ニュースの難語の言い換えを用いた例文提示手法

11471036 義尚晃（灘本研究室）

あらまし：インターネット上には日々様々なニュースが流れ様々な年齢層が利用している．しかしながら小学生は語彙が少ないため，ニュースは理解困難である問題がある．そこで，本研究ではニュース記事の中で使われている難しい言葉に対して易しい言葉に言い換えると共に，その易しい言葉を用いたわかりやすい例文を提示する手法を提案する．

1. はじめに

近年，スマートフォンなどの普及により小学生がインターネットを利用する機会は年々増加傾向にある．それに伴い，インターネットでの情報の取得が多くなってきている．小学生のインターネット利用目的の大半はゲームなどの娯楽である．それに対してニュースをみている人は少数である[1]．その要因として娯楽のコンテンツが楽しいという点もあるが，小学生は語彙が少ないため，一般的な Web ニュースを読むことが難しいと考えられる．また，文章の読解力にも個人差があるため，文章を理解できたとしても，その理解度に差が生じる．そこで Web ニュースの読解を支援できれば，小学生がニュースにふれる機会を高めることができると考えた．

現在，小学生の Web ニュース読解を支援するツールの一例として，NHK が Web 上で公開している「NEWS WEB EASY」[2]や Yahoo!が公開している「Yahoo! きっず」[3]がある．例えばNHKのサイトの機能として，漢字にはふりがなが付与され，難しい漢字には辞書の説明を付与されている．そして小学生にもわかる易しいことばでニュースが記載されている．しかしながら，日々のニュース記事は膨大であり，その中で子供に見せたいニュース記事が存在しているのに対し，これらのニュースサイトにおける単語の説明の付与や提示する情報の整備は人手で行われている．そのため，更新頻度は 1 日に 5 記事のみで，利用できる記事数がかなり少ない．

そこで本研究では，Web ニュースに対してニュース記事内で使用されている難解な語に対し，小学生（低学年）にわかりやすい例文を自動で提示すると便利であると考え，この手法の提案を行う．具体的には，ニュースから難語を抽出し，抽出した難語を易しい語に言い換え，その語を用いた例文を抽出し提示する手法の提案を行う．

2. 難語の抽出

本研究では，難語をニュースの中に出てくる一般語で難解なもの（人名、地域、固有名詞は除く）と定義する．以下に難語の条件を記す．

1. 小学生（高学年以上）で学ぶ漢字が入っている単語．

ニュースに使用されている言葉の中には，小学生高学年以上で学ぶ漢字が含まれている語がある．その為，低学年生は理解できないと考え，高学年以上で学ぶ漢字で構成されている語を難語の候補とする．

2. 単語の親密度の値が 5.2 より低いもの

難語は親しみにくい単語であると考え,天野ら[4]の提案する親密度を用いて，実験により親密度が 5.2 以下の単語を，難語の候補とする．

3. 子供辞書に記載されていない単語

小学生を対象とした子供国語辞典[5]に掲載されていない単語は小学生にとって難語であると考え，難語の候補とする．

3. 難語の言い換え

3.1 類語候補の抽出

抽出した難語の類語候補の抽出にはWeblio [6]のシソーラスを用いる．この時，掲載されているすべてのシソーラスを類語候補とする．

3.2 類語候補群から言い換え候補の決定

抽出された類語候補の中から，以下の条件1～3の優先順位に該当する語を言い換え語として決定する．1～3の条件で言い換え語の候補が複数ある場合は4の条件を使う．

1. 類語候補にひらがなで構成されている語がある

小学生低学年は漢字で書かれている言葉よりもひらがなで書かれている言葉に触れる機会が多く，より親しみがあると考え，ひらがなで表現されている言葉を言い換え語とする．

2. 類語候補に小学生（低学年）で学ぶ漢字が入っている．

本研究の対象が小学生低学年のため，言い換え語に含まれる漢字は低学年で学ぶ漢字のみとする．

3. 類語候補が子供国語辞書に掲載されている．

子供国語辞書に掲載されている単語は易しい単語であると考え，子供国語辞書に掲載されている単語を言い換え語とするに置き換える．

4. 類語候補の内，検索結果の多い語．

検索結果の多い語は少ない語に比べ，一般的であると考える．その為，類語候補が複数ある場合は検索結果の多い語を言い換え語とする．

4. 例文決定

3 章で決定した言い換え語＋例文をクエリとし検索結果の上位 10 件のページをスクレイピングする．抽出した候補の中から佐藤らが 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」[7]を使用し一番難易度が易しいものを例文とする．

