

# 公的文書に対する「やさしい日本語」換言辞書作成のための調査

李 真奈見<sup>†</sup> 山本 和英<sup>†</sup>

<sup>†</sup> 長岡技術科学大学 {moku, yamamoto}@jnlp.org

## 1 はじめに

現在、日本に在住する外国人は 200 万人を超え、その中でも日常生活に必要な日本語能力を持たない外国人は数十万人に及ぶ。しかし、一般的に日本社会で日本語以外は使用されない。よって、外国人が日本で生きていくためには、日本語の知識は必要不可欠である。現状では、一部の Web や案内における情報は多言語で表記される傾向にある。しかし、世界には多くの言語が存在するため、用意しなければならない言語数は多い。しかし、言語ごとに資料を作成するためには、時間や金銭面で提供側の負担が大きいので、多言語化は行き届いていない。以上のことから、全ての外国人に対する情報提供の支援策として、多言語化は最善とは言えない。

こうした外国人のために、必要最低限の日本語を提示する「やさしい日本語」が必要である。「やさしい日本語」とは、日本語母語話者が日本語の文法や語彙に制限をかけて難しい日本語から「やさしい日本語」へ変換を行ったものである。ここでの難しいとは、最低限の文法と語彙を習得した日本語初學者の外国人が理解できないこと、やさしいとは日本語初學者でも理解できることである。「やさしい日本語」を用いることで、日本語初學者に最低限の文法と語彙を習得してもらうこと、日本語初學者でも読むことのできる文書への書き換えを自動化することの 2 つを目的としている。そして、この 2 つを合わせることで、日本語母語話者と日本語初學者がお互いに歩み寄ることを目指している。日本語初學者を対象としている理由は、「やさしい日本語」が日常生活に必要な日本語能力を持たない外国人に最低限の文法と語彙を習得してもらった後に言い換えた文書を読むことを想定しているからである。

我々は、後述する関連研究により公的文書を用いた「やさしい日本語」のコーパスを活用するためには換言辞書の作成が有効であると考えた。本研究では、日本語初學者でも読むことのできる文書への換言のため、人手により換言辞書を作成した。換言辞書の評価実験は日本語初學者である留学生を対象として行った。また、書き換えを行う文書は公的文書を対象としている。公的文書とは、市役所や病院、学校等の公共施設で配られる文書を指しており、これらの文書は生活するために重要な情報を多く含んでいる。しかし、日本語初學者が学習する文に比べ理解が困難であり、特有な表現も含むため、「やさしい日本語」へ換言する必要がある。

## 2 関連研究

関連研究として、災害時のための「やさしい日本語」[1]がある。これは、外国人が学習する機会が少ない災害時特有の表現を理解しやすくするために文構造の規則や換言対を作成した研究である。以下に災害時特有の表現を、「やさしい日本語」へ換言した例を示す。

例 1)

- a. 火の元を確認する。
- b. 火を消す。

例 1 の a は災害時に使用される一般的な指示である。しかし、この言語表現は災害時特有であるため、学習範囲外となってしまうことが多い。このことが原因で、災害時の危機的状況下においても、理解することができない恐れがある。これを b のように換言することで、必要最低限な日本語のみを学んだ外国人でも指示を理解することができる。よってこの研究では特殊な状況下における言語表現と「やさしい日本語」の文構造の規則や換言対が作成された。この結果、外国人に対する日本語の理解に有効であると考えられる。

松田ら[2]は、統計的機械翻訳を用いて公的文書の日本語を「やさしい日本語」へ換言する研究を行っている。テスト文 200 文について換言実験を行い、翻訳結果と正解データの一致度を BLEU 値を用いて評価した結果、20%と低くとどまっている。原文と難しい日本語から「やさしい日本語」に訳した逐語訳、そして機械翻訳による換言の 3 種類の文を手で①比較的良質、②解読不能・翻訳誤り、③変化せずの 3 段階に評価したところ、①比較的良質な換言フレーズは 7.5%と少なかった。そのうちのほとんどが見出しのような短い句であった。これは、コーパスの形式や質、量が機械翻訳に適していないことが原因としてあげられている。質の問題では、「①」、「●」、「TEL」、「→」等の記号等がノイズとなり、翻訳誤りが起きている例が多くあげられている。他にも、コーパスは人手でつくられたので誤字・脱字、表現順序の入れ替わり等も含まれており、これらが翻訳誤りの原因と考えられる。

