

長い文は常にわかりにくいか？
～わかりにくさの要因とその依存関係～

乾 裕子^{1&2} 岡田直之¹

¹{h_inui,okada}@pluto.ai.kyutech.ac.jp ²hinui@ibs.or.jp

¹九州工業大学大学院情報工学研究科

²計量計画研究所

概要

推敲とは、推敲課題の検出と検出された問題を解決する過程である。推敲課題の検出とは誤りやわかりにくさの検出であり、問題解決とは検出された問題をなんらかの方法で取り除くことである。本研究では、この定義に基づいて推敲過程の実装を目指す。推敲過程を推敲支援システムとして実装するために、推敲課題として検出すべき問題の検出規則および検出された問題の修正規則の作成を進めている。本稿では、検出規則作成のために行った推敲文書と元の文書の分析結果について報告する。

従来の認知心理学的な推敲研究では、問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、言語処理の立場では、わかりにくさの要因を踏まえた問題検出でなく表層的な処理が行われている。

われわれは、これらの問題を踏まえ、推敲すべき問題を係り受けの問題に絞ってわかりにくさの真の要因、表層上の構造、検出条件、推敲方法という観点に基づいて整理し、それぞれの対応関係について分析した。その結果、検出のための規則を作成した。

Is a Long Sentence Always Incomprehensible? :
A Structural Analysis of Readability Factors

INUI Hiroko^{1&2} OKADA Naoyuki¹

¹Graduate School of Computer Science and System Engineering,
Kyushu Institute of Technology

²The Institute of Behavioral Sciences

Abstract

Revision is an intelligent activity carried out in the cognitive process of text production. Our research is computational realization of this revision process. In order to realize revision system, we make rules for detecting errors or incomprehensibility of the text and applying a tactics to solve each.

Cognitive psychological works on the revision model have never given us any concrete scheme of either problem detection or solving. The field of NLP have produced various RSS, the performance of which is, however, still far unsatisfied because of shallow processing such as character/word counting and string pattern matching. There still remains a significant gap between what has been revealed by cognitive works and what has been implemented on practical systems.

Among many open issues, one of the most essential we should attack next is the understanding of the comprehensibility of texts. Given this background, we manually analyzed our text corpus, which we called a revision corpus, to explore what problems should be detected and how they can be solved in text revision. This paper gives an overview of our corpus analysis and the preliminary results of it.

1. はじめに

推敲は、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である。推敲において書き手は、文産出のゴールやプランに基づいて自己の文章を評価する。書き下ろした文章を評価しながら、推敲課題を検出し、検出/診断された問題を解決するための方法を検討する(図1)。

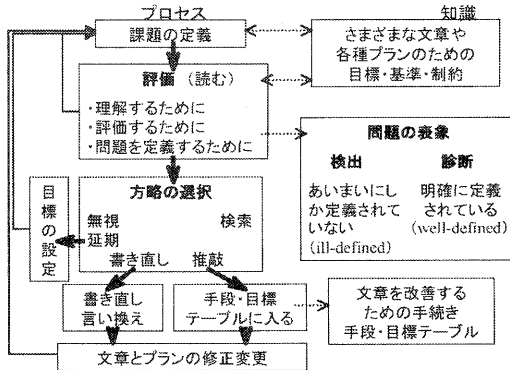


図1 Flower & Hayes の推敲過程の認知モデル([6]より)

これらを踏まえ、本研究では「推敲課題の検出/診断とは、誤りやわかりにくさの検出/診断であり、問題解決の方法すなわち推敲方法とは誤りや意味をわかりにくくする不適切さを取り除くこと」と定義する。また、わかりにくさとは意味の捉えにくさを指す。これは文中に構文的曖昧さを含んでいる場合に限らず、広く意味的に捉えにくくわかりにくさである。

