

# 「わかりにくい」表現の検出規則作成 —推敲支援システムの実装をめざして—

乾 裕子<sup>1&2</sup>

岡田直之<sup>1</sup>

<sup>1</sup>{h\_inui,okada}@pluto.ai.kyutech.ac.jp <sup>2</sup>hinui@ibs.or.jp

<sup>1</sup>九州工業大学大学院情報工学研究科

<sup>2</sup>計量計画研究所

## 1. 背景と目的

認知科学の分野では、推敲を文産出の認知的プロセスの中で行われている知的活動のひとつと捉える。書き手は文産出のゴールやプランに基づいて自己の文章を「読む」ことによって評価し、書き下ろした文章の推敲課題を定義する。定義した課題について、文産出に必要なさまざまな知識を参照しながら書き直しや言い換えを行うのか、推敲目標を作成するのか、あるいは修正方法や推敲目標を見つけられずに書き下ろした文章を読み直すのかといった方略を繰り返し選択することで推敲を進めていく (図 1)。

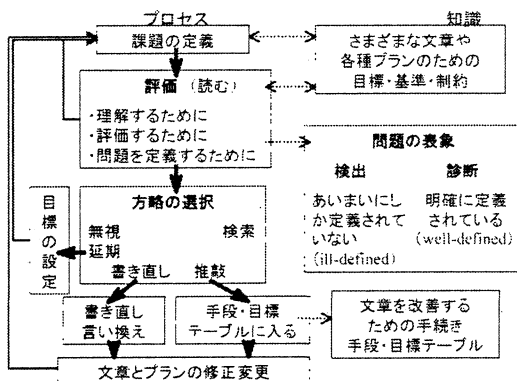


図1 Flower & Hayes の推敲過程の認知モデル[3]

われわれは、これらの推敲に関する定義に基づいて推敲過程の研究を進め、その実装を目指している[4][6]。本研究では、推敲課題の検出部と書き直し部から成る推敲過程の実装モデルをもとにシステムを作成する (図 2)。推敲課題の検出

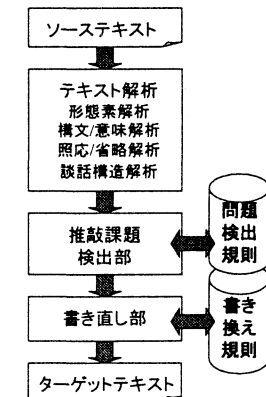


図2 推敲過程の実装モデル

とは、誤りや意味のとらえにくい表現や構文を指摘し、そのわかりにくさの要因を検出することである。書き直しとは、検出された要因に対する問題解決方法として文字列や記号の挿入・削除・置換えを含めた書き換え (paraphrase) を行うことである。

従来の認知心理学的な推敲研究では、推敲課題の検出方法や修正方法について具体的かつ網羅的な研究はなされていない。また、わかりにくい表現に対する言語学的な研究では、現象の説明とそのわかりにくさの要因、およびわかりにくさに対する修正方法とが陽には区別されておらず、また具体的に一般化もされていない。例えば下記の例について；

引用例) この協力の欠如のために、ことがうまく運ばず、よく調整のとれている共産主義陣営に比べて、資本主義体制全体が弱体化しているという事態を克服する必要があるからである。  
(雑誌)

「『ことがうまく運ばない』のは資本主義陣営なのであるが、『資本主義体制』という語がだいたい後にあるので、初め『うまく運ばずー調整のとれている』のようにとれ、読みなおしが必要となる。」と説明されている[1]。これは、わかりにくさを持つ文の現象説明であり、これらの説明からは直接規則を取り出すことはできない。

この文のわかりにくさを検出するためには次のような一般化が必要であると考える。

仮説1: 連用修飾節に後続する要素の中に係り

先候補の述語が複数ある (「調整のとれている」「弱体化している」)

仮説2: 連用修飾節に後続する要素の中に係り

先候補述語が複数ある場合、その述語の補足語 (格) がそれぞれ類義語である (「共産主義陣営」「資本主義体制全体」)

また、他の類似の事例からもこれらの規則が一般化されなければならない。したがって、問題検出や修正方法を規則化するためにはわかりにくさの要因とその問題解決方法について、できるだけ多くの事例を調査し規則化する必要がある。問題検出については聾者のための可読

