

# 日本語学習者にとって読みやすい文章について

## — 日本語教科書における書き換えの分析から —

クリスティーナ・フメリャク・寒川 (リュブリャーナ大学文学部) †

### On the Readability of Texts for Japanese Language Learners -- From an Analysis of Text Adaptations in a Japanese Textbook

Kristina Hmeljak Sangawa (University of Ljubljana, Faculty of Arts)

#### 1. はじめに

リーダビリティ (読みやすさ) の判定は、弱読者のための文章を執筆・選別するとき役に立つ判定で、日本語母語話者のための読みやすさに関する研究は近年大きな成果をあげている。しかし、母語話者にとって読みやすいものは、必ずしも日本語学習者にとっても読みやすいとは限らない。そこで本稿では、日本語学習者にとっての読みやすさを探るために、中級読解教科書に掲載された文章と、その原文からなる小規模コーパスを分析し、教科書著者が文章を読みやすくするために、どのようなストラテジーを使い、テキスト内のどの要素を言い換えたかを考察する。

#### 2. 読みやすさに関する研究

日本語母語話者のための読みやすさに関する研究は、英語のリーダビリティ研究(Flesch 1948 など) の影響も受け、文章の測定可能な要素 (文の長さ、高頻度語彙の割合など) から、その文章の難しさを判定する難易度算定公式を開発してきた(森岡 1952、阪本 1964、建石ら 1988、佐野・丸山 2008、柴崎ら 2008、柴崎・玉岡 2011 など)。柴崎らのリーダビリティ測定公式を用いたシステムはインターネット上公開されており<sup>1</sup>、実用化されている。近年、コーパスに基づいた統計的な言語モデルを用いたテキスト難易度推定システム (近藤ら 2008、佐藤 2011) も構築、公開されており<sup>2</sup>、日本語母語話者 (特に若年者) にとって、特定の文章がどのぐらい読みやすいか、日本の学校のどの学年の文章に近いかを測ることが簡単にできるようになった。しかし、Hmeljak-Sangawa(2009)において指摘されているように、これらの測定・推定システムは、特にやさしい文章を測定する際、その表記、特に文章に含まれている教育漢字とひらがなの割合に大きく左右されるもので、標準表記で書かれている日本語学習者用のテキスト、特に初級・中級用の文章レベルを測定・推定するには限界がある。

日本語学習者にとっての読みやすさに関しても、これまでに多くの研究が行われてきたが、研究の焦点は文章の一側面に限定され、その一側面における文章難易度を測定・推定したものが多く。例えば、川村(1999)、川村ら(2008)、北村ら(2009)などの一連の研究は語彙に着目し、日本語能力試験出題規準、語彙親密度などの語彙難易尺度を用い、その尺度にそって文章に含まれる語彙の割合を基に文章の読みやすさを判定する。難構文に着目した研究もあり、水嶋ら(2011)は中止法、名詞修飾節、主格省略の検出を、竹井ら(2005)はゼロ代名詞 (語句の省略) の検出と明示化を、中村ら(2012)は主格省略文の検出を提案している。また、日本語能力試験読解問題における日本語能力試験出題規準の語彙含有率を測定した研究(柴崎・李 2010)、および学習辞書のための分かりやすい用例を選別するために、語彙、文型、構文 (連体修飾説、複雑な係り受け構造など) の難易度測定システムを提案した研究 (小林ら 2007、水野ら 2008、吉橋ら 2007、Hazelbeck ら 2009) もあげられるが、日本語学習者にとっての読みやすさ、読みにくさの要因はまだ総合的に十分明らかにされて

† kristina.hmeljak@ff.uni-lj.si

<sup>1</sup> <http://readability.nagaokaut.ac.jp/readability>

<sup>2</sup> <http://kotoba.nuee.nagoya-u.ac.jp/sc/obi2/>

いない。

そこで本研究では、日本語学習者にとって難読と思われる要素を探るために、日本語ベテランである日本語教科書の著者が自らの日本語教育の経験に基づき書き換えた文章を分析・調査した。

