

## 文章の読みやすさに影響する要因の分析\*\*

乾 裕子<sup>1</sup> 江村ひとみ<sup>2</sup> 石井啓子<sup>3</sup> 末広秀子<sup>3</sup>

1) 計量計画研究所 2) 東京女子大学 3) 富士通

## 1. はじめに

我々は、ワープロソフト「OASYS」の『推敲』における文の読みやすさ評価について研究を進めている[13,14](図1)。文章の読みにくさの要因になる現象について、その詳細を明らかにするために実際に業務で使用されたビジネス文書を材料にして調査を行った。国語学・言語処理に携わる専門家を対象に、文書中の読みにくい箇所およびその理由について記述する調査である。この調査で収集した資料を修正案と呼び、修正案に分析者の判断を加えた資料を要因分析データと呼ぶ(表1)。要因分析データは、延べ 1630 件のデータで、指摘箇所の異なり数は約 1000 件である。指摘箇所を現象ごとに、ボトムアップに整理している。これらの現象を文の読みやすさ評価の測定指標とみなす(図2)。

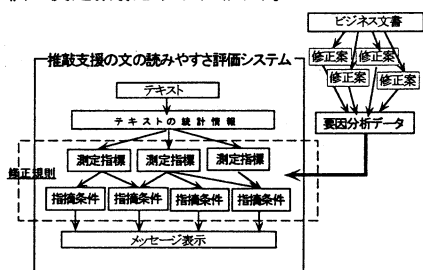


図1 読みやすさ評価システムの構成

これらの測定指標のうち長い文について、1) 文章が長くなる要因は何か、2) 短くするためにはどこに着目すべきかという観点から、それぞれ推敲支援機能の、1) 指摘条件、2) 修正方法を提案した[14]。

本稿では、測定指標として要求度の高い上位項目 1) 不適切な述語、2) 読点の不足、3) 繰り返し表現に着目し、これらの現象について分析を進め、実装化に向けて作成した読みにくさの指摘条件、およびその作成過程について報告する。

## 2. 関連研究

読みやすさ、あるいはその尺度化に関しては、Flesch[1]を始めいくつかの研究がある[2][3][5][6][9]。これらの研究では、基本的に文長、およびシラブル(数日本語では漢字含有率)や語彙数が測定指標とされている。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。表層情報に加えて、心理学実験から求めた理解度との相関で読みやすさを測る研究もある[10]。

しかし、表層的な指標の組み合わせだけでは様々な要因の因果関係から生じる読みにくさを捉えきれず効果的でない。言語処理の分野では、測定指標を文の構造に求め、人の記憶量から検討している研究もあるが、指標を作成するのが研究者自身であるためその妥当性を評価するのが難しい[12]。近年、実験にもとづく評価研究が行われている[8][11]。これらの研究では、いずれも読みや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仮説が必要であるが、多様な測定指標を得られない可能性がある。

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り、ボトムアップに測定指標を取り出す。また、ある測定指標に対して他の測定指標と関連づけられた指摘条件を取り出すことにより、様々な要因の相関関係から生じる読みにくさを考慮した修正規則を作成する。

## 3. 分析と考察

読点については、不要な読点(37件)の指摘を含む 167 件を対象とする。不要な読点を削除する条件および読点を補う際の条件について、要因分析データを基に修正規則を作成する。また、不適切な述語については述語の構成要素に、繰り返し表現については繰り返される語の種類と距離に着目し、指摘条件および過剰指摘抑制条件の候補を取り出す。

## 3.1 読点に関する指摘条件とその効果

指摘の内容を、修正案で挿入された読点に前接する語の品詞で分類すると表2のとおりになる。しかし、この結果で得られた項目をそのまま指摘条件にする

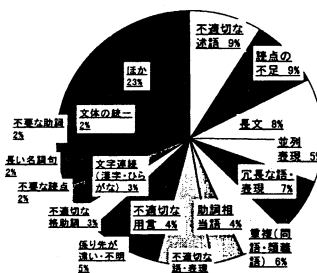


図2 読みにくさの要因となる現象

表1 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			
指摘箇所	修正例	コメント	修正方法	修正対象	現象1	現象2
変換することで、したがって	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	連用中止
		接続詞「したがって」の後には読点を入れる。	挿入	読点	読点の不足	接続詞

