２０２３年度

修 士 学 位 請 求 論 文 要 旨

ニュース記事における「やさしい日本語」の考察
―NEWS WEB EASY を一例として―

国際日本学研究科　国際日本学専攻

日本語学・日本語教育学領域

4911214002

新井　智大

やさしい日本語は，「難しい言葉を言い換えるなど，相手に配慮したわかりやすい日本語のこと」（出入国在留管理庁・文化庁，2020）とあるように，日本語非母語話者に対して行われる言語調整のことである。減災時の使用に端を発するやさしい日本語は，書き換え方針が公開（弘前大学人文学部社会言語学研究室，2005）されたり，書き換えの参考情報が提案（庵，2009）されたりしているものの，特定の使用に留まっているのが現状である。日本語教育場面や日本語非母語話者の生活支援にも有効に活用していくためには，やさしい日本語の言語的特徴を把握し，書き換え方針を精緻化する必要があると考える。そこで，本研究ではやさしい日本語の言語的特徴を知るために，やさしい日本語の１例として，「NEWS WEB EASY」（以下，NWE）とその書き換え前の記事である「NHK NEWS WEB」（以下，元記事）を分析し，リーダビリティ研究の知見を援用しながら，その言語的特徴を整理する。具体的な研究課題は以下の４つである。（１）元記事と NWE では，言語計量的にどのような違いがあるか，（２）NWE は元記事より文章の難易度がどの程度下がったか，（３）NWE は元記事より語彙難易度がどの程度下がったか，（４）NWE に使用されている中級以上の難易度の高い語にはどのようなものがあるか，である。

研究課題に解答するために，本研究では NWE と元記事のパラレルコーパスを独自に作成した。具体的には，WEB 茶まめ（https://chamame.ninjal.ac.jp/）を使用し，形態素解析，及び形態論情報のアノテーションを行い，*jReadability*（https://jreadability.net/sys/ja）でリーダビリティ値のアノテーション，*jReadability* に搭載されている「日本語教育語彙表」による語彙難易度の同定を行った。さらに，記事ごと，文ごとにアライメントを行い，書き換え前後を比較可能にした。そのコーパスを利用した分析結果が５章から７章である。

５章では，WEB 茶まめと *jReadability* を使用した，計量的な分析を行った。その結果，述べ語数は約４分の１となっており，情報の量が減らされていること，平均文長は書き換えによって，16.4 語減り，文長が短くなっていることがわかった。また，品詞構成は，文や文章の硬さについて述べた先行研究の結果とほぼ一致し，NWE では書き換えによって，軟らかくなっていることが確認できた。さらに，やさしい日本語の文の硬さの指標である文名詞密度を計算した結果，NWE の数値は 8.4 であり，依然として，アカデミックな自然科学の文章と同程度の硬さを保っていることがわかった。

jReadability で文章難易度を判定したところ，リーダビリティ値が 1.40 から，2.98 と増加し，中級後半語レベルまで難易度が下がっていた。文章難易度を下げた要因は，和語率，漢語率，平均文長であった。しかし，NWE の想定読者は中級前半の学習者であることから，文章難易度が想定読者のレベルまで十分に下がっていないこと，平均文長は短くなっていたが，依然として，複文形式は散見されることなど，課題がみられた。さらに，「日本語教育語彙表」で語彙難易度を同定し，語彙難易度構成を確認したところ，中級後半語が初級語に書き換えられ，難易度が下がっていることが確認された。しかし，NWE の初級語の比率は約 56％であり，読者のレベルが中級前半であることから，読者の文章中の既知語率は初級語の比率と考えられる。そのため，小森他（2004）の知見を参考にすると，想定読者がテキストの言語情報だけでは読むことが難しいことがわかった。

５章で，中級後半語が初級語に書き換えられ，語彙難易度がやさしくなっていることがわかったため，６章では，パラレルコーパスのアライメントを利用し，中級後半語の名詞の書き換えの実態を確認した。具体的には，書き換えのレベルを，語・名詞句レベル，動詞句レベル，節以上レベルの３つのレベルに分け，実態を確認した。その結果，書き換えレベルの中で，語・名詞句レベルでの書き換えが最も多く，全体の約 53%が語・名詞句レベルで書き換えられていた。しかし，語･名詞句レベルでは，すべての語が初級の範囲で書き換えられているわけではなかった。中級後半語や，中級後半語を含む派生語，複合語の類義語が初級に存在しなかった場合，書き換えが困難であった。そのような場合，文章中から語の意味を具体化し，その具体化した意味で書き換えを行ったり，中級以上の語である類義語の中で，構成漢字が旧日本語能力試験３・４級である語を使用したりして，書き換えを行っていた。動詞句レベルでは，文の構造を理解しやすくするために，実質的な動作を含む語を述語に移動させた書き換え，抽象的な被修飾名詞を動詞とともに１つの単位と考える書き換え，事実や情景を描写するために必要である語を，文構造の核に入れるような書き換えが行われていた。節以上レベルでは，書き換え前の文に記述されているイベントを，時間的隣接性を利用して，読者の行動や経験しうる状況に近づけるように書き換えていた。また，このような隣接性を利用した書き換え以外の傾向は，取り上げることができなかった。このことから，節以上レベルでは，書き手によって書き換えが左右されていることが示唆された。

また，中級後半語の名詞の約 22%は，そもそも書き換えられていなかった。語・名詞句レベルの中級以上の語を使用し，書き換えられたものと合計すると，約 40%が初級語に書き換えられていなかった。これらのことから，初級語のみで記述することは，やはり難しいことが確認された。

６章で中級以上の語を使用する際に，想定読者が漢字知識を考慮し，構成漢字が簡単な語を使用している可能性が示唆された。そこで，７章では，NWE に出現している中級以上の２字漢字熟語について，構成部分と語の意味の関係性を検討している意味論的透明性の研究の知見を援用し，探索的に検討を行った。その結果，NWE に出現した中級以上の語は，漢字と語の意味が結び付いている透明な語が最も多いことが確認された。延べ語数では，構成漢字と，語の意味との結びつきがある透明，片透明の数値を合わせると，全体の 90%を超えていた。しかし，透明性の議論は，両方の構成漢字を知っていること，語の意味を知っており，語の意味から分析的につながりがあるかを判定していることの２点を前提としている。したがって，NWE の想定読者を考慮すると，判定結果をそのまま使用することはできなかった。そこで，本節では，想定読者の漢字知識を旧日本語能力試験の３・４級とし，漢字を既知，未知に分け，文の意味をとれるほどに語の意味推測が可能かどうかを検討した。その結果，文の意味をとるために十分なほど，意味推測できる可能性のある考えられた語は，512 語中 80 語となった。結果として，透明な語は使用されているが，想定読者が語の透明性という性質を活かし，意味推測可能であると考えられる語はあまり多くないことがわかった。