推敲過程は、言語処理における文章作成支援技術の推敲機能として実装が期待されている。しかし、推敲課題として検出されるべき問題と、具体的な対象および推敲方法が実際の文章中どのように検出されるべきかは明らかになっていないため、実装に際しては問題が山積している。現在、言語処理の分野で実装されている推敲支援システムでは形態素解析や単純なパターンマッチングの技術が用いられており、単語の意味や構文・文脈を考慮した推敲支援にいたっていない。また、検出/診断すべき問題は何かということが明確になっていないため、決められた文字数を超える文や読点の数のチェックなど推敲すべき可能性のある表現を取り出す処理を行っている[4]。

Faigley らは、推敲には意味内容に変化を与えるレベルと与えないレベルがあるとして両者を区別し、それぞれのレベルに追加、削除、置き換えなどの具体的な推敲方法があると提案した[2]。

本研究では、what-to-say を決める処理と how-to-say を決める処理の二つの階層を持つ言語

生成モデル[10]に対応した推敲過程があると考えている(図2)。what-to-say を決める処理では、言語化されていない概念やイメージなどを生成する非言語的意味レベルから、概念の単語化・意味的な談話構造の生成を行う言語的意味レベルへの変換を行う。how-to-say を決める処理では、言語的意味レベルで生成された意味構造に対応した構文形式や機能語を言語的構文レベルで選択する。非言語的意味レベルから言語的意味レベルへと生成される意味内容が適切に構造化されない場合—例えば並列や従属の関係で列挙されない、適切に分割されない—は表層表現レベルから言語的意味レベルへ、言語的意味レベルから非言語的意味レベルへの推敲が必要となる。これは、複数の命題を含むだらだら長い文の分割や概念的に對をなしていない並列表現の推敲、同じ語が何度も繰り返されることに対する推敲に相当する。同様に、言語的構文レベルで生成される文がだらだら長い文を分割する際の不適切な接続詞や、副詞節の接続形式の不適切さを含む場合、これらは言語的意味レベルで生成された意味構造を捉えにくいため、問題を取り除くには表層表現レベルから構文レベルへ、構文レベルから言語的意味レベルへの推敲が必要である。

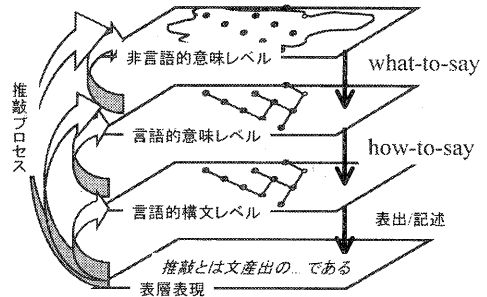


図2 推敲支援システムの実装モデル

このように、不適切さを取り除いた意味構造や構文から生成された文が表層表現としてわかりやすい文であると考えられる。したがって、表層表現の書き換えによる意味の変化は問題にしない。不適切さを取り除く過程で、意味に変化が生じるかもしれないことを前提とするモデルである。

このようなモデルのもとで、推敲過程を推敲支援システムとして実装するという立場から、

- 1) わかりにくい文には推敲すべきわかりにくさの要因があり、それを検出することができる(問題の検出)
- 2) 検出された問題をなんらかの方法で解決することによって文がわかりやすくなる(推敲)

方法の選択)

上記を仮説として推敲過程を分析し、その結果に基づいてわかりにくさの検出規則と検出された問題表現を修正するための書き換え規則を作成する。これらを推敲支援システムとして実装することが目標である。

本稿では、検出規則の作成を踏まえて推敲過程を分析した結果について報告する。具体的には文書とその推敲文書を用意し、どのような推敲を文のどの部分に対して行っているか整理したうえで、その推敲の要因を分析した。また、その要因を実装可能にするための構文的・意味的手がかりを調べた。

2. わかりにくさの真の要因

ワープロソフトに内蔵された推敲支援の問題点として、推敲課題でなく推敲すべき課題が指摘されることについては先に述べた。これは、可読性の基準を表層的な現象の字数や形態素数で判断するところに問題がある。同じ文字数や形態素数であっても、わかりやすさすなわち可読性に差があることを考慮していないためである。この考え方において可読性基準を明らかにするということは、「長い」「多い」「離れている」と示されたわかりにくさは何字あるいは何個の形態素か、何回使われるといけなさを求めることになる。表面的な現象だけに着目した可読性基準は不十分である。

