

推敲支援における読みやすさの評価**

乾 裕子

丸元聡子

hinui@ibs.or.jp

smarumot@ibs.or.jp

(財) 計量計画研究所

1. はじめに

近年、ワードプロセッサには推敲支援および読みやすさ評価のツールが搭載されている。現在、評価結果の多くは、文長・漢字含有率、あるいは読点間の文字割合(句読点間距離)といった測定指標を数値で示したものである。現状ツールの問題点を下記に挙げる。

- ・測定指標が少なく、読みやすさを評価するには不十分である。
- ・測定指標と評価値との関連、あるいは測定指標そのものに対する説明が十分でない。
- ・評価結果を示されたあと、ユーザは何をするべきかわからない。

単に一文の文字数が表示されるだけでなく、1) 文が長くなる要因は何か、2) 文意を変えずに短くするには、どのように修正すれば良いか、この二点を提示することによりユーザにとって使いやすいツールが実現できる。推敲支援システムは、評価だけでなく修正方法を提示することによって効用がある。

以上の背景を踏まえ、本研究の目的は文を読みにくくする表層の現象を分析し、その結果から測定指標および修正規則を抽出することである(図1)。また、分析の際に複数の観点で現象を捉えることにより、測定指標間の関係から修正規則を作成する。本稿では、それらの測定指標の一つである文長を中心に報告する。

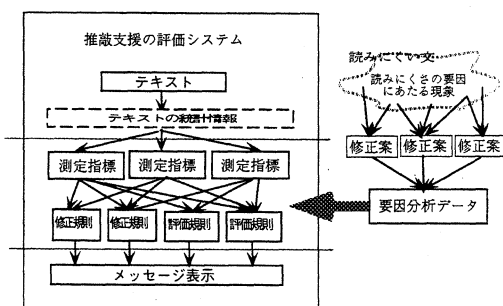


図1 読みやすさ評価システムの構成

尚、本研究での読みやすさには、文字のポイント・行間の幅などを指標にする見易さ(legibility)を含まない。また、段落間の関係に関する読みやすさも原則として扱わない。

2. 関連研究との比較

1940~50年代、Fleschの研究から読みやすさの尺度化

が盛んになった。Fleschの公式では、文長・シラブル数から難度を、人称語の数から易度を測定することで読みやすさをモデル化している[1]。シラブル数を数える根拠は、-ity, -tionなどの接尾辞によって、単語が長くなり抽象的で難解な意味になるという仮説に基づく。人称という指標は、人に関して書かれた文には興味を覚えやすいというFlesch独自の仮説による。Fleschの与えた影響は強く、従来研究でも文長および漢字含有率(シラブル数に相当する)が測定指標とされている[2][3][4][6]。しかし、読み手のレベルや、評価対象テキストの分野を特定しないため、読み手の立場によっては直観的な読みやすさと難易のスコアが必ずしも一致しない。

日本では、堀川が大規模な調査実験に基づき読みやすさの尺度化を試みた[2]。これは、義務教育修了者ならば共通にわかりやすいと感じるレベルがあるという仮説をもとに行われている。1) 文長 2) 漢字含有率 3) 構文の複雑さの数、以上三点の指標に関して読みやすい文の条件を示している(1:一文50~200字, 2:20~40%, 3:100字あたり2.5個程度)。文章の種類(事実的文章・文学的文章)によって読みやすさが変わることに着目した点が重要である。

堀川の仮説に対し森岡は、読み手・読まれる対象・場の違いによって、読みやすさは相対的に変化すると説いた[3]。尺度化の際に一般的な読みやすさの基準を設定することは難しいと考え、読みやすさを可読性とみなした。学力レベルが低くても読むことができる文を読みやすいと定義している。雑誌調査では、外来語・主語となる抽象名詞・翻訳漢文体などの指標にも着目したが、実際の測定指標は文長および漢字含有率である。

言語処理の分野では、実験にもとづく評価研究が行われ始めた。高橋らは、対象テキストをマニュアルに限定して評価モデル式を作成している[6]。また、富士は機械翻訳システムを評価するために、心理学実験から翻訳精度向上のための指標を取り出している[7]。これら二つの研究態度は、対象を限定する森岡の立場に近い。読みやすさの尺度化を言語処理に応用するとき、堀川の仮説は粗すぎるためである。しかし、森岡の可読性という立場だけではアプリケーションとしての効果が薄い。したがって、言語処理には限定した対象範囲での読みやすさ評価が適当である。本研究でも対象テキストをビジネス文書に限定する。対象分野の決定理由については後述する。

