

文章推敲支援における読み手に誤解される文の抽出

菅沼 明† 小野 貴博‡

†九州大学大学院システム情報科学研究所

‡九州大学大学院システム情報科学府

〒819-0395 福岡市西区元岡 744 番地

E-mail: suga@limu.is.kyushu-u.ac.jp

あらまし 読み手に誤解されるような文はマニュアルとしては不適切である。このような文は推敲する際に書き手に提示して書き直しを促す必要がある。本稿では、読み手に誤解される恐れのある文を書き手に指摘することについて述べている。係り受け解析により複数の係り受け解が求められる文を、誤解される文の第1次近似とする。この中には、不自然な係り受けが含まれているので、言語情報を利用してそれらを振り落とす。本研究では、文節のランク付けによる制約や名詞句に関する絞込み、共起情報による絞込みを導入した。さらに、抽出した文について、誤解を生む原因となる文節を推定する手法を構築した。最後に、この手法を実際の文章に適用して評価実験を行った。

Extraction of the Sentences with Misunderstood Modification Relation for a Writing Tool

Akira Suganuma, Takahiro Ono

Department of Intelligent Systems, Kyushu University

744, Motoooka, Nishi-ku, Fukuoka, 819-0395, Japan

E-mail: suga@limu.is.kyushu-u.ac.jp

Abstract It is infelicitous to write the sentences misunderstood by some readers in a manual. It is necessary to inform the writer of these sentences in his document when he polishes it. In this paper, a method extracting the sentences with misunderstood modification relation is described. The proposal method considers that the sentence with a number of modification relations is misunderstood by the readers. The basic rule for the modification relation of Japanese estimates lots of solutions including the unnatural relations. To find natural modification relations for readers, we have adopted some rules: the restriction about the predicate phrase, the refinement of the modification relations in noun phrase and the collocation of the words. Furthermore, the proposal method estimates the origins of the misunderstood relation. We have applied, furthermore, our method to a writing tool and done an experiment to evaluate it.

1. はじめに

推敲は文章を見直し、より適切な文に書き直す作業である。しかし、書き手による推敲では、

書き手の主観や思い込みにより書き直すべき文を見逃す場合が多々ある。そのため、コンピュータを用いて主観等を除いた視点で文を判断する

ことで、書き手の推敲作業を支援することができると考える。書き手が書いた文章をコンピュータが解析し、文章内の書き直すべき箇所を抽出し、書き手に提示する。書き手は、提示された箇所を見直して、書き直すかどうかを判断する。このような協調系で文章の推敲を行うとすると、コンピュータが指摘する箇所の中には、書き直さなくても良い箇所が含まれても良いことになる。しかし、書き直すべき箇所はほぼすべて指摘して欲しい。

マニュアルなどの文章は、内容が読み手に一意に伝わる文を書くべきである。文中に現れる単語の意味もさることながら、文節の係り受けが元で意味が曖昧になることもある²¹。係り受けは読み手が文を解釈する際の文の構造を示すものであるため、一つの文に複数の係り受けが存在すると、複数の解釈が可能になる。複数の解釈が可能な文では、書き手と読み手の解釈が一致せず、書き手の意図が読み手に正しく伝わらない場合がある。本稿では、複数の係り受けをもつ文を抽出する手法について述べる。

2. 誤解される文

文節の係り受け関係が異なると、同じ文であっても、読み手が異なる意味に捉えることがある。例えば、例文 1 に対しては図 1 に示す係り受けが考えられる。(a) のような係り受け関係だと見れば、眼鏡を「海でかけていた」ものだと説明しているが、「落とした」場所については言及されていない。また、(b) のように捉えるのならば、かけていた眼鏡についての詳細は不明だが、「落とした」場所は「海」だと捉えられる。解釈の仕方によって、同じ文でもその内容は大きく違ってくる。

このように、複数の解釈ができる文では、書き手の意図が必ずしも読み手に伝わるとは限らない。書き手が考える解釈とは別の解釈を読み手が行う可能性があるからである。係り受け関係は文の解釈に影響を与えるので、一つの文に対して複数の係り受け関係が存在するならば、その文に対して複数の解釈が可能であると言える。