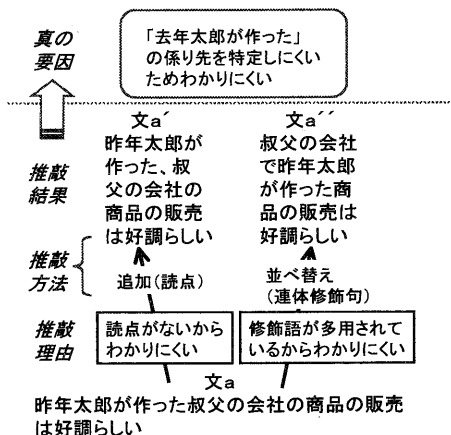


図3 わかりにくさの真の要因と推敲方法の対応

作文技術書のコメントや推敲文書での推敲理由を見ると表層的な現象をわかりにくさの要因とみなしていることがある[3]。

- a 昨年度太郎が作った東京の会社の商品の販売は好調らしい

(推敲例1) 昨年太郎が作った、東京の会社の商品の販売は好調らしい

(推敲理由) 読点が不足している

(推敲例2) 叔父の会社で昨年太郎が作った商品の販売は好調らしい

(推敲理由) 修飾語が多usedされている

文aからa'およびa'への書き換えは言語的構文レベルから言語的意味レベルへの推敲の結果である。意味的な捉えにくさの問題を解決するために、構文に修正操作を加えた例である(図3)。

推敲支援では、書き下ろされた文章の表層表現を手がかりにわかりにくさの要因を検出するため、「修飾語が多い」「読点がない」などの表層上の特徴は注目する必要がある。しかし、わかりにくさは修飾語の数や読点の有無に依存するのではない。表現の現象でなく、その背後にあるわかりにくさの要因を取り出さなければ本質的に問題を解決したことにはならない。本研究では、意味の捉えにくさを考慮しないわかりにくさ、すなわち真の要因を設定しない場合は単なるボタンからボタンへのマッチングとなり、可読性を向上させるための問題検出にならないと考えている。

わかりにくさの要因と推敲方法を対応付けた検出規則および修正規則を作成するため、先行研究を踏まえた問題検出と推敲方法の選択に関する仮説(前掲1,2)に下記の仮説を加えて分析を行う。

- 3) わかりにくさには必ず真の要因が存在する。真の要因は意味的捉えにくさに起因する
- 4) 観察される具体的表現に、わかりにくさの真の要因が検出される場合、その表層上の構造と真の要因の間には因果関係を含む依存関係があるため、問題検出を規則化できる

3. 推敲テキストの作成

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。

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。

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。言語処理や言語学に携わる22人を対象に、一人あたり3~4文書の推敲調査を行った。22人の推敲担当者に配布し

たビジネス文書は、実際に筆者の職場で作成された企画書や報告書など14文書（281文）である。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（表1：推敲文は明朝体、推敲理由はゴシック体）。この調査の結果、1文書につき平均6.4文書、合計90例の推敲文書（約1800文）が集まった。

表1 推敲テキストの作成例

原文	ニュータウン開発の影響に伴う商業系の開発については、消費者の自動車利用ニーズにこたえるべく、少なくとも、 <u>A町周辺の市町村の現況の大型店で整備水準レベル</u> では、駐車場が整備されるものと予想されるため、これを基本的前提として検討を行う。
推敲文A	ニュータウン開発に伴う商業系の開発は、消費者の自動車利用ニーズに応える必要がある。 \$ （長文なので、文を分割。格代用表現は分かりづらい。冗長な表現を削除。動詞「こたえる」を漢字表記にする。） \$ <u>A町周辺の市町村にある大型店と同程度</u> には、駐車場を整備すべきである。そこで、これを基本的前提として検討を行う。 \$ （「整備」が重複するので削除。「少なくとも」を「～すべき」と同意と考えた。連体修飾が続くと分かりづらい。冗長な表現を削除。デ格の意味が取れない。） \$