\*\* 本研究は、富士通株式会社からの平成9年度委託研究「文章の可読性に関する調査研究」の成果に基づく。

のは不適切である。「主題を示す助詞ハの後は読

表2 読点の不足を指摘する箇所

内訳	個数	内訳	個数
主題	34	助詞相当	16
格助詞	14	接続詞	13
連用中止	12	テ形中止	13
接続助詞	9	副詞	6
連用・並列	4	副助詞	3
他	6		
		計	130

めである[7]。そこで、テキストの統計情報および要因分析データに基づき指摘条件と、その過剰指摘抑制条件を作成した。

いくつかの特徴的な傾向を挙げる。一文が 25 字以上で、かつ読点が含まれない場合には読点不足の指摘が必要であると感じられる[13]。今回の調査で、その指摘位置は、1)文頭の接続詞・副詞の直後 2)文の始めに出た係助詞ハの直後が妥当であるとわかった。また、一文文中の読点間距離が 27 字を下回ると不要な読点を削除する修正要求が出る。文中の助詞相当語(ニオイ等)の読点を不要とする例である。読点間距離 12 字~23 字では、文中の副詞の直後の読点、文頭の副詞の直後の読点を削除する案が示される。テキストから得るこれらの統計情報と、表2の指摘項目との関係から下記の指摘条件を提案する。

#### 0) 読点に関する過剰指摘を防ぐ前提

0.1 並列表現の読点(例:普通名詞、普通名詞)は読点間距離のカウントに含めない。

0.2 文でない(句点がない)場合は読点のチェック対象にしない。

#### 1) 不要な読点を削除するための指摘条件

1.0 読点間距離が 15 字以下の場合、文中に現れる副詞の直後の読点は不要である(但し、文中に読点が0もしくは1個の場合は指摘しない)。

この条件が再現されたのは、全テキストの中で文中の副詞を含む8例のうち、15 字以下の文に現れた文中の副詞に読点不要の指摘があった4例である。この4例には、削除要求があるにもかかわらず 15 字以下に適合しないため指摘されない2例を含む。

#### 2) 読点の不足を補うための指摘条件

2.0 一文が 25 文字以上で、かつ読点が多々含まれていない場合には読点を打つよう指摘する。

例) とくに雑貨部門の商品回転率が著しく低下して損益分岐点を下回る月もあった。

この条件が再現されるのは、条件に該当する 10 文のうち 6 文である。

#### 2. 1) 連用中止

2.1 連用中止が含まれている文では、連用中止形の直後で読点を打つ。但し、長い文の分割指摘[14]は、連用中止の読点付与指摘より優先度が高い。

#### ◆連用中止の過剰指摘抑制条件

機能使用時の過剰な指摘を抑制するため、下記に

点を打つ」「動詞の連用中止形の後には読点を打つ」といった文章読本型の指摘を言語処理ツールで実装すると過剰指摘になるた

挙げた場合は連用中止であっても読点挿入を指摘しない。

2.11 読点がつかない連用中止形に連用中止形の形態素が接続した場合 → 繰り返し表現あるいは慣用表現 例) 後ろを振り返り振り返り、歩いていった。

2.12 読点がつかない連用中止形に次の形態素並びが接続した場合

1) {(名詞句<sup>1</sup>\_\_助詞)/(数詞)}\_\_動詞連用形\_\_ 読点 例) 各月の客1人あたりの金額はほぼ一定しているが、本期の客数は前期に比べ15%ダウンし、このことが直接的な業績低下になっている。

2) 名詞句<sup>2</sup>\_\_助詞\_\_{動詞句/形容詞句<sup>3</sup>\_\_句点 例) 道路行政の担当者がマルチメディア社会について理解を深めることは、今後の「情報ハイウェイの整備推進」に向け重要なことと考えられる。

対象としたビジネス文書に出現したすべての連用中止のうち、実際に打たれていた読点と修正案で打つべきとされていた読点を合わせた割合は 90% (88/96)である。したがって、上記の抑制条件は、不要と指摘されている読点が後接する転用中止、および挿入指摘のない連用中止 11%を参考に作成した。上記 2.12 の抑制条件が再現できるのは8例中5例である。

