

令和5年度 卒業論文

文章の読みやすさに影響する言語的特徴について

広島大学文学部人文学科
日本・中国文学語学コース
日本文学語学専攻
B202151 宮原由依

目次

1. はじめに
2. 先行研究
 - 2.1 文章の読みやすさについて
 - 2.2 文章の計量的分析について
 - 2.3 教材の難易度を計る研究について
3. 調査
 - 3.1 調査対象
 - 3.2 調査方法
4. 結果
5. 考察
 - 5.1 文の長さと言点の頻度について
 - 5.2 語種について
 - 5.3 品詞について
 - 5.4 指示語について
 - 5.5 終助詞「か」で終わる文について
 - 5.6 本研究で処理できなかった要素
6. おわりに

※引用部分の傍線は筆者によるものである。

1. はじめに

私たちの身の回りにある文章には、「初級」や「中学生向け」というように、読み手のレベルを設定しているものがある。では、読みやすい文章とはどのようなものなのだろうか。もちろん、文章の主題が難しければ、読者は読みにくいと感じるだろう。しかし、同じ主題を扱っていても、それをどのように書くかは何通りも存在する。含まれる語彙や扱う主題といった文章の内容的なこと以外にも、文章の読みやすさに影響を与える要素はあるのではないだろうか。

そこで筆者は、比較的読みやすいとされている文章と比較的読みにくいとされている文章を統計的に分析し、比較すれば、文章の読みやすさに影響する言語的特徴が見えてくるのではないだろうかと考えた。

3年次の中間発表では、中学校と高等学校の国語科教科書に掲載された説明的文章を合わせて12作品調査し、文の長さ、和語率、漢語率、指示語率の4項目について分析を行った。本研究では、さらに調査対象を広げるだけでなく、調査項目も見直し、より詳しい調査を行う。その結果をもとに、文章の読みやすさにはどのような言語的特徴が影響しているのか、また、文章の読みやすさを計るにはどのようにすればよいのかを考察していく。

2. 先行研究

2.1 文章の読みやすさについて

日本語の文章の読みやすさについての研究の代表例としては、堀川（1957）¹が挙げられる。この研究は、朝日新聞の記事をわかりやすくすることを発端としたものであり、大規模な社会調査をもとに行われた。以下にその結果の要約部分（p.196）を引用する。

- ①内容が同じだとすれば、その文章をわかりやすくしたり、わかりにくくしたりする要因は、漢字の多少、センテンスの長短、構文の複雑単純によつて（原文ママ）作用される。
- ②この要因のうちで、漢字の多少が最もウェイトが重い。
- ③事実を報道する文章と、文学的文章では要因の働き方が違う。
- ④一定の極限で、事実文は

¹ 同書の「はしがき」にて、同時期に研究への動きがあったという国立国語研究所、NHKの放送文化研究所、東京女子大学の森岡健二助教授らの研究はそれぞれの理由で中絶したことを説明し、「この種の研究を継続しているのは、私の知る限りでは私だけとなってしまった。」と述べていることから、堀川の研究が日本における文章の読みやすさについての原初の研究だと思われる。

漢字の多い方がわかりよい

(センテンスは長い方がよい) 有意差なし

構文は複雑な方がわかりよい

⑤文学文では

漢字が少ない方がわかりよい

センテンスは短い方がよい

(構文は単純な方がわかりよい) 有意差なし

これは、当時の東京都区内の義務教育終了時(中学3年生の2月、3月)の男女を対象としたテストによって得られた結果である。このテストは、Aの文章はセンテンスが長く構文が複雑であり漢字が多いもの、Bの文章はセンテンスが長く構文が単純であり漢字が少ないもの、Cの文章はセンテンスが短く構文が単純であり漢字が多いもの……といったように、文章の読みやすさの要因と思われるものについて少しずつ条件を変えた文章をいくつかのテーマで作成し、東京都内の中学三年生全体を代表できるよう厳密に抽出した対象者に読み比べさせるというものである。

ここでの「構文の複雑さ」とは、複雑な構文の型としていくつかのタイプを想定し、そのうちいくつを含んでいるかで判定される。その型とは、①複数の文に分けられるものを1文にしているもの、②1つの主語に対し複数の述語があるもの、③述語が文頭に来るもの、④主語のないもの、⑤主語b-主語bが主語a-主語aの中に挿入されているもの、⑥⑤の型以外で主語と述語が非常に遠いもの、⑦形容詞が長いもの、⑧受身表現のもの、⑧の8つを想定されている。そのため、これらの複雑な構文の中の具体的にどの要素が文章の読みやすさに影響したのかについては明らかにされていない。

2.2 文章の計量的分析について

樺島・寿岳(1965)は、作家の文体を客観的にとらえる方法として、統計的把握を行っている。ここでは、現代語で書かれた短編小説100作品から各作品80文を無作為抽出し、分析している。調査項目は①名詞の比率(%)、②MVR²、③指示詞の比率(%)、④字音語³の比率(%)、⑤文の長さ(自立語数)、⑥引用文の比率(%)、⑦接続詞をもつ文の比率(%)、⑧現在どめの文の比率(%)、⑨色彩語の比率(%)、⑩表情語の比率(%)である。本研究では、これらを参考にしつつ、文章の読みやすさに影響を与えると考えられる項目をさらに追加して調査する。

樺島・寿岳(1965)は文章の計量的分析を行った研究の代表的なものであり、教材の分析

² 形・形動・副・連体詞の数/動詞の数*100の計算式で求められる値のこと。この値が大きいほど様態を表す語が多く、小さいほど動作を表す語が多いことを示す。

³ 漢語のこと。

にも活用されている。遠藤・大谷(2015)は、小学校2年生の説明文教材12作品に対して、樺島・寿岳(1965)にならい統計的な観点から分析を行っている。この分析の結果は以下の通りであった。

したがって、2年生の説明的文章は、「論理的展開より、むしろ自明の順序、時間的な順序にしたがって淡々と事柄を述べる平易な語り口の文章」というあたりなのだろう。もともと和語的表現を基調とするため語感は一層柔らかく、修飾語の値も高くないので、詳述的な文章とはいえ、原色豊かな油絵のような映像が喚起されるわけでもない。

このような結果になったのは、小学校2年生は教材に使用できる漢字や語の数がまだ少ないことが大きな原因であると考えられる。そのため、本研究では、学習していない漢字による語の選択の制限が起こりにくくなると思われる、中学校以上の教材を対象として調査を行う。

2.3 教材の難易度を計る研究について

文章の難易度を計る研究には、教科書を分析対象として語彙の分析を行っている研究が多い。そのような研究についてまとめた似内(2020)によると、教材の難易度についての指標には「(1) 語彙そのものの難易度、(2) 文章の長さ、(3) 文法的な要素、(4) 児童・生徒・学生が日常で使用する語彙か、(5) 説明されている「概念」を理解するために必要な語彙の抽出、などが有効である」と述べられている。(2)や(3)のように内容以外の面に着目した研究もあるが、いずれも日本語を母語としない日本語学習者を想定したものや、教材の教科を限定したもの、特定の語の文法的な特徴を取り上げたものなど、部分的なものにとどまっている。

教科書は読み手の対象年齢が明確であり、比較的読みやすい文章と比較的読みにくい文章を客観的に位置づけることができると考えたため、本研究でも教科書の教材を分析対象とするが、内容以外の面に着目する。本研究は教材研究のために行うのではなく、比較的読みやすい文章と比較的読みにくい文章を計量的に分析し比較することで、文章の読みやすさに影響を与える言語的特徴とはどのようなものか、また、実際に数値としてどのように違いが現れるのかを調査するものである。

3. 調査

3.1 調査対象

中学校・高等学校の国語科教科書に教材として掲載された説明的文章を調査対象とする。本研究における説明的文章とは、筆者の意見や主張を何らかの根拠や例をもとに述べる文章のことである。小説・詩・随筆のような文学的文章は、今回は対象としない。また前述の

通り、小学校の教科書に掲載された作品については、学習していない漢字が多いことによって語の選択が制限されると考えられるため、今回は取り上げない。

本発表で取り扱うのは以下の資料1にまとめた30作品（中学校15作品、高等学校15作品）である。中学校・高等学校ごとに著者名の五十音順で並べ、①から⑮の作品番号を付した。また、教科書に示された出典と、筆者の主観的なものになるが、おおよその内容を簡潔に示した。

中学校の部は、入手できた最新の中学2年生の教科書に教材として掲載された説明的文章のうち、引用部分が複数行に渡っているもの以外全て調査した。高等学校の部は、入手できた最新の論理国語の教科書に教材として掲載された説明的文章のうち、引用部分が複数行に渡っているもの以外から無作為に15作品選んで調査対象とした。これは、中学校の作品数と揃えるためである。なお、無作為抽出にはExcelのRAND関数を利用した。