4. わかりにくさの要因と推敲方法の対応に着目した分析

先にも述べたとおり、推敲文書を作成する際に記述された理由からは単純に真の要因を取り出すことはできない。なぜわかりにくいのか、文のどの部分がどのように推敲されているかを整理し、どのように推敲されるとわかりやすくなるのかといった真の要因、その表層上の構造、推敲方法との対応を前述の仮説に基づいて分析する。

今回は、ビジネス文書 281 文の中から 50 形態素以上の文 41 文と、その推敲文 219 文を対象に分析を行った。この対象範囲の限定について簡単に述べる。わかりにくさの要因として「長い」という現象が問題であるとき、文を短くすることは推敲方法のひとつの選択である。短くする方法には、不要な語句を削除して簡潔にする方法、文を分けて短くする方法がある。[8][9]の研究では、今回と同じデータを使って分析した結果、一文が 40 形態素以上で、かつ連用形修飾節・連用形並列節を含む文であると分割されやすいという長い文の判断基準と推敲方法を得た。今回は、選択された推敲方法の要因は何かといった深い分析を行うため、長い文を一文 50 形態素以上としてさらに対象範囲を絞っている。

本研究では、言語研究者・言語処理研究者を対象に推敲文を作成することで原文よりも推敲文がわかりやすくなっている、すなわち可読性が上がっているとの前提のもとに分析を進めている。

どのように推敲されているかについては、Faigleyらの六つの操作（追加・削除・置き換え・並べ替え・分割・統合）を基準に記述した[2]。さらに、推敲方法には一次的な操作と付随的な操作があると考え区別している。例えば、わかりにくさの要因を解消するために文を分けた場合、分割にもなって述語を整え、接続詞や連用修飾節などの接続表現が追加されることを付随的な操作とした。

対象となる語句・文の構文的属性については田窪・益岡の体系を基準に記述した[5]。これによると、長い文の構文属性として名詞句の並列、補足節の並列、連用形並列節、連体修飾節、副詞節（理由、目的、付帯状況、時、前提）を挙げることができる。構文属性のうち、推敲対象となっている属性を取り出しただけでも一文につき平均 2.7 個含まれていた。推敲文および推敲理由と併せて上記の情報を書き加えていくことにより、要因分析データベースを構築している（表2）。

4.1 長い文における係り受けの問題

分析対象とする 50 形態素以上の 41 文の推敲文 219 文の中で、推敲対象となっている部分は 644 件ある。これらの推敲動機を調べたところ、(1) 誤り（誤字・脱字、助詞落ちなど）(2) 語彙的冗長性 (3) ビジネス文書という文書のスタイルに合わない表記・語彙・文体 (4) 一文の中に複数の命題が含まれる (5) 語と語の関係がわかりにくい (6) 係り受け関係がわかりにくい等に大別することができた。これらについては、まだ不十分な点も多く、現在も分析を進めている。

このうち、本稿では係り受けのわかりにくさを中心に報告する。係り受けのわかりにくさについては、認知的な記憶量という観点から連体修飾節の構造を分析した研究がある[7]。しかし、推敲支援を前提にしていないため、問題解決に対しては言及されていない。対象とするデータのうち、係り受けのわかりにくさは 644 件中 95 件で全体の 15%程度を占めていた。

推敲理由に「係り先が遠い」「修飾しているものが不明」など、推敲の動機が係り受け関係であることを明示したものは 20 件程度である。そこで、推敲方法とその対象および書き換えられた結果から、同様の動機によって推敲された文を調べた。推敲の対象となっているのは、読点や格助

詞・接続助詞から連体修飾節・連用修飾節までの25種類の語句や分節である。推敲における操作は、追加(36件)、削除(12件)、置き換え(30件)、並べ替え(16件)である。また、操作と、推敲対象およびその書き換え結果の組み合わせは58通りになった。それぞれの例について詳しく見る。

