

読解文章の難易度の比較

— JReadability を使って —

Comparison of Reading Texts Difficulty.
- Using JReadability -

藤田裕一郎

要旨

本稿では、「日本語文章難易度判別システム」を使用し、日本語能力試験 N1 と N2、日本留学試験の読解の文章、そして NHK NEWS WEB の記事の文章の難易度を調べた。その結果、おおむね、NHK NEWS WEB、日本語能力試験 N1、日本留学試験、日本語能力試験 N2 の順で難しいが、統計学上で差が出るほどの違いはないことが分かった。そして、この結果からこれらの試験の読解問題を目標とする場合、ニュースの記事程度の難易度の文章を扱えばいいことが示唆された。

キーワード 日本語能力試験 日本留学試験 読解 難易度 JReadability

1. はじめに

進学を目的とする日本語学校や大学留学生別科では、進学に向け、日本語能力試験 (Japanese-Language Proficiency Test) や日本留学試験 (Japanese-Language Proficiency Test) を受験する学習者が多い。また近年、非漢字圏出身の学習者の増加に伴い、これらの試験のとりわけ読解教育に力を入れる機関も多いと思われる。

そこで、本稿では「日本語文章難易度判別システム」を使用し、日本語能力試験や日本留学試験などの文章の難易度を客観的に示す。これによって、これらの試験の読解問題の文章の難易度が把握できるようになり、今後の日本語指導の一助になると考えられる。

2. 先行研究

2.1 日本語能力試験

日本語能力試験は、公式ホームページ (<https://www.jlpt.jp/index.html>) によると、日本語を母語としない人の日本語能力を測定し認定する試験である。課題遂行のための言語コミュニケーション能力を言語知識 (文字・語彙・文法)、読解、聴解の3つの要素から総合的に測る。日本語能力試験には、N1、N2、N3、N4、N5の5段階のレベルがあり、本稿では、このうち上位のN1とN2の読解の文章を対象にする。N1とN2は現実の生活の幅広い場面での日本語がどのくらい理解できるかを測るとされる。

2.2 日本留学試験

日本留学試験は、公式ホームページ (<https://www.jlpt.jp/index.html>) によると、外国人留学生として、日本の大学 (学部) 等に入学を希望する者について、日本の大学等で必要とする日本語力及び基礎学力の評価を行うことを目的に実施される試験である。日本語、理科 (物理・化学・生物)、総合科目及び数学の出題科目があるが、このうち日本語科目は、記述、読解、聴解・聴読解の3領域から構成され、読解は17問程度で200点満点で採点される。

2.3 文章の難易度と文章の難易度判断システム

文章の難易度について英語では、文章に含まれる1文あたりの平均語数、1語あたりの音節数、文字数などから難易度が算出されることが多い (Flesch, 1949; Smith & Kinkaid, 1970)。一方、日本語では文の長さ、文字数に加え、語種別の割合などから文章の難易度が算出される (建石他, 1998; 柴崎・玉岡, 2010; Sato, Matsuyoshi, & Kondoh, 2008; 李, 2016)。また、現在は「日本語文章難易度判別システム」 (<http://jreadability.net/>)、「リーディングチュウ太」 (<http://language.tiu.ac.jp/>)、「リーダビリティ・リサーチ・ラボ」 (<http://readability.nagaokaut.ac.jp/readability/>) などによって、ウェブ上で文章の難易度が判定できる。

3. 本研究の目的

進学や就職、ビザ所得、また自身の学習成果の測定などを目的に日本語能力試験を受ける学習者は多い。また、大学や専門学校進学を目的に日本留学試験を受ける学生も少なくない。そのため、国内の日本語学校や大学留学生別科などでは、これらの試験のための対策クラスを設

置したり、特別なクラスを設けなくても、教師・学習者ともに授業の中で少なからず意識をしている部分はあるだろう。

他方、近年、東南アジアを中心とした非漢字圏出身の学習者の増加に伴い、読解教育に力を入れる教育機関も少なくないようである。これは、非漢字圏出身の学習者は漢字圏出身の学習者に比べて漢字語彙が不得手であること、言語体系の違いにより文章理解がより困難である傾向にあることなどが理由だと推察される。しかし、これまでのところ、目標とする試験の読解文章の難易度や、これらの試験を目標とした授業の際にどの程度の難しさの読解文章を扱えばいいかは教師の経験による主観に委ねられているのが実情だろう。そこで、これらの試験の文章の難易度を明示的に示すことで、目標となる試験の文章の難易度を客観的に把握したり、今後授業で読解を扱う際の指針にしたりできるようになると考える。

