

# 日本語教育における文章研究の新展開

## —文章の読みやすさを測る—

李在鎬（早稲田大学）

lee@waseda.jp

### 【要約】

自然言語の文章データが持つ潜在的な難しさを測定する研究領域として、リーダビリティ研究がある。リーダビリティ研究では一文あたりの語数などの表層情報をもとに文章の難しさをランクづけすることを目指している。とりわけ英語を対象とするリーダビリティ研究は、たくさんの先行研究があり、回帰式による読みやすさの計算式が提案されてきた。本研究は、日本語研究におけるリーダビリティ研究を発展させ、日本語教育のためのリーダビリティの構築を目指すものである。

### 1. はじめに

リーダビリティ (Readability) の研究は、今から 100 年前にアメリカで誕生したとされている (野本 2016)。その研究背景としては、1920 年代にヨーロッパや南アメリカなどから多数の移民がアメリカに入ったことにより、アメリカの学校に教科書が読めない児童が増えたことがあげられている。こうしたことから、1920 年当時は児童の読解力にあった教科書選定が喫緊の課題として認識されていた。

こうした社会的課題を解決するものとして、リーダビリティ研究が登場し、文章の適正学年を推定する技術として注目された。このタスクでは、当該文章がリーダブル(readable) であるというのは、あるコホート (特定年齢の児童集団) が適度な早さで読むことができ、またその内容を十分に理解できる状態という前提で研究が行われる。

アメリカにおける初期の研究の特徴としては、語彙リストに基づく語彙の出現数をもとにその文章のリーダビリティをはかるという考え方が主であった。例えば、エドワード・ソーンダイク (Edward L. Thorndike) が行った次のところみがある。児童書、聖書、新聞等 41 の資料から抽出した約 4,565,000 語の中から頻度が最も高い 10,000 語を選び、語彙リストとして公開された。この語彙リストをもとに、難解語が定義され、その難解語の出現頻度をもとに、文章の難易度が計算されていた。こうした作られたリーダビリティ値が児童の読解力テストの得点と一定程度の相関が見られ、教育的な有効性が示された。

さて、リーダビリティの有効性が認知されたものの、研究成果を活用するステップにおいて、次の問題が発生した。1920 年代はコンピュータが普及しておらず、リーダビリティ値は手計算で求めている。文章を単語にわけて、その単語を 10,000 語の語彙リストから探し、頻度を計算するというのは、一般の教師にとっては負担の大きい作業であった。こうした理由から、リーダビリティの活用のために、より単純な計算方法が求められるようになった。これを受け、語彙リストを使わない単純な回帰式によってリーダビリティを計算する手法が提案された。具体的には、Flesch, R. (1948) によって提

案された以下の公式である。

$$y = (-1.015 \times \text{文の平均単語長}) + (-84.6 \times \text{単語の平均音節数})$$

Flesch, R. (1948)が提案する方式は、文の平均単語長と単語の平均音節数に回帰係数を組み合わせるだけで、ソーンダイクの語彙リストを使った計算方法と同等の計算ができることで、注目されるようになり、50年代、60年代へと時代の進化とともに、様々な計算式が提案されるようになる。

英語における100年ほどの研究の蓄積によって明らかになったこととして、文章の読みやすさに関連する要因は様々であるが、とりわけ重要なものとして、1)文字長と難解語が重要、2)語彙と文の長さによってリーダビリティがかわる、3)異なり語数の多さはリーダビリティに関係する、4)高頻度が入っていると記憶へのアクセスが早く、読みやすくなる、5)語彙の難解性は語の内在的な性質ではなく、学習によってかわるもの、だということが分かってきている。

さて、日本語に関するリーダビリティ研究としては、坂本(1971)を始めとして、「読書科学」の分野で注目すべき研究があった。ただ、その多くが手分析によるものであったため、大量のテキストを扱うには限界があった。しかし、近年、計算機の普及やWebの進化に伴い、ネットサービスの一つとして一般に公開されているものもあるなど、リーダビリティ研究は多様化しつつある。具体的には、建石(他)(1988)、佐藤(2011)、柴崎・原(2010)、Lee&Hasebe(2016)などの研究があり、ウェブサービスとしてリーダビリティの計算式を提供する試みがなされてきている。