4.1.1 係り受けの問題における追加

係り受けの問題を解決する手段として、もっとも多く取られた推敲操作が読点の追加(32/36件)である。追加された箇所は、最多のテ形修飾節・テ形修飾句(14件)に続き、提題表現、連用形修飾節、連体修飾節・連体修飾句、補足語(名詞+格助詞)、副詞節である。

#1 社会全体のマルチメディア化の進展により、地域間・個人間の情報交換量は飛躍的に増加する方向にあり、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。

(推敲例A) (略) こうした高度情報化社会を支える情報基盤として、建設省は平成×年度の重点施策の一つに「情報ハイウェイの整備推進」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推進しようとしている。

この例では、「読点を追加」「長く読んでみると読点を入れる」といったことが推敲理由になっている。しかし、問題の本質は、推敲するつもりで読まないでテ形連用修飾節「こうした高度情報化社会を支える情報基盤として」が、「掲げ」に係るのか、「支える」「推進しようとしている」に係るのか、すぐには理解できないことにある。テ形修飾節の後に、係り先述語の候補が複数あるということが問題検出の条件になっている。したがって、#1の事例から得られる要因と推敲の関係を一般化すると下記のとおりになる。

推敲過程：表層表現→構文レベル→言語的意味レベル
わかりにくさの要因：係り先を特定しにくい
表層上の構造：テ形修飾節、読点が不足
検出条件：テ形修飾節の後に係り先述語の候補が複数ある
推敲操作：読点の追加

この推敲操作は、検出された問題を十分に解決できているだろうか？読点を追加したことによって、テ形修飾節の直後の述語である「掲げ」に係

っているのではないことを示した結果になった。しかし、「支える」「推進しようとしている」のどちらに係っているかは構文からは判断しにくい。次に、同じ例文に対する別の推敲例を示す。

4.1.2 係り受けの問題における並べ替え

「修飾語は被修飾語の近くに置く」という原則は、作文技術書で一般に見られる。なぜ被修飾語の近くに置く必要があるのか、それは前節でも述べたとおり係り先の候補が複数できるため曖昧さが生じるからである。下記の推敲例では、節の並べ替えによって係り先を「支える」に特定している。

(推敲例B) 社会全体のマルチメディア化の進展により、地域間・個人間の情報交換量は飛躍的に増加している。建設省では平成×年度の重点施策の一つに「情報ハイウェイの整備推進」を掲げ、高度情報化社会を支える情報基盤として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の推進を計画している。

推敲過程：表層表現→構文レベル→言語的意味レベル
わかりにくさの要因：係り先を特定しにくい
表層上の構造：テ形修飾節
検出条件：テ形修飾節の後に係り先述語の候補が複数ある
推敲操作：テ形修飾節の並べ換え。係り先述語の前に移動

並べ替えによる推敲では、上記の「(テ形修飾節の後に)係り先述語の候補が複数ある」という検出条件の他に、連体修飾における条件も見られる。

#2 今後の開発途上国における都市圏規模での都市交通計画手法の調査・分析手法に関する1つの方向性
(推敲例) 開発途上国における都市圏規模での都市交通計画手法の調査・分析手法に関する今後の1つの方向性

#2は、「長い修飾語は短い修飾語の外側に置く」という一般的な原則に基づくものとなった。

推敲過程：表層表現→構文レベル→言語的意味レベル
わかりにくさの要因：係り先を特定しにくい
表層上の構造：長い名詞句
検出条件：名詞句内の連体修飾句(節)において同じ名詞に係る分節が複数ある場合、短い分節が最左にある
推敲操作：短い連体修飾句の並べ換え。係り先名詞の前に移動

4.1.3 係り受けの問題における置き換え

置き換えによる推敲で対象になるのは連体修飾語句・節(19/30件)、述語、連用修飾節、助詞、副詞節である。連体修飾節は連用修飾節(#3)か副詞節(#4)に、連体修飾句は連体修飾節(#5、6)に置き換えている。