5. 評価実験

提案手法の有用性を示すために，言い換え語の難易度評価実験と例文の難易度評価実験の 2 つの実験を行った．

5.1 言い換え語の難易度評価実験

実験条件

システムが決定した 50 語の言い換え語の適合率を求め，提案手法の有用性を示す．正解データは，クラウドソーシングを用いて 10 名の被験者により決定した．具体的には，被験者に言い換え語とその類語 4 単語，計 5 単語を提示し最も易しい単語を選択してもらった．これを 50 セット行った．

結果と考察

実験の結果，適合率は 28%であった．表 1 に被験者に提示した単語と正解データとシステムが易しいと判定した単語を示す．

表 1：5.1 実験結果

提示した単語	被験者	システム
圧巻・ピーク・すごい・山場・名場面	すごい	ピーク
遺棄・捨てる・断念・見切り・放っておく	捨てる	捨てる
拡充・拡大・拡張・広げる・膨張	広げる	拡大

適合率が低い値になった要因としては，アンケートでは，「ニックネーム」⇒「あだ名」，「ピーク」⇒「すごい」，などカタカナ語が含まれていない語が選択されていることが見られる．これはカタカナよりもひらがなが含まれている語の方がより易しくなったためであると考える．また「工作」⇒「考え」等辞書には載っているが言い回しが違い，選ばれなかった語も存在していた．しかしながら，正解データの傾向を鑑みると，本研究のひらがなで構成されている語を言い換え語の候補にするという条件は，適切であるといえる．今後の課題として，子供辞書の語彙の拡張が必要であると考える．

5.2 例文の難易度評価実験

実験条件

本研究での提案手法で推薦された例文に対し，その文が易しい文であるかの実験をクラウドソーシングを用いて行った．被験者は 10 名である．表 2 に示すように，まず被験者にランダムに選択した難語を含む例文と言い換え語を含む別の例文合計 50 文を提示し，難しい/易しい いずれかを判定してもらった．また別の日に，先に示した例文に対して，難語を含む例文は言い換え語を含む例文に，言い換え語を含む別の例文はその難語を含む例文に変更して，被験者に提示し同様の判定をしてもらった．どちらの実験においても，ある例文に対して難しい/易しい いずれかの回答が過半数を超えた場合その難易度を例文の難易度とした．

表 2：例文を用いた実験結果

難語を含む文	被験者の回答	言い換え語を含む文	被験者の回答
あなたに異名があれば教えてください	難しい	あなたにあだ名があれば教えてください	易しい
算段をして百円だけこしらえた	難しい	やりくりをして百円だけこしらえた	易しい
参加者については異説もある	難しい	参加者については異議もある	難しい

結果と考察

難しい語を含む例文より言い換えた語を含む例文の方が易しくなったとなった文が 50 文中 31 文あり適合率は 62%であった．5.1 の実験結果よりひらがなが多く含まれている言葉のほうがより易しいと判断されやすいとわかった．
漢字のみで構成される熟語をひらがなが含まれている語に言い換えられたため，例文も易しくすることができたのではないかと考える．一方で，漢字のみで構成されている熟語を別の熟語に言い換えしまったため，例文も難しくなったという場合もあった．これは，子供辞書の語彙数が少ないためであるといえる．

上記二つの実験により，例文での精度が高く単語レベルでの語の言い換えの精度が低いことがわかった．このことにより，言い換えの候補をより簡単にすることができれば言い換えの例文の精度を上げることができるとわかった．

6. まとめと今後の課題

本研究では難語を易しい語に言い換え，その語を用いた例文を抽出し提示する手法の提案を行った．今後の課題としては，被験者を増やすこと．難語の正確な言い換え候補の決定の精度の向上．ニュースの文脈にあった例文の提示も考えていきたい。

参考資料

[1] 内閣府，“平成 28 年度青少年のインターネット利用環境実態調査結果（速報）”
[2]NEWS WEB EASY, http://www3.nhk.or.jp/news/easy
[3] Yahoo! きっず，https://kids.yahoo.co.jp/
[4] 天野成昭，近藤公久，“NTT データベースシリーズ　日本語語彙特性　第 1 期　CD-ROM 版”，三省堂.
[5] 小学館国語辞典編集部，『ドラえもん　はじめての国語辞典』，(2013)，小学館.
[6] Weblio，https://thesaurus.weblio.jp/
[7] 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/