### 3. データ

今回の調査で使用したデータは、産能短期大学日本語教育研究室編の『日本語を楽しく読む本・中級』におさめられている2種類の文章、すなわち、想定されている中級日本語学習者のために書き換えられた読解用のテキスト（以下これを書き換え文と記述）と、その原文である。この教科書を調査対象にしたのは、原文と書き換えた文章の両方を掲載した数少ない資料だからである。また、中級の教科書に着眼したのは、初級教科書の読み物には学習者用の書き下ろし（語彙・文型・構文を極端に制限した文章で、場合によって母語話者に違和感さえ与える文章）、上級教科書の読み物には日本語母語話者用に書かれた生教材が多く使われるのに対し、中級教科書では、日本語母語話者のために書かれた文章を日本語学習者に読みやすいように書き換えられた文章がよく利用されるからである。このような日本語学習者用の書き換え文と母語話者用の原文の比較からは、日本語学習者に特有な難読要素が得られると考え、その比較を行った。

調査した教科書は9つの章に分かれており、以下の読み物が掲載されている。

表1：分析資料の概略

章	テキスト題名	文字数		出典
		書換文	原文	
1	禁酒	449	352	『月刊アサヒ』1989年6月号、朝日新聞社
2	いつもとちがう	682	527	『ポケットジョーク9トラベル』植松黎編訳、角川書店
3	賢い農夫／百姓	1327	1285	矢崎源九郎『世界の民話』社会思想社
5	振り向き賃	526	909	深田祐介『ちょっといい旅いい話』ジャルパック出版
6	ママ、手が凍るんだよ 行康君、手すり変わるよ	707	702	朝日新聞 1985年2月20日朝刊
		829	922	朝日新聞 1985年3月1日朝刊
7	みんなって何人？	1277	2640	斉藤勇『人間関係の分解図』誠信書房
8	握手1 握手2	1147	2408	阿部謹也『逆光のなかの中世』日本エディタースクール
		851	1199	樋口清之『日本の風俗の謎』大和書房
9	趣味	1219	1972	守屋毅『日本文明77の鍵』（梅悼忠夫編）創元社
	合計：	9014	12916	

教科書の前書きによると、450～600時間ぐらい勉強し3500～5000語程度の語彙を学習している中級レベルの日本語学習者を想定し、自作の語彙リストに基づいて3500語程度のレベルを標準として、原作を書き換えたということである。語彙以外の書き換えの基準について言及していないということからは、この教科書の著者は、多くのリーダビリティ研究と同様に、語彙が読みやすさの最も重要な要素だと考えていることが窺える。

### 4. 分析方法

原文と書き換え文を電子化し、JDiff X (Matsumoto 2010)というソフトウェアを用いてテキストの対を比較した上で、検出された相違を表計算ソフトに入力し、それぞれの書き換えの種類、レベル（語彙、構文、談話など）と書き換え内容の記述を付与した。一つの文の中で複数の書き換え（相違）が確認された場合、それぞれの書き換えを別項目として

記述した。例えば、第1章の読み物「禁酒」には次の原文(1)と書き換え文(2)があった。

(1-原文) 「亡くなった飲み友だちと約束してね、僕が飲みに行くときは、必ずオレの分も注文して飲んでくれという遺言を実践しているんだ」

(2-書き換え文) 「先週、僕の親友が亡くなったんだが、彼が亡くなる前に約束してね」  
「はあ」

「僕が飲みに行く時は、必ず、彼の分も注文して飲むということになったんだ。それで、その約束を実行しているってわけさ」

上記の原文と書き換え文を比較した際、次の相違(書き換え項目)が確認された。

- ・統語構造の単純化：「亡くなった飲み友だち」という連体修飾節から「親友が亡くなった」という単文への書き換え、「オレの分も注文して飲んでくれという遺言」という修飾節から「…彼の分も注文して飲むということになったんだ。…その約束を…」という文への分解

- ・談話構造の単純化：「僕が飲みに行くときは、必ずオレの分も注文して飲んでくれという遺言を[僕が]実践している」という短い複文の中で、主語が「僕」から「オレ」、そしてまた「僕」へと変わることを避け、「僕」への統一

- ・談話構造の明示化：会話場面を意識させる相づち「はあ」の挿入

- ・意味的な明示化：「先週」の挿入による時間設定の明示化、「飲み友だち/親友」の属性を明示化する「僕の」の追加

- ・語彙的な単純化：口語的な代名詞「オレ」の代わりに「彼」、低頻度語彙の「飲み友だち」の代わりにより広い意味の「親友」、「遺言を実践」の代わりに「約束を実行」への書き換え