性基準作成について同様に研究を進めている[7]。

本稿では、推敲過程実装モデルの推敲課題検出部で適用される問題検出規則を作成するために作成した推敲テキストの作成方法および分析方法について、また、問題検出規則の作成方法とその検証方法について報告する。

## 2. 規則作成およびテキスト分析の方針

本節では、まず規則作成の方針について述べ、そのためにに行ったテキスト分析の方法について述べる。

### 2.1. 問題検出規則の作成方針

本研究で作成する問題検出規則とは、図2に示したように入力されたテキストから文をわかりにくくする要因を取り除くための規則であり、下記のとおり一般的なIF-THEN ルールとして記述できる。

IF condition<sub>1</sub>, ..., n THEN factor X

このとき左辺の条件文は問題検出の検出条件に相当し、右辺は検出される要因Xである。「着目する文が条件nを満たす場合、その文にはわかりにくさの要因Xがある」と読む。この検出条件は、条件文がひとつの場合も複数組み合わせる場合もある。

前掲の仮説1と仮説2は、下記のとおり問題検出規則の検出条件に相当する。

**規則候補1:** 連用修飾節に後続する要素の中に係

り先候補の述語が複数あれば、連用修飾節の係り受けにわかりにくさがある

**規則候補2:** 連用修飾節に後続する要素の中に係

り先候補述語がある場合、その述語の補足語（ガ格）がそれぞれ類義語であれば、連用修飾節の係り受けにわかりにくさがある

これらの規則候補では、規則候補2の中に規則候補1が包摂されているため、切り離して考えることは難しいかもしれない。しかし、検出されるわかりにくさの種類によっては規則候補1だけで検出できる場合と、規則候補1と2の両方を組み合わせなければ検出できない場合とがあると考えられる。

また、この組み合わせの有無は、検出条件には含まれる。例えば、規則候補1では、a) 注目する構成要素、b) aと他の要素との間に成り立つ関係によって検出条件が成り立っている。

例) a: 連用修飾節

b: aに後続する要素の中に係り先候補の述語が複数ある

このように、ある文の構文属性やその属性の環境などの依存関係によって、文のわかりにくさは定義できるのである。

### 2.2. テキストの分析方針

2.1で例示した検出規則を作成するために次のような方針で分析を進める（図3）。

1) テキストと、その元テキストに対する推敲テキストを用意する

2) テキストと推敲テキストを観察し、わかりにくさを指摘された文s<sub>1</sub>と書き換えられた文t<sub>1</sub>から特徴要因aと書き換えの動機となっている要因Xを取り出す（例：仮説1もの「連用修飾節」）

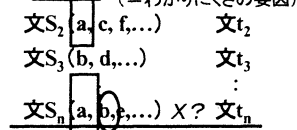


図3 テキストの分析方針

3) 特徴要因aと共起関係にある特徴bを探す

（例：仮説1の「後続する複数の係り先候補」）

4) 特徴要因aとbが共起するs<sub>n</sub>でも、わかりにくさの要因Xが成り立つかどうか調べる。

これにより「着目する文が検出条件a,bを満たせばわかりにくさの要因Xがある」という規則候補が記述できる。また、上記手順を繰り返すことにより、問題検出規則の候補を蓄積する。

これらの手順で分析を進めるためには、推敲テキストを作成する必要がある。また、テキストと推敲テキストを比較できる推敲データベースも必要である。次節では、推敲テキストと推敲データベースの作成方法について簡単に述べる。

### 3. 推敲テキストと推敲データベースの作成

推敲支援システムにおいて、テキスト中の誤りだけを検出するのは不十分である。読み手にとってわかりにくい表現や構文をわかりやすくすることが重要である。しかし、実際には、文書の目的や読み手の知識によってわかりにくさは変動する。そこで、われわれは、推敲対象とする文書をビジネス文書に絞り、わかりにくさの要因を調べることにした。

#### 3.1. 推敲テキストの作成方法

分析対象とした推敲テキストの作成方法について説明する。文章のわかりにくさを明らかにするために、次の方法で推敲テキストを収集した。

実際に筆者の職場で作成された企画書や報告書など13文書(195文)を対象に、言語処理や言語学に携わる20人が一人あたり3~4文書の推敲テキストを作成した。これらの推敲に際して、①テキスト中の問題箇所(推敲対象となる語句・文)、②問題箇所を修正したもの(推敲文)、③推敲した理由、以上の三点を自由記述してもらった。この調査の結果、1文書につき平均6文書、合計79例の推敲テキスト(1218文)が集まった。