言語処理における上記二つの研究では、いずれも読みや

**本研究は、平成8年度委託研究(委託元:富士通株式会社)「日本語文書可読性に関する調査研究」の成果に基づく。

すさに関する内省的仮説をもとに実験を計画している。精度のよい実験を行うためには必要だが、多様な測定指標を得られない可能性がある。そこで、本研究では仮説を立てずに読みにくい文の修正作業を行うことによって、ボトムアップに測定指標・修正規則を作成する。修正作業は、校正者および言語学・言語処理の研究者に依頼した。これについても後述する。

3. 作業手順

下記の手順で作業を進める。

- 1) 言語学および言語処理の専門家に依頼して文書修正作業を行う。
- 2) 1の作業結果から要因分析データの作成方法を検討し、作成する。
- 3) 分析結果から、測定指標およびパラメータを抽出する。
- 4) 3の測定指標・パラメータをもとに実験用文書を作成し、測定指標やパラメータ値の妥当性を検証するための調査実験を行う。

以下、各作業の手順について述べる。

3. 1 修正案の作成

1) 対象テキストの限定

われわれは、読みやすさの感じ方が次の三点、1) 読み手のレベル、2) 対象の内容、3) 表現手段（新聞、レポート、絵本の違い）によって決まると考える[8]。一方、書き手側から見ると、小説・随筆のように読みにくさが文章の個性になる場合がある。本研究では、読みやすさを評価されるべき文章のうち、ワープロで作成される機会の多いビジネス文書を対象とする。対象としたテキストは、ビジネス文書作成例[5]からの2文例、実際の報告書からの12文例である。これらの規模は、14文書 281文（12372字、6067形態素）である。

2) 対象文書の修正担当者を限定する

修正担当者は、校正者（7名）および言語学（20名）・言語処理（6名）の研究者に限った。理由は、1) 構文の読みにくさを指摘できること、2) 指摘の際に、適切な用語で説明できること、の二点に拠る。

3) 修正の方法

担当者に、①修正手順、および修正対象テキスト（各4文書）の②電子化版（修正用）③印刷版（参照用）を送付し、下記の注意点にしたがって作業をしてもらった。

- 1) 対象文書の印刷版を読んで、修正すべき箇所を指摘する。
- 2) 指摘箇所に対し、内容を変えない程度で読みやすいと感じる文章に直す。

表2 要因分析データ

修正案から取り出したデータ			分析者による付加情報			
1	2	3	4	5	6~n	
指摘箇所	修正例	コメント	修正方法	修正操作対象	現象1	現象2
はかるための、	はかるために、	ははかり先が曖昧なので言い換え。	置換	連体修飾	係り先が不明	連体修飾
建築物、敷地	建築物・敷地	単語の並置は中点が読みやすい	置換	読点	並列表現	
しかしながら	しかしながら、	接続詞の後には読点があると読みやすい	挿入	読点	読点の不足	接続詞
ここで別途	ここで	用語選択が適切でない	削除	副詞	冗長な副詞	
したがって	したがって、	接続詞の後には読点があると読みやすい	挿入	読点	読点の不足	接続詞
変換することで、	変換することである。	一文⇒二文：長文は読みにくい。	分割・挿入	連用中止・指示語	長文	
権利変換とは	権利変換とは、	権利変換とは一権利変換は、：主題の後には読点があると読みやすい	挿入	読点	読点の不足	主題
基づいて	基づいて、	係り先が遠いので読点により後件句を括る	挿入	読点	係り先が遠い（テ形中止）	
権利を 図解したものであるが、	権利を、 図解したものである。	係り先が遠いので読点により後件句を括る	挿入	読点	係り先が遠い	
所有者が保有する土地に対する権利	所有者が土地に対して保有する権利	所有者が保有する土地に対する権利-所有者が土地に対して保有する権利：連体修飾の係り先が遠いと解釈しにく	順序変更	句	係り先が遠い	連体修飾

- ・文数などを修正対象の原文と一致させる必要はない。
 - ・原則として、段落構成に関する修正は行わない。
 - ・専門用語を一般向けに書き改める必要は必ずしもないが、わかりにくい・適切でないと思われる語については指摘する。
- 3) 修正もしくは挿入した文章の文末に、変えた理由（コメント）を書く。
- 修正案は、一文書につき平均7.1例集まった。