## 2. リーダビリティ研究

### 2.1 3つの問題意識

本節では、リーダビリティ研究における具体的な問題意識について確認しておく。リーダビリティ研究では、次の3つの研究課題が盛んに議論されてきた。1)文章の難易度を決定する要因は何か、2)文章の難易度を決定する複数の要因をどのように重み付けし、公式化するか、3)どのような難易度のスケールを使うかである。

まず、1)に関しては次の事実を考慮する必要がある。文章の難しさは、いくつもの要因が複雑に絡み合って決まっていく。マクロな要素としては、どのような話題かという問題や文章としてのまとまり具合などが考えられる。ミクロな要素としては、語彙の難しさ、文法構造の難しさ、さらには、計量文体論で問題視されてきた語の長さ、文の長さなどが考えられる。

次に、2)の問題として、文章の難易度を決める要素が複数であることが明らかになった場合、個々の要素が持つ強さの度合いをどう表現するのかということが考えられる。つまり、文章に含まれる語彙の難しさの要因、文法項目の難しさの要因、文長などの長さの要因などを同等に扱ってよいかという問題に帰結する。当然ながら、これらの要因の強さは異なるものであり、その異なり具合は指標の重みとして明らかにする必要がある。

最後に、3)として、文章の難しさを表現するスケールをどう設定するかの問題が考えられる。日本語教育の文脈で言えば、日本語能力試験の1級から4級、またはN1からN5が代表的な難易度のスケールになるであろう。国語教育の文脈で言えば、小1～高3までの学年が代表的なスケールになるであろう。

## 2.2 リーダビリティ研究の方法論

2.1節で示した3つの研究課題は、内省で明らかにできるものでもなければ、個々の事例をもとに短編的な考察を行ったところで、明らかになる問題でもない。こうした理由からリーダビリティ研究では、大規模なデータ（基準コーパス）を用いて、それを計算論的な手法で分析し、公式化するというアプローチが採用されている。計算論的手法とは、多変量解析などの統計モデルや文字の連続からなる言語モデルなどを用いる。具体的な研究例として3つの研究をとりあげる。

まず、柴崎・原（2010）は小学校1年から高校3年までの12段階を難易度のスケールとして設定し、重回帰分析によるリーダビリティ公式を提案している。難易度を決定する要因としては、①文章中の平仮名の割合、②1文の平均述語数、③1文の平均文字数、④文の平均文節数の4つの要素をとりあげている。

次に、佐藤（2011）では bigram という文字の連続をもとに各スケールの難易度を定義している。難易度スケールとしては、9段階のもの（とてもやさしい、やさしい、かなりやさしめ、やややさしめ、ふつう、ややむずかしめ、かなりむずかしめ、むずかしい、とてもむずかしい）と13段階（小学校1年～高校3年の12スケール+大学レベル）のものを設定している。

最後に、Lee&Hasebe（2016）は、柴崎・原（2010）と同様に重回帰分析によるリーダビリティ公式を提案している。難易度を決定する要因としては、①平均文長、②漢語率、③和語率、④動詞率、⑤助詞率の5つの要素をとりあげている。難易度のスケールとしては、6段階のもの（初級前半、初級後半、中級前半、中級後半、上級前半、上級後半）を設定している。

重回帰分析とは、1つの目的変数を複数の説明変数で予測するもので、例えば、説明変数として「身長」や「腹囲」や「胸囲」を使って、目的変数の「体重」を予測するという分析、あるいは「喫煙率」や「週あたりの運動時間」や「毎日の平均摂取カロリー」を使って、「健康寿命」を予測するという分析が考えられる。リーダビリティ研究の文脈で言えば、説明変数とは文の長さや文章内の語の性質といった要素になり、それをもとに、目的変数になる文章の難しさを予測するということになる。佐藤（2011）が行っている言語モデルを使った例では、文字の連続だけを手がかりにしているため、テキストの種類に関係なく、判別できるメリットがあるが、大量のテキストデータが必要になる。そのため、作業負担が多いこと、そして、判別の基準となる要素を言語分析的立場で明示化するのが難しいということが考えられる。一方の重回帰分析を使った場合は、1文の長さや述語の数や漢字の多さといった（難易度に関連が深いと思われる）説明要因の中から、どの要素が重要であるかを判定し、公式化するため、判別の基準となる要素が比較的クリアである。ただし、この方法に関しては、難易度に関わると思われる要素が事前に分かっている場合にしか使えないという問題がある。また、分析のためのデータ量に関しても、佐藤（2011）の方式ほどの大量のテキストを要求するものではないが、分析テキストが少ないと予測精度が低くなるという問題がある。