【資料1】調査対象作品一覧（著者五十音順）

【中学校】

- ①井上恭介（2021）「壁に残された伝言」『現代の国語』、三省堂
出典：（2003）『ヒロシマ——壁に残された伝言』を書き改めたもの
内容：原爆直後に残された文字が伝えるもの
- ②内田樹（2021）「学ぶ力」『伝え合う言葉 中学国語2』、教育出版
出典：書き下ろし
内容：学ぶことができる力を伸ばすために必要なこと
- ③奥山英登（2021）「動物園でできること」『現代の国語』、三省堂
出典：書き下ろし
内容：動物園の役割と野生動物の素晴らしさ
- ④小野里公成（2021）「日本の花火の楽しみ」『伝え合う言葉 中学国語2』、教育出版
出典：書き下ろし
内容：日本の花火の素晴らしさ
- ⑤熊谷晋一郎（2021）「自立とは『依存先を増やすこと』」『現代の国語』、三省堂
出典：全国大学生協同組合連合会ウェブサイト「大学生協の保障制度」の記事を書き改めたもの
内容：脳性まひの筆者自身の経験からわかった「自立」とは何か
- ⑥千住博（2021）「サハラ砂漠の茶会」『新しい国語2』、東京書籍
出典：（2004）『美は時を超える 千住博の美術の授業Ⅱ』
内容：美の感覚からわかる人間の共通性
- ⑦龍村仁（2021）「ガイアの知性」『伝え合う言葉 中学国語2』、教育出版
出典：書き下ろし
内容：象や鯨の「知性」から人間が学ぶこと

- ⑧田中真知 (2021)「黄金の扇風機」『新しい国語 2』、東京書籍
 出典：(2009)『美しいをさがす旅にしよう』
 内容：グローバル化とメディアの普及による美の感覚の均質化
- ⑨橋本淳司 (2021)「一〇〇年後の水を守る」『現代の国語』、三省堂
 出典：書き下ろし
 内容：水不足の現状とその解決のために人間ができること
- ⑩坂茂 (2021)「紙の建築」『伝え合う言葉 中学国語 2』、教育出版
 出典：書き下ろし
 内容：紙管を利用した避難所の間仕切り
- ⑪藤田祐樹 (2021)「ハトはなぜ首を振って歩くのか」『新しい国語 2』、東京書籍
 出典：書き下ろし
 内容：ハトの歩行時の首振りの原因
- ⑫布施英利 (2021)「君は『最後の晚餐』を知っているか」『国語 2』、光村図書
 出典：書き下ろし
 内容：「最後の晚餐」の素晴らしさ
- ⑬丸井敦尚 (2021)「水の山 富士山」『伝え合う言葉 中学国語 2』、教育出版
 出典：書き下ろし
 内容：湧水に着目してわかる富士山が私たちにもたらす恩恵
- ⑭安田喜憲 (2021)「モアイは語る——地球の未来」『国語 2』、光村図書
 出典：書き下ろし
 内容：イースター島の例をもとにした人間の環境破壊への批判
- ⑮渡部潤一 (2021)「人間は他の星に住むことができるのか」『現代の国語』、三省堂
 出典：書き下ろし
 内容：地球以外の星での生命活動の可能性と地球のかけがえのなさ

【高等学校】

- ①稲垣栄洋 (2023)「進化が導き出した答え」『標準論理国語』、第一学習社
 出典：(2019)『敗者の生命史 38 億年』
 内容：生物の進化による多様化
- ②大森荘蔵 (2023)「真実の百面相」『論理国語』、筑摩書房
 出典：(1981)『流れとよどみ』
 内容：「真実」とは一面的なものではない
- ③小田原のどか (2023)「『展示』が伝えるもの」『新 論理国語』、三省堂
 出典：「美術手帖」ウェブサイト
 内容：広島市平和記念資料館のリニューアル展示から得られる教訓
- ④菅野仁 (2023)「共同性の幻想」『標準論理国語』、第一学習社

- 出典：(2008)『友だち幻想一人と人の〈つながり〉を考える』
内容：「同質性から並存性へ」という提言
- ⑤齋藤純一 (2023)「デモクラシーと平等」『精選 論理国語』、明治書院
出典：(2017)『不平等を考える 政治理論入門』
内容：平等な政治的自由の保障の重要性
- ⑥関陽子 (2023)「野生の『クジラ』と人間の『鯨』」『論理国語』、数研出版
出典：尾関周二編 (2016)『「環境を守る」とはどういうことか』
内容：環境問題の解決のために尊重されるべき価値
- ⑦瀬地山角 (2023)「女性/男性だから〇〇」『新 論理国語』、三省堂
出典：(2020)『炎上 CM でよみとくジェンダー論』
内容：異質平等論の問題点と性別からの自由
- ⑧多木浩二 (2023)「消費されるスポーツ」『論理国語』、数研出版
出典：(1995)『スポーツを考える』
内容：メディアによるスポーツの社会的な現象への変化
- ⑨土井隆義 (2023)「メディアの変容」『精選論理国語』、東京書籍
出典：(2014)『現代文化を学ぶ人のために』
内容：現代の若者たちの人間関係への依存とその営みの難しさ
- ⑩野矢茂樹 (2023)「ロボットが隣人になるとき」『新編論理国語』、東京書籍
出典：書き下ろし
内容：「意志」「自由」とは何か
- ⑪平田オリザ (2023)「対話の精神」『論理国語』、大修館書店
出典：(2012)『わかりあえないことから』
内容：対話の大切さ
- ⑫前田健太郎 (2023)「多様な視点から考える」『新 論理国語』、三省堂
出典：(2019)『女性のいない民主主義』
内容：ジェンダーの視点の提示と、視点の多様性の重要性
- ⑬丸山圭三郎 (2023)「言語と記号」『精選論理国語』、東京書籍
出典：(1985) 雑誌『書齋の窓』
内容：言語と他の一般記号との違い
- ⑭森博嗣 (2023)「『具体』から『抽象』へ」『論理国語』、数研出版
出典：(2013)『人間はいろいろな問題についてどう考えていけばいいのか』
内容：抽象的思考のメリット
- ⑮鷺谷いづみ (2023)「人類による環境への影響」『精選 論理国語』、明治書院
出典：(2004)『自然再生 持続可能な生態系のために』
内容：人類による地球生態系の不健全化

本研究では、OCRを利用して教科書を電子テキスト化し、読み取りの誤りを目視で確認して修正したものを利用する。OCRとは、手書きあるいは印刷された文字を読み取り、デジタルの文字コード化する技術のことである。本研究では「LINE」の「文字認識」機能を利用した。

3.2 調査方法

本発表では、まず「Web 茶まめ」を利用してテキストの形態素解析を行う。「Web 茶まめ」とは、「複数の UniDic⁴辞書による形態素解析を行える Web アプリケーション」⁵である。ここでやっている形態素解析とは、テキストを言語上で意味を持つ最小単位である形態素に分け、その品詞などを判別することである。今回は「現代語」辞書を利用し、出力項目は語彙素、語彙素読み、品詞、活用型、活用形、発音形、仮名形、語種とした。

次に、その解析結果を Excel に出力し（図 1）、明らかな判定ミスを目視で確認し修正したあと、Excel と Python⁶を利用して各項目を集計する。

辞書	文境界	書字形	語彙素	語彙素読み	品詞	活用型	活用形	発音形	仮名形	語種
現代語	B	あなた	貴方	アナタ	代名詞			アナタ	アナタ	和
現代語		は	は	ハ	助詞-係助詞			ワ	ハ	和
現代語		今	今	イマ	名詞-普通名詞-副詞可能			イマ	イマ	和
現代語		,	,		補助記号-読点					記号
現代語		広島	ヒロシマ	ヒロシマ	名詞-固有名詞-地名-一般			ヒロシマ	ヒロシマ	固
現代語		の	の	ノ	助詞-格助詞			ノ	ノ	和
現代語		雑踏	雑踏	ザットウ	名詞-普通名詞-サ変可能			ザットー	ザットウ	漢
現代語		に	に	ニ	助詞-格助詞			ニ	ニ	和
現代語		立つ	立つ	タツ	動詞-一般 五段-タ行 連用形-促			タツ	タツ	和
現代語		て	て	テ	助詞-接続助詞			テ	テ	和
現代語		,	,		補助記号-読点					記号
現代語		半	半	ハン	名詞-普通名詞-一般			ハン	ハン	漢
現代語		世紀	世紀	セイキ	名詞-普通名詞-一般			セーキ	セイキ	漢
現代語		以上	以上	イジョウ	名詞-普通名詞-副詞可能			イジョー	イジョウ	漢
現代語		も	も	モ	助詞-係助詞			モ	モ	和
現代語		前	前	マエ	名詞-普通名詞-副詞可能			マエ	マエ	和
現代語		の	の	ノ	助詞-格助詞			ノ	ノ	和

【図 1】形態素解析の結果の例（「壁に残された伝言」冒頭）

集計する項目は、延べ語数（すべての語数の合計）、自立語数（延べ語数から助詞、助動詞、接頭辞、接尾辞の数を除いたもの）、文の数、読点の数、モーラ数、和語数、漢語数、外来語数、動詞数、接続助詞数、接続詞数、指示語数、文末の終助詞「か」の数の 13 項目

⁴ 形態素解析器である MeCab 用の解析用辞書のこと。

⁵ 堤・小木曾（2023）のプレプリント（https://chamame.ninjal.ac.jp/docs/IPSJ_64-3_749-757.pdf）より引用。

⁶ プログラミング言語の 1 つ。本研究では、openpyxl という Excel ファイルを読み書きするための外部ライブラリを使い、集計に利用した。

である。詳しい数え方は、本稿の末尾に例を挙げて説明した。(資料2)