#### 2. 2) 係助詞ハ

2.2 文の始めに現れる係助詞ハの後に読点を打つ。

係助詞ハの場合、指摘条件だけでは過剰指摘になり、実際に上記条件を満たしているのは47%である。これは、他の指摘条件に比べて優先度が低いことによる。したがって、下記の条件に該当する場合は読点の挿入指摘をしない。

#### ◆係助詞ハの過剰指摘抑制条件

2.21 文頭に接続詞あるいは副詞(これらに相当する表現)があり、かつその直後に現れるハを含む助詞句がある場合。→ 文頭の接続詞・副詞に対する読点の要求の方が強い。 例)これに対して、A店では調査からの推定ではあるが、着実に客数を増やしている。

2.22 読点間距離が20字以下で、かつ文中に連用中止形、接続助詞(ナガラ、テ形は除く)がある場合。→ 接続助詞の前までが係り先である可能性が高い。 例) 第2種事業は用地買収方式であるので、施行者が収用権を持ち、開発区域を全面買収する。

2.21 および 2.22 の抑制条件が再現されるのは、条件に該当する 49 文のうち 32 文、65%である。

#### 2. 3) 接続詞・副詞

2.3 文頭の接続詞・副詞では読点を打つ。

<sup>1</sup> ここでは名詞あるいは短い名詞句とする。名詞は、一形態素の品詞が名詞であるもの。短い名詞句は、名詞+名詞、連体詞+名詞、形容詞連体形+名詞、形容動詞連体形+名詞

<sup>2</sup> 上記1同様。

<sup>3</sup> 述語にあたる文末の動詞および形容詞。助動詞のついたものも取った方がよい。ため「動詞句/形容詞句」とした。形態素解析システムの文法の枠組みに依存する。

例) 走った → 走つた 動詞語幹/助動詞

尚、テキスト中に現れた文頭の接続詞・副詞のうち実際に打たれていた読点と修正案で打つべきとされていた読点を合わせた割合は 80%である。接続助詞については 76%に同様の傾向がみられた。接続表現は、その種類によってさらに検討する必要があるため[4]、過剰指摘抑制条件については今後の課題とする。テ形中止の場合、係り先が遠い場合の「～テ」には読点が必要だが、それ以外で使用されているときには不要である可能性が高く、その判定は難しい。

### 3. 2 不適切な述語に関する指摘条件

不適切な述語に対する指摘は調査結果の中でもっとも多く、要因分析データの約1割を占める。その内容を見ると、サ変名詞+機能動詞(検討7行ナリ)、二重受身(ラレルコト考エラレル)など特徴的な表現がある。不適切な述語の詳細を明らかにし、指摘条件および過剰指摘抑制条件を取り出すために指摘された述語を含む 279 文を形態素解析システム juman で解析した。解析結果に下記の各要素がある場合、それらを含む文字列を集めて分析した(以下は juman が依拠する文法の枠組みによる)。

サ変名詞+機能動詞:検討7行ナリ  
受身・使役:レル・ラレル、セル・サセル  
形式名詞:モノ、コト  
アスペクト表現:テイル、テオル、テオク、テクル  
形容詞性接尾辞:タイ、ヤスイ  
判定詞:ダ、デアル  
助動詞:ラシイ、ノダ

不適切な述語の過剰指摘抑制条件として、上記項目を構成要素に持つ文字列の構造に着目した。

#### ◆適切な述語の過剰指摘抑制条件(候補)

a1)サ変: サ変名詞+格助詞(副助詞モ・ハも含む)+動詞(形容詞も含む) 例) 改善:サ変名詞/を:格助詞/する:動詞

次の例は上記に含めていない。

例) 出現:サ変名詞/に:格助詞/よる:動詞/事情:普通名詞  
(格助詞相当とみなす)

b)受身使役: 動詞性接尾辞のうち「れる／られる」の活用形・「せる／させる」の活用形 例) ・求め:動詞/られて:動詞性接尾辞/いる:動詞性接尾辞/。:句点

c)形式名詞: A(動詞性接尾辞／形容詞／動詞／判定詞／助詞:の)+B(形式名詞:もの／こと(ノは除く))+C(判定詞)| (格助詞+動詞)) 例) ・する:動詞/こと:形式名詞/に:格助詞/する:動詞/。:句点