- ・文型の明示化：「しているんだ」の中で、広い用法範囲をもつ「んだ」の代わりに、説明に使う「ってわけ」を使い「しているってわけさ」への書き換え

- ・文字レベルにおける明示化：単語の境界が明確でないひらがなの連鎖「行くときは」から、初級レベルで習得される漢字「時」への書き換え

- ・句読点による統語構造(境界点)の明示化：「必ず」という副詞の後に読点の挿入

このように原文と書き換え文の相違を記述した結果、815項目の相違リストが得られた

## 5. 書き換え項目の分析

以上のような分析によって得られた相違項目は、三つのストラテジー、すなわち単純化、明示化、標準化に分類できる。これらのストラテジーは、複数レベルで確認された。

### 5.1 単純化

このストラテジーはもっとも多く確認され、96の削除を含み、472の単純化の例が確認された。

- ・文字表記においては、低頻度漢字から平仮名への書き換え(尋ねた→たずねた、土産→おみやげ、嫌がります→いやがります、石鹸→せっけん、挨拶→あいさつ、など)

- ・語彙においては、低頻度内容語の書き換え(誰も→多くの人が、奇怪/珍奇な→珍しい、育種学→植物学、事柄→こと、食糧→食べ物、受容する→受け入れる、大衆化し→広がり、一覧表→リスト、かたっぱしから→何でも、絆→関係、帰宅した→家に帰った、など)または定義や説明の追加(古典→昔の文学作品、利己的遺伝子の乗り物→利己的遺伝子[ふりがな：セルフイッシュジーン]の乗り物[ふりがな：ヴィークル])；また、機能語の書き換え(のみ→だけ、かならずしも...というわけではない→ばかりではない、...にせよ...にせよ→例えば、すら→も、...Vられっぱなしで→その間...Vられていて、など)

- ・文法においては、口語的な文型、または硬い口調から、初級で学習される文型への書き換え(会社を出→会社を出て、言わずに→言わないで、同じく→同じように、みせねばならない→みせなければならない、など)

- ・統語においては、長い文、複文の分解が多く見られた(彼女はあわてて手の甲をぬぐって私に右手の甲をさし出し、お互いの右手の甲を合わせて握手の代わりとしたことがあ

った。→彼女は急いで手の甲をふいて右手の甲を差し出した。お互いの右手の甲を合わせて握手の代わりにしたことがあった。) (この実験は、人が集団の圧力を感じ、集団に従おう、あるいは従わざるをえないというような気持ちになるのは、三人以上であることを明らかにしています。→この実験から、次のようなことがわかります。私たちが、集団の圧力を感じて他の人と同じ意見を言うようになるのは、3人以上の人が同じことを言った時であること)

・ 談話構造において主題移行を避けた主題の統一 (王さまは…、大臣たちが集まったとき、王さまは…→…お城に帰ると、大臣たちを集めて、王さまはこう言いました)

単純化の極端な例である削除も、さまざまなレベルにおいて確認された。

・ モダリティー表現の削除 (どうも土産が足りない→おみやげが足りない)

・ アスペクト形式の削除 (捨ててしまいます→捨てます)

・ 意味上、付随的な情報の削除 (林野庁業務部販売推進室長羽賀正雄さん→林野庁業務部の羽賀正雄さん、越後からやってくるたび東京の街なかで→東京に出てくると、法則の基礎知識をもっていた→法則を知っていた、など)

## 5.2 明示化

明示化の例も多く、203の例がさまざまなレベルで確認された。

・ 意味的な明示化、例えば設定の明示 (ある男が、バーのカウンターにすわった。→ある男が、バーのカウンターにすわった。はじめての男だった)、より具体的な表現の使用 (とうとう→次の日、連絡をとった→電話をした、こいつ→この客、やっている→経営している)

・ 漢字表記による曖昧さ解消と単語境界の明示化 (いっている→言っている、もっている→持っている、よんで→呼んで、など)