係り受け関係には、条件を満たす文節のうちで最も近いものに対してのみ係るといった制約はない。条件を満たす文節であればどの文節に

例文 1) 海でかけていた眼鏡を落とした。

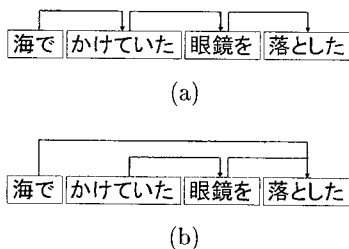


図 1: 複数の解釈ができる文

も係り得る。図 1 の例文でも、「海で」という文節は「かけていた」と「落とした」のどちらの文節にも係ることができる。ただし、係り受け関係を求める際には、それら候補のうちからただ一つのみを選択することになる。よって、異なる候補を選んだ場合では文の係り受け関係もまた異なる。それゆえ、複数の係り受け関係が存在することになる。

このことから、係り受け関係を解析し、一つの文に対して複数の係り受け関係が求まるならば、読み手が書き手の意図を誤解する可能性がある。複数の係り受け関係が存在する原因は、複数の係り先候補を持つ文節が文中に存在するからである。よって、そのような文節を含んでいれば、その文は誤解される可能性があると言える。

文単独では誤解される可能性があっても、文脈を用いた判断で一意に解釈できる文も存在する。しかし、この判断は読み手の読解能力に依存する。文脈を正しく把握できなければ、読み手はこの判断を行うことはできず、文を誤解する可能性がある。このことから、文脈を用いると一意に解釈できる文でも文単独では誤解される可能性があるれば、本研究では誤解される文として扱う。

3. 係り受けを誤解される文の抽出

3.1 従属句の階層構造に基づく係り受け制約

日本語文には、物事を客観的に表現する部分(叙述)と、それに対する書き手の心情を表現する部分(陳述)が存在する。これらの分類は主に従属句に重点を置いているので、従属句に注目することで、日本語文の構造を分析できる。

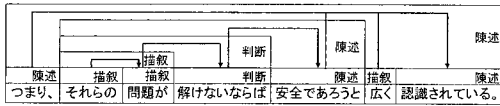


図 2: 階層構造と係り受け

南^[3]は文の構造解析の手がかりとして従属句の内部構造に注目し、以下に示す3段階の分類を行っている。日本語文の構成要素はこれらのいずれかの段階に属する。この分類は文中での出現順序や位置に関わらず、含まれる要素の違いのみによって分類されたものである。この3つの段階には描叙段階を最下位、陳述段階を最上位とする位が設けられる。

陳述段階 推量、意志などを含んだ表現で、文末や事実的逆接条件を要素とする。

例)「彼の薦めで参加したが」「晴れるだろう」

判断段階 打ち消し、過去、丁寧などを含んだ表現で、主語、場所や時間の修飾句、事実的順接条件、仮定的条件を要素とする。

例)「風邪をひいたので」「京都には泊まらずに」

描叙段階 対象の客観的記述のみの表現で、格助詞、状態や程度の副詞などを要素とする。

例)「手をつないで」「かきまぜながら」

陳述段階は判断段階の要素を自身の要素として持つことができる。同様に、判断段階も描叙段階の要素を自身の要素として持つことができる。このように、ある段階はそれより下位の段階の要素を自身の要素として持つことができる。さらに、上位段階は自身の前に出現する下位段階を、まとめて要素として扱う性質を持つ。複数の要素をまとめて一つの要素群として扱う場合には、この要素群の段階は、要素群中で最高位の要素と同じものになる。

この「AがBを要素として持つ」という構造は、係り受け関係では「BがAに係ることができる」ということを意味する。反対に、AがBを要素として持てない場合には、BはAに係る

ことができない。このことから、係り受け関係に(1)上位の段階から下位の段階への係り受け関係は存在しない、(2)自身より上位の段階を越えて係ることはない、の2つの制約を与えることができる。