以上のことから本研究では、自動で処理が難しい翻訳誤りやノイズの除去、多くの換言対を得るために人手で換言辞書を作成する。これによって記号や人為的な誤字・脱字、表現順序の入れ替わり等を省くことができ、換言の精度が上昇すると考える。

### 3 使用するコーパス

本研究では、2名の日本語教師が公的文書の日本語を逐語訳、意識、要約の3段階の「やさしい日本語」に訳したものをコーパスとした。

公的文書は日本語初学者が学習する文に比べ理解が困難であり、特有な表現も含むため、「やさしい日本語」へ換言する必要がある。以下に「やさしい日本語」へ換言した例を示す。

例 2)

- a. 予防接種
- b. 予防注射

「予防接種」は重要な情報だが、日本語の学習内容として一般的ではないため、理解できない外国人が多い。しかし、「接種」を一般的な語彙である「注射」に換言することによって、意味を理解できる可能性が高くなる。

このように、公的文書に含まれる特有の文または単語を「やさしい日本語」に換言したコーパスが作成された。その際、一定の文法基準[3]と「日本語能力試験 2 級レベル」の語彙に制限した。

本研究では「やさしい日本語」の3段階の訳である逐語訳、意識、要約を以下のように定義する。

逐語訳:日本語文の「難しい」表現を「やさしい」表現に、忠実に訳したもの。

意 訳:文意等を損なわないように可能な限り「やさしい日本語」に書き換えたもの。

要 約:可能な限り文を簡約化したもの。

「難しい」、「やさしい」の基準は日本語教師の主観である。以下にそれぞれの訳の具体的を示す。

例 3)

原 文:ニュース等で報道されておりますように、世界的に新型(豚)インフルエンザの流行が危惧されています。

逐語訳:ニュースなどにもあるように、世界中で新型インフルエンザの流行が心配されています。

意 訳:ニュースでもありますが、世界中で新型インフルエンザが増えています。

要 約:世界中で新型インフルエンザが増えています。

このコーパスは、日本語教師の直感を引き出すことを目的として作成したため、多くの基準や制限を設けるとこれを阻害する可能性がある。よって、「やさしい日本

語」の研究では、文法の制限は最低限の基準のみが定められた。文法基準は、例えば「～は…です」「たぶん～です/ます」「～たほうがいいです」等の基本的な文法が使用できる。語彙の制限において、日本語能力試験 2 級レベルとは 6000 語程度である。

筒井ら[4]の研究により、通知等の理解に役立つ言語表現は公的文書内では頻出する言葉であり、学習する必要のある言語表現と考えられたため訳さないことにした。以下にそれら言語表現の例を示す。

例 4)

手続き, 登録, 申請, 機関, 場合, 証明, 緊急, 書類, 地域, 年度, 保護者までに, について, ~先(問合せ先, 振込先など), とは

コーパスは 3700 文の原文とそれに対応した 3 段階の訳で構成される。翻訳では多くの翻訳対が必要であると考えられることから比較すると、本研究で使用したコーパスの量はとても少ない。しかし文構造が類似する言語同士では少ない換言対でも統計的機械翻訳の枠組みによる換言は可能と考える。

### 4 換言辞書の作成

本研究では、松田ら[2]と同様に「やさしい日本語」の原文と逐語訳に着目する。換言辞書は原文と逐語訳の対を比較し、相違部分を人手で抽出することで作成した。以下に原文と逐語訳から抽出した換言対の例を示す。

例 5)

原 文:今年度より「保健調査票」が変更になります。

逐語訳:今年から「保健調査票」が変わります。

→換言対:今年度 ⇒ 今年

より ⇒ から

変更になり ⇒ 変わり

換言対を増やすため、同一の換言箇所であっても、有効と判断した場合は複数の表現対に対して抽出した。以下にその例を示す。

例 6)