本研究では、より速く正確に集計するために、Python を利用して以上の項目を集計した。その際に用いた Python のコードを以下の資料3に示す。

【資料3】集計用プログラム

```
import openpyxl
import re
wb = openpyxl.load_workbook("ファイル名.xlsx")
ws = wb.active
nobego = 0
bunto = 0
dosisu = 0
setujosu = 0
setusisu = 0
jiritugo = 0
wagosu = 0
kangosu = 0
gaigosu = 0
sijisu = 0
gimon = 0
haku = 0
dokuten = 0
sijilist = ["この", "これ", "こんな", "それ", "その", "そんな", "あれ", "あの", "あんな"]
for row in ws.iter_cols(min_row=2):
    for cell in row:
        if cell.col_idx == 2:
            if "B" == cell.value:
                bunto += 1
        if cell.col_idx == 3:
            if cell.value in sijilist:
                sijisu += 1
            elif "か" == cell.value:
                b = cell.offset(2,-1)
                f = cell.offset(0,3)
                if "B" == b.value and "助詞-終助詞" == f.value:
                    gimon += 1
```

```

elif ", " == cell.value:
    dokuten += 1
if cell.col_idx == 6:
    if cell.value is None:
        pass
    elif "記号" not in cell.value:
        nobego += 1
    if "動詞-一般" == cell.value:
        dosisu += 1
    elif "助詞-接続助詞" == cell.value:
        setujosu += 1
    elif "接続詞" == cell.value:
        setusisu += 1
    elif "助詞" not in cell.value and "助動詞" != cell.value and "接頭辞" not in cell.value
and "接尾辞" not in cell.value:
        jiritugo += 1
if cell.col_idx == 9:
    a = re.sub(r'ャ|ユ|ヨ|ァ|ィ|ウ|エ|ォ', "", cell.value or "")
    haku = len(a) + haku
if cell.col_idx == 11:
    if "和" == cell.value:
        wagosu += 1
    elif "漢" == cell.value:
        kangosu += 1
    elif "外" == cell.value:
        gaigosu += 1
jiritugo = jiritugo + dosisu + setusisu
ws2 = wb.create_sheet(title="集計")
ws2["A1"] = nobego
ws2["A2"] = jiritugo
ws2["A3"] = bunto
ws2["A4"] = dokuten
ws2["A5"] = haku
ws2["A6"] = wagosu
ws2["A7"] = kangosu
ws2["A8"] = gaigosu

```

```
ws2["A9"] = dosisu
ws2["A10"] = setujosu
ws2["A11"] = setusisu
ws2["A12"] = sijisu
ws2["A13"] = gimon
wb.save("ファイル名.xlsx")
wb.close()
```

集計した 13 項目を利用して、文の長さ(一文当たりの自立語数、一文当たりのモーラ数)、読点の頻度、語種率(和語率、漢語率、外来語率)、一文当たりの動詞数、一文当たりの接続助詞数、接続詞率、指示語率、文末「か」率の 11 項目を求め、分析する。以下にそれぞれの分析項目の計算式⁷を示す。

- a. 文の長さ(自立語) = 自立語数/文の数
- b. 文の長さ(モーラ) = モーラ数/文の数
- c. 読点の頻度 = モーラ数/読点の数

文の長さは二通りの計算式で計る。a は、一文当たりの自立語数を求める。これが多くなると、一文に含まれる情報量が多くなるため、読みにくくなると予想される。b は、一文当たりのモーラ数を求める。モーラとは、日本語の音を数える単位である。例えば、「チューリップ」は「チュ|ー|リ|ッ|プ」となり、5 モーラとなる。二通りの計算方法を採用した理由は、黙読するときだけでなく、音読するときの感覚的な文の長さも計るためである。また、c は、読点が入る頻度が少ないほど数値が大きくなる。b、c の値が大きくなると、読みにくく感じると予想される。

- d. 和語率 = 和語数/延べ語数 * 100 (%)
- e. 漢語率 = 漢語数/延べ語数 * 100 (%)
- f. 外来語率 = 外来語数/延べ語数 * 100 (%)

和語とは「山(やま)」「川(かわ)」のように古くから日本で使われていた言葉で、漢語とは、「山岳(さんがく)」「河川(かせん)」のように中国から入ってきた言葉を指す。和語には柔らかい印象があり、訓読みのため意味がわかりやすい。漢語には硬い印象があり、音読みのため意味がわかりにくい。したがって、読みやすい文章であるほど、d は上がって e は下がることが予想される。

⁷ a~c の計算式は大川(2019)、1 は樺島・寿岳(1965)による。

外来語とは「ニュース (news)」「コーヒー (coffee)」のように外国から入ってきた言葉のことである。広義には漢語も含めるが、本研究では主に欧米から入ってきた狭義の外来語を指す。外来語には一般に定着していない語も多いため、fが上がると読みにくくなることが予想される。

- h. 一文当たりの動詞数=動詞数/文の数
- j. 一文当たりの接続助詞数=接続助詞数/文の数
- k. 接続詞率=接続詞数/文の数*100 (%)

h、j、kの値が大きいほど、文章に含まれる情報量は多く、構造は複雑になり、読みにくくなると予想される。

- l. 指示語率=指示語数/自立語数*100 (%)

本研究での「指示語」は、前に出てきた語を指し示す「この、これ、こんな、それ、その、そんな、あれ、あの、あんな」に限定する。このような語が多くなると、文脈を理解して読む必要性が高くなるため、読みにくくなると予想される。なお、延べ語数ではなく自立語数で割っているのは、指示語は自立語を置き換えるものであるからである。つまりlは、自立語がどの程度指示語に置き換えられているかの目安になる。

- m. 文末「か」率=文末の終助詞「か」の数/文の数*100 (%)

疑問の終助詞「か」が文末に現れるとき、「では、○○のときはどうなるのだろうか。」のような、読者に問いかける文となる。このような文は読者に考えさせ、興味を引く効果があるため、より対象年齢の低い、読みやすい文章に多くなると予想される。

4. 調査結果

調査結果をまとめたものが次のページの表1、表2である。なお、以降の表において小数点以下は全て1桁のみ表示し、作品名は簡略化して表記し、資料1で示した作品番号を付している。

【表 1】中学校の調査結果

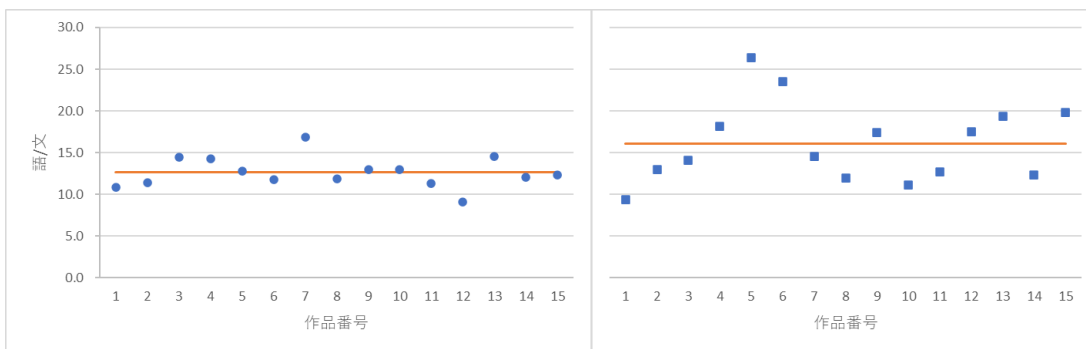
	①伝言	②学ぶ力	③動物園	④花火	⑤自立	⑥サハラ	⑦ガイア	⑧黄金	⑨一〇〇	⑩紙	⑪ハト	⑫晩餐	⑬富士山	⑭モアイ	⑮人間	平均	標準偏差
述べ語数	2157	1995	2547	1267	868	1210	1995	1155	1926	1625	2142	1749	1377	1753	1284		
自立語数	1139	1029	1388	700	433	612	1076	593	1132	845	1133	937	770	953	692		
文数	105	90	96	49	34	52	64	50	87	65	100	103	53	79	56		
読点数	123	131	171	92	57	66	168	83	139	109	161	150	95	105	87		
拍数	4281	3887	5300	2617	1677	2431	4010	2327	3963	3359	4015	3477	2726	3539	2614		
和語数	1690	1651	1971	1027	721	984	1524	914	1331	1149	1851	1351	1024	1256	969		
漢語数	416	320	450	225	132	171	419	168	507	412	265	311	295	404	276		
外来語数	28	11	72	3	6	17	26	31	56	29	6	23	11	50	9		
動詞数	197	181	189	121	76	91	136	104	133	103	201	130	100	102	88		
接続助詞数	84	92	132	55	52	54	87	52	101	64	166	83	89	81	58		
接辞詞数	13	5	14	6	2	8	15	10	6	6	11	10	11	9	11		
指示語数	44	54	45	23	14	28	53	17	26	15	25	52	19	32	20		
文末「か」数	12	4	3	1	0	4	2	1	1	3	3	3	2	8	6		
文長 (自立語)	10.8	11.4	14.5	14.3	12.7	11.8	16.8	11.9	13.0	13.0	11.3	9.1	14.5	12.1	12.4	12.6	1.8
文長 (モーラ)	40.8	43.2	55.2	53.4	49.3	46.8	62.7	46.5	45.6	51.7	40.2	33.8	51.4	44.8	46.7	47.5	6.7
読点の頻度	34.8	29.7	31.0	28.4	29.4	36.8	23.9	28.0	28.5	30.8	24.9	23.2	28.7	33.7	30.0	29.5	3.7
和語率	78.3	82.8	77.4	81.1	83.1	81.3	76.4	79.1	69.1	70.7	86.4	77.2	74.4	71.6	75.5	77.6	4.7
漢語率	19.3	16.0	17.7	17.8	15.2	14.1	21.0	14.5	26.3	25.4	12.4	17.8	21.4	23.0	21.5	18.9	4.0
外来語率	1.3	0.6	2.8	0.2	0.7	1.4	1.3	2.7	2.9	1.8	0.3	1.3	0.8	2.9	0.7	1.4	0.9
1文あたりの動詞数	1.9	2.0	2.0	2.5	2.2	1.8	2.1	2.1	1.5	1.6	2.0	1.3	1.9	1.3	1.6	1.8	0.3
1文あたりの接続助詞数	0.8	1.0	1.4	1.1	1.1	1.5	1.0	1.4	1.2	1.0	1.7	0.8	1.7	1.0	1.0	1.2	0.3
接続詞率	12.4	5.6	14.6	12.2	5.9	15.4	23.4	20.0	6.9	9.2	11.0	9.7	20.8	11.4	19.6	13.2	5.5
指示語率	3.9	5.2	3.2	3.3	3.2	4.6	4.9	2.9	2.3	1.8	2.2	5.5	2.5	3.4	2.9	3.5	1.1
文末「か」率	11.4	4.4	3.1	2.0	0.0	7.7	3.1	2.0	1.1	4.6	3.0	8.7	3.8	10.1	10.7	5.1	3.6