図2の文では、係り受けの基本則だけで判断すると、「問題が」が「安全であろうと」や「認識されている」にも係ることができるが、不自然な解釈だと感じられる。しかし、この制約によってこの2文節は「問題が」の係り先から除かれる。このように、従属句の構造に合わず、読み手が不自然に感じる係り受けを、この制約によって取り除くことができる。

3.1.1 文節のランクと係り受け制約

上で述べた階層構造と係り受け制約を、係り受け解析に導入するために、福本らのランク付け^[4]を参考にし、南の分類に準じる形で構造に応じたランクを文節に与える。本研究で用いるランクを表1に示す。Cが下位、Aが上位のランクである。

読点は、書き手が文の区切りを明確にしたい所に付けられる^[5]。この読点の性質に着目し、文節のランクをさらに細かく分け各文節に与える。A(B, C)ランクの文節に読点が付いたものにはA'(B', C')ランクを与える。読点が文の区切りを示し、読点を含む文節が遠方に係る傾向があることから、これらのランクはCを最下位とし、C', B, B', A, A'の順に上位にあるものとする。前節で示したものと同様、これらのランクに下に示す制約を導入する。

制約1 上位ランクの文節から下位ランクの文節への係り受け関係は存在しない

制約2 自身より上位ランクの文節を越えて係ることはない

3.1.2 文読解モデルによる解析

係り受け解析は以下の手順で行う。解析対象文はCaboCha^[6]により形態素情報を与えられ、文節で区切られているものとする。文頭の文節を係り候補として手順を開始する。

(1) 係り候補の直後の文節を受け候補にする。直後に文節がなければ解析終了。

表 1: 文節のランク

ランク	構成要素
C	用言連体形, 連体詞, 体言接続助詞, 程度・様態の副詞, 形容詞・形容動詞の連用形, 格助詞など
B	動詞連用形, 仮定的条件句(なら, ば, たら, ても), 事實的順接条件句(ので, ため)
A	事實的逆接条件句(のに, けれど), 接続詞, 用言終止形

- (2) 候補の 2 文節間に係り受け関係が存在するか判定。存在しなければ現在の受け候補を新しい係り候補にして手順 1 へ。
- (3) 係り候補に係る文節がなければ, 係り候補の直前の文節を新しい係り候補にして手順 2 へ。直前の文節が存在しなければ現在の受け候補を新しい係り候補にして手順 1 へ。
- (4) 係り候補に係る最も前方の文節を新しい係り候補にする。
- (5) 候補の 2 文節間に係り受け関係が存在するか判定。存在しなければ手順 3 へ。
- (6) 文節 B に係っていた文節 A の係り先が文節 C に変更される場合, B に係る文節のうち, A の次に前方にある文節 D が BC 間の文節に係るかどうかが調べる。この処理は係り先の変更が生じなくなるまで再帰的に呼び出される。処理が終了したら手順 3 へ。

この手順は読み手の読解の流れを模して、取りうる全ての係り受け解を求める。手順 2, 5 では前述のランクによる制約も含めて係り受け関係の存在判定を行う。

3.2 従属句の名詞句化規則を用いた絞込み

前節のランクによる制約を受けないが、人が読むと複数の候補から係り受けを一意に決定できる文がある。下の文はそのような文の例である。

例文 2) 時系列上での楕円球の動きを求める

上の文では、「時系列上での」が「楕円球の」と「動きを」の二つの係り先を持つ。しかし、人が読む際には「時系列上での」は「動きを」だけに係ると捉えられる。

この例文は、「時系列上での楕円球の動き」と「動きを求める」の 2 文を 1 文にまとめたもので

ある。このうち、前者は従属句「時系列上で楕円球が動く」を名詞句化したものである。このように、名詞句の中には従属句を変化させてできるものが存在する。

従属句から変化した名詞句は複数の係り受けを持つことがあるが、読み手はこの名詞句を一意に解釈できる。これは、読み手が名詞句から元の従属句を推測しているためである。元の従属句における係り受けは名詞句になっても変化しないため、読み手は従属句での係り受けを基に名詞句の係り受けを求めている。