・図解:サ変名詞/した:動詞/もの:形式名詞/である:判定詞/が:述語接続助詞/。:読点

d)アスペクト: 動詞性接尾辞のうち「いる／おる／いく／くる」の活用形 例) ・入り:動詞/やすく:形容詞性述語接/なつて:動詞性接尾辞/おり:動詞性接尾辞/。:読点

e)形容詞性接尾辞:

例) ・改善:サ変名詞/いたし:動詞/たく:形容詞性述語接  
・保存:サ変名詞/し:動詞/やすく:形容詞性述語接/  
し:動詞性接尾辞

f)判定詞: 「だ／である」の活用形。ただし、「だ」の

連用形「で」は助詞「で」との区別がつきにくいため除外。「ではある／でもある」のように、副助詞ハ・モが割って入る場合だけ取る。例) 調査:サ変名詞/から:格助詞/の:名詞接続助詞/推定:サ変名詞/で:判定詞/は:副助詞/ある:動詞性接尾辞/が:述語接続助詞

#### g)助動詞:

例) ひどい:形容詞/らしい:助動詞

ここに挙げた品詞列は過剰指摘を抑制するための候補に過ぎないが、形態素解析の表層情報によって不適切な述語の構造を明らかにできる可能性を示している。文字列や形態素の組み合わせの絞り込み、また a～g では何が一番大きな要因になるか等を明らかにすることが今後の課題である。

尚、a～g の文字列を含むにも関わらず読みにくさを指摘されない要因を明らかにすることは、過剰指摘抑制条件の抽出につながると考える。述語を作る要素が同じ組み合わせである場合、指摘の有無が生じるのは以下のとおりである。

- ・「ている」の多用よりも、「ネットワークする」などの誤用が指摘されやすい
- ・同じ要素の組み合わせであっても、1文に複数現れた場合は指摘されやすい
- ・一つの述語表現の中に様々な要素が現れたときには指摘されやすい
- ・他の読みにくさの現象(文長、重複)の影響が強いために指摘されない
- ・要素の組み合わせを指摘しているのではなく、重複あるいは動詞自体が不適切な場合に対する指摘がある

### 3. 3 繰り返し表現に関する指摘条件

繰り返し表現は、過剰指摘になりやすい指標である。多くの文章読本では、文章の書き方として「同じ言葉の繰り返しを避ける」と述べている。しかし、これを規則として実装化した場合、同じ話題について書かれた文書では毎行ごとに指摘される可能性がある。指摘しなくてもよい条件、すなわち過剰指摘抑制条件を多数取り出すことが繰り返し表現の適切な指摘につながる。したがって、問題ごとの分析を精緻に行う必要がある。本稿では、予備的調査の結果とその考察から今後の課題を中心に述べる。繰り返し表現には、同語重複と類義重複がある。ここでは、94 件の繰り返し表現のうち 73%を占める同語重複について述べる。

繰り返し表現の指摘に関しては大きく分けて以下の問題がある。

(1)一文の中で指摘される繰り返し

- (a)重複する語の種類
- (b)重複する語の距離
- (c)重複する回数

(2)複数文に及んで指摘される繰り返し

- (a) 重複する語の種類
- (b) 重複する語の距離
- (c) 重複する回数

本稿では、(1)の(a)名詞の重複について(b)の距離に着目した分析結果を報告する。分析は、jumanによる形態素解析結果に基づく。表3は重複する名詞の間の形態素数(以後重複距離と呼ぶ)と、その際の繰り返し表現の読みにくさに対する指摘率との関係を示す(結果1)。予想できる結果であるが、重複距離が短いほど指摘率が高くなる。しかし、逆に

表3 繰り返し表現における重複距離と指摘率  
結果1

距離 (形態素数)	指摘率 ①～⑥を含む
2-5	21.4(9/42)
6-10	13.2(7/53)
11-15	13.4(7/52)
16-20	14.2(4/28)
21-25	10.0(2/20)
26-30	0(0/23)
31-35	7.1(1/14)
36-40	0(0/9)
41-45	0(0/10)
45以上	0(0/6)

→

指摘率 ①～⑥を含めず
75(9/12)
23.3(7/30)
16.2(6/37)
20(4/20)
6.3(1/16)
0(0/22)
14.3(1/7)
0
0
0