・ 境界の明示化も多く (70項目)、もっとも多いのが句読点や括弧の挿入による単語境界の明示化 (二十年間うちのスープ→もう20年も、うちのスープ; 男はそれからもときどき→男は、それから、ときどき; 実際手は→実際、手は; 三人というのはみんななので→3人というのは「みんな」なのです)、または句読点挿入による文節境界の明示化 (いったいどうしたというんですか?→いったい、どうなさったんですか)、次に多いのが句読点の挿入や段落分割による統語構造の明示化

・ 統語的な明示化は、省略されていた主語等の挿入 (手袋は、滑ってしまうから、つけてはいかれない→手袋をすると滑ってしまうから、手袋をすることはできない)、自動詞から動作主語が明確な他動詞 (一杯目が終わったら→一杯目をお飲みになったら)、「の」を使った複合名詞の分割 (現代日本→現代の日本、遺伝法則→遺伝の法則、など)

・ 結束性の明示化、指示詞の追加 (その結果が図です→その結果が、上の図です)

## 5.3 標準化

3つめのストラテジーは、単純化に類似するストラテジーだが、低頻度要素 (語彙・文型) を高頻度要素に書き換えた単純化項目のに対し、標準化と分類した項目は、初級で学習しない口語的、または硬い、あるいは特殊な文体要素を、中性的な要素に書き換えた例である。

・ 表記において、標準的な句読点や括弧への書き換え (ですか?→ですか。;《皆様の…いただきます》→『みなさんの…いただきます』)

・ 時制の統一 (過去形で書かれた文章の中で: されるのは→されたのは)

・ 文体の統一 (すべての文を「です・ます体」、または全ての文を普通体へ)

・ 口語表現や方言などの特殊な文体から標準表現への書き換え (飲まなかったって→飲まなくても、話さん→話さない、ようがす→わかりました、など)

・ 漢数字からアラビア数字への統一 (百枚→100枚、十二時→12時、十七世紀→17世紀 六、など); これは、原文の縦書きから教科書の横書きへの変換によるものと思われる。

## 5.4 その他

数例においては、書き換えの目的が確認できなかった。例えば、タイプミスと思われる

例（追っかけてきます→追っかけてきてます）、非標準的な句読点（そうですか。わかりました。→そうですか、わかりました。）のような書き換えが数例あった。

### 5.5 結果のまとめ

このように、書き換えを分析した結果、さまざまなレベルにおいて、複数のストラテジーが確認された。それぞれの書き換えの例をレベルでまとめると、表2のようになり、ストラテジー別にまとめると、表3のようになる。

表1 レベル別書き換え例の数と割合

レベル	例数	割合
表記	175	21%
句読点	56	7%
内容語	193	24%
機能語	44	5%
語彙+統語	167	20%
活用形	20	2%
モダリティー	18	2%
統語	42	5%
意味	45	6%
結束性	10	1%
談話	33	4%
文体	12	1%
合計	815	100%

表3 ストラテジー別書き換え例の数と割合

ストラテジー	例数	割合
単純化（その内、 削除の例が 96）	472	58%
明示化	203	25%
標準化（その内、 数字表記統一例が 80）	128	16%
その他	12	1%
合計	815	100%

## 6. 考察

中級日本語教科書において学習者用に書き換えられた文章をその原文と比較した結果、単純化の例、特に教科書の前書きにも明記されており、先行研究でも頻繁に取り上げられている低頻度語彙から高頻度語彙への置き換えという単純化の例は多かったが、他のストラテジー、他のレベルにおける書き換えも多く確認され、それらの書き換えは日本語学習者にとって難読だと思われる要素を暗示している。

## 7. まとめ

本研究では、日本語母語話者のリーダビリティ研究に対して、日本語学習者にとっての読みやすさの要因を探るために、中級日本語教科書のために書き換えられた文章と、その原文となる日本語母語話者用の文章を比較し、その相違を分析した。その結果、母語話者にとっても読みにくいとしばしば指摘される低頻度語彙、長くて複雑な文が書き換えられたことが確認され、この要素に関しては母語話者と日本語学習者の共通難点が認められたが、同時に、日本語母語話者の幼児にとって簡易であろうと思われる口語的な要素、漢