【表2】高等学校の調査結果

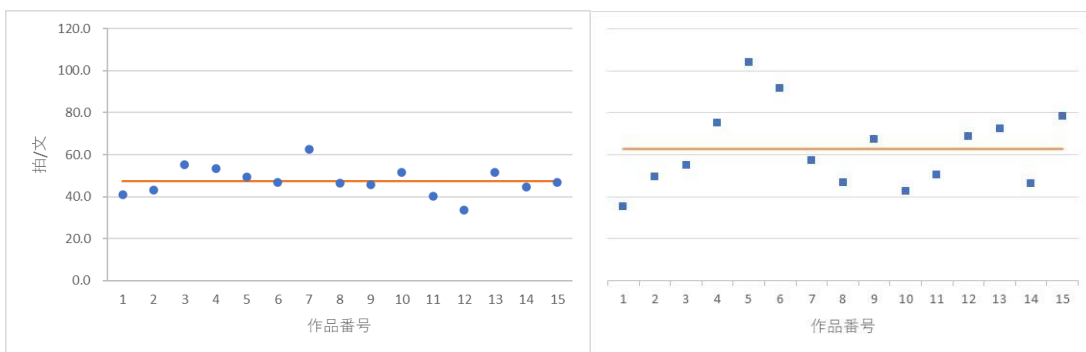
	①進化	②真実	③展示	④共同	⑤子モ	⑥野生	⑦女性	⑧消費	⑨メディア	⑩ロボット	⑪対話	⑫多様	⑬言語	⑭具体	⑮人類	平均	標準偏差
述べて語数	1746	2150	2054	1694	1556	2538	1403	2417	2252	1925	1514	1130	1862	1899	2558		
自立語数	940	1116	1099	855	844	1387	726	1304	1199	1004	799	612	1005	1010	1407		
文数	100	86	78	47	32	59	50	109	69	90	63	35	52	82	71		
読点数	128	62	107	106	91	132	74	96	158	89	125	74	122	167	158		
拍数	3538	4246	4308	3534	3331	5396	2865	5095	4642	3837	3191	2408	3765	3805	5562		
和語数	1330	1770	1383	1225	967	1691	1028	1686	1649	1533	1194	796	1383	1520	1663		
漢語数	283	329	574	439	540	804	351	498	518	336	275	252	416	323	834		
外来語数	115	16	40	19	22	26	3	175	60	40	12	48	24	44	30		
動詞数	116	147	137	123	82	141	115	144	118	159	142	82	149	160	150		
接続助詞数	98	82	61	82	67	97	87	136	113	97	91	45	95	101	93		
接続詞数	16	13	17	9	4	12	8	14	21	11	15	4	10	11	19		
指示語数	30	73	32	12	32	18	23	33	37	52	28	16	43	44	44		
文末「か」数	5	12	0	3	0	8	0	7	3	12	3	3	5	2	1		
文長(自立語)	9.4	13.0	14.1	18.2	26.4	23.5	14.5	12.0	17.4	11.2	12.7	17.5	19.3	12.3	19.8	16.1	4.6
文長(モーラ)	35.4	49.4	55.2	75.2	104.1	91.5	57.3	46.7	67.3	42.6	50.7	68.8	72.4	46.4	78.3	62.8	18.6
読点の頻度	27.6	68.5	40.3	33.3	36.6	40.9	38.7	53.1	29.4	43.1	25.5	32.5	30.9	22.8	35.2	37.2	11.2
和語率	76.2	82.3	67.3	72.3	62.1	66.6	73.3	69.8	73.2	79.6	78.9	70.4	74.3	80.0	65.0	72.8	5.8
漢語率	16.2	15.3	27.9	25.9	34.7	31.7	25.0	20.6	23.0	17.5	18.2	22.3	22.3	17.0	32.6	23.4	6.0
外来語率	6.6	0.7	1.9	1.1	1.4	1.0	0.2	7.2	2.7	2.1	0.8	4.2	1.3	2.3	1.2	2.3	2.0
1文あたりの動詞数	1.2	1.7	1.8	2.6	2.6	2.4	2.3	1.3	1.7	1.8	2.3	2.3	2.9	2.0	2.1	2.1	0.5
1文あたりの接続助詞数	1.0	1.0	0.8	1.7	2.1	1.6	1.7	1.2	1.6	1.1	1.4	1.3	1.8	1.2	1.3	1.4	0.4
接続詞率	16.0	15.1	21.8	19.1	12.5	20.3	16.0	12.8	30.4	12.2	23.8	11.4	19.2	13.4	26.8	18.1	5.5
指示語率	3.2	6.5	2.9	1.4	3.8	1.3	3.2	2.5	3.1	5.2	3.5	2.6	4.3	4.4	3.1	3.4	1.3
文末「か」率	5.0	14.0	0.0	6.4	0.0	13.6	0.0	6.4	4.3	13.3	4.8	8.6	9.6	2.4	1.4	6.0	4.8

5. 考察

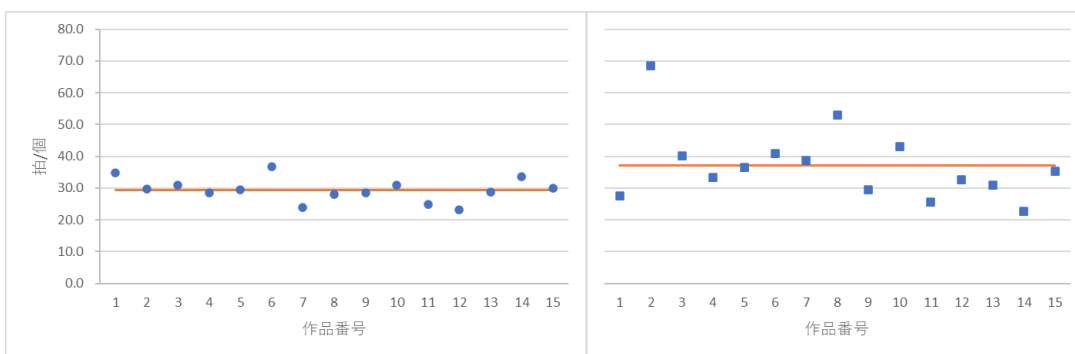
5.1 文の長さと言点の頻度について



【図2】文の長さ（自立語）の結果



【図3】文の長さ（モーラ）の結果



【図4】読点の頻度の結果

結果をグラフに表すと図2、3、4のようになる。なお、これ以降のグラフでは、特に断りのない場合、左側の丸が中学校の部、右側の四角形が高等学校の部の結果を、直線がそれぞれの部の平均値を示している。横軸の作品番号とは、資料1で示したものである。

3項目とも、平均値は高等学校の教科書に掲載された作品のほうが大きかった。例えば、

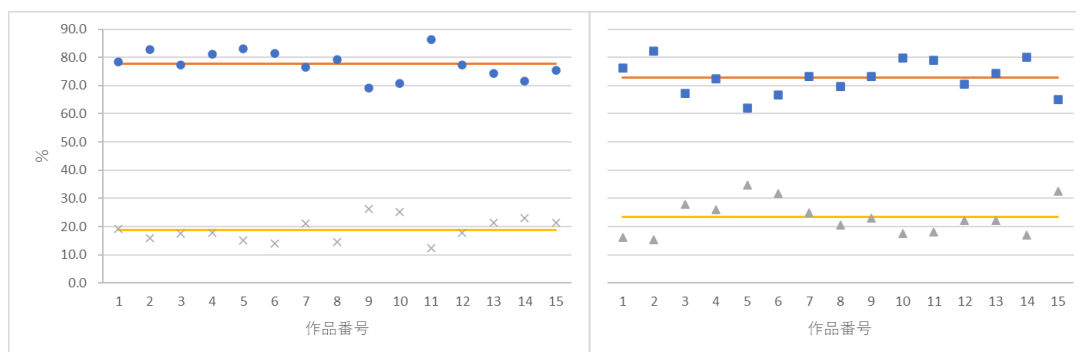
高等学校⑤「デモクラシーと平等」は、文の長さにおいて平均値を大きく上回っている。この作品の中でも特に長い文には以下のようなものがある。

