

教育課程において早い段階で習う漢字はより難易度の低い漢字か、より重要な漢字と考えられる。そこで、より早い段階で習う漢字のみを用いている語は易しい単語であると考え、今回は、文部科学省の指導要領 [11] に掲載されている小学 3 年生までに習う漢字を易しい漢字であるとする。そして、この条件を満たさない漢字を含む語を難語とする。また、ひらがなとカタカナのみを用いた語は易しい漢字を用いた語と同じく易しい単語であることが多く、また、ひらがなは送り仮名としても用いられるため、易しい漢字と同様に扱う。

条件 1-2

例えば、「五月雨」や「波止場」のように、たとえ早期に習う文字のみで構成されたとしても、普段用いられない語も存在する。そこで、実際にはあまり用いない語は難しい語と考え、難語とする。今回は、三省堂の「日本語の語彙特性」[12] に収録された文字単語親密度を用い、予備実験より親密度が 5.2 以下のものを難語とする。

条件 1-3

小学生低学年向けに作成された国語辞書は日常的に用いられる語が多数収録されている。そこで、この国語辞典に収録されている語を除外することで難語以外の語を除外し、残りを難語として抽出する。今回は小学館の子供向け国語辞典 [13] を用いる。

4.2 言い換え語の抽出

言い換え語は、難語に似た意味を持ち、かつ、難語より易しい語である。言い換え語候補の抽出には Weblio のシソーラスを用いる。抽出された候補の中から、以下の条件 2-1~2-3 の優先順位順に該当する語を言い換え語として決定する。2-1~2-3 の条件で言い換え語の候補が複数ある場合は、2-4 の条件により言い換え語を決定する。

条件 2-1 全て仮名文字で構成されている

条件 2-2 仮名文字が小学 3 年生までに習った漢字が含まれている

条件 2-3 子供用の国語辞典に記載されている

条件 2-4 検索結果数の多い語

ボケ	協定を <u>締結</u> すると発表した
ツッコミ	ところで「 <u>締結</u> 」ってどういう意味かわかるとるか？
ボケ	「路面が凍る」の「凍る」みたいな意味やろ？
ツッコミ	ちゃうわ、それは「凍結」や
ツッコミ	「 <u>締結</u> 」は「 <u>契約を結ぶ</u> 」の「 <u>結ぶ</u> 」みたいな感じや
ボケ	細かいな、おんなじやん
ツッコミ	全然ちゃうわ

図 1 説明ボケの例

以下、条件の詳細を述べる。

条件 2-1

全て仮名文字で構成されている単語はより触れる機会の多い単語が多く存在すると考えられる。そこで、すべて仮名文字で書かれている単語を言い換え語の候補とする。

条件 2-2

難語の抽出条件と同様に、教育課程において早い段階で習う漢字はより難易度の低い漢字か、より重要な漢字と考えられる。そこで、より早い段階で習う漢字のみを用いている語の一部は、より早い段階で意味を理解する必要のある語であると考え、今回は、文部科学省の指導要領 [11] に掲載されている小学 3 年生までに習う漢字とし、この漢字で構成されている語を言い換え語の候補とする。

条件 2-3

難語の抽出条件と同様に、小学生低学年向けに作成された国語辞書は日常的に用いられる語が多数収録されている。そこで、この国語辞典に収録されている語を類義語の抽出条件とすることで、日常的に用いられている語とみなし、言い換え語の候補とする。今回は 4.1 節の条件 3 と同じ小学館の子供向け国語辞典 [13] を用いる。

条件 2-4

検索結果数の多い語は、少ない単語と比較して、より一般的な単語であると考えられる。そのため、検索結果数が最も多い単語を言い換え語として抽出を行う。

4.3 例文の抽出手法

例文の抽出手法として、はじめに Web の検索結果を用い抽出を行う。この際用いるクエリは言い換え語と“例文”を用いる。これにより得られた検索結果となるページの上位 10 ページに対しスクレイピングを行い、文を取得する。取得された文から、以下の条件を満たす文を例文候補として抽出する。

