

文章の読みやすさと漢字含有率の関係

名前 川瀬皓暉 学生番号 23B30300
東京工業大学情報理工学院

1. はじめに

読みにくく理解に時間がかかる文章ほど文章中の漢字の割合(漢字含有率)が多いのかという疑問に関して取り上げた。文章を読みやすいものと読みにくいものに分け、漢字含有率を調べ、文章の読みやすさと漢字含有率の関係について考察した。

2. 方法

青空文庫からいくつかの小説を取り上げpythonで文字種ごとに数を数えるプログラムを組み、プログラムを実行して文章中の漢字含有率を調べた。また、比較的読みやすい文章として掲示板(5ch)における書き込みを取り上げ、小説と比較した。漢字含有率を求める際には句点や読点などの記号の数を無視し、漢字数を平仮名と片仮名、漢字の合計数で割って調べた。

3. 結果

表1に文章ごとの漢字含有率を示した。掲示板の漢字含有率は比較的読みやすいと個人的に感じる作品のものと同程度になった。小説の中でも吾輩は猫であるや山月記などの個人的に読みづらいと感じた文章の方が痴人の愛などの個人的に読みやすいと感じた文章に比べて、漢字含有率が高くなった。憲法やドグラ・マグラなどの特に解とされている文章や小説に関してはほかの文章に比べても漢字含有率が高くなっていた。

表1: 文章ごとの漢字含有率

題名	文字数	漢字数	漢字含有率(%)
吾輩は猫である	269,292	89,630	33.3
日本国憲法	8,680	4,275	48.3
痴人の愛	151,319	41,840	27.7
こころ	149,769	44,368	29.6
三四郎	154,747	44,610	28.8
学問のすすめ	81,400	28,610	35.1
ドグラ・マグラ	381,902	145,136	38.0
掲示板	706413	210042	29.7

4. 考察

掲示板の文章の漢字含有率が想定よりも高くなったが、掲示板の書き込みは口語体によって記述されているものが多く、完全な文ではなく名詞だけの文が多かったため、漢字の割合が高くなったと考えられる。

難しいとされる文章ほど漢字含有率が高かったため、漢字含有率と文章の読みやすさには関係があるといえる。一般的に読みやすいとされている文章においては平易な文で表現されるような単語を、難しいとされる文章においてはより厳密に読み手に伝えるために漢字を用いた表現で書いているためであると考えられる。

しかしながら、漢字含有率を下げれば下げるほど文章が読みやすくなるというわけではない。平仮名だけの文章では同音異義語の区別や文節、単語の区切りの解釈が難しくなるため、かえって読みにくいものとなってしまふ。

そのため、読みやすい文章を書くには、単語の区別を考えることなくスムーズに行えるように最低限度は漢字を用いることが大切であると考えられる。

5. おわりに

問題点は文章の読みやすさは主観でしかないという点である。文章の読みやすさは個人の感覚に依存しているため、一般的な傾向を調べるためにアンケートを行い調べる必要がある。また、解析に用いた小説は著作権切れの古いものであり、古い文章のほうが漢字含有率が高い傾向にあるので、現代の小説との比較も必要である。読みやすい文章として掲示板の書き込みを選んだが、より大衆向けの新聞などの媒体から文章を選ぶほうが良いと考えられる。

文献:

梶原滉太郎. (1982). 新聞の漢字含有率の変遷——明治・大正・昭和を通じて——. 国立国語研究所報告, 71, 209-236.
菅野倫匡. (2021). 語種の観点から見る漢字含有率の安定要因 芥川賞作品の分析を通して. 計量国語学, 32(8), 479-495.