3. 2 分析データ作成

下記の手順で、修正案から要因分析データを作成した（表2）。

- 1) 修正案から指摘箇所・修正例・コメントを取り出す（表2, 1~3）。
- 2) 表2, 1~3の情報をもとに、指摘箇所の修正方法（表2, 4）と修正の対象になった構文要素（表2, 5）を記入する。
- 3) コメントをもとに、読みにくさの要因となる現象を分類した（表2, 6）。コメントがない場合は、指摘箇所と修正例から適宜判断する。
- 4) 読みにくさの要因を現象1に分類したあと、さらに指摘できる要因がある場合は、現象nを立てて記入する（表2, n）。

4. 結果と考察

本節では、作成したデータに基づく調査結果の概要と、読みにくい要因の現象を分析した結果を示す。

4. 1 調査概要

要因分析データの概要を示す（表3）。データ総数は、異なる修正者が同一の指摘箇所を同様に修正した場合もカウントしている。重複を除くと、異なりは約1000件程度である。

表3に示すとおり、表2の現象1にあたる項目数は

表3 結果の概要

データ 総数	延べ1630件 (異なり約1000件)		
現象数	71項目		
指摘の多い現象 (11項目) の件数			
不適切な 用語	137	読点の不足	132
長文	121	重複 (同語・類 義語)	94
並列 表現	85	係り先 (遠い・ 不明)	77
助詞 相語	61	同種文字連続 (ひらがな・か か)	51

71である。この項目をもとに現象nの記入作業を進めている。現象1の71項目中、読みにくい箇所として指摘の多かったものを表3に示す。現在、各項目からの指標抽出を進めている。

次節では、そのうちの指標である文長について述べる。本稿の冒頭で述べたように、長文はなぜ読みにくいか、また文が長くなる要因は何かという観点で分析した結果を示す。

4.2 文が長くなる要因

修正担当者のコメントに「長文」「一文が長すぎる」と書かれた場合、また、長い一文を句点で区切る修正例が示された場合は長文の項目に分類した。要因分析データから長文項目を検索し、一部抜粋したものが表4である。

長文の指摘箇所と修正方法を分析すると、121件中102件が連用中止および接続助詞の箇所では分割している(表5 網掛け太字)。また、図2のグラフは、一文ごとの形態素数と、分割時指摘される文の構文要素の関係を示す。連用中止・接続助詞を含む文であっても、40形態素以下であれば指摘されにくいことがわかる。

一方、1) 40以上の形態素から成る、かつ2) 連用中止や接続助詞を含む、二点の条件を満たすにも関わらず、指

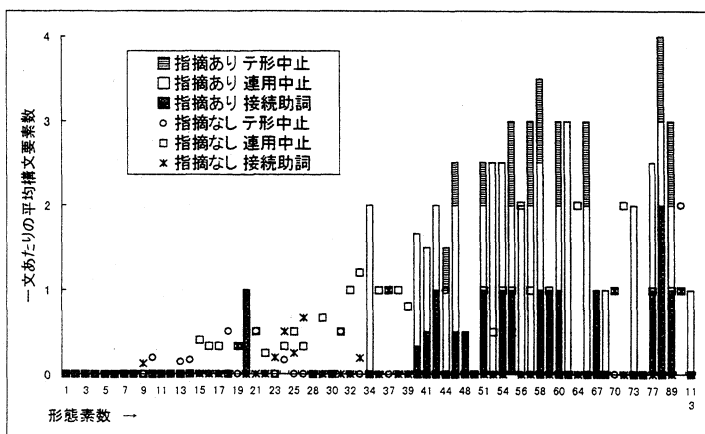


図2 分割時指摘される文と形態素数の関係

摘されない文がある(58文中14文)。これらを調べると、a) 専門性の高い文であるため、担当者が修正できない、b) 連用中止や接続助詞を2箇所以上含む読みにくい構文であるため、適切な修正ができない、c) 文中の括弧内の連用中止であるため分割できない、という傾向がある。

ここで、前述の40形態素以上を長い文とみなそう。この範囲で構文要素を調べると、分割の指摘を受けない文には格助詞・助詞相当語句(二関等)・並列助詞が多いことがわかる(図3)。これらの要素が含まれる文は、長い連体修飾を作る可能性が高い。図3は、1) 連用中止・接続助詞でつながれた文、2) 長い連体修飾による文、以上