原 文:保護者の方

逐語訳:家族の人

→換言対:保護者の方 ⇒ 家族の人

保護者 ⇒ 家族

作業の結果、換言辞書として、2500 対の換言対を作成した。以下に内容の一例を示す。

- 単語の換言  
接種 ⇒ 注射
- 複合名詞の分割  
任意接種 ⇒ 任意の注射
- 年号を西暦に換言  
平成 21 年 ⇒ 2009 年
- 余分な情報の省略  
等、を対象、各
- 省略された部分の補完  
(月)⇒(月曜日)
- 丁寧語・尊敬語・謙譲語の一部削除  
ご理解 ⇒ 理解  
たまりまして ⇒ いただいて
- 補助動詞の削除  
お読みいただき ⇒ 読んで
- カッコ内の説明の削除・選択・付与  
新型(豚)インフルエンザ ⇒ 新型インフルエンザ  
低学年(3 年生まで) ⇒ 3 年生まで  
学年積立金 ⇒ 学年積立金(学校のお金)
- その他、定型文や補完されているもの  
助成措置が講じられております ⇒ 少し安くなります  
が流れるたびに ⇒ を見ると  
惨状が思い出され ⇒ を思い出して  
皆様におかれましては ⇒ 皆さんは  
臨時休業 ⇒ 学校は休み

換言対における原文の形態素<sup>1)</sup>に着目すると約半数が名詞であって、複合名詞を名詞+助詞+名詞等の分割した形に換言するものが多かった。また、名詞以外の約半数は動詞であって、そのほとんどが敬語に関するものであることがわかった。

## 5 評価実験

作成した換言辞書が日本語初学者に有効であるかを評価するために小規模の実験を行った。公的文書特有の単語を用いて 20 文の評価データを作成した。これを換言辞書により人手で換言し、元の文と換言後の文のどちらが「やさしい」かについて人手で評価した。以下に元の文と換言後の文の例を示す。

例 7)

- a.元の文: 平成 22 年 8 月 16 日(月)~20 日(金)に市立病院にて予防接種を行います。
- 換言後の文:2010 年 8 月 16 日(月曜日)~20 日(金曜日)に市立病院で予防注射を行います。

- b.元の文: ご理解のうえ、ご協力くださいますようお願い申し上げます。

換言後の文:理解して協力をお願いします。

- c.元の文: 緊急時のために、かかりつけ医の連絡を。

換言後の文:緊急の時のために、いつも行く病院の電話をしてください。

本研究は留学生や外国人労働者等を対象にしているため、「やさしい」の判断は留学生 2 名が行った。留学生に元の文と換言後の文を提示し、①元の文の方がやさしく感じる、②換言後の方がやさしく感じる、③どちらも同程度の難易度を感じる、④どちらも難しいという 4 種類で評価してもらった。

評価実験は日本語歴が 7 年と 4 年である 2 名のマレーシアからの留学生を被験者として行った。評価実験の結果、被験者が換言後の文をやさしくなったと判断した文は、評価用データ 20 文のうち平均 15.5 文であった。この結果から、日本語初学者に公的文書を用いた「やさしい日本語」の換言辞書は有効であるといえる。

しかし、評価者のうち 1 名の評価は②換言後の方がやさしく感じると判断した文が 11 文に対し、①元の文の方がやさしく感じると判断した文が 8 文と多かった。

## 6 考察

本研究で作成した換言辞書の動詞のほとんどが敬語に関するものであった。このことから、「れる・られる」や謙譲語・尊敬語・丁寧語が頻出することが公的文書の特徴の 1 つであると考えられる。これらを日本語初学者は理解できないため、彼らは情報を手に入れることができないと考えられる。また、半数を占める名詞の換言には複合名詞を名詞+助詞+名詞等の分割した形に換言するものが多かった。以下にその例を示す。

例 8)

- a.複合名詞:予定人数  
換言の後:予定の人数
- b.複合名詞:必要書類  
換言の後:必要な書類
- c.複合名詞:地区住民  
換言の後:地区に住んでいる人

複合名詞が多いことは漢字圏以外の外国人が意味を理解できない原因と考えられる。複合名詞は区切る場所を提示することで、意味を理解しやすくなる。しかし、

複合名詞を全て換言するわけではなく、例 2 に示したように「予防接種」は「予防注射」となり、「予防の注射」とは換言しないという場合もあった。「予防の注射」は一般的ではなく、不自然さを感じるためである。このように、全ての複合名詞を換言するわけではないということが、「やさしい日本語」において日本語教師の直感を引き出している部分と言える。