連体修飾節が連用修飾節に置き換えられるのは、修飾節が隣接する名詞(斜字体)に係るように読めてしまう場合である。実際には、係り先である名詞(ゴシック体)は離れている。

#3 その中で、Thomson は開発途上国の不確実性を考慮するため、長期フレームと整合性のとれた短期計画が設定できるような長期の「方向付け計画(Directional Planning)」と短期の「設計計画(Design Planning)」の2段階に分け、それぞれ別個の分析手法を用いた戦略的な計画手法の提案している。

(推敲例) (略) 不確実性を考慮して、長期フレームと整合性のとれた短期計画が設定できるように、長期の(略)の2段階に分け、

#4 また、太田は、Thomson の提案を受けて、開発途上国の現実を踏まえた上で計画の長期整合性・総合性を高める計画手法やモデルの改良・開発の必要性(モデル体系を簡略化し、短・中期を目標とした数多くの代替的仮定と計画代替案の分析手法の開発)を提案している。

(推敲例) (略) 開発途上国の現実を踏まえた上で計画の長期整合性・総合性を高める計画手法やモデルの改良・開発の必要性(略)

#3では「設定できるような～2段階に分け」を「設定できるように、～(2段階に)分け」という構文に、#4では「踏まえた上で～改良・開発」を「踏まえた上で～高める計画手法やモデルの改良・開発」という構文に置き換えている。これらは連体修飾節を連用修飾節・副詞節に置き換えることで係り先を換え、意味の捉えにくさを構文レベルで解消しようとする how-to-say の推敲である。ここでも、係り先名詞の候補が複数あることが問題検出の条件となっている。しかし、係り先名詞の候補がテ形修飾や連用修飾の係り先述語の候補よりも多いため、これだけでは条件にならない。規則として一般化するためには、置き換えがおこっている名詞句の詳細な分析が必要である。

「AのBのC」といった構造の名詞句では、#4のように「Aの」の係り先を特定しにくい場合と、係り先は特定できるがAとBの関係を捉えにくい場合がある。前者の推敲はわかりにくさの要因が構文にあるため、表層表現→構文レベル→言

語の意味レベルという推敲過程で行われる。しかし、後者については、構文レベルの不適切さではなくA、B、Cの概念の関係を適切に捉えられていないところに問題がある。したがって、推敲過程は表層表現→言語の意味レベル→非言語の意味レベルの推敲過程で行われる。これらの例について見てみよう。

#5 社会全体のマルチメディア化の進展により、(推敲例) 社会全体のマルチメディア化が進むなかで

#6 本研究では、開発途上国における都市交通計画手法としての非集計モデルの適用の必要性を考察し、(推敲例) 本研究では、発展途上国に対し、非集計モデルを都市交通計画手法として適用する必要性について考察し、

#7 開発途上国特有の社会・経済情勢のダイナミックな変化による計画の不確実性や計画を実施する上での財政、法制度、組織体制等の制約

(推敲例) 開発途上国特有の社会・経済情勢のダイナミックな変化が計画を不確実にすることや、計画を実施する上で財政、法制度、組織体制等の制約があるといった

名詞句における問題であり、置き換えという推敲操作は共通していても、真の要因が語と語の関係の不適切さにある点で、上記の事例から作成される規則は#3～#4と異なるものになる。

推敲操作としては、連体修飾句を連体修飾節に置きかえる、すなわち述語化する例では、1) 連体修飾節を受ける名詞がサ変名詞の場合、それを和語動詞(#5)サ変名詞+する(#6)に置き換える、2) 連体修飾節を受ける名詞を補足語とする述語を加えて置き換える(#7)事例がある。

4.1.4 係り受けの問題における削除

削除による推敲は、同語や類義語の重複や、なくても意味が変わらないような不要な副詞・指示詞などの語彙的な冗長性に対して行われる推敲方法である。係り受けの問題解決として使われる際は、読点を削除する場合が多い(5/11件)。この推敲では、係り先が特定できるのに不要な読点がある、すなわち隣接する分節に係っているのに読点の打たれた分節が検出条件になっている。