表4 長文の指摘と修正

指摘箇所	修正例	コメント	修正方法	修正操作対	現象1	現象2
変換することで、変換することである。	変換することである。	文が長いので分短文に	分割	連用中止	長文	
変換することで、変換することである。	変換することである。	また、文が長いので、短く切る。	分割	連用中止	長文	
変換することで、変換方法は、～	変換することである。その変換方法は、～	一文⇒二文:長文は読みにくい。	分割・挿入	連用中止・指示語	長文	
変換することで、変換方法は、～	変換することである。その変換方法は、～		分割・挿入	連用中止・指示語	長文	
変換することで、変換方法は、～	～変換することである。その方法は、～	文が長いので、読点を加え、2文に分割	分割・挿入	長文・指示語	長文	
当てはまり、	当てはまる。	文が長すぎる	分割	連用中止	長文	
当てはまり、	当てはまる。	長文	分割	連用中止	長文	
当てはまり、	当てはまる。	文が長いので切る	分割	連用中止	長文	

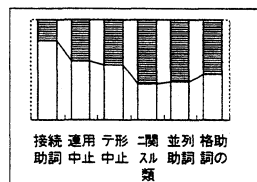


図3 要因別の指摘の割合

注) 40形態素以上(58文)
白色部: 分割時指摘あり
横線部: 分割時指摘なし

表5 長文の指摘箇所と修正方法

指摘箇所／修正方法	修正方法								
修正操作対象	順序 変更	挿入	分割	分割 (3)	分割(3)・ 順序変	分割・順 序変更	分割・ 挿入	分割’	総計
(空白)	1	0	7	2	1	2	1	4	18
接続詞	0	0	0	0	0	0	1	0	1
接続助詞	0	0	25	0	0	0	0	0	25
接続助詞・接続詞	0	0	0	0	0	0	14	0	14
接続助詞・動詞	0	0	0	0	0	0	1	0	1
読点	0	1	0	0	0	0	0	0	1
連用中止	0	0	51	0	0	0	0	0	51
連用中止・指示語	0	0	0	0	0	0	3	0	3
連用中止・接続詞	0	0	0	0	0	0	4	0	4
連用中止・副詞	0	0	0	0	0	0	2	0	2
連用中止・連用修飾	0	0	0	0	0	0	1	0	1
総計	1	1	83	2	1	2	27	4	121

注) 空白: 分割時指摘箇所の構文要素を分類することが困難なもの。

分割(3): 一文を3分割したもの。分割: 全体的な書き換えが行われているもの。順序変更: 句の順序を変えたもの。

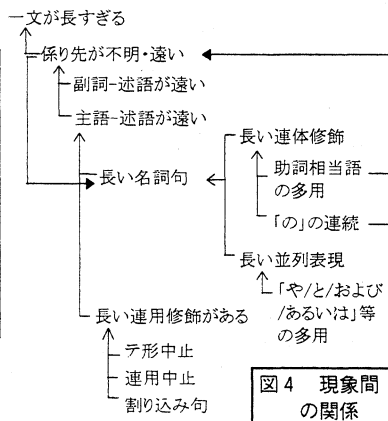


図4 現象間の関係

二つの文で修正方法を変える必要があることを示す。長い連体修飾については、さらに分析が必要である。

現在までの分析結果から、文が長くなる要因を図4のように関係づけることができた。また、次のヒューリスティクスを得ることができた。

規則1. 一文が40形態素以上の場合、文中に1) 連用中止形または2) 接続助詞がある文は、長すぎて読みにくく感じられる。

規則2. 文中に1) 連用中止または2) 接続助詞がある長い文は、1) 述語の連用形を終止形に換える、2) 接続助詞を取って終止形にする、という操作によって文を分割できる。

規則3. 接続助詞でつながれた文を分割する際は、必要に応じて後続する文の文頭に接続詞を補う。

これらの妥当性は、今後、アンケート調査によって検証する予定である。

4. 3 他の要因について

既存の読みやすさ評価ツールで、測定指標になっているひらがな連続、漢字連続について考察する。

1) ひらがな連続

読みやすさ評価では、ひらがな連続の値がデフォルトで10字前後を超えると指摘される。しかし、調査結果では「ひらがな文字列が長い」という理由で読みにくさを指摘したものはなかった。分析結果によると、ひらがなによる読みにくさは、1) 動詞のひらがな表記、2) 助詞とひらがな連続、3) 冗長な文末表現（しようとしている／していこうとしている／されているところである／等）の三点に起因する。