- 対象となる語を含んでいる
- “例文”の文字列が含まれていない
- 20 文字以下である

抽出された例文候補の内、佐藤らが Web 上で公開している日本語の文章の難易度を測る Web サービス“帯” [15] により最も難易度が易しいと判別したものを例文として提示する。

5. 説明ボケの手法

我々が提案する説明ボケの例を図 1 に示す。図中の 1 行目は、台本生成に用いたニュース記事の一部分、下線部は、今回提案する掛け合いに用いる難語や、掛け合いのために抽出された単

ボケ	XXXXX難語 AXXXXXXX
ツッコミ	ところで「難語 A」ってどういう意味かわかるとるか？
ボケ	「例文 B」の「言い換え語 B」みたいな意味やろ？
ツッコミ	ちゃうわ、それは「難語 B」や
ツッコミ	「難語 A」は「例文 A」の「言い換え語 A」みたいな感じや
ボケ	細かいな、おんなじやん
ツッコミ	全然ちゃうわ

図 2 説明ボケのテンプレート

語や例文であり、それ以外の部分はテンプレートである。図 2 に説明ボケのテンプレートの例を示す。ここで最初のボケ役の台詞は難語 A を含むニュースの記事の 1 文である。

説明ボケの流れとしては、ボケ役がニュース記事の一部を読み上げた後、ツッコミ役がその文中に存在するある難語に対しその語の意味がわかっているかと問う。これに対し、ボケ役は別の語と取り違い、取り違った語に対する意味の解説を例文を用いて行う。ツッコミ役はそれに対しツッコミを行い、正しい例文とともに解説を行う。

生成の手順を図 2 に示す語を用いて説明する。はじめに掛け合いに用いるための難語 A を抽出する。その後、難語 A に対して言い換え語 A と、それを用いた例文 A を抽出する。また、難語 A に対し取り違える難語 B を抽出し、難語 B に対して言い換え語 B と、それを用いた例文 B を抽出する。これらの単語を用いることで掛け合いを生成する。

5.1 取り違える語の基準

難語 B は以下の条件により抽出を行う。

条件 3-1 難語 A と難語 B は一部の読みが位置を含め同じである

条件 3-2 難語 A と難語 B は読みの字数が同じである

条件 3-3 難語 A と難語 B の意味は類似していない

以下、条件の詳細を述べる。

条件 3-1

難語 A と難語 B は互いにある程度の関連性を持たなければ掛け合いが成立しない。そこで、単語の読みの類似性を用いることで掛け合いができる類似性を持たせる。具体例としては、「大體」という語に対しては「大小」や「物体」などは取り違える語の候補となるが、「大袈裟」などは候補とならない。ただし、同音異義語は読みが全て同じになってしまう、掛け合いが成立しなくなるため除外する。

条件 3-2

難語 A と難語 B の読みの長さに大きな差があると聞き間違いや読み間違いのような掛け合いとならない。そこで、読みの字数が同じという条件を付け加えることでこの問題を回避する。具体的には「確認」という語に対し、「確保」などの語を除外する。

条件 3-3

難語 A と難語 B が互いに意味が類似していると掛け合いが成立しない。そこで、今回は 4.2 節と同じく、Weblio のシソーラスを用いる。難語 A に対し、Weblio のシソーラスから取得した単語群を類語群 A、難語 B に対し、Weblio のシソーラス

から取得した単語群を類語群 B とし、以下の条件を満たす語を用いる。

- 難語 A が類語群 B に含まれていない
- 難語 B が類語群 A に含まれていない
- 類語群 A と類語群 B に共通の語が存在しない

これらの条件により難語 B の候補が単数抽出された場合は、その語を提示する。複数抽出された場合には、難語 A と難語 B で一部の文字が位置を含め同じである語を抽出する。具体的には「確認」という語に対しては「確定」や「承認」などは取り違える語の候補となるが、「認可」などは候補とならない。ただし、この際に文字数は問わないため、「大小」に対し「大規模」なども対象となる。共通の文字を持つ単語が複数存在する場合は、その中から条件 3-1 による読みの共通部分が最も多いものを抽出する。