#8 特に、前述の非集計分析の特徴であるセグメンテーションに基づく選択の異質性とそれに対する同質性に着目し、マニラ首都圏をケーススタディとして、データ収集の効率化を図っていく上で、より有効な調査・分析手法を実証的に検討する。

(推敲例) (略) データ収集の効率化を図っていく上で

より有効な調査・分析手法を実証的に検討する。

＃8の例で「データ収集の効率化を図っていく上で」という副詞節は「有効な」に係っているのだが、原文の読点の位置では「検討する」に係っているように読めてしまう。この事例から規則として一般化すると下記のとおりになる。

推敲過程：表層表現→構文レベル→言語的意味レベル
わかりにくさの要因：係り先を特定しにくい

表層上の構造：副詞節

検出条件：連用修飾語句・節や副詞節の後に係り先述語の候補が複数ある場合、直後の述語に係っているのに読点が打ってある

推敲操作：読点の削除

他に、連用修飾節の係り先が見つからないため削除してしまう例もあった。これは「係り先がない」という問題は検出できて、推敲操作の選択が適切でない例といえる。また、問題検出ができていない場合、係り先のない修飾節に対して読点を追加する例も見られた。一方、問題がきちんと検出できて推敲操作の選択も適切である例では、修飾節に対して受け語となる述語を追加するという操作も見られた。

このように、同じ問題を検出した場合でもその推敲操作は異なる。次節では、検出規則の条件について述べる。

4.2 係り受けの問題検出

4.1節では事例に基づいて、係り受けの問題を検出する規則について報告した。現在までのところ「係り先を特定しにくい」という真の要因に対し、19件の表層上の構造、14件の検出条件、4件の推敲方法の組み合わせから35件の検出規則を作成している。下記は「係り先を特定しにくい」という真の要因に対する表層上の構造と検出条件の例である。

係り先を特定しにくいという真の要因に対する表層上の構造には次の種類があった(19件)：

述語の不足/冗長な表現/読点の不足/の連続/不要な係助詞/不要な読点/不要な指示語/主題の位置/長文/格助詞の不足/不適切な述語/連用修飾の位置/助詞の不足/副詞節/不適切な格助詞/不要な格助詞/長い名詞句/名詞句の構造/助詞相当語

また、検出条件としては以下を取り出すことができた(14件)：

係り先が見つからない/係り先が特定できるのに不要な読点がある/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/主題であることがわかりにくい/主題と述語の呼応関係が適当で

ない/主題の係り先述語候補が複数ある/テ形連用形の後に複数の係り先述語候補がある/副詞節の係り先候補の述語が複数ある/補足語の後に複数の係り先候補がある/連体修飾節の係り先名詞の候補が複数ある/連体修飾句の係り先名詞の候補が複数ある/連体修飾節が冗長/連用形修飾節の後に複数の係り先述語候補がある/連用修飾句の後に複数の係り先述語候補がある
推敲の方法：追加/削除/置き換え/並べ替え

5. おわりに

推敲過程における問題の検出と推敲方法を推敲支援システムとして実装するために、検出と修正の規則化を進めている。推敲元の文と推敲文を比較し何をどのような方法で推敲しているか、どのような動機で推敲しているかを分析した結果、真の要因、表層上の構造、検出条件、推敲方法といった観点から問題を整理する必要があるとわかった。現在は、係り受けの問題を中心に35個の検出規則を作成した。

しかし、実装可能な規則とするためには、推敲方法の一時的な操作・付随的な操作も含めてさらに検出条件を詳細化する必要がある。また、今後の課題として、(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)作成された規則については評価実験を行う、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] Rudolf Flesch, A New Readability Yardstick, Journal of Applied Psychology, vol.32 No3, 1948
- [2] Lester Faigley and Stephen P. Witte, Measuring the effects of revisions on text structure, New Directions in Composition Research, pp.95-108, 1984.
- [3] 大隈秀夫, 「短くてうまい文章の書き方」, 三笠書房
- [4] 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
- [5] 益岡・田窪, 「基礎日本語文法」くろしお出版, 1992
- [6] 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
- [7] 松岡正男, 文構造に着目した日本語文の理解しやすさ・しにくさの指標について, 京都大学大学院修士論文, 1996
- [8] 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」, 1997
- [9] 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, pp 1997
- [10] 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

