

やさしい日本語書換えシステムの開発と評価

川村 よし子

東京国際大学言語コミュニケーション学部

キーワード：日本語教育 やさしい日本語 日本語弱者 共生社会 eラーニング 語彙

1. はじめに

本研究の目的は、やさしい日本語への書き換えを自動で行うことのできるシステムの開発にある。日本における共生社会の実現が望まれる中、防災や医療関係等、様々な分野で、「やさしい日本語」の必要性が高まっている。防災関係の分野では、佐藤(2004)が共生社会の実現のためにやさしい日本語による情報発信が必要であると指摘している。さらに、弘前大学社会言語学研究室(2009)は『「やさしい日本語」の構造——社会的ニーズへの適用に向けて』として、やさしい日本語のあるべき姿を提示している。これを受けて伊藤(2009)は、防災分野に特化したやさしい日本語への書き換え支援システムを開発し公開している。

本研究では、分野を特定せず、日本人向けに書かれた一般的な日本語の文章を自動でやさしく書き換えるシステムの構築に取り組み、Web上で利用可能なツールとして提供することを目指している。「やさしい日本語」で書かれた文章は、日本語非母語話者のみならず、年少者や高齢者等、いわゆる「日本語弱者」にとっても、理解しやすいものになるはずである。筆者らはこれまでインターネット上で自由に利用できる日本語読解支援システムを開発し、無償で公開してきた。今回開発するシステムは、その読解支援ツールの一つに加えて公開していく予定である。このシステムが完成すれば、日本語弱者ばかりでなく、海外で日本語を教えている非日本語母語話者の日本語教師や「やさしい日本語」で情報発信をする必要のある人々への手助けにもなるはずである。

2. 先行研究

1999年から公開を始めた日本語読解学習支援システム『リーディング・チュウ太』(<http://language.tiu.ac.jp/>)には、入力文中の単語の難易度を自動判定する語彙チェッカーが組み込まれ、学習者のレベルにあった教材作成のための支援ツールとして、多くの日本語教育関係者によって活用されている。ところが、難易度が高い語をやさしく書き直す作業は、一般の日本人や日本語非母語話者の日本語教師にとって、必ずしも簡単ではない。そこで、筆者らは日本語の文章を自動でやさしく書き換えるシステムの開発を進めてきた。まず、川村・北村(2012)において、やさしい日本語への書き換えリスト(以下「書き換えリスト」)の作成を行った。書き換えリストは、次のようなルールをもとに作成した。

- a. 中級以上の単語を可能な限り初級の単語に書き換える
- b. 初級の単語に書き換えられない場合、当該級以下の単語に書き換える
- c. 語義が複数あり、書き換え候補が複数になる場合には、「/」(半角スラッシュ)で併記する
- d. 単純な書き換えが難しい語については、別コラムに書き換え表現を記載する

e. 複数の品詞用法がある語については、名詞形とその他の形について別コラムに記載する

以上のルールに基づいて作成された書き換えリストが、図1である。

| 見出し語 | 読み | やさしい日本語 | 名詞形 | 級 | @ | 異表記 |
|-------|---------|----------|-------|---|-----------|-----------|
| 華奢 | きゃしゃ | @ | | 1 | 細くて弱そうな様子 | 華奢 |
| 内乱 | ないらん | @ | | 1 | 国の中の戦争状態 | 内乱 |
| 匹敵 | ひってき | 同じ程度の | 同じ程度 | 1 | | 匹敵 |
| 日頃 | ひごろ | いつも | | 1 | | 日頃 |
| ぼう然 | ぼうぜん | ぼう然 | | 1 | | 呆然/茫然/ぼう然 |
| 解除 | かいじょ | 中止する | 中止 | 1 | | 解除 |
| 値 | あたい/ね | 値段 | | 1 | | 値 |
| あらかじめ | あらかじめ | 前から | | 1 | | 予め/あらかじめ |
| 悲惨 | ひさん | 悲しい | | 1 | | 悲惨 |
| 兼業 | けんぎょう | かねる | かねること | 1 | | 兼業 |
| 上演 | じょうえん | 行う | 行うこと | 1 | | 上演 |
| 転回 | てんかい | 回る | 回ること | 1 | | 転回 |
| 襲う | おそう | おそう | | 1 | | 襲う/おそう |
| 新入生 | しんにゅうせい | 新しく入った生徒 | | 1 | | 新入生 |
| あわせる | あわせる | 一緒にする | | 1 | | 合わせる |
| かなう | かなう | 実現する | | 1 | | かなう/叶う |
| 繕う | つくろう | 直す | | 1 | | つくろう/繕う |
| かねて | かねて | 前から | | 1 | | かねて/予め |
| 原書 | げんしょ | 元の本 | | 1 | | 原書 |
| 了承 | りょうしょう | 認める | | 1 | | 了承 |
| とかく | とかく | よく | | 1 | | とかく/兎角 |
| びっしょり | びっしょり | @ | | 1 | ぬれている様子 | びっしょり |
| 概説 | がいせつ | 説明する | 説明 | 1 | | 概説 |