6. 実 験

提案した言い換えと例文の手法の有用性を示すため、評価実験を行った。

6.1 言い換え語の難易度の評価実験

条件

システムが決定した言い換え語の難易度の評価実験を行った。被験者は 10 人で、クラウドソーシングを用い実験を行った。実験は 50 語の難語に対し難語と類語 4 単語の計 5 単語を提示し、被験者は最も易しい単語を選択する。選ばれた数が最も多いものを言い換え語の正解データとした。

結果と考察

結果の一部を表 1 に示す。言い換え語の適合率は 31%であった。なお、最も多いものが複数存在したデータが 8 個存在したが、これらは正解データの特定ができないため、除外した。また、すべての設問に対し、難語が選ばれたものはなかった。適合率が向上しなかった原因として、仮名文字のみを用いた語を優先して抽出したためと考えられる。例えば、被験者は「異名」に対し「ニックネーム」ではなく「あだ名」を選択したが、システムは「異名」に対しカタカナで構成された「ニックネーム」を抽出したため、「あだ名」のような語が抽出されなかった。しかし、被験者が外来語など仮名文字で構成された語を選択する場合もあるため、仮名文字で構成された語を抽出するという条件は間違いとはいえない。また、「拡充」に対し検索結果数の多い「拡大」が抽出され、検索結果数がそれより少ない「広げる」のような語が抽出されなかった。これは、複数言い換え語の候補が抽出された場合に、検索結果数の多い方を言い換え語候補としていたためであり、適合率が向上しなかった原因として挙げられる。

6.2 例文の難易度の評価実験

条件

システムが提示した例文が易しい例文に置き換えることができたかを評価した。被験者は 10 人で、クラウドソーシングを用いて評価を行った。実験手順は、50 語の難語を用いた例文と、その難語に対する言い換え語を用いた例文をそれぞれ 50 文ずつ用意し、これらを 25 ペアずつ 2 つの集合 A, B に分割する。

表 1 提示した語群と結果の例

提示した難語	提示した類語				被験者	システム
圧巻	ピーク	すごい	山場	名場面	すごい	すごい
異名	通り名	二つ名	ニックネーム	あだ名	あだ名	ニックネーム
遺棄	捨てる	断念	放っておく	見切り	捨てる	捨てる
閲覧	検査	ご覧	読み物	リーディング	読み物	リーディング
拡充	拡大	拡張	広げる	膨張	広げる	拡大

表 2 提示した例文と結果の例

用いた語	例文	評価
言い換え語	今が夏の <u>ピーク</u> です。	易しい
言い換え語	火をもって火を <u>制する</u>	難しい
言い換え語	あなたの <u>存在</u> に疑問を感じます	難しい
難語	星が降ってくるような、 <u>圧巻</u> の夜空	難しい
難語	川は、夏には <u>干上がる</u>	易しい

表 3 例文の評価実験結果 (例文単体)

		評価			合計
		易しい	難しい	同数	
使用した単語	言い換え語	27	17	6	50
	難語	4	46	0	50
合計		31	63	6	100

表 4 例文の評価実験結果 (言い換えの前後)

		言い換え語の例文			合計
		易しい	難しい	同数	
難語の例文	易しい	3	0	1	4
	難しい	24	17	5	46
	同数	0	0	0	0
合計		27	17	6	50

はじめに、集合 A の難語を用いた例文 25 文と集合 B の言い換え語を用いた例文 25 文をランダムに被験者に提示し、評価を行った。後日、集合 A の言い換え語を用いた例文 25 文と集合 B の難語を用いた例文 25 文をランダムに被験者に提示し、評価を行った。評価は「易しい」と「難しい」の 2 つで行った。評価が過半数となった例文を、その評価を例文の評価とした。提示した例文の一部を表 2 に示す。例文中に下線が引いてあるものが使用した言い換え語または難語である。

結果と考察

結果を表 3 と表 4 に示す。表 3 中の数字は例文の数を、表 4 中の数字は難語を用いた例文と、その難語に対する言い換え語の例文の対の数を示す。言い換え語を用いた例文の適合率は 54% であった。言い換え語を用いた例文が難しいと評価された要因としては、「火をもって火を制する」のように、言い回しは難しいが用いている語そのものは難しくなく、かつ難易度の測定ツールが易しい文として抽出しているものが挙げられる。また、「あなたの存在に疑問を感じます」のように、難語「既存」を言い換え語「存在」に言い換えて用いた例文ではあるものの、例文中の別の単語が難しい語である、という例文も言い換え語を用いた例文が難しいと評価された原因として挙げられる。難語を用いた例文が易しいと評価された要因としては、「川は、夏