4. 調査

4.1 調査対象

今回の調査では、日本語能力試験のN1とN2の読解の文章（以下、それぞれN1、N2）、日本留学試験の読解の文章（以下、日留試）、NHK NEWS WEB (<https://www3.nhk.or.jp/news/>) のニュース記事（以下、NHK）を対象にした。

日本語能力試験については2010年の改定以降、問題が非公開になったため、2018年に作成された公式問題集第二集 (<https://www.jlpt.jp/samples/sampleindex.html>) のN1とN2の読解文章それぞれ10題ずつを使用した。日本留学試験は2022年度第2回の日本語の読解16題^{注1}を使用した。NHK NEWS WEBは2月4日から2月9日付けの記事からランダムに10編（参考資料1参照）を選び、使用した。

調査にあたり、文章の難易度を明示的に示すことを目的にするため、問題の説明やタイトル、問い、選択肢の文、注釈、出典情報は除き、問題の文章部分のみを調査対象にした。

4.2 調査方法

現在、WEB上でできる文章の難易度判定システムはいくつかあるが、本調査では文章の難易度判定に「日本語文章難易度判別システム」(JReadability)を使用した。このシステムは計算式を考えるにあたり、複数の公式を試して人間による主観判定に一番近いものを採用したとしており（李, 2016）、読み手側の主観に近いこと。また、外国人学習者を読み手に想定して開発されており、本稿の目的と合致していることが主な理由である。

5. 結果

N1とN2の10編ずつ、日留試の16編、NHKの10編を「日本語文章難易度判別システム」で分析した結果の平均は以下のとおりである（表1）。

表1 文章難易度

		リーダビリ ティ・スコア	総文数	総形態素数 (延)	総形態素数 (異)	総文字数	一文の平均 語数
N1	AV	2.78	14.50	346.00	133.40	566.00	26.99
	SD	1.16	10.50	205.18	63.32	344.88	8.86
N2	AV	2.55	9.40	303.40	111.70	479.70	35.22
	SD	1.27	5.15	144.19	36.87	228.52	12.14
日留試	AV	2.56	11.31	316.75	125.75	513.88	30.00
	SD	0.72	4.45	81.60	24.99	132.63	8.05
NHK	AV	1.98	11.10	423.90	152.20	667.10	38.98
	SD	0.67	4.15	143.70	37.14	227.29	6.41

※AVは平均、SDは標準偏差を表す。

リーダビリティ・スコアは文の難易度を示す値で、数値が大きいほど相対的に易しいテキスト、数値が小さいほど相対的に難しいテキストである。今回の調査では、NHK、N2、日留試、N1の順で難しかった。総文数はそれぞれの文章における文の数であり、N1、日留試、NHK、N2の順で多かった。総形態素数(延)は各文章における延べ総形態素数であり、NHK、N1、日留試、N2の順で多かった。総形態素数(異)は各文章における異なり総形態素数であり、NHK、N1、日留試、N2の順で多かった。総文字数は各文章における文字数であり、NHK、N1、日留試、N2の順で多かった。一文の平均語数は各文章の各文の平均数であり、NHK、N2、日留試、N1の順で多かった。

6 考察

6.1 文章の難易度

日本語文章難易度判別システムによって算出された結果を分析項目ごとにまとめたのが表2である。

表2 文章難易度結果のまとめ

分析項目	順序
リーダビリティ・スコア	NHK > N2 > 日留試 > N1
総文数	N1 > 日留試 > NHK > N2
総形態素数(延)	NHK > N1 > 日留試 > N2
総形態素数(異)	NHK > N1 > 日留試 > N2
総文字数	NHK > N1 > 日留試 > N2
一文の平均語数	NHK > N2 > 日留試 > N1

全ての結果が一致して瞭然としているとは言えないが、全体としては、NHK > N1 > 日留試 > N2の順で難しいと考えられる。N1に関して、総文数は調査対象の中で最も多い一方で、一文の平均語数は最も少なくなっており、短い文が多く使われていることが推察される。