見ると、重複距離が2～5と短いにもかかわらず読みにくさを指摘されない表現が80%もあることを示す結果である。そこで、よみにくさを指摘されない繰り返しについて分析し、結果を表4に示す。表中の条件について下記に説明する。

表4 指摘されない繰り返しの内訳

条件		出現回数	割合 (%)
①固有名詞		5	14.3
②数詞と結合		3	8.6
③括弧		3	8.6
④接辞		6	17.1
⑤並列		9	25.7
⑥語句説明		6	17.1
⑦その他	優先順位が低い	1	2.9
	その他	2	5.7

①重複している語が別の語と結合して固有名詞を形成している場合。  
例) A市とB市との交流を通じ

②接頭辞+数詞+重複語または重複後+数詞+接頭辞+数詞+重複語>～<接

頭辞+数詞+重複語> 例) 1980年～1993年

③重複している語同士の間に関節括弧が開き括弧、もしくは両方が含まれている場合:～<重複語>

～<重複語>～ 例) 用地買収方式(第2種事業)の2種類の事業方式があり

④重複語の片方に接辞がついている場合:<(接辞+)  
重複語+接辞>～<重複語>or<重複語>～<(接辞+)  
重複語+接辞> 例) 住宅改善に関しては

地域性が強いので、地域に密着した情報を求めています  
⑤重複している語同士の間、並列点や並列の接続詞(および)、助詞(と・や)などが含まれている場合:～<重複語>～および～<重複語>～ 例) 航空旅客輸送および航空貨物輸送

⑥語句説明に含まれる語句が、説明される語句と重複している場合:

・語句1(とは)～～だ(である) 例) 市街地再開発事

業とは、[市街地の土地の合理的で健全な高度利用と都市機能の更新を計画的に

・～重複語句 例) [全国の発生量を予測する]発生量予測モデル

⑦その他 例1) 国土全体をネットワークしている道路空間の活用が着目されており、道路空間を情報空間として整備していくことが求められている。

例2) 分析手法を用いた戦略的な計画手法の提案している都市圏規模での都市交通計画手法の調査・分析手法

読みにくさを指摘されない繰り返し表現で①～⑥の条件に当てはまるのは91.4%であった。そこで、表3結果1から①～⑥までを省いたところ重複距離と指摘率の関係が明確化した(結果2)。今後は、重複距離が長くても指摘される場合、名詞以外の重複、複数文に渡る重複についても調べる予定である。

#### 4.おわりに

文章の読みにくさの要因になる現象について、1)読点、2)不適切な述語、3)繰り返し表現のそれぞれ1)出現位置、2)構成要素、3)種類と距離を中心に実験的調査を行い、測定指標の指摘条件(ヒューリスティクス)を取り出す分析・考察を行った。分析を経て、これらが読みやすさを測る指標として有効であることが明らかになった。今後、読点については本研究で作成した指摘条件の精度を上げ、評価実験によって有効性を検証する。また、不適切な述語・繰り返し表現については、分析を進め過剰指摘抑制条件を取り出し修正規則にする。

謝辞: 本研究を進めるにあたり、適切なアドバイスをくださった富士通研究所の富士秀氏に、また、討論に参加してくださった富士通株式会社の諸星氏、角田氏、富士通研究所の伊吹氏に感謝申し上げます。

#### 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] 南不二男, 『現代日本語の構造』大修館書店 (1974)
- [5] 建石ほか, 読みやすさの指標を作る基礎としての日本語の統計解析, ヒューマンインタフェース'87 講演論文集, pp15-22 (1987)
- [6] 小鶴康浩, 日本語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)
- [7] 小泉保, 句読法概説『日本語学』vol. 8 (1989)
- [8] 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, 情報処理学会論文集, vol. 32, No. 4 (1991)
- [9] Tuldava, J. The statistical structure of a text and its readability. pp215-227 (1993)
- [10] Mikk, J. & Elts, J. Comparison of texts with familiar and unfamiliar subject matter. pp228-238
- [11] 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)
- [12] 松岡正男, 文構造に着目した日本語の理解しやすさ・しにくさの指標について, 京都大学大学院修士論文 (1996)
- [13] 計量計画研究所, 富士通株式会社平成7年度委託研究報告「日本語文書可読性評価の研究」(1996)
- [14] 乾・丸元, 推敲支援における読みやすさの評価, 平成8年度言語処理学会第3回年次大会論文集 (1997)