3.2.1 名詞句化の変化規則

従属句を名詞句に変化させるには、まず文末の用言を体言化する。用言が一般動詞の場合には、連用形止めが名詞の形になる。サ変接続動詞は、語尾の「する」を取り除けば体言化できる。また、形容詞、形容動詞は語幹に「さ」を加えることで体言化できる。

次に、この用言に係っていた文節を連体修飾化する。この文節が用言連用形であれば連体形にする。機能語が格助詞の場合には以下の 3 通りに分けられる。

- **ガ, ヲ格** 体言接続助詞「の」に置き換え
- **デ, ヘ, カラ, ト, マデ格** 元の格助詞 + 「の」
- **二格** 用途に応じて(「で」「へ」「から」+) 「の」

これらの変化規則を用いることで名詞句から元の従属句を推測することが可能である。

3.2.2 係り受け解の絞込み

従属句が名詞句に変化しても、文節間の係り受け関係は変化しない。よって、名詞句内に複数の係り受け解が存在する時、従属句の形を推測

できれば、従属句における係り受け関係を用いて冗長な解を取り除くことができる。

従属句から変化した名詞句を探すには、形態素解析で得られる名詞のうち、(1) サ変名詞であるもの、(2) 末尾がイ段またはエ段のひらがなである一般名詞で、それをウ段に置き換えると動詞になるもの、(3) 末尾が「さ」である一般名詞で、「い」に置き換えると形容詞になるものの3種類を手がかりにする。

(1) 格助詞＋「の」＋用言の体言化 格助詞＋「の」は元の形が格助詞のみであり、従属句では用言に接続する。よって、係り先候補に体言化した用言の文節があれば、その文節に優先的に係ると考えられ、他の文節を係り先とする係り受け解を取り除く。

(2) 「の」＋用言の体言化＋「の」＋用言の体言化 このような構造の名詞句には以下のようなものがある。

例文 3) エンドホスト同士の接続先の選択

従属句) エンドホスト同士を接続する ＋ 接続先を選択する

「エンドホスト同士の」は「接続先の」と「選択」の二つの係り先を持つ。ただし、元の従属句の形を推測すると、前方の体言化した用言「接続」までの従属句と、それ以降の要素による従属句の二つが求められる。この二つの従属句の中で、「エンドホスト同士」は「接続」までの従属句に含まれ、この従属句では「接続する」に係る。よって、元の名詞句でも「接続する」が体言化した「接続先の」にのみ係ると判断できる。

体言化した用言が連続して出現する場合、その前にある連体修飾文節は前方の体言化した用言とともに従属句を構成する要素であると捉えられる。よって、このような構造の場合には第1文節は第2文節に係ると考えられ、他の文節を係り先とする係り受け解を取り除く。

3.3 知識情報を用いた絞り込み

前節までは、構文情報のみを用いた文の係り受け解析である。しかし、我々は実際に文を読む際、構文情報のみを用いて読解を行っているわけではない。持っている知識や経験で判断し、複数

の解釈候補の中から不自然と思われる解釈を取り除いている。よって、知識や経験といった知識情報を用いて文の解析を行い、誤解される文かどうかを判定する。これらの知識情報は読み手によって異なるが、多くの読み手が共通して持つ知識情報のみを用いれば、不特定の読み手を想定した解析だと言える。

共起とは複数の要素が一定の範囲内で同時に出現することである。複数の語が共起する時、これらの語の間には何らかの関連性があると考えられる。よって、共起関係にある2語間には共起関係にない2語間よりも係り受け関係が存在しやすい。ある文節に対し共起関係にある係り先と共起関係にない係り先が存在した場合、後者との係り受けは前者との係り受けと比較して不自然に感じられる。そこで、この場合には共起関係にない係り先とは係らないとして解析を行うことで冗長な解を除去することができる。

共起関係の知識を獲得するために、京都大学テキストコーパス⁷⁾を用いる。このコーパス内で係り受け関係にある2文節の主辞を共起関係にある単語対として学習する。ただし、この単語対は名詞と用言の対のみとし、格情報も同時に獲得する。

獲得した共起情報は、2文節間に係り受け関係が存在するという情報である。しかし、2文節間に係り受け関係が存在し得ないという情報はコーパスから学習できない。よって、共起情報に含まれない2文節間にも係り受けが存在する可能性があると言える。そこで、共起関係が学習された係り受けを、そうでない係り受けより信頼度が高い係り受けとして扱う。