したがって、ひらがな連続値を測定指標にするのではなく、ひらがな文字列がどのような要素から構成されているかをもとに測定指標を抽出する必要がある。本研究では、上記三点の指標の可能性を今後検討する。

2) 漢字連続

漢字連続の場合も、漢字列を構成している要素に着目することが重要である。今回の調査では、1) 「トイレ等水回りの」「需要推計上」に見られる「等／上」、2) 「一昨年九月開店当初は」「今回新開店の」に見られる副詞に接続した表現の二点に指摘があった。

表7 漢字連続の修正パターン

修正パターン	挿入語句	修正例(ゴシックの箇所)
ひらがな表記にする	—	トイレなど水回り関係の
適当な表現を挿入する	格助詞ノ	一昨年9月の開店当初 今回の新開店の
	接続点	今回、新開店の
	述語	今回行った新開店の
	他の格助詞	1月に3回、2月に2回、3月に2回
	他	計8回(1月3回、2月2回、3月2回)調査した。

これらの修正方法は、表7のようにまとめられる。

4. 4 ビジネス文書の評価指標

「ビジネス文書にふさわしくない」とコメントされた指摘箇所を見ると、内容を表8のとおりに分けることができ

表8 ビジネス文書に不適切な表現例

要因	指摘箇所	修正例	コメント
形式名詞ノ	打ち出すのを	打ち出すことを	ビジネス文書にそぐわない感じがするから
冗長な表現	何が必要であるのか		冗長。ビジネス文書らしく
括弧の使い方	別紙に～示しました。	～(別紙参照)	冗長。ビジネス文書の定型表現。ポイントの提示が好ましい
不適切な係助詞	も	は	係助詞「も」の使用は、主観が入るためビジネス文書としては好ましくない
不適切な述語	企画するものである。	企画した	ビジネス文書では特に冗長に感じられるので削除した。
不適切な用言	大変である	難しい面もある	ビジネス文書らしく
不適切な量詞	できれば	可能であれば	提案でなく願望に近くなり、ビジネス文書に好ましくない
不適切な表現	というよりは	でなく／のみならず	「というよりは」ビジネス文書に適さない
文体の統一	ご講演いただきたいこと	講演内容	一般にビジネス文書に待遇表現を使う必要はない

る。「形式名詞ノ」については、対象文書中の出現箇所5例のうち3例が修正者の半数以上によって指摘された。

「こと」への書き換えが多い。しかし、指摘されなかった2例のうち「こと」に書き換えると文意が変わる文があるため、書き換え可能な条件を検討する必要がある。

例) 現状の交通問題を対症療法的に対応していくのではなく、このような問題を短・中期的な計画で位置づけて早期に問題を解決していく都市交通計画の確立が必要となる。

5. おわりに

本研究では、読みやすさ評価の測定指標の抽出を進めている。今回は、長い文の読みにくさに関する分析結果と、ひらがな連続・漢字連続・ビジネス文書の評価指標に関する検討結果を報告した。

今後は、上記要因に関する詳細な検討、および係り受け関係や読点の不足をはじめとする他の要因の分析も進める予定である。さらに、得られた測定指標を心理学実験によって再評価する。

謝辞： 本研究を進めるうえで、富士通関係者の方々と議論が非常に有益でした。感謝いたします。あわせて、修正案の作成にあたりご協力くださった皆様に深く感謝いたします。

参考文献：

- [1] Rudolf Flesch, A New Readability Yardstick, Journal of Applied Psychology, vol. 32, No.3(1948)
- [2] 堀川直義『文章のわかりやすさの研究』朝日新聞調査研究室報告, 社内用 65(1957)
- [3] 森岡健二「リーダビリティ」『コトバの美学』中山書店(1958)
- [4] 小鶴康浩, 日本語文の読みやすさの評価に関する研究, 情報処理学会第34回全国大会論文集(1987)
- [5] 安田賀計『ビジネス文書基本文例130』PHP(1989)
- [6] 高橋・牛島, 計算機マニュアルの分かりやすさの定量的評価方法, 情報処理学会論文誌, vol. 32, No.4(1991)
- [7] 富士 秀, 英日機会翻訳文の読解に関する評価実験, 言語処理学会第2回年次大会発表論文集(1996)
- [8] 計量計画研究所『日本語文書可読性評価に関する研究』平成7年度富士通株式会社委託研究報告(1996)