表2 わかりにくさの要因と推敲方法の対応にもとづく分析

No	原文	推敲文	推敲理由	構文属性	推敲対象	書き換え	一次的推敲	付随的推敲	推敲対象	書き換え	真の要因	検出規則
1	社会全体のマルチメディア化の進展により、地域間・個人間の情報交換量は飛躍的に増加する方向にあり、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。	社会全体のマルチメディア化の進展により、地域間・個人間の情報交換量は飛躍的に増加する方向にある。建設省では、平成×年度の重点施策の一つに「情報ハイウェイの整備推進」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	「こうした高度情報化社会を支える情報基盤として」のかわかる場所がない	連用形並列節	テ形連用修飾(こうした高度情報化社会を支える情報基盤として)		削除				係り先を特定しにくい	係り先が見つからない
2	そこで、道路行政の遂行に当たって考慮しなければならない「道路をとりまく諸情勢の変化」を展望するため、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。	そこで、道路行政の遂行に当たって考慮しなければならない「道路をとりまく諸情勢の変化」を展望するため、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。	読点挿入。	副詞節(目的)	読点	読点(講師として、)	追加				係り先を特定しにくい	係り先が見つからない
3	特に、前述の非集計分析の特徴であるセグメンテーションに基づく選択の異質性とそれに対する同質性に着目し、マニラ首都圏をケーススタディとして、データ収集の効率化を図っていく上で、より有効な調査・分析手法を実証的に検討する。	特に、前述の非集計分析の特徴である、セグメンテーションに基づく選択の異質性とそれに対する同質性に着目し、データ収集の効率化を図っていく上でより有効な調査・分析手法をマニラ首都圏をケーススタディとして実証的に検討する。	「、」→の削除→「、」がないほうが「～いく上で」が「有効」にかかることがわかりやすい。	連体修飾節、名詞句並列、連用形並列節、副詞節(前提)	読点(効率化を図っていく上で、より有効な～検討する)	副詞節:前提(効率化を図っていく上でより有効な～検討する)	削除				係り先を特定しにくい	係り先が特定できるのに不要な読点がある
4	そこで、特にA市と国際姉妹都市であるB市との交流を通じ、両市のこれまでの経験を共有しながら今日の都市行政課題の解消をめざし、新たな都市づくりのあり方を展望することを目的に本調査を実施する。	そこで、特にA市と国際姉妹都市であるB市との交流を通じて両市の経験を共有しながら、今日の都市行政課題の解消をめざし、新たな都市づくりのあり方を展望することを目的に本調査を実施する。	意味的に考えると、「…ながら」までまっすぐに感じるので、「、」をうっ箇所を変えた。	連用形並列節	連用形修飾節(通じ、)	テ形修飾(通じて)	置き換え	挿入	読点(共有しながら、)		係り先を特定しにくい	係り先が特定できるのに不要な読点がある
5	A市の成長は、明治以降の国土政策や高速交通等体系の整備による首都圏との近接性だけではなく、こうした都市の中に存在する自然環境が都市に「ゆとり」や「やすらぎ」を与え、他都市との立地環境面での比較優位をもたらしたという側面もある。	明治以降の国土政策や高速交通等体系の整備による首都圏との近接性だけではなく、こうした都市の中に存在する自然環境が都市に「ゆとり」や「やすらぎ」を与え、他都市との立地環境面での比較優位をもたらしたという側面もA市の成長要因となっている。		連用形並列節	主題・述語(A市の成長は、～もたらしたという側面もある。)	述語(もたらしたという側面もA市の成長要因となっている。)	並べ換え	順序変更			係り先を特定しにくい	主題と述語の呼応関係が適当でない