語からカタカナ語への書き換えなど、日本語学習者に特有な制限に起因すると思われる書き換えの例も確認された。これは、日本語学習者用のリーダビリティ研究の必要性を暗示している。今回の分析は 1 冊の教科書における書き換えに限定したので、より一般的に有効な日本語学習者のための読みやすさの要因と、リーダビリティ判定への応用の可能性については、更なる研究が必要である。

## 文 献

- Flesch, R. (1948) A new readability yardstick, *Journal of Applied Psychology* 32:3, pp. 221-233.
- Hmeljak-Sangawa, Kristina (2009) A Corpus for Readability Measurement for Non-Native Learners of Japanese, *Technical report of IEICE. Thought and language*, 109:84, pp.19-24.
- Hazelbeck, Gregory and Saito, Hiroaki. 2009: A Corpus-based E-learning System for Japanese Vocabulary, *Information and Media Technologies* 4:4. pp.1104-1128.
- 川村よし子(1999)「語彙チェッカーを用いた読解テキストの分析」『講座日本語教育 34』 pp.1-22
- 川村よし子、北村達也、富岡洋介、林真一(2008)「単語親密度と頻度情報を活用した難易度判定システム」『日本語教育方法研究会誌』 15:1.
- 北村達也、富岡洋介、川村よし子(2009)「IDFを用いた単語レベル判定システムの構築と検証」『日本語教育方法研究会誌』 16:1、 pp. 52-53.
- 近藤陽介、松吉俊、佐藤理史(2008)「教科書コーパスを用いた日本語テキストの難易度推定」『言語処理学会第 14 回年次大会発表論文集』 pp. 63-66.
- Matsumoto, Satoshi (2010) JDiff X Document Comparison Plug-in for Jedit X.  
([http://www.artman21.com/en/jdiff\\_x/](http://www.artman21.com/en/jdiff_x/)よりダウンロード可能)
- 水嶋博志、田聖也、北村達也、川村よし子(2011)「学習者にとって難解な構文の自動検出」『日本語教育方法研究会誌』 18:1、 pp. 64-65.
- 森岡健二(1952)『「読みやすさ」の基礎的研究』昭和 26 年度国立国語研究所年報、 pp. 91-108.
- 中村慶太、北村達也、川村よし子(2012)「検索エンジンを用いた主格省略文の自動判定」『日本語教育方法研究会誌』 19:1.
- 阪本一郎(1964)「文の長さの比重の測定法-Readability 研究の試み」『読書科学』 8:1、 pp. 2-6.
- 産能短期大学日本語教育研究室編(1991)『日本語を学ぶ人たちのための 日本語を楽しく読む本・中級』産能短期大学国際交流センター／凡人社
- 佐野大樹、丸山武彦(2008)「システミック文法に基づく書きことばの複雑さ測定」『言語処理学会第 14 回年次大会発表論文集』 pp. 1097-1100.
- 佐藤理史(2011)「均衡コーパスを規範とするテキスト難易度測定」『情報処理学会論文誌』 52:4、 pp. 1777-1789.
- 柴崎秀子、玉岡賀津雄、山本和英、加納満、原信一郎、李在鎬(2008)「日本語コーパスを応用した文章の難易度測定の研究」特定領域研究「日本語コーパス」平成 19 年度公開ワークショップ予稿集、 pp. 125-130.
- 柴崎秀子、玉岡賀津雄(2010)「国語教科書を基にした小・中学校の文章難易学年判定式の構築」『日本教育工学会論文誌』 33:4、 pp. 449-458.
- 柴崎秀子、李在鎬(2010)「日本語能力試験読解問題を土台にした文章の難易尺度の構築—日本語教育リーダビリティの基礎研究—」『日本語教育学会春季大会予稿集』 pp. 1-5.
- 建石由佳、小野芳彦、山田尚勇(1988)「日本文の読みやすさの評価式」情報処理学会研究報告 1988-HI-018/ pp. 1-8.
- Yamura-Takei, Mitsuko, Aizawa, Teruaki, and Fujiwara, Miho (2005) Diversity of zeros in Japanese discourse: a corpus analysis and a tool for language teachers. In: *Proceedings of Pacific Associations for Computational Linguistics (PACLING 2005)*, pp. 358-367.
- 吉橋健治、傅亮、仁科喜久子(2007)「学習者に合わせた例文表示ツール」Proceedings of CASTEL-J in Hawaii 2007、 pp. 223-226.