しかし、総形態素数（延／異）と総文字数は他との比較では少なくないので、量的に少ないということではない。また、N1はリーダビリティ・スコア、総形態素数（延）、総形態素数（異）、総文字数において、SDの値が高く、文章ごとの差が大きいことが分かる。N1のリーダビリティ・スコアが最も高く、特にN2よりも高い（値が高いほど易しいテキストである）ことについて、このような点が影響している可能性も考えられるが、解釈はなかなか難しい。

他方、NHKが最も難しいことについて、今回の結果から、総形態素数（延）、総形態素数（異）、総文字数の値が他の文章よりも高く、文章の中には長い文章があることが影響していることが考えられる。これは、他の文章は試験問題であり、回答時間などから文章の長さ制限があることや、一定程度の難易度の文章の一部分を抽出し、内容を問うなどの目的があるのに対し、NHKはこのような制約がなく、必要な内容をほぼ制限なく記述できる違いことによる違いから考えると考えられる。

また、本調査では、データのランダムサンプリングができていないこと。また、N1やN2など特定のレベルの文章群が正規性を成しているか疑問であることから、分散分析を行うことの適切性が疑われた。一方で、正規性と分散の等質性に関しては、ある程度分散分析は頑健であるとの指摘もある（栗田, 1999）。そこで、参考までに分析項目について一元配置の分散分析を行ったところ、リーダビリティ・スコア、すなわち文章の難易度に差は見られなかった ($F(3, 42)=1.30, p=.28, \eta^2=.09$)。総文字数すなわち各文章における文の数について有意差は見られなかった ($F(3, 42)=1.13, p=.08, \eta^2=.08$)。総形態素数（延）、すなわち各文章における延べ総形態素数について有意差は見られなかった ($F(3, 42)=1.52, p=.22, \eta^2=.09$)。総形態素数（異）、すなわち各文章における異なり総形態素数について有意差は見られなかった ($F(3, 42)=1.73, p=.18, \eta^2=.11$)。総文字数、すなわち各文章における文字数について有意差は見られなかった ($F(3, 42)=1.30, p=.29, \eta^2=.09$)。一文の平均語数、すなわち各文章における各文の平均語数について有意差が見られた ($F(3, 42)=3.72, p=.02, \eta^2=.21$)。Post Hoc Testsによって下位検定を行った結果、NHKとN1の間に有意差が見られた ($NHK > N1, p=.02$)。この結果は、各分析項目の分析結果によって順序付けはできるものの、その差はごくわずかで、統計学上は差がないということになる。

以上をまとめると、NHK NEWS WEBにある記事は日本語能力試験N1、N2、また日本語能力試験の文章と同等かそれ以上に難易度が高いと考えられる。NHKのニュース記事と他のニュース記事の難易度の差は分かりかねるが、これらの試験の読解問題の文章を扱うにあたって、ニュース記事を選択することは難易度からいって十分であることが示唆される。

6.2 調査から得られたその他の示唆

本稿の目的とは直接関係ないが、興味深い点として2つ挙げたい。

1つ目は、文章の難易度の幅についてである。N1、N2、日留試は試験問題であり、各試験の目的に合わせて、日本語の難易度が設定されていると考えられる。しかし、リーダビリティ・スコアのSDを見ると、NHKが最も低く、他の試験のほうが高くなっている。したがって、これらの試験問題は特定の難易度の文章ばかりを選択しているというよりは、ある程度の難易度の幅を持った文章を選択していることが考えられる。

2つ目は使用されている語彙の幅広さについてである。学習者の場合、一般的にある文章に

において幅広くさまざまな語彙が使われていれば、読みにくいと考えられる。それだけ幅広い知識を必要とするからである。このことについて、文章ごとに総文字数を総形態素数（異）で割って異なり率を算出したところ、N1が26.3%、N2が14.6%、日留試が25.1%、NHKが23.5%だった。文章の中で使われている語彙の難易度は考慮されていないが、異なり具合からすると、他の文章よりもN2はやや低いため、繰り返しが多く、読みやすいと考えられる。