表1 推敲テキストの作成

元テキスト(ビジネス文書)	13 文書(195 文)
推敲テキスト	79 文書(1218 文)
一文書あたりの推敲テキスト	4~7 文書(平均 6)
書きかえられた文	643 文
書き換え箇所	約 1,400 箇所

## 3.2. 推敲データベースの作成

推敲データベースでは、推敲テキストの担当者に記述してもらった推敲箇所と推敲文、および推敲理由に加えて、1) 推敲対象のわかりにくさの要因、2) 着目している構文要素、3) 書き換え対象の構文要素、4) 書き換え後の構文要素、5) 修正の方法、6) 検出条件を記述している(表2)。構文的な情報については[2]に基づいて記述する。

表2 推敲データベース

構文要素	構文	要因	着目要素	書き換え対象	書き換え後	修正方法	検出条件
連体修飾節、名詞句並列、連用形並列節	AハB、CハD	係り受け	提題表現	読点	主題	挿入	主題の係り先述語候補が複数ある
			補足語	読点	補足語	挿入	補足語(ヲ格)の後に複数の係り先候補がある
			テ形連用修飾	読点	テ形連用形	挿入	テ形連用形の後に複数の係り先述語候補がある
名詞句の並列	AハB	係り受け	副詞句	読点	助詞(に)	置換(削除)	係り先が特定できるのに不要な読点がある
			照応表現	読点	指示語	削除	係り先が特定できるのに不要な読点がある、指示する必要のない指示語がある

上記のとおり情報を付加した結果、約 1400 の書き換え箇所の内訳を文献[5]を参考に分類した。これらは、それぞれの関係を完全に排他的な関係では捉えられないものである。したがって、対象としたデータの中で観察できた割合も概ねの値である。

## ◆表層表現レベル (40%) :

誤字・脱字 | 表記

語彙的不適切さ | 語彙的冗長性

例) 行うものとする→行う/この際→これを機に/進めていく→推進する/しかしながら→しかし

## ◆言語的構文レベル (37%) :

・係り受け (15%)

提題ー述語関係 | 補足語ー述語 (格) 関係

修飾ー被修飾関係

連用修飾 (語/句/節) | 連体修飾 (語/句/節) | 副詞 (語/句)

並列関係 (語/句)

例) ××市の再開発事業や新市街地計画とそれらを取りまく交通整備体系の方向

・文の境界 (10%)

従属節 (連用修飾節/副詞節)

並列節

・談話 (5%)

接続表現 (接続詞/副詞節)

例) この結果→しかし/しかし、この結果/その一方

・照応 (7%)

指示詞 | 代名詞 | 同義語・類義語重複

例) 市外地再開発事業には→この事業には/変換方法は→その方法は

## ◆言語的意味レベル :

領域知識 (専門用語) | 書き手の意図 | 論旨の展開

上記のうち、本稿では並列表現を除いた係り受け関係の問題検出規則の作成について述べる。

## 4. 問題検出規則の作成と考察

前掲2節で

述べた方針に従って、文章のわかりにくさを検出する

規則作成を進

めている。図4に示したように、規則作成の分析対象となるのは推敲対象となった643文である。これらと推敲対象にならなかった575文との違いを取り出していく。現在のところ係り受けに関して指摘された131文に注目し、わかりにくさの検出条件を記述している。これらは、まだ規則として利用できる段階にないが、ここでは分析過程で明らかになってきたことを報告する。

規則候補: ある文が検出条件aかつbを満たす場合、係り受けにわかりにくさがある

検出条件a (着目する構文要素): | 補足語 | 提題表副詞

| テ形連用修飾 | 連用形連用修飾 | 連体修飾 |

(以下、下線は係り要素、二重線は受け語、点線は係り先候補)

検出条件b (aと他の要素との間に成り立つ関係):

1) 係り先がない | 条件b1: 係り先候補の語選択が受け語として適切でない |

例) 各分野における第一人者を講師として下記のとおり講演会を開催いたしますので多数ご参加下さい。

2) 係り先が特定できない

[条件b2-1: 係り先候補の述語や名詞が複数あり、それらが類義である]