図1 やさしい日本語書き換えリスト (川村・北村 2012版 一部)

図1の「やさしい日本語」欄には、見出し語をやさしく書き換えたものが記載されている。この欄の@印は、単純な書き換えがむずかしい語であることを示している。例えば「華奢」は一語で書き換えることが難しい。そのため、「やさしい日本語」欄に@印を記載し、「@」欄にその単語をやさしい日本語で説明している。ただし、「新入生」のように複数の語で構成される書き換えであっても単純な書き換えで問題が生じない場合は、「やさしい日本語」欄に記載した。一方、「名詞形」の欄は、形容動詞あるいはサ変動詞が名詞として使われている場合の書き換え候補である。また、「異表記」欄には、見出し語を含む異表記を網羅した。

さらに、川村・北村 (2012) では、このリストの有効性を検証するために、やさしい日本語への書き換えツールのトライアル版を開発し、運用実験を行った。運用実験の結果、以下のような対応が必要であることが明らかになった。

a. 複合語への対応

複合語としてリスト化し、複合語全体をやさしい日本語に書き換えられるようにする。

例：「共生社会」⇒「人々が一緒に生きていく社会」

b. 慣用句への対応

慣用句および慣用表現はリスト化し、各々適切にやさしく言い換えられるようにする。

例：「身につける」⇒（「身に」の部分省略）「つける」

c. 接尾辞への対応

文中で該当語が接尾辞として機能しているかどうかは形態素解析によって分析可能な

ため、接尾辞の場合には前接名詞の品詞によって書き換え方法を変える等の対応が可能な仕組みにする。

例：「利用者」⇒「利用している人」

d. サ変動詞への対応

サ変動詞を動詞に書き換える場合、必要な助詞を補う形の対応を行う。

例：「自立支援が必要だ」⇒「自立を助けることが必要だ」

e. 書き換え後の前後の語への対応

書き換えリストには、見出し語の前後の語への対応が必要かどうかに関する情報を記載することにして、システムもそれに対応できるような仕組みにする。

例：変化に富む⇒変化がたくさんある

f. 多義語への対応

多義語は単純に置き換えをすることは難しい。そこで多義語に関しては、「やさしい日本語」欄の@印の語への対応と同様に、元の語をそのまま残し、バールン表示機能等を用いてやさしい日本語で書かれた意味説明を参照できる仕組みにしていく必要がある。

以上の結果を受けて、本研究では、書き換えリストの改良を行うとともに、複合語・慣用句・接尾辞等にも対応可能なやさしい日本語書き換えシステムの開発を行った。以下に、今回新たに開発したやさしい日本語書き換えシステムの開発とその評価実験の結果について報告する。