さて、次の課題として公式を作成する際に使用する「基準コーパス」について考えてみたい。リーダビリティ研究では、分析に使用する「基準コーパス」によって得られる公式が決まるため、どのようなテキストをどれだけ用いるかが研究の要になる。「基準コーパス」と統計分析の手法はワンセットとして考えられている。柴崎・原（2010）の場合、読解教育に役立つリーダビリティシステム構築を目標にしていたため、公教育の場で使用される国語科の教科書を使用している。佐藤（2011）は、「平易な日本語表現への工学的アプローチ」という観点から、汎用性の高い解析システムを作ることを目指していたため、公教育の場で使用される全教科の教科書を「基準コーパス」として使用して

いる。そして、Lee&Hasebe (2016) では、日本語教育のためのリーダビリティシステム構築を目標にしていたため、基本的には日本語教科書を使用しているが、上級前半と上級後半レベルを定義づけるために、例外的に「現代日本語書き言葉均衡コーパス」を使用している。

### 3. リーダビリティの公式ができるまで

本節では、リーダビリティシステムが持つ難易度を判別する公式ができるまで、どのような分析が必要かを示す。具体的には、「jReadability」を例に、使用したデータと統計分析の方法について述べる。

「jReadability」では、日本語教育のためのリーダビリティ構築のために、2種類のデータセットを構築している。1つ目は、初級から上級までの日本語教科書 83 冊と李 (2011) で使用した「現代日本語書き言葉均衡コーパス」のデータで構成した「学習データ」、2つ目は、旧日本語能力試験の 25 年分の読解テキストで構成した「評価データ」である。「学習データ」はリーダビリティ公式を開発するためであり、「評価データ」はリーダビリティ公式の妥当性を確認するために使用している。

分析は、次の3ステップで行われた。第1ステップとして「学習データ」をもとに「基準コーパス」を構築する作業、第2ステップとして「基準コーパス」をもとにリーダビリティ公式を作成する作業、第3ステップとして「評価データ」をもとにリーダビリティ公式の妥当性を確認する作業である。それぞれの作業における詳細な結果は、李 (2016) で紹介しているため、ここでは省略するが、分析の結果として、以下のような公式が提案されている。

$$X = \{ \text{平均文長} * -0.056 \} + \{ \text{漢語率} * -0.126 \} + \{ \text{和語率} * -0.042 \} + \{ \text{動詞率} * -0.145 \} + \{ \text{助詞率} * -0.044 \} + 11.724$$

上述の公式は、日本語学習者のための日本語文章の読みやすさを6段階で判定する。1) 初級前半、2) 初級後半、3) 中級前半、4) 中級後半、5) 上級前半、6) 上級後半である。公式としては、入力文の平均文長（平均的な文の長さ）と漢語の使用率と和語の使用率、さらに動詞の使用率、最後に助詞の使用率をもとに文章の難しさを計算する。漢語と和語の使用率は、文章に含まれる語彙的性質を測定する変数で、動詞と助詞の使用率は、文法の難しさを測定する変数である。そして、平均文長は、文章の長さを測定する変数で、リーダビリティ研究では、長ければ長いほどワーキングメモリに対して負荷がかかるため、難しくなる、ということが広く言われており、その性質を反映しているものと捉えられる。

さて、上記のこの公式は、個人が手計算で使うのは、容易なことではないため、私たちの研究プロジェクトにおいてはウェブサービスとして提供し、研究成果の普及に努めている。具体的には、<http://jreadability.net/>に実装されている。



図 1. jReadability の分析例

図 1 では、jReadability による分析例を示すため、本稿の冒頭にある要旨をコピー&ペーストで測定してみた。その結果、上級前半（むずかしい）文章であると判定された。なお、<http://jreadability.net/sample> において各レベルのサンプルテキストが公開されている。上級前半の文章は、以下のようなもので、本稿の要旨と比較しても、同じくらいの難しさを持っていることがわかる。

動物の動きにしてもそうで、ネズミはちょこまかしているし、ゾウはゆっくりと足を運んでいく。体のサイズと時間との間に、何か関係があるのではないかと、古来、いろいろな人が調べてきた。例えば、心臓がドキン、ドキンと打つ時間間隔を、ネズミで測り、ネコで測り、イヌで測り、ウマで測り、ゾウで測り、と計測して、おのおのの動物の体重と時間との関係を求めてみたのである。サイズを体重で表わすのは、体重なら、はかりにポイと載せればすぐ測れるが、体長でサイズを表わすと、しっぽは計測値に入れるのか、背伸びした長さか丸まったときの長さかなどと、難しい問題がいろいろ出てくるからだ。