例) こうした高度情報化社会を支える情報基盤として建設省では平成×年度の重点施策の一つに「情報ハイ

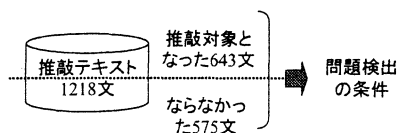


図4 問題検出規則の作成

ウェイの整備推進」を掲げ、マルチメディア社会を支える全国的な光ファイバーネットワーク整備を今後推進しようとしている。

〔条件 b2-2: 係り先候補の述語が複数あり、それぞれの補足語が類語である (前掲 2 節)〕

1) で検出すべき問題は、係り語と受け語の意味的依存関係が成り立っていないことである。したがって、実際に検出する際には「講師として開催する」「講師として参加する」という依存関係が意味的に共起しないことを書き手に知らせることが問題の検出である。2) では、係り先候補である各節の述語「整備推進を掲げる」「整備を推進しようとする」が意味的に類似していることにより、読み手にとってわかりにくいものとなっている。したがって、書き手にこれらの類似性を示すことが問題の検出となる。次に示す問題は、係り先が特定できている場合に構文要素の配置や読点の打ち方によって意味がとらえにくくなる例である。

3) 係り先を特定しにくい

3-1. 近くの述語に係るのに遠くに係るように読める  
〔条件 b3-1: 係り要素の直後に係るのに、係り要素に読点がある〕

例) 北ブロックおよび西ブロックにおいて、需要のアンバランスがみられたが、これは商店街が活性化した場合、自動車利用に十分に対応できないことを示しているものとも考えられる。

3-2. 遠くの述語に係るのに近くに係るように読める  
〔条件 b3-2: ノ連体修飾節「AのBのC」では、「AのB/BのC」が自然であるが、ここでは「AのC/BのC」という係り受けになっている〕

例) 第1は長期を目標とした都市交通M/Pとは別の、  
短: 中期を目標とする都市交通計画の必要性である。

〔条件 b3-3: 副詞節の後に連体修飾節 (係り先でない述語) がある〕

例) また同時に、質の高い地域づくりを進めるため、道路空間の景観形成に果たす重要性も強く認識されているところである。

条件 b3-1 は、読点があることによって遠くに係るように読めてしまう例である。一方、b3-2 では、読点があるにも関わらず近くに係るように読めてしまう文であることを示している。これは、前者が遠くに係りやすい連用修飾であるのに対し、後者は後接する要素に係る読みが優先される構文であることによる違いである。このように、読点の打ち方に共通点はあってもどのような構文要素に打たれているかによって読みに違い

が生じることがわかった。

## 5. 今後の課題

推敲過程における問題検出と検出された問題の書き換えを推敲支援システムとして実装するために、検出規則と修正規則を作成している。作成にあたっては、一定の手続きにしたがって収集した推敲テキストを分析することで検出の条件を取り出している。今回は、係り受けに関する検出条件と規則の候補について報告した。規則の候補は、まだ精緻化する余地がおおいにあり、同時に推敲対象とならなかった文には適用されないことを検証しなくてはならない。これらは今後の課題である。また、(1) 係り受け以外のわかりにくさの要因を整理し、それぞれについて検出規則を進める、(2) 書き直し部の修正規則を作成する、以上を行う予定である。

謝辞 本研究を進めるにあたり、議論に参加してくださった富士通株式会社の石井氏、奥畑氏、西澤氏、角田氏に感謝いたします。

## 参考文献

- [1] 岩淵悦太郎編『悪文』第3版、日本評論社、1960
- [2] 益岡・田窪、『基礎日本語文法』くろしお出版、1992
- [3] 伊東昌子、文章の推敲における認知過程とその支援システム、『認知科学』、1(1)、pp121-134、1994
- [4] 乾・九元、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
- [5] 乾健太郎、「文章生成」、田中穂積監修『自然言語処理-基礎と応用-』第4章、電子情報通信学会、1999
- [6] 乾・岡田、長い文は常にわかりにくいから～わかりにくさの要因とその依存関係～、情報処理学会研究報告 2000-NL-135、pp63-70、2000.
- [7] 山本・乾・藤田・野上・乾、筆者向け文章読解支援のための文可読性基準の調査、情報処理学会研究報告 2000-NL-135、pp127-134、2000.