3. やさしい日本語書き換えシステムの開発

やさしい日本語書き換えシステムの書き換えは、川村・北村（2012）で作成した「やさしい日本語書き換えリスト」の修正版（以下「修正版」）をもとに行う。

| 見出し語 | 読み | 前接助詞 | やさしい日本語 | 名詞形 | 級 | @ | 後接助詞 | 異表記 |
|-------|-------|------|---------|------|---|-----------------|------|-------------|
| 売出し | うりだし | | @ | | 1 | 売りはじめる/売っている | | 売り出し/売出し/売出 |
| 潤う | うるおう | | @ | | 1 | 水分が多くなる/お金が多くなる | | うるおう/潤う |
| 上回る | うわまわる | を>より | 多くなる | | 1 | | | 上回る |
| 植わる | うわる | | 植えられる | | 1 | | | 植わる |
| 運営 | うんえい | | 動かす | 進行 | 1 | | | 運営 |
| うんざり | うんざり | | @ | | 1 | 本当にいやになる気持ち | | うんざり |
| 運送 | うんそう | | 運ぶ | 運ぶこと | 1 | | | 運送 |
| 運賃 | うんちん | | 料金 | | 1 | | | 運賃 |
| うんぬん | うんぬん | | あれこれ言う | あれこれ | 1 | | | うんぬん/云々/云云 |
| 運搬 | うんぱん | | 運ぶ | 運ぶこと | 1 | | | 運搬 |
| 運命 | うんめい | | 運 | | 1 | | | 運命 |
| 運輸 | うんゆ | | @ | | 1 | 荷物を送ること | | 運輸 |
| 運用 | うんよう | | 活用する | 活用 | 1 | | | 運用 |
| 柄 | え/から | | @ | | 1 | 模様/持つところ | | 柄 |
| エアメール | えあめーる | | 航空便 | | 1 | | | エアメール |
| 英字 | えいじ | | 英語の文字 | | 1 | | | 英字 |
| 映写 | えいしゃ | | 映す | 映すこと | 1 | | | 映写 |
| 閲覧 | えつらん | | 見る | 見ること | 1 | | | 閲覧 |
| エレガント | えれがんと | | 上品 | 上品 | 1 | | | エレガント |
| 演じる | えんじる | | 演技する | | 1 | | | 演じる |
| 演ずる | えんずる | | 演技する | | 1 | | | 演ずる |
| 沿線 | えんせん | | @ | | 1 | 線路に沿っている部分 | | 沿線 |
| 縁談 | えんだん | | 結婚の話 | | 1 | | | 縁談 |
| 遠方 | えんぼう | | 遠く | | 1 | | | 遠方 |
| 円満 | えんまん | | @ | | 1 | うまくいっている | | 円満 |

図2 やさしい日本語書き換えリストの修正版（一部）

図2が、修正版である。書き換えによって、見出し語に前接する助詞等の語句を変更する必要がある場合は「前接助詞」欄に、後接する助詞等を変更する必要がある場合には「後接助詞」欄に、「(変更前) > (変更後)」の形で「>」の前後に変更前の語句と変更後の語句（助詞等の削除の場合には「#」）を記載する。例えば、図2の「上回る」を「やさしい日本語」欄の「多くなる」と書き換える場合には、「～を上回る」は「～より多くなる」と書き換える必要がある。そこで、「前接助詞」欄に「を>より」と記載する。また、修正版作成の際には、各見出し語の多義性について、留意することにした。そのため、修正版では、以前のリストに比べて「やさしい日本語」欄の「@」印が多くなっている。

以上のような、改良を行った修正版をもとに、やさしい日本語への書き換えシステム『チュウ太のやさしくな一れ』（以下「システム」）を開発した。システムは、先行研究の結果明らかになった複合語・慣用句・多義語等への対応を行い、日本語弱者のための読解支援ツールとして開発した。形態素解析には、MeCab（工藤 2006）、解析や書き換え後の動詞・形容詞等の活用に用いる辞書としては、ipadic2.7.0 を利用している。システムの基本設計については、渡邊・川村（2013）で詳しく報告しているため、以下、書き換えの手順のみを述べることにする。

システムは、まず入力された文章の形態素解析を行う。次に、修正版をもとに語句の書き換えを行う。MeCab が切り出した語句が、修正版の「見出し語」あるいは「異表記」欄にあり、かつ、品詞が一致した場合に「やさしい日本語」欄の語句への書き換えを行う。システムが行うやさしい日本語への書き換えルールは、次のとおりである。