#### 4. 終わりに

本稿では、語彙や文法項目に比べ、難易度という観点からの分析が難しい「文章」データに関して、リーダビリティの観点から考察を行った。特に、日本語教育のためのリーダビリティ研究として李 (2016) の研究成果を紹介してきたが、本節では、考察の締めくくりとして、難易度判別の公式を使って難易度を推定することの限界についても述べておきたい。

本研究の基本的態度として、文章の難易度を決定するための唯一無二の公式が存在するとは考えていない。実際の難易度というのは、文章によっても異なるが、読み手の属性によっても異なる。それ

を示す具体的な成果として、柴崎（2010）の研究があげられる。柴崎（2010）では、仮名だけで表記されたテキストを小学生と大学生にそれぞれ読ませ、文正誤判断課題を実施し、反応時間を測定した場合、大学生のほうが反応時間が長かったと報告している。この実験結果は、「仮名で書いてある文章＝やさしい文章」という前提がすべての日本語話者にとって成立するわけではないことを示している。柴崎（2010）の結果を日本語学習の文脈で考えてみた場合、漢字で書いてあるテキストであれば、すべて難しいかということ、そうではないということを示唆しているのではないだろうか。漢字圏の学習者にとっては、漢字で書いてあるほうが読みやすいと考えられる反面、非漢字圏の学習者にとっては、仮名表記で書かれてあったほうが読みやすいと考えられる。こうしたことから考えた場合、難易度判別の公式というのは、学習者の属性の数だけ存在するものと見るべきであろう。なぜなら、漢字圏学習者にとってのリーダビリティと非漢字圏学習者にとってのリーダビリティ、教室学習者にとってのリーダビリティと自然習得者にとってのリーダビリティ、さらには、成人学習者にとってのリーダビリティと年少者日本語学習者にとってのリーダビリティが同じであるという保証はどこにもないからである。こうした問題から考えてみた場合、筆者の研究グループは、日本語教科書という比較的安定した資料を利用することで、現状として可能な最大公約数としての難易度判定を試みたと見るべきであろう。そして、日本語学習者と言っている集団に関しても理想化された均質な集団として定義をしているが、この前提そのものが妥当かどうかについては、今後問い直す必要があるかもしれない。

謝辞：本研究は JSPS 科研費 25370573 および 16K02794 の助成を受け、行ったものである。また、本稿で紹介したリーダビリティシステムは、長谷部陽一郎氏、久保圭氏との共同研究によって開発したものである。

## 参考文献

- 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、情報処理学会、pp. 1777-1789
- 柴崎秀子（2010）「文字種による文の認知処理速度の差異—日本語テキストの難易尺度構築のための基礎研究—」、『実験音声学・言語学研究』2巻、日本実験言語学会、pp. 18-31.
- 柴崎秀子・原信一郎（2010）「12 学年を難易尺度とする日本語リーダビリティ判定式」『計量国語学』27-6、pp. 215-232、計量国語学会
- 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」『文書処理とニューマンインターフェース』18-4、pp. 1-8、情報処理学会
- 野本忠司（2016）「リーダビリティ研究の 100 年」『情報処理学会 SIG Technical Reports 2016-DC-101』、pp. 1-7.
- 李在鎬（2011）「大規模テストの読解問題作成過程へのコーパス利用の可能性」『日本語教育』148、pp. 84-98.
- 李在鎬（2016）「日本語教育のための文章難易度研究」『早稲田日本語教育学』Vol. 21、pp. 1-16.
- 李在鎬・長谷部陽一郎・久保圭(2016)「日本語コーパスの文章難易度に関する大規模調査の報告」『2016 年度日本語教育学会春季大会予稿集』、pp. 152-157.
- Lee, Jae-ho and Yoichiro Hasebe (2016 forthcoming) "Readability Measurement for Japanese Text Based on Leveled Corpora," *Papers on Japanese Language from Empirical Perspective*.
- Flesch, Rudolph (1948) "A new readability yardstick," *Journal of Applied Psychology*, Vol. 32. pp. 221-233