こうした負の循環を脱していくためには、労働分配率を上げれば国際競争力が失われる、所得税における累進制を強化すれば成長を導くインセンティブが損なわれるといった、よく提起される(実用的な)理由が妥当なのかどうかもあらためて問い返しながらか、政治的影響力がより分散し、より対称的に行使されるためにはどのような制度の再編が必要かを真剣に考える必要がある。(高等学校⑤)

この文は、「理由」にかかる部分が非常に大きく、構造が複雑で意味を追いくい。長い文は、ただ単に感覚的に読みにくいというだけではなく、長く書くほど文の構造が複雑になりやすいから読みにくくなるのではないだろうか。

また、3項目とも、中学校は平均値付近に集まって分布しているのに対し、高等学校はばらつきが大きい結果となった。ここから、読みやすい文章の規範はおよそ中学校の平均と近いところにあり、高等学校の場合はそれを超えるものも許容されているのではないかと考えられる。

5.2 語種について

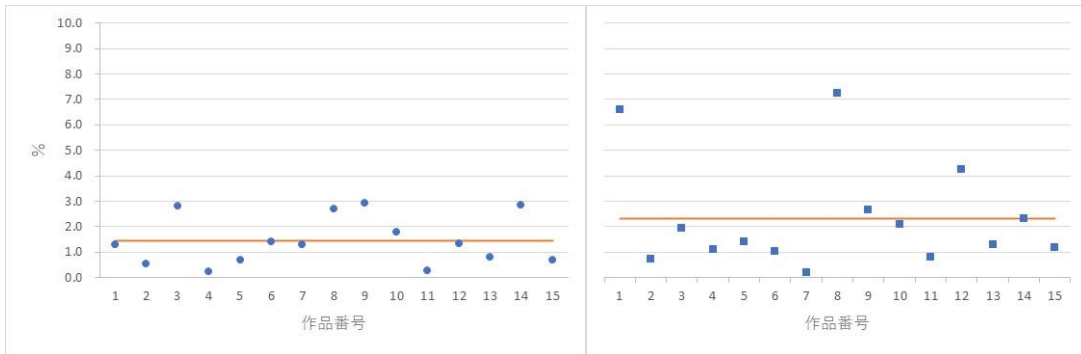


【図 5】和語率・漢語率の結果

和語率と漢語率は、どの作品も和語率のほうが高く、漢語率が高くなれば和語率は低くなる関係にある。よって、2項目をまとめてグラフに表した。丸が中学校の和語率、バツ印が中学校の漢語率、四角形が高等学校の和語率、三角形が高等学校の漢語率である。高等学校の教科書に掲載された作品のほうが読んでいて硬い印象を受けるものが多く、数値的にも高等学校のほうが漢語率の平均値が高い結果となった。遠藤・大谷(2015)は、小学校2年生の説明文教材を分析するにあたって、「もともと児童向け文章が漢語を避け、和語的表現による自明の理由」があるとして漢語率の調査を行っていないが、これは小学校2年生では学習している漢字が少ないために和語を選択せざるを得ないことと関連すると考え

られる。

今回の調査で扱った教材の対象である中学 2 年生は、少なくとも中学 1 年生の国語科の内容として示されている「小学校学習指導要領第 2 章第 1 節国語の学年別漢字配当表（以下『学年別漢字配当表』という。）に示されている漢字⁸に加え、その他の常用漢字のうち 300 字程度から 400 字程度までの漢字を読むこと。」⁹を達成していると考えられるため、常用漢字 2136 文字のうち 6 割以上はすでに読めると見込まれるが、やはり高等学校に比べて中学校のほうが優先的に和語が使用されていると言えるだろう。



【図 6】 外来語率の結果

外来語率については、どの作品においても数値が小さく、その割にばらつきが大きい結果となった。最も数値が高くなった高等学校⑧「消費されるスポーツ」を例として取り上げる。以下に外来語に下線を引いてその一部を示す。

ベーブ・ルースのようなバッターはビジネスとしてのスポーツから現れた個人としてのヒーローであった。(高等学校⑧)

メジャーなスポーツとマイナーなスポーツは、競技人口の大小も影響しているが、メディアによる階層化の方が強力である。(高等学校⑧)

「消費されるスポーツ」はスポーツとメディアについて論じた文章であり、「メディア」は 23 例、題名にも含まれている「スポーツ」は 60 例登場する。また、2 番目に数値が高い高等学校①「進化が導き出した答え」はニッチ¹⁰について論じた文章であり、「ナンバー

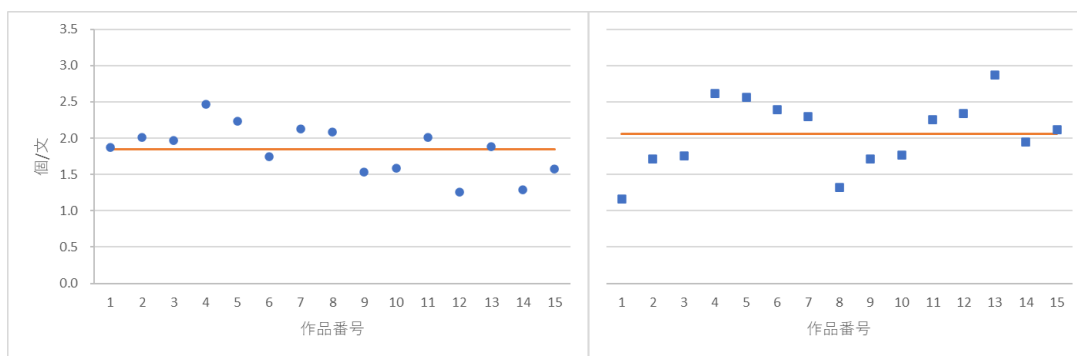
⁸ 小学校までに学習する漢字は 1026 字である。

⁹ 中学校学習指導要領（平成 29 年告示）pp.29-30

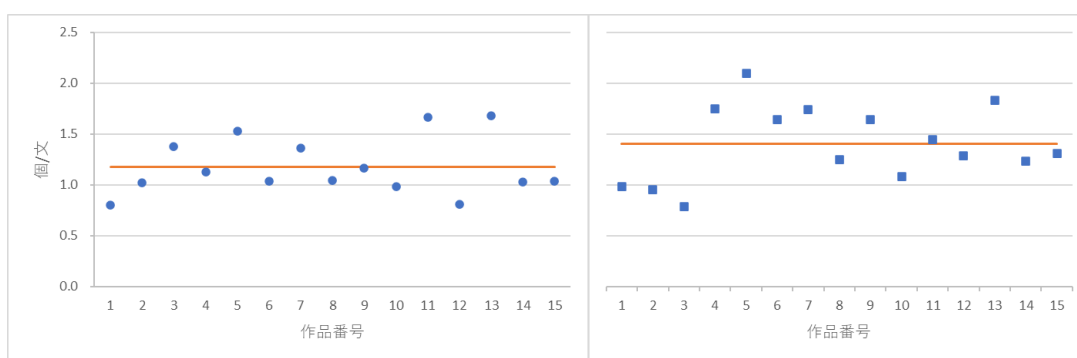
¹⁰ この作品では、「ある生物が生態系の中で占める位置。生態的地位。」のことを指す。（ジャパンナレッジ『日本国語大辞典』より引用）

1]「オンリー1」「ニッチ」という語が頻出する。つまり、この外来語率の高さは内容によるところが大きい。したがって、内容に関わらない文章の読みやすさを計る指標にはなりにくいと考えられる。

5.3 品詞について



【図7】1文あたりの動詞数の結果

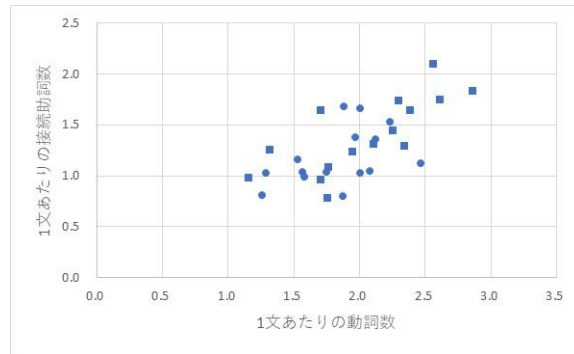


【図8】1文あたりの接続助詞数の結果

1文あたりの動詞数と1文あたりの接続助詞数については、どちらも平均値としては高等学校の教科書に掲載された作品のほうがわずかに高い値を示したが、ほぼ似たような結果となった。動詞数については、中学校と高等学校の平均値である、1文につき2つ以下程度が読みやすい文章の規範になるのではないだろうか。接続助詞については、そのほとんどが以下のように「動詞+接続助詞『て』」の形で現れているため、動詞数の影響を受けたと考えられる。