7. まとめと今後の課題

本稿では、「日本語文章難易度判別システム」を使用し、日本語能力試験N1とN2、日本留学試験の読解の文章、そしてNHK NEWS WEBの記事の文章の難易度を調べた。その結果、おおむね、NHK NEWS WEB、日本語能力試験N1、日本留学試験、日本語能力試験N2の順で難しいが、統計学上で差が出るほどの違いはないことが分かった。そして、この結果からこれらの試験の読解問題を目標とする場合、ニュースの記事程度の難易度の文章を扱えばいいことが示唆された。ただ、日本語能力試験は幅広い話題を扱うのに対し、日本留学試験はアカデミックな内容を扱うと考えられる一方、今回はそのような内容面には全く触れていないこと。また、各学習者によって持っている背景知識が異なるなど学習者側の観点を考慮していないなどの問題もある。そして、非漢字圏の学習者に対する教育の観点からすると、今回対象にした非常に高度なレベルの文章よりも、より低いレベルの文章を検討するほうが有益だろう。

今後はこれらの課題を含め、包括的に検討していくことが実際の授業につなげるためには必要だと考えられる。

注

- 1 2022年度第2回の日本語の読解問題は17題であったが、そのうち1題は「出版上の都合により掲載しない」とのことで掲載がなく、掲載されている16題を調査対象とした。

参考文献

- 栗田佳代子（1999）「実際のデータを用いた t 検定および検定力分析の「観測値の独立性」からの逸脱に対する頑健性の検討」『教育心理学研究』, 47, pp.263-272.
- 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.
- 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』, pp. 1-8, 情報処理学会.
- 李在鎬（2016）「日本語教育のための文章難易度に関する研究」『早稲田日本語教育学』, 21, pp.1-16.
- Flesch, Rudolph (1949) "A new readability yardstick," *Journal of Applied Psychology*, 32, pp.221-233.
- Sato, S., Matsuyoshi, S., and Kondoh, Y. (2008). Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. *LREC-08*.
- Smith A. Edgar and Peter Kincaid (1970) "Derivation and Validation of the Automated Readability Index for Use with Technical Materials," *The Journal of the Human*

参考資料

- 1 NHK NEWS WEB の記事 (全て 2023 年 2 月 13 日にアクセス)
- 1) 「寒い日でもジャンパー着て登校 「校則」で認めず 広島の中学校」(2023 年 2 月 8 日)
<https://www3.nhk.or.jp/news/html/20230208/k10013974921000.html>
- 2) 「商品渡すのは「クマの手」 対面せず接客のカフェ できた理由は」(2023 年 2 月 8 日)
<https://www3.nhk.or.jp/news/html/20230208/k10013974761000.html>
- 3) 「フィリピンから容疑者 4 人送還完了 広域強盗事件捜査本格化へ」(2023 年 2 月 9 日)
<https://www3.nhk.or.jp/news/html/20230208/k10013975281000.html>
- 4) 「途上国の新生児 死亡防ぐワクチン履歴把握 世界初のシステム」(2023 年 2 月 6 日)
<https://www3.nhk.or.jp/news/html/20230206/k10013972631000.html>
- 5) 「北海道 紋別 この冬初めて流氷が沖合に確認 平年より 12 日遅く」(2023 年 2 月 5 日)
<https://www3.nhk.or.jp/news/html/20230205/k10013971451000.html>
- 6) 「囲碁 仲邑菫三段 女流タイトル初獲得 最年少記録 13 歳 11 か月」(2023 年 2 月 6 日)
<https://www3.nhk.or.jp/news/html/20230206/k10013972191000.html>
- 7) 「アフガニスタンから避難の元留学生 家族と日本で生活へ 宮崎」(2023 年 2 月 5 日)
<https://www3.nhk.or.jp/news/html/20230205/k10013971291000.html>
- 8) 「話したことばを自動で文字に デフリンピックへ実証実験 東京」(2023 年 2 月 5 日)
<https://www3.nhk.or.jp/news/html/20230205/k10013971251000.html>
- 9) 「米「グラミー賞」に宅見将典さんのアルバム『SAKURA』」(2023 年 2 月 6 日)
<https://www3.nhk.or.jp/news/html/20230206/k10013971911000.html>
- 10) 「さっぽろ雪まつり 3 年ぶりに会場設け開幕 観光客でにぎわう」(2023 年 2 月 4 日)
<https://www3.nhk.or.jp/news/html/20230204/k10013970581000.html>

朝日大学留学生別科