実験結果において、①元の文の方がやさしく感じると判断した文が多い被験者がいた。この原因は被験者の日本滞在歴が 5 年と長いため、経験により公的文書特有の表現を理解できるが、換言後の文は一般的に使用されないの、理解しづらかったのだと考える。

換言しても読めない語彙については、例 4 で示した学習する必要がある言語表現であると考えられ、換言しなかった。そのため、学習していない留学生は読むことができなかった。このことから、現時点で実際に換言システムを作成する際は、これらの表現を注釈等で説明する必要がある。

本実験では被験者を留学生としたが「やさしい日本語」の対象は留学生だけでなく、外国人労働者も含まれる。しかし、外国人労働者は留学生と生活環境が異なる。留学生は主に日本語を学び、話すことを目的に日本に滞在しているが、外国人労働者は日本で働くために滞在しているので、一般的に日本語を学ぶことは目的に含まれていない。そのため、彼らの言語知識は働くために必要最低限な日本語に偏っていると考えられる。そういった外国人に「やさしい日本語」を習得してもらい、評価実験を行うことで換言辞書の有効性がさらに高まる。このような問題を解決するためにも「やさしい日本語」学習の普及が求められる。

現コーパスの逐語訳では、表現順序が入れ替わる、表現の削除や補完が行われるといった問題があった。以下にその例を示す。

例 9)

原文: 万が一現地でインフルエンザが発症した・・・

逐語訳: 旅行のとき、もしインフルエンザが出た・・・

例 9 の下線部を見ると、意味としては「万が一」が「もし」、「現地で」が「旅行のとき」へ換言されていると考えられる。しかし、これらは表現順序が変わっており、また、場所の情報である「現地で」が補完され、状況の情報である「旅行のとき」へ換言されている。これらの対を用いて機械的に換言を行うと、表現順序が変わったまま登録した語の内、1 つの単語のみを換言に求められた場合、対応できない可能性が高い。また、補完は前後の文章から推測することが必要なので他の文章においてその

まま利用できることは稀である。これらのことから、表現順序の入れ替わりや補完は、原文を機械的に逐語訳に変換する際に問題となると考える。

また、本研究で使用したコーパスは規模が小さいため、ある程度の換言は可能だが被覆率は低い。

そこで複数の日本語教師により、さらに大規模なコーパスが作成されている。新コーパスでは例 9 のようにより逐語性を高く、すなわち表現順序を維持して過不足のない文とする予定である。コーパスの質の向上が換言精度の向上に、量の増加がより多くの文の換言につながると考える。

## 7 おわりに

日本語初学者のために公的文書を「やさしい日本語」に自動的に換言する研究を進めている。本研究では「やさしい日本語」コーパスの逐語訳から人手で換言辞書を作成し、小規模な実験によってその有効性を確認した。公的文書には複合名詞や敬語が多く、これらが日本語初学者の読解を妨げていることを実験で明らかにした。今後は換言辞書を用いた換言の自動生成に取り組んでいきたい。

## 参考文献

- [1]弘前大学人文学部国語学研究室『新版・災害が起こったときに外国人を助けるためのマニュアル』  
<http://human.cc.hirosaki-u.ac.jp/kokugo/newmanual/top.html>
- [2]松田真希子.やさしい日本語への自動言い換えシステムの開発.日本語教育学会大会 2009(平成 21)年度春季大会予稿集, pp.91-93 (2010)
- [3]庵功雄.「やさしい日本語」をめぐる.多文化共生社会における日本語教育研究会 第4回研究会(2008)
- [4]筒井千絵.試用版書き換えコーパスの作成.日本語教育学会大会 2009(平成 21)年度春季大会予稿集, pp.86-87 (2010)

## 使用した言語資源およびツール

- 1) 形態素解析器 ChaSen, Ver.2.3.3,  
奈良先端科学技術大学院大学松本研究室,  
<http://chasen.naist.jp/hiki/ChaSen/>,