また、読解の流れや短期記憶などといった要素により、係り受けは文節間の距離が短いほど係り受けしやすい性質がある¹⁶⁾。このことから、以下のように係り受け解を絞り込む。

絞り込み条件 複数の係り先候補がある時、文節間距離が短い候補との共起情報が存在し、文節間距離が長い候補との共起情報がなければ前者の候補にのみ係り受けを許可する。

3.4 三連続名詞句に対する処理

従属句からの変化形を含まない名詞句では、一意に解釈できる場合でも、複数の係り先を持つ

文節に対し、構文情報を用いて係り先を決定することができない。これは、複数の係り先候補から係り先を決定する構文的特徴が見つからないためである。

このような名詞句の中に、『体言』の+「体言」の+「体言』の形で表現できる三連続名詞句がある。この表現において「+」で区切られている部分は、全て単独の文節である。この三連続名詞句は、先頭文節の係り先候補が二つ存在し、構文情報では係り先を一意に決定できない。しかし、三連続名詞句は、一意に解釈できる構造だと考えられる。そこで、三連続名詞句の後方2文節をまとめて1文節とみなすことにし、三連続名詞句は係り受けが一意に定まるものとして処理を行う。

4. 誤解される原因文節の指摘

4.1 誤解される原因

文中に複数の係り先候補を持つ文節(以下、原因文節と呼ぶ)が存在すると、文に複数の係り受けが存在することになる。この原因文節は1文中に1つとは限らない。原因文節を指摘する場合、誤解を与える影響度や誤解の深刻度に基づいて順位を定め、この順位が高いものから書き手に指摘することでより効果的な指摘が行える。本研究では、下の基準で原因文節を分類し、この3つの要素の誤解に対する重要度を(1)>(2)>(3)としている。

(1) 述語に近いところに係りうる原因文節

述語に近いところに係ることは、文の主題に近い要素であると捉えられる。主題に近い要素が別の係り先候補を持てば、主題近くの解釈を変えることになる。よって、複数の解釈間には大きな相違が生じる可能性がある。

(2) 最初に複数の係り先が現れる原因文節

読み手が文を読み進めていく中で、最初に複数の解釈の可能性に気付く部分である。それまで一意だった文の解釈が複数に分かれるため、以降の解釈に対し疑念を抱ききっかけになると考えられる。

(3) 多くの係り先候補を持つ原因文節

多くの係り先候補を持つことで多くの解釈が可能になる。そのため、他の原因文節よりも書き手と読み手の解釈が異なりやすい。

4.2 原因文節の順位付け

前節で述べた要素の重要度を基に、次のような手順で原因文節に順位を付ける。

手順1 原因文節の高さ比較

複数の係り先を持つ文節に対し、文節の高さを求める。この高さは係り受け木における根(述語)からの距離である。係り受け木は文の係り受け構造によって形成されるので、文の係り受け解の数だけ木ができる。原因文節の高さを複数の木の中で最も低いもので定義する。こうして求めた文節の高さを比較し、低い文節から順位を付ける。これにより、述語の近くに係りうる原因文節ほど高い順位が与えられる。

手順1では文節の高さが等しい場合も考えられる。高さが等しい文節間の順位付けは以下の手順2を用いる。

手順2 文読解モデルの利用

3.1.2節で述べた文読解モデルでの解析時に、複数の係り先候補が存在する文節に番号を与える。この番号は、複数の係り先候補が出現した順番とする。各原因文節の番号を比較し、番号が小さい順に順位を付ける。

4.3 複数の原因文節

前節までは誤解される原因が複数の係り先候補を持つ文節のみにあるとして、原因指摘の手法を述べた。しかし、自身の係り先は一意に決定されるものの、誤解に関わる文節も存在する。

例文4) 教授が学生に図書館司書が貸した珍しい古文書を見せた。

上の例文を読むと、「学生に」が「貸した」ではなく「見せた」に係ると判断するのが自然だと考えられる。これは、「教授が」の係り先が不定のまま、同じ格助詞を持つ「図書館司書が」が現れたため、「図書館司書が」より前の部分をひと