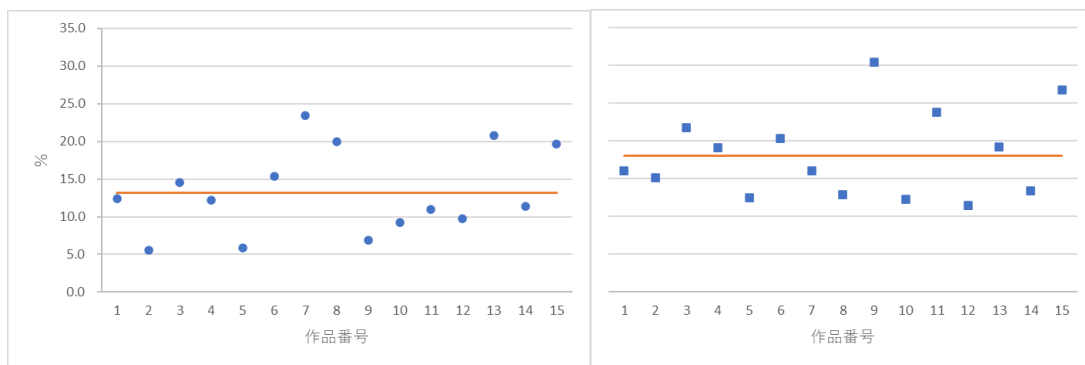
茶道という美的体験が時空を超えて伝わってきた瞬間でした。(中学校⑥)

接続助詞と動詞の関連を確認するために、横軸に1文あたりの動詞数、縦軸に1文あたりの接続助詞数を示したグラフを作成した。結果は以下ようになる。



【図 9】 1 文あたりの動詞数と 1 文あたりの接続助詞数の相関

ここで、1 文あたりの動詞数と 1 文あたりの接続助詞数の相関係数を計算すると、中学校が約 0.46、高等学校が約 0.74 となった。相関係数は、2 つのデータの関係の強さを -1 から 1 の数値で表すものだが、一般に、0.4 以上¹¹で正の相関があると見なされる。したがって、1 文あたりの動詞数と 1 文あたりの接続助詞数には正の相関があると言える。



【図 10】 接続詞率の結果

接続詞率については、高等学校の部のほうが平均値がやや高いものの、作品ごとのばらつきが大きい結果となった。これには、今回利用した形態素解析に一因があると考えられる。Web 茶まめでは主に「さらに」「しかし」「しかも」「そして」「また」などが接続詞として判定されているが、一般に文と文を繋ぐ言葉として使われている「このように」「はじめに」「一つ目は」「逆に」「だから」などといった言葉は接続詞として判定されない。このため、

¹¹ この相関係数の目安についての明確な根拠を示した論文等は見つけることができなかったが、総務省による平成 30 年度政策評価に関する統一研修で使用されたデータ活用についての資料 (https://www.soumu.go.jp/main_content/000607574.pdf) では、0.2~0.4 は「弱い正の相関がある」0.4~0.7 は「正の相関がある」0.7~1.0 は「かなり強い正の相関がある」とされている。

仮に2つの作品の筆者が同じくらい文と文を繋ぐ言葉を入れる意識を持っていたとしても、語の選択によってはそれが見えにくくなってしまった。

そこで、追加調査を行った。追加調査では、集計の範囲を接続詞のみに限定していたところを接続表現にまで広げ、接続表現率を求める。以下に計算式を示す。

n. 接続表現率 = 接続表現数 / 文の数 * 100 (%)

接続表現とは、「後続部の内容が談話の先行部や談話全体とどのような関係にあるかを示す言語的要素」¹²のことである。本研究では、日本語記述文法研究会編（2009）に示された代表的な接続表現（表3）に含まれており、かつ文頭に表れたものを集計することとした。

【表3】代表的な接続表現の一覧¹³

論理的展開 を表示する 接続表現	確定条件	だから/それで/そのため/したがって/ゆえに/よって/そこで
	仮定条件	それなら/なら/それでは/では/そうしたら/すると/そうすると/とすると/してみると/そうなると/こうなると/となると
	否定条件	さもないと/さもなければ/そうでなければ/そうでないと/そうしないと/そうしなければ
	理由	なぜなら(ば)/なぜかという/どうしてかという/というの/というの/だって
	逆接	しかし/だけど/でも/ところが/それが/それなのに/そのわりに/そのくせ/と(は) いても/とはいえ/そうは いても
加算的關係 を表示する 接続表現	添加	そして/それで/それと/あと
	累加	それに/そのうえ(に)/おまけに/そればかりか/ひいては/しかも/それも
	換言	すなわち/つまり/いいかえると/いわば/いってみれば
	例示	実際/事実/たとえば
	卓立	特に/とりわけ/なかでも
	代替	代わりに/その代わり(に)/いな/そうではなく(て)/そうじゃなくて/というより(も)/どうか/むしろ
対等な關係 を表示する 接続表現	並列的提示	およみ/ならびに/かつ/なおかつ
	選択的提示	または/ないし(は)/もしくは/あるいは/それとも
話題の展開 を表示する 接続表現	転換	ところで/さて/それにしても/話は飛ぶけど/話は変わるけど/それはそうと/それより
	列挙	第1に/第2に/1つめに/2つめに/まず/はじめに/最初に/つぎに/つづいて/ついで/それから/また/さらに/ほかに/最後に
	対比	それに対して/反対に/反面/逆に/一方(で)/他方(で)
	まとめ	このように/以上(のように)/要するに/こうして/結局/このようにして
	補足	ただし/ただ/もっとも/ちなみに/なお
	無視	いずれにしても/どっちみち/どっちにしても/何にしても/何はともあれ/とにかく

¹² 日本語記述文法研究会編（2009）p.49

¹³ 日本語記述文法研究会編（2009）p.58

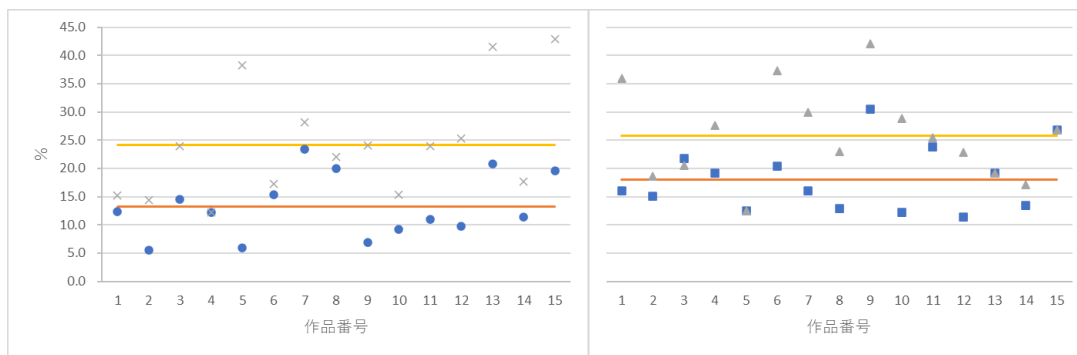
前述の通り、一部の接続表現は Web 茶まめでは分解されてしまうため、この調査には、Web 茶まめや Python、Excel を用いた自動的な処理を行うことはできなかった。そのため、OCR で作成した電子テキスト内で検索を行い、目視で確認して集計した。なお、「たとえば/例えば」「いいかえると/言い換えると」などのように、日本語記述文法研究会編（2009）による一覧ではひらがなで示されていたが、漢字表記の可能性もあるものについては、ひらがな表記と漢字表記の両方が検索に該当するようにした。また、「第 1 に」のように数字が入るものは、数字が何であっても検索に該当するようにした。

集計結果は以下の通りである。

【表 4】 接続表現数と接続表現率の結果

中学校	①伝言	②学ふ力	③動物園	④花火	⑤自立	⑥サハラ	⑦ガイア	⑧黄金	⑨一〇〇	⑩紙	⑪ハト	⑫晚餐	⑬富士山	⑭モアイ	⑮人間		平均	標準偏差
接続表現数	16.0	13.0	23.0	6.0	13.0	9.0	18.0	11.0	21.0	10.0	24.0	26.0	22.0	14.0	24.0			
接続表現率	15.2	14.4	24.0	12.2	38.2	17.3	28.1	22.0	24.1	15.4	24.0	25.2	41.5	17.7	42.9		24.2	9.5
高等学校	①進化	②真実	③展示	④共同	⑤デモ	⑥野生	⑦女性	⑧消費	⑨メディア	⑩ロボット	⑪対話	⑫多様	⑬言語	⑭具体	⑮人類		平均	標準偏差
接続表現数	36.0	16.0	16.0	13.0	4.0	22.0	15.0	25.0	29.0	26.0	16.0	8.0	10.0	14.0	19.0			
接続表現率	36.0	18.6	20.5	27.7	12.5	37.3	30.0	22.9	42.0	28.9	25.4	22.9	19.2	17.1	26.8		25.8	7.8

接続表現は基本的に接続詞を含むものであるため、先ほどの接続詞率のグラフに接続表現率の結果を重ねて示すと以下ようになる。なお、バツ印が中学校の部の接続表現率、三角形が高等学校の部の接続表現率である。



