

## やさしい日本語ニュースのやさしさの分析

李 品其<sup>†</sup> 黎明かん<sup>†</sup> 杉本 徹<sup>‡</sup>芝浦工業大学大学院 理工学研究科<sup>†</sup> 芝浦工業大学 工学部<sup>‡</sup>

## 1. 研究背景と目的

2019年6月末の在留外国人数は282万9416人で、前年末に比べ約10万人(3.6%)増加となり過去最高であった[1]。日本で生活している外国人の中には日本語が十分理解できない人も多くいるため、普通の日本語よりも簡単で外国人にも分かりやすい「やさしい日本語」でニュースなどの情報提供を行うことが望まれている。そこで、やさしい日本語に関して多くの研究やガイドラインの作成が行われている[2]。やさしい日本語における1文の文字数は平均24文字、長くても30文字以内が適切であるとされており、旧日本語能力試験3級程度が基準とされている。

NHKでは、国内在住の外国人のためにNHKオンラインの通常ニュースの「NHK NEWS WEB」をやさしい日本語に書き換えたニュースである「NEWS WEB EASY」を2012年4月から提供している[3]。

我々は、日本で生活する外国人留学生が日常目にする日本語の実用的な文章を理解しやすいやさしい日本語に書き換えるシステムの開発を目指している。文章中で書き換えが必要な所と書き換えの方法を検討するため、本研究では、やさしい日本語ニュース「NEWS WEB EASY」のやさしさを、記事文の分析とインタビューの2つの観点から分析する。記事文の分析では、やさしい日本語の基準に基づき、文の構成と単語の難易度に関して分析を行う。インタビューでは、外国人留学生に記事を読んでもらい、難しいと感じた点を答えてもらう。それらの結果から、実用的な文章をやさしい日本語に書き換える際の課題について考察する。

## 2. コーパスの作成

記事文を分析するために、NHKの「NEWS WEB EASY」(<https://www3.nhk.or.jp/news/easy>)に2019年6月から9月の間に掲載されたやさしいニュースと、それに対応する元のニュース各100記事を人手で集めてコーパスを作成した。

## 3. 記事文の分析

## 3. 1. 記事ごとの文数、字数と単語数

作成したコーパスを用いて、記事ごとの文数、字数と単語数を計算した。単語数を求めるために、ニュース記事を1文ごとにMeCab[4]を使って形態素解析した。やさしいニュースと元のニュースの記事ごとの文数、字数と単語数の平均を求めた結果を表1に示す。

表1: 記事ごとの文数、字数と単語数

	やさしい ニュース	元のニュース
平均文数	9.0	16.4
平均字数	321.9	944.2
平均単語数	199.3	562.4

## 3. 2. 旧日本語能力試験の級別の単語数

次に、記事ごとの旧日本語能力試験の級別の単語数を計算した。旧日本語能力試験は、日本語を母語としない人を対象として日本語能力を測定し、認定することを目的としている[5]。日本語能力を入門レベルの4級から最上級の1級までの4段階で認定する。また旧日本語能力試験の出題基準には各級の出題の目安となる語彙や文法事項のリストが公開されている[6]。

「NEWS WEB EASY」に掲載されたやさしいニュースと元のニュースに含まれる旧日本語能力試験の級別の語彙を求めて、記事ごとの単語数の平均を計算した。結果を表2に示す。

やさしい日本語では、旧日本語能力試験の3級程度の単語が基準とされているが、今回の調査ではやさしいニュースの中に1級の単語が0.8%、2級の単語数が6.6%含まれていた。

Analysis of understandability of simplified Japanese news

<sup>†</sup>Li Pinqi, <sup>†</sup>Li Minghan, <sup>‡</sup>Toru Sugimoto

<sup>†</sup>Graduate School of Engineering and Science, Shibaura Institute of Technology

<sup>‡</sup>College of Engineering, Shibaura Institute of Technology

表 2 : 旧日本語能力試験の級別の単語数

平均単語数	やさしい ニュース	元のニュース
1 級	1.6	27.4
2 級	13.1	75.5
3 級	21.1	37.4
4 級	43.8	15.9

#### 4. 留学生へのインタビュー

実際に外国人がやさしい日本語ニュースをどのくらい理解しているか調べるために、日本で生活している留学生へのインタビューを行った。実験では6月5日と6月10日のニュースを使用し、各ニュースの最初の5つの文を用いる。被験者の半数には、6月5日のやさしいニュースと6月10日の元のニュースを読んでもらう。残りの半数には、6月10日のやさしいニュースと6月5日の元のニュースを読んでもらう。そして読んだ記事をどの程度理解できたか答えてもらう。また、記事の中で理解できなかった単語を選択してもらう。

被験者は日本で生活している留学生10人で、内訳はインドネシア人2人、タイ人1人と中国人7人である。また、日本語能力試験のN1合格者が1人、N2が3人、N3が4人、N4以下が2人である。インタビューの結果、被験者の過半数が理解できないと答えた単語を旧日本語能力試験の出題基準ごとに分類したものを表3と表4に示す。

表 3 : 6月5日のニュースで留学生が理解できない語

	やさしいニュース	元のニュース
1 級		
2 級	突っ込む、ぶつかる	突っ込む、衝突、複数、死亡
3 級	反対側、けが	
4 級	交差点	
その他	ながら	同乗、逆走、反対車線、乗用車、猛、乗り上げる

表 4 : 6月10日のニュースで留学生が理解できない語

	やさしいニュース	元のニュース
1 級		決勝、ファン
2 級		委員会、消費
3 級	試合	試合
4 級		
その他	ラグビー、ワールドカップ	準々、開幕、飲食店、増設、サーバー、ラグビー、ワールドカップ、イングランド

また、やさしいニュースと元のニュースの理解程度の分布を表5に示す。

表 5 : やさしいニュースと元のニュースの理解程度

	やさしい ニュース	元のニュース
0%-20%	0 人	0 人
20%-40%	0 人	1 人
40%-60%	0 人	2 人
60%-80%	7 人	6 人
80%-100%	3 人	1 人

#### 5. 考察

留学生ができなかった単語は、旧日本語能力試験出題基準における1級と2級の単語、および基準に掲載されていない単語が多かった。

基準に掲載されていない単語にもやさしい単語と難しい単語があるので、文章をやさしい日本語に書き換える際には任意の単語の難易度を自動推定できることが望ましい。また、カタカナの単語が理解できないことが多いので、外来語には原語のつづりを付けるなどの配慮が必要である。

#### 6. 結論

本研究では、やさしい日本語ニュース「NEWS WEB EASY」の記事文の分析と、留学生への理解程度のインタビューを行った。

今後は、実用的な文章をやさしい日本語に自動的に書き換える方法、特に単語の難易度推定と平易化の研究を進めていきたい。

#### 参考文献

- [1] 法務省：“令和元年6月末現在における在留外国人数について(速報値)”，2019年10月25日
- [2] 弘前大学人文学部社会言語学研究室：“減災のための「やさしい日本語」”  
<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ1a.htm> (2020年1月9日確認)
- [3] 田中英輝, 熊野正, 後藤功雄, 美野秀弥：“やさしい日本語ニュースの制作支援システム”，自然言語処理, Vol.25, No.1, pp.81-117, 2018
- [4] MeCab <https://taku910.github.io/mecab/>
- [5] 日本語能力試験 <http://info.jees-jlpt.jp/>
- [6] 日本語能力試験 出題基準, 国際交流基金, 日本国際教育協会, 凡人社, 1994