には干上がる」の「干上がる」のように日常的には用いるが小学校では習わない語が存在したため、これが難語として抽出されたと考えられる。しかし、システムが難語とした語を用いた例文が易しいと評価された数は難しいと評価された数と比較して少数であるため、難語の例文が難しいと評価されていると言える。また、言い換え語を用いた例文は全体の半数以上が易しいと評価されているため、易しい例文が抽出されていると言える。

言い換え前後の例文の難易度の変化の例を、表 5 に示す。下線が引かれている語がそれぞれの例文に用いられている難語または言い換え語である。言い換え前後の例文の難易度の変化は、難しいと評価された例文が言い換えにより易しいに評価が変化したものが 48% であった。また、言い換える前後がどちらも難しいと評価されたものは 34% 存在した。また、言い換え語によりかえって難しい例文となった例はなかった。難語の例文も言い換え語の例文も難しいと評価された要因としては、「参加者については異説もある」から「異議なしです」のように、難語である「異説」をニュースではよく用いられる語ではあるものの、日常会話では用いない言い換え語の「異義」に言い換え、例文を抽出したため、難しい例文が抽出されたと考えられる。よって、難語の例文が難しくなった語が言い換え語の例文に置き換えたなら易しいに変わったものが全体の約半数存在し、言い換え語によりかえって難しい例文となった例が存在しないため、難語を言い換え語に置き換え、例文を提示することで全体としては難易度が下がっていると言える。

6.3 言い換え語と例文の評価

条件

システムが提示したある難語に対する言い換え語とその言い換え語を用いた例文が、元の難語の解説として適しているかを評価した。被験者は 10 人で、クラウドソーシングを用いて実験を行った。実験は、6.1 節の被験者実験により難語に対し言い換え語を正しく抽出されたものを用いて行った。難語と言い換え語、例文は『「難語」は「例文」の「言い換え語」みたいなもの』という形で行った。例えば『「凍結」は『路面が凍る』の『凍る』みたいなもの』の様なものである。評価は「正しい」と「正しくない」の 2 段階で行い、過半数を超えたものを各項目の評価とした。提示した文の一部を図 6 に示す。

結果と考察

被験者の評価が「正しい」と「正しくない」で同数となったものは評価を「正しくない」とし、不正解とした。適合率は 45% であった。適合率が向上しなかった原因として、『「複合」は「合わせる顔がない」の「合わせる」みたいなもの』の「合わせる」

表5 提示した例文と結果の例

難語を用いた例文	判定	言い換え語を用いた例文	判定
たなばた姫につけられた7種類の異名	難しい	あなたに <u>あだ名</u> があれば教えてください	易しい
どのように生き延びたのか	易しい	子供は <u>成長</u> が速い	易しい
参加者については <u>異説</u> もある	難しい	<u>異議</u> なしです	難しい

表6 提示した文の例

文	評価
「巡視」は「防犯パトロール」の「パトロール」みたいなもの	正しい
「補修」は「メンテナンス特集を公開しました」の「メンテナンス」みたいなもの	正しい
「複合」は「合わせる顔がない」の「合わせる」みたいなもの	正しくない
「試用」は「リーディングテスト」の「テスト」みたいなもの	正しくない

のように、難語と言い換え語のみでは正しく言い換えが行われているが、例文中の言い換え語の意味と元の難語の意味にずれがある。ここで、複数の意味を持つ言い換え語に対して考慮する手法の提案が必要であることが分かった。

7. まとめと今後の課題

本研究では、ニュース記事中に存在する難語を抽出し、それをわかりやすい語に言い換え、例文を提示する手法を提案し、それに対する実験を行った。これにより、記事中に含まれる語に対し難語であるかを判定し、それに対する言い換え語と言い換え語を用いた例文を抽出することができることを確認した。また、この言い換えた語と例文を用いた新たな掛け合いである説明ボケを提案した。