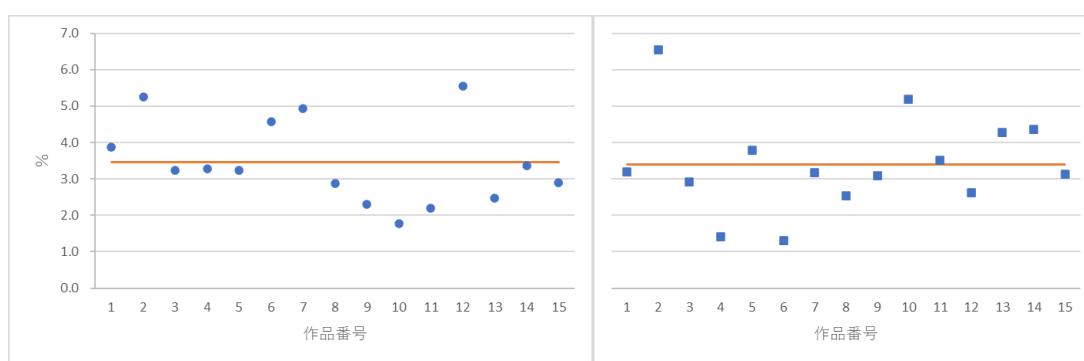
【図 11】 接続詞率と接続表現率の結果

やはり作品ごとのばらつきが大きい結果となった。さらに、接続詞率の段階ではあった中学校の部と高等学校の部の平均値の差が、接続表現率になるとほとんどなくなった。また、この接続表現率と接続詞率の差は、接続詞には含まれない、「このように」「はじめに」「一つ目は」「逆に」「だから」といった比較的緩やかなつながりがどのくらいあるかを示すことになるが、それも中学校の部と高等学校の部でまとまった傾向は見られなかった。したがって、接続詞や接続表現をどのように使うのかということは、書き手の癖によるところが大きいと考えられる。

筆者の所感になるが、接続詞や接続表現が多い文章は論理的で硬い印象を受けた。その一

因としては、接続詞や接続表現は書き言葉的であることが考えられる。しかし、仮に接続詞や接続表現が多い文章からそれらをすべて取り除いたとしても、より読みやすい文章にはならないだろう。逆に、子ども向けの易しい文章に接続詞や接続表現をむやみに入れても、柔らかい文章と硬い印象の接続詞や接続表現が馴染まず、より読みやすくなることはないだろう。つまり、接続詞率や接続表現率は、その値が大きいから読みにくくなるというのではなく、ある程度難しい内容を論理的に説明しようとしたときに高くなるのではないだろうか。

5.4 指示語について



【図 12】 指示語率の結果

この項目も作品ごとにばらつきがある結果となった。最も指示語率が大きい値となった高等学校②「真実の百面相」の一部を以下に抜粋する。

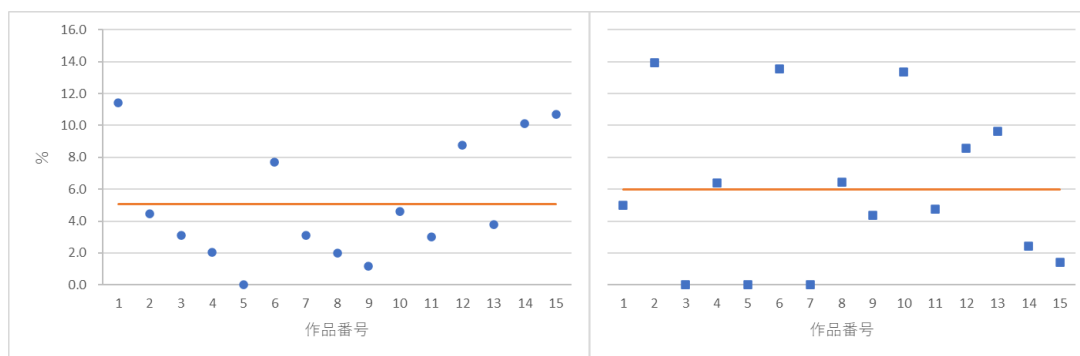
だからこの比喩の入り口に立ち戻ってみよう。山道の人影はそれが私に見えたとき真実そこにあったのではないか。そのとき世界は真実そのような姿であらわれたのではあるまいか。奇妙な形の岩は白昼近よって見ればまごうことなく岩の姿であられる、それと同様に薄暗がりの遠目には時に人の姿であられる、そういった種類のものではないか。その岩もまた、そしてその岩を含む世界もまた百面相であられるのである。(高等学校②)

この作品は内容も哲学的で難しいが、このように指示語が多用されていることによって、前後の文脈を注意深く追う必要があり、読み取るのが難しくなっていることがわかる。また、指示語が多いだけでなく、指示語とそれが示す語が離れているときも、読みにくくなる。これは、文中の指示語が今何を指しているのかを短期的に記憶しながら読む必要があるからだと考えられる。この要素については、人間の言語処理能力に関わってくる問題だと思われる。

しかし、全体の平均値としては中学校が 3.5%、高等学校が 3.4%と近い値になり、明確

な差は見られなかった。そもそも、自立語を指示語に置き換えすぎると文意が通じなくなるため、指示語率には上限があると考えられる。そのため、全体の値が小さくなり、あまり差が出なかったのではないだろうか。

5.5 終助詞「か」で終わる文について



【図 13】文末「か」率の結果

この項目も作品ごとのばらつきがある結果となった。最も大きい値となった高等学校②「真実の百面相」の一部を引用する。

青磁の壺や翡翠の帯留めは、どの向きからどのような近さで見たときにその本当の色を見せると言うのだろうか。いやこういう場合にもその折々のさまざまな色のすべてが本当の色なのであって、特定の一つの色が他を差しおいて真実の色になるわけではあるまい。(高等学校②)

そこでこの座席で聞く音こそ本当の音なのだ、といえるような座席があるだろうか。座席によって料金が違うのは高い席ほどより真実の音が聞けるからだろうか。そうではあるまい。(高等学校②)

筆者がこの項目を集計したのは、「では、いったい何のためにハトは頭を静止させているのだろうか。」(中学校⑪)のように、読者に問いかけて興味を誘い、説明の導入とする文が多いほど読みやすい文章であると考えたからであった。しかし実際には、上に引用したように、「座席があるだろうか。(いや、ない。)」というような、反語的な表現に含まれる終助詞「か」も集計に含まれていた。反語的な表現には硬い印象があり、文意を注意深く読み取る必要があるため、読みやすいとは言いがたい。このように、読みやすさに影響する要素と読みにくさに影響する要素が混在してしまったため、読みやすさを計る指標としては適切とは言えなかった。

5.6 本研究で処理できなかった要素

本研究では、比較的読みやすい文章と比較的読みにくい文章を計量的に分析し比較することで、文章の読みやすさに影響を与える言語的特徴とはどのようなものか数値的に見ることができるとはならないかというところを出発点としていた。ここまで、文章の読みやすさに影響を及ぼすと考えた項目ごとに考察してきたが、本項では、この手法では処理できなかった要素と、この手法そのものが適切だったかということについて述べる。

分析した11項目のうち、文の長さ（一文当たりの自立語数、一文当たりのモーラ数）、読点の頻度、和語率、漢語率については、中学校の部と高等学校の部それぞれでまとまった傾向を示したため、比較的読みやすい文章と比較的読みにくい文章で数値的にどの程度差があるのか説明することができた。しかし、外来語率、一文当たりの動詞数、一文当たりの接続助詞数、接続詞率、接続表現率、指示語率、文末「か」率については、中学校の部と高等学校の部それぞれでまとまった傾向を示さず、作品ごとのばらつきが大きい結果となった。

外来語率については、内容によるところが大きかったのは前項で述べた通りだが、外来語の有無が読みやすさに全く影響がないとは言えないと筆者は考えている。「インセンティブ」（高等学校⑤）や「シノニム」（高等学校⑬）のような、一般的にあまり馴染みがない外来語は、高等学校に掲載された作品に比較的多かったように感じられた。今回の手法では、専門的な外来語と「スポーツ」のような既に一般的に使われている外来語、つまり、読みやすさに影響する外来語と影響しない外来語を区別して集計できなかったが、分けて集計することができれば、傾向が見えてくる可能性がある。

文末「か」率については、今回の手法では、想定していた要素を抽出しきれなかった。また、導入のための疑問文以外にも、読者に興味を持たせる文として、「しかし、この機会に、少しでも気を鎮めて、『学力が低下した』とはどういうことなのか、考えてみましょう。」（中学校②）や、「たくさんの手が描かれているが、試しに、その一つ一つのポーズを君もまねてみよう。」（中学校⑫）のように、読者に誘いかける文があることが分かった。このような文は高等学校の部より中学校の部に多いと感じたため、読みやすさに影響する要素であるかもしれない。

また、他にも今回の手法では処理できなかったが、調査の過程で読みやすさに影響すると感じた要素としては、語の特殊な用法と、比喩表現である。

語の特殊な用法とは、以下のようなものである。

そして、この動きは、自分たちだけの「国民」を再び取り戻すという排他的で後ろ向きプロジェクトを伴っている。（高等学校⑤）

このことを、普遍的価値を示す「クジラ」と、人間にとっての価値を示す「鯨」に置き換えつつ考えてみたい。（高等学校⑥）

「国民」「クジラ」「鯨」はどれも一般的な語だが、ここでは鍵括弧が付けられ、この文章の中だけの意味を与えられている。そのため、語そのものの辞書的な意味だけではなく、文脈上の意味も読み取る必要があり、読みにくい。このような書き方は高等学校の部に多いように感じられた。

次に、比喩表現について述べる。比喩表現は、読者にとって未知のものや理解しにくい抽象的なものを、既知のものやわかりやすい具体的なものに例えるものである。つまり、基本的には文章をわかりやすくするためのものだが、中にはその比喩が読みにくい文章もあった。実際、筆者が「広告の形而上学」¹⁴を教材とした高校1年生の授業を見学した際には、多くの高校生が比喩の読み取りに苦戦していた。例えば、以下のような比喩である。