①書き換え前後の単語が活用のない単語の場合：

見出し語の単語を「やさしい日本語」欄の単語に置き換える。

②書き換え前後の単語が活用のある語の場合：

必要な助詞・助動詞の追加削除を行った上で書き換え前の語の活用形に準じて活用させる。

（例：書き換え前が形容動詞で書き換え後が形容詞の場合書き換え後は「だ/である」を削除し、形容詞を「だ/である」の活用形に活用させる）

ただし、いずれか一方が活用のない語の場合には、後接語の書き換えが必要な場合もある。

また、書き換え前の語がテ形・タ形・可能形・使役形等の場合には、書き換え後の単語によって後接する助詞・助動詞等を適切に選択する必要がある。

③書き換え後が、「名詞+の+動詞で始まる語句（あるいはサ変動詞+する）」の場合：

動詞あるいはサ変動詞+するの前の「の」を「を」に置き換える。

④該当語句の前後の語の変更が必要な場合：

「前接助詞」あるいは「後接助詞」欄に記載された書き換え条件に従って書き換える。

⑤「見出し語」が多義の場合（「やさしい日本語」欄に@印がある）：

書き換えは行わず、@欄の語句をバルーン形式で表示する。

⑥書き換えが困難な語句の場合（「やさしい日本語」欄に@印がある）：

書き換えは行わず、@欄の語句をバルーン形式で表示する。

⑦名詞が連続している場合：

書き換えは行わず、「名詞形」欄に記載があればその語句、なければ、「やさしい日本語」欄の語句をバルーン形式で表示する。

システムは、以上のような規則に対応して、入力文を自動でやさしい日本語に書き換える。

図3が、システムの入力画面である。

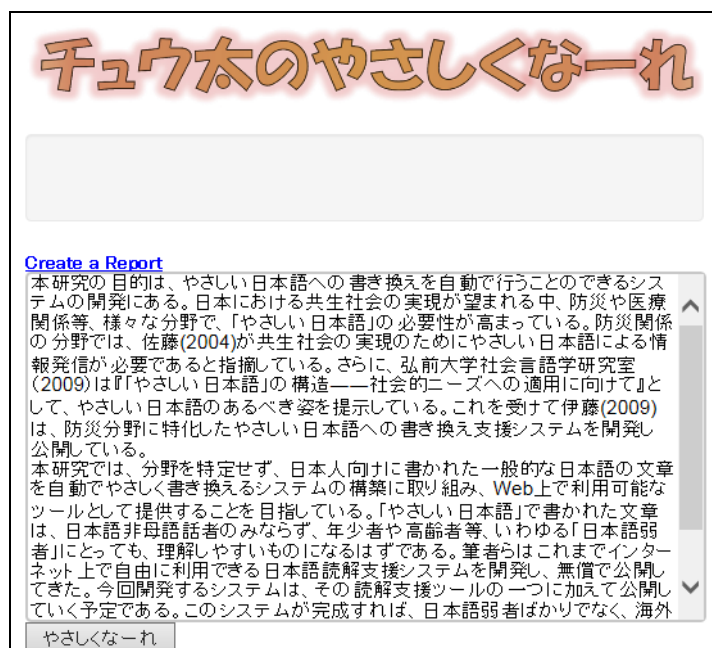


図3 システムの入力画面

システムを利用するには、テキストボックスに日本語の文章を入力し、テキストボックス下の「やさしくな一れ」のボタンを押す。ここでは、本論文の「はじめに」の文章をコピー&ペーストして入力した。図4が、システムがやさしく書き直した結果画面である。



図4 システムの結果画面

結果画面の左上が書き換え後のやさしい日本語の文章であり、画面左下は入力文である。画面右は MeCab の分析結果を示している。また、図 4 は、書き換え後の文の 1 行目の「共生」にカーソルをあてた状態で、「共生」の意味がバルーン表示されている。さらに、書き換え箇所がわかるように、書き換えた語（緑）、バルーン表示の語（青）、書き換えていない語（黒）、リストに未登録の語（赤）という色分け表示を行っている。

さらに、このシステムには、書き換え結果に問題があった場合に報告レポートが作成できる仕組みを組み入れた。これによって、運用実験で明らかになった問題点を文脈情報とともに記録することが可能になり、リストの修正やシステムの改良が容易になった。

4. システムの評価と今後の対応

システムの評価は、入力した文によって大きく異なってくる。そこで、ここでは、図 4 に示した書き換え結果をもとにして、現時点で明らかになっているシステムの問題点と各々の問題に対する解決方法を報告する。

入力した文は、622 文字であり、単語数は延べ語数で 297 語、異なり語数で 130 語である。

誤解析は、2 例のみで、「特化する」を「特+化する」、「日本語弱者」を「日本語+弱+者」と区切っていた。前者は、「特化する」を MeCab 辞書に登録することによって、対応可能である。一方、後者に関しては、結果の下から 7 行目のように「」内にある場合にはきちんと解析できているので、重みづけの問題で生じた誤解析のようであり、現時点での対応は難しい。

やさしい日本語への書き換えは、22 例（異なり）中 17 例が適切に行われていた。品詞が異なった場合の対応（「本発表」⇒「この報告」、「高まって」⇒「高くなって」、「完成すれば」⇒「出来上がれば」等）や、前接の助詞等を含めた書き換え（「共生社会+の+実現」⇒「共生社会+が+本当になること」「システム+の+開発」⇒「システム+を+作ること」等）も問題なく行っていた。