言い換え語の抽出に対する今後の課題として、各条件の優先順位が挙げられる。これは、仮名文字のみの語を優先したために解りにくい外来語が抽出されたため、そうでない語も含めた上で抽出対象を決定するかの改善が必要となるためである。また、複数の言い換え語が抽出された場合の順位付けの変更も今後の課題となる。順位に検索結果数のみを用いているため、易しい語より説明文などでよく用いられる語が多く抽出される。しかし、その中でも難語が存在するためこれらの改善が必要となる。また、難語や言い換え語で用いている漢字を習う学年の閾値の調整も今後の課題とする。これは、難語の評価に対する実験で、今回設定した小学3年生までに習った漢字ではない漢字を用いた単語が易しいとなった場合が存在するためである。これと同時に、外来語のように仮名文字だけで構成された難語も存在するため、これらを難語とするための手法も今後の課題とする。

例文の抽出に対する今後の課題として、文中に含まれる語の難易度の考慮が挙げられる。これは、言い換え語の例文中に別の難語が入り、それによって例文が難しくなっている場合が存在しているためである。また、言い換え語と例文中の言い換え語の意味にずれがある場合があるため、これに対する手法の検討も今後の課題とする。

説明ボケは難語の抽出の他に、言い換え語の抽出やそれに対する例文の抽出が必須となる。しかし、難語に対する言い換え語の抽出手法や、言い換え語に対する例文の抽出手法に対する改善が必要であるため、説明ボケそのものの評価実験も今後の

課題とする。

謝 辞

本論文の一部は JSPS 科研費 17K00430, 16K07973 及び、私学助成金（大学間連携研究補助金）の助成によるものである。ここに記して謹んで感謝の意を表する。

文 献

- [1] 総務省, “情報通信白書平成 28 年版”, <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/>
- [2] 青木哲, 梅谷智弘, 北村達也, 灘本明代, “Word2Vec を用いた対立語に基づく漫才台本の自動生成”, DEIM Forum 2017 F7-3
- [3] 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
- [4] 岡崎直観, 石塚満, “言い換え可能な括弧表現の抽出法”, 言語処理学会第 13 回年次大会, pp.911-914, 2007.
- [5] 内田 ゆず, 荒木 健治, 米山 淳, “ブログ記事からのオノマトベ用例の自動抽出手法”, 知能と情報 (日本知能情報フェジィ学会誌) Vol. 24 No. 3, pp 811-820, 2012
- [6] 下川尚亮, Rafal Rzepka, 荒木健治, “対話文生成のための Web を用いた話題語の抽出”, 自然言語処理研究会報告, Vol.2009, No. 2, pp. 121-126, 2009
- [7] 福岡知隆, 税田竜一, 久保村千明, 服部峻, 亀田弘之, “文の類似性を用いた未知語処理手法の提案とそれに基づく円滑な対話応答システムの作成”, 全国大会講演論文集 第 72 回 (人工知能と認知科学), pp 619-620
- [8] 塚原裕史, 内海慶, “対話行為と話題推定によるラベル伝搬を利用した雑談生成方法の改良”, 第 30 回人工知能学会年次大会 2016
- [9] 竹越智也, 萩原将文, “ロボット漫才自動生成システム:動作が漫才に与える影響の考察-”, 日本感性工学会論文誌 Vol. 15, No. 1, pp 47-54
- [10] 元祖爆笑王, “漫才入門 ウケる笑いの作り方全部教えます”, リッターミュージック, 2011
- [11] 文部科学省, “小学校学習指導要領 別表 学年別漢字配当表”, http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/koku/001.htm
- [12] 天野成昭, 近藤久久, “日本語の語彙特性 第一期”, 三省堂 2003 年
- [13] 小学館, “ドラえもんはじめての国語辞典”
- [14] 文化庁, “常用漢字表”, <http://www.bunka.go.jp/kokugo-nihongo/sisaku/joho/joho/kakuki/14/tosin02/>
- [15] 佐藤理史, “日本語テキストの難易度を測る - ことば不思議箱”, <http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/>