マルクスはどこかで、商品世界の中における貨幣の存在は、動物世界の中でライオンやトラやウサギやその他全ての現実の動物たちと相並んで「動物」なるものが闊歩しているように奇妙なものだ、と書いている。

この部分は、「『動物』なるもの」というように、語の特殊な用法も含まれた比喩表現であるため非常に難しい。この比喩は文章の冒頭にあるため、あえて難しい比喩を最初に提示して興味を持たせ、それを説明することで筆者の意見に繋げていくという工夫だと思われるが、興味を持たせるといっても、「読みにくい」「難しい」という感覚を抱かせる原因になっているように思われた。今回の手法では処理することができなかったが、文章の読みやすさを考える上では必要な要素であると考えられる。

最後に、「比較的読みやすい文章と比較的読みにくい文章を計量的に分析し比較する」という今回の手法そのものの是非について検討する。個々の項目の処理については以上に述べた通りだが、全体として、文章の読みやすさについて計るにはやはりこの手法だけでは限界があると感じた。読みにくい文章と比較して、より読みやすい文章は文の長さが短く、読点の頻度が多く、和語率が高く漢語率が低い、ということは今回の手法で見ることができた。しかし、それ以外の要素は、それぞれの作品によるところが大きい。外来語率、一文当たりの動詞数、一文当たりの接続助詞数、接続詞率、接続表現率、指示語率、文末「か」率が読みやすさに全く影響しないというわけではないと思われるが、全ての読みやすい文章で一貫した傾向が見られる要素ではないようである。調査を進めていく中で感じたことだが、読みにくい文章は読みにくさに影響する要素を全て持っているわけではない。複数の要因が絡み合って文章の読みやすさを決めているため、今回の手法だけでは処理できない部分がある。

¹⁴ 岩井克人 (2022) 「広告の形而上学」『現代の国語』、東京書籍

6. おわりに

本研究は、文章の読みやすさに影響を与える言語的特徴とはどのようなものか、また、実際に数値としてどのように違いが現れるのかを明らかにすることを目的とした。その上で、比較的読みやすい文章として中学校の教科書に掲載された作品を、比較的読みにくい文章として高等学校の教科書に掲載された作品を設定し、文章の読みやすさに影響すると考えられる要素について両者を統計的に分析し、比較を行った。

結果として、読みやすい文章は読みにくい文章と比較して文の長さが短く、読点の頻度が多く、和語率が高く漢語率が低いということは共通して言えるが、その他の要素については、読みやすさへの影響はあると考えられるものの、一貫した傾向は見られないということが明らかになった。また、語の特殊な用法や比喻表現など、今回の手法では処理できなかったが、文章の読みやすさに影響する要素は他にもあることがわかった。

文章の読みやすさにはさまざまな要素が複雑に絡み合っており、統計的な処理が難しいものもある。処理の方法を再検討するだけでなく、感覚的な「読みやすさ」「読みにくさ」をどのように数値的に表せるか、更なる要素を考えることも今後の課題である。

参考文献

- ・遠藤仁・大谷航（2014）「小学校説明文教材の構造と表現に関する基礎的研究（1）」『宮城教育大学紀要』48巻、宮城教育大学
- ・遠藤仁・大谷航（2015）「小学校説明文教材の構造と表現に関する基礎的研究（2）」『宮城教育大学紀要』49巻、宮城教育大学
- ・大川孔明（2019）「平安鎌倉時代の文学作品の文体類型—多変量解析を用いて—」『計量国語学』31巻8号、計量国語学会
- ・樺島忠夫・寿岳章子（1965）『文体の科学』、文功社
- ・計量国語学会編（2009）『計量国語学事典』、朝倉書店
- ・嶋中道則ほか76名（2022）『現代の国語』、東京書籍
- ・堤智昭・小木曾智信（2023）「複数の UniDic 辞書による形態素解析支援ツール『Web茶まめ』の実装と運用」『情報処理学会誌』64巻3号、情報処理学会
- ・似内寛（2020）「教科書等、教材に使用される語彙の研究動向について」『東北福祉大学研究紀要』44巻、東北福祉大学
- ・日本語記述文法研究会編（2009）『現代日本語文法7 第12部談話 第13部待遇表現』、くろしお出版
- ・堀川直義（1957）『文章のわかりやすさの研究』、朝日新聞社

参考 URL

- ・言葉の Q&A 009、文化庁 広報誌 ぶんかる（最終閲覧日：2023年11月6日）
https://www.bunka.go.jp/prmagazine/rensai/kotoba/kotoba_009.html

- ・ ジャパンナレッジ Lib（最終閲覧日：2024年1月21日）
<https://japanknowledge.com/library/>
- ・ 小学校学習指導要領（平成29年告示）、文部科学省（最終閲覧日：2023年11月6日）
https://www.mext.go.jp/content/20230120-mxt_kyoiku02-100002604_01.pdf
- ・ 政策評価に関する統一研修（地方研修）の概要、総務省（最終閲覧日：2024年1月21日）
https://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/h30chiho.html
- ・ 中学校学習指導要領（平成29年告示）、文部科学省（最終閲覧日：2023年11月6日）
https://www.mext.go.jp/content/20230120-mxt_kyoiku02-100002604_02.pdf
- ・ 「文字認識」機能の使い方、LINE みんなの使い方ガイド（最終閲覧日：2024年1月14日）
<https://guide.line.me/ja/services/scantext.html>
- ・ OCRとは、メディアドライブ（最終閲覧日：2023年11月5日）
<https://mediadrive.jp/technology/ocr.html>
- ・ openpyxl（最終閲覧日：2023年11月5日）
<https://openpyxl.readthedocs.io/en/stable/>
- ・ UniDicとは、「UniDic」国語研短単位自動解析用辞書（最終閲覧日：2024年1月19日）
https://clrd.ninjal.ac.jp/unidic/about_unidic.html
- ・ Web茶まめ（最終閲覧日：2023年11月4日）
<https://chamame.ninjal.ac.jp/>

調査資料

- ・ 相澤秀夫・野矢茂樹ほか76名（2021）『新しい国語2』、東京書籍
- ・ 安藤宏・門屋敦・紅野謙介・河野龍也・五味淵典嗣・清水良典・関口隆一・橘直弥・仲島ひとみ・服部徹也・マツダ顕子・吉田光（2023）『論理国語』、筑摩書房
- ・ 岩崎昇一・三浦和尚ほか14名（2023）『新 論理国語』、三省堂
- ・ 大塚常樹・難波博孝ほか20名（2023）『標準論理国語』、第一学習社
- ・ 甲斐睦朗ほか29名（2021）『国語2』、光村図書
- ・ 北原保雄ほか（2023）『論理国語』、大修館書店
- ・ 児玉忠・植山俊博・丹藤博文ほか45名『伝え合う言葉 中学国語2』、教育出版
- ・ 嶋中道則ほか76名（2023）『新編論理国語』、東京書籍
- ・ 嶋中道則ほか77名（2023）『新編論理国語』、東京書籍
- ・ 坪内稔典ほか16名（2023）『論理国語』、数研出版
- ・ 中島国彦・渡部泰明・安藤信廣ほか12名（2023）『精選 論理国語』、明治書院
- ・ 中洩正堯ほか36名（2021）『現代の国語2』、三省堂

【資料2】詳しい集計方法

・延べ語数

記号を除いた全ての語の総数を数える。

で/は/、/その/よう/な/ロボット/は/作れる/だろう/か/。(高等学校⑩)

・自立語数

自立語とは、単独でも文節を構成できる語のことを指す。本発表では、延べ語数から助詞、助動詞、接頭辞、接尾辞の数を除いたものとしている。

で/は/、/その/よう/な/ロボット/は/作れる/だろう/か/。(高等学校⑩)

・文の数

引用文の中にあるもの以外の「。」「?」「!」で区切る。例えば、「○○。『○○。』○○。」のような場合、「○○。 / 『○○。』○○。」と区切り、2文と数える。

「私はもう知るべきことはみな知っているのです、これ以上学ぶことはない。」と
思っている人には「学ぶ力」がありません。 / このような人が、本来の意味での「学
力がない人」だと私は思います。(中学校②)

・読点の数

文章中のすべての「、」の数を数える。

・モーラ数

Python 上では、「Web 茶まめ」出力結果の「発音形出現形」のセルに含まれる文字列か
ら「ヤ」「ユ」「ヨ」「ア」「イ」「ウ」「エ」「オ」を削除し、文字数を数えるという処理を行
っている。

しかしこの問いは、単純な技術の問題ではない。(高等学校⑩)

→シカシコノトイワ タンジュンナギジュツノモンダイデワナイ

→シカシコノトイワタンジンナギジツノモンダイデワナイ (25 拍)

・和語数、漢語数、外来語数

「Web 茶まめ」出力結果の「語種」のセルがそれぞれ「和」「漢」「外」である数を数え
る。

- ・動詞数、接続助詞数、接続詞数

「Web 茶まめ」出力結果の「品詞」のセルを参照し、それぞれの数を数える。

- ・指示語数

「この、これ、こんな、それ、その、そんな、あれ、あの、あんな」の総数を数える。

- ・文末の終助詞「か」の数

引用文の中のものを含めない。例えば、「○○は『○○だろうか。』と言うだろうか。」という文があった場合、文末の終助詞「か」の数は1となる。