一方、問題があった 5 例は、次のとおりである。以下に、各々の問題の原因を明らかにするとともに、今後の対応方法についても検討する。

a. 「分野」⇒「範囲」

「分野」も「範囲」も活用のない語であり、一般には書き換えリストの修正によって対応可能なケースだが、「分野」という単語をやさしい日本語で説明するのは難しく、言い換えるとかえってわかりにくくなってしまう。こうした語については、基本語として扱う等、何らかの対応を考えていく必要がある。

b. 「実現+の+ために」⇒「本当になること+の+ために」

この例では、「実現」を「本当になること」と書き換えること自体には問題がないものの、その後ろに「のために」が続くと、文が冗漫になってしまう。このあたりがやさしい日本語への書き換えの難しい点とも言えよう。

c. 「やさしい+日本語+の+構造」⇒「やさしい+日本語+の+作り」

「構造」については、「やさしい日本語」の候補として「作り」が記載されていたが、今回入力した文では、抽象的に用いられている。この場合、「仕組み」という意味であるが、「仕組み」は 1 級の単語であるので使えない。強いてやさしい語で説明するならば「作られ方」となる。いずれにせよ、@扱いにして、バルーン表示で説明すべき語である。

d. 「社会的ニーズへの適用」⇒「社会的+ニーズ+への+当てはめること」

この例のように、間に入る助詞が複数並んでいる場合には、単純な置き換えが難しい。前接助詞の置き換え指示をさらに細かく設定する等、今後、対応方法を考えていく必要がある。

e. 「分野を特定せず、」⇒「範囲+を+それだと決めない、」

文語的な用法の助動詞「ず」を「ない」に置き換えたことによって発生した問題である。「後接助詞」欄に「、>で」と記載することによって、「ず」の後ろに「、」がある場合には、「ない+で」とすることで対応可能である。ただし、あらゆる文脈においてその対応でいいかどうかについては、今後検証していく必要がある。

以上、やさしい日本語への書き換えでおきた問題とその対応方法を詳しく見てきたが、いずれも単純な置き換えが難しい例である。今後、こうした事例に対応できるように、前後関係を考慮しながら適切な書き換えが行える仕組みに、システムを整備していく必要がある。

また、「ツール」「母語話者」「弱者」「読解」等、難しい語でありながら書き換えリストに未登録なため対応できなかった語が、延べ語数で8語(3%)、異なり語数で5語(4%)あった。これは、今回のシステムが、旧日本語能力試験の出題基準の語(約8500語)と介護用語(約2000語)を基に作成した「やさしい日本語書き換えリスト」を用いていることによるものである。書き換えリストの語を増やしていくことで対応は可能である。

また、前述した通り、複数の置き換え候補がある語句については、現在バールン表示で対応している。今後の課題として、文脈を判断して複数の置き換え候補の中から正しいものを選択できるようにする仕組みを作り上げる等、よりよい対応方法を考えていく必要がある。

謝辞：本研究の一部は、科学研究費基盤研究(B)課題番号24320096の助成を得て行われた。システムの開発には甲南大学の北村達也氏および筑波大学大学院生の渡邊飛雄馬氏、書き換えリストの編集にはチュウ太プロジェクトチームの協力を得た。ここに記して感謝の意を表する。

参考文献

- 伊藤彰則(2009)『やさしい日本語』作成支援システム『やさしい日本語』の構造——社会的ニーズへの適用に向けて』弘前大学人文学部社会言語学研究室。
- 川村よし子(2011)「文章の難易度判定システム構築のための基礎調査」『ヨーロッパ日本語教育』15, pp.171-178, ヨーロッパ日本語教師会。
- 川村よし子・北村達也(2012)「やさしい日本語への書き換えリストの作成とその評価」『日本語教育学会秋季大会予稿集』pp.123-128, 日本語教育学会。
- 工藤拓(2006) MeCab: Yet Another Part-of-Speech and Morphological Analyzer ,
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> (2013年11月25日閲覧)
- 佐藤和之(2004)「災害時の言語表現を考える——やさしい日本語: 言語研究者たちの災害研究」『日本語学』23巻10号, pp.34-45, 明治書院。
- 日本国際教育支援協会(2006)『日本語能力試験出題基準改訂第2版』凡人社。
- 弘前大学社会言語学研究室(2009)『やさしい日本語』の構造——社会的ニーズへの適用に向けて』弘前大学人文学部社会言語学研究室。
- 渡邊飛雄馬・川村よし子(2013)「やさしい日本語書き換えシステムの基本設計」『日本語教育方法研究会誌』Vol.20 No.2, pp.48-49, 日本語教育方法研究会。