表 2: 予稿 1000 文に対する抽出実験結果

用いた手法	誤解されると感じる文		誤解されないと感じる文	
	抽出数	非抽出数	抽出数	非抽出数
係り受けの規則のみ	213	0	331	456
従属句の制約導入	213	0	305	482
名詞句の絞込み導入	213	0	261	526
知識情報, 三連続名詞句の処理の導入	196	17	155	632

かたまりとして捉え、「教授が」と「学生に」の 2 文節が、ともに「見せた」に係ると判断するためである。

このように、ある格助詞を含む文節 (A と呼ぶ) に対し、A の係り先が不定のまま同じ格助詞を持つ文節が出現すると、2 文節の間にある連用修飾文節 (B と呼ぶ) は文節 A の働きによって最近傍の係り先候補に係りにくいと判断される⁸⁾。書き手が文節 B の係り先を最近傍の候補と考えて文を書いた場合には、文節 A によってより誤解されやすくなっていると言える。よって、誤解される原因文節として文節 B を指摘する際には、B による誤解に影響を与えている文節として、文節 A も同時に指摘する。

5. 実験

5.1 NAIST コーパスからの抽出

本手法には、誤解される文の抽出漏れを小さくすることが求められるため、抽出漏れを犯しにくいことを調べる実験を行った。この実験には、NAIST コーパス⁹⁾中の文のうち、複数の係り受け情報を与えられた文を用いる。これらの文は、読み手から見て複数の解釈がある文だと考えられる。NAIST コーパスでは、関係性を持つと思われる主辞対に対して、格情報を含めたタグを付与している。このタグ付けは係り受けの基本則に関係なく行われており、文前方への係り受け、他の文中との係り受け、係り受けの交差が存在する。しかし、本実験では、係り受けの基本則を満たす 31 文のみを対象文として用いている。

実験の結果、対象の 31 文全てが誤解される文として抽出された。これにより、誤解される文の抽出漏れはなく、抽出手法は目的に合った、適切

表 3: 予稿 9639 文に対する抽出結果

抽出数	3271
非抽出数	6368

なものであると言える。

5.2 演習予稿からの抽出

次に、誤解される文とされない文の集合に対して誤解される文の抽出を行い、抽出精度について調べた。まず、文単独で誤解されるかどうかを書き手以外の人間が判断した文を対象にする。この文は、九州大学大学院システム情報科学府で行われている演習において提出された予稿から、1000 文を無作為に抽出したものである。各手法による抽出数の変化を表 2 に示す。

誤解されると感じる文の抽出数、誤解されないと感じる文の非抽出数に値が集まることが望ましい。それぞれの手法を導入することで、誤解されると感じないが抽出されている誤抽出文の数が減少している。この結果から、それぞれの手法が不自然な係り受け解だけを取り除いていると言える。しかし、抽出数の 44% が誤抽出であり、誤抽出文の割合が大きい。また、知識情報を導入することで誤解されると感じない文が抽出されなくなる。

また、同演習予稿 9639 文に対し、誤解される文の抽出を行った。結果を表 3 に示す。全体の 33.9% の文が誤解される文として抽出される。この抽出によって、全体の三分の一の文に対して誤解されるかどうかを見直す必要がある。書き手の推敲を支援することを考えると、この抽出数を減少させて、見直しにかかる作業量を小さくさせることが望ましい。

表 4: 第一位原因文節の評価

読み手と一致	18
読み手と差異	2

5.3 原因文節の順位付け実験

4節で述べた原因文節の順位付け方法に対し、その妥当性を検証するために実験を行った。本実験では、前節の1000文からの抽出で用いた文のうち、誤解されると感じ、かつ抽出された文を用いる。この中から20文を無作為に取り出し、対象文とする。対象文のどの部分が原因で読み手が誤解されると感じるかを予め調べる。その後、対象文に対し原因文節の順位付けを行う。順位付けで第一位原因文節と判断される文節と、読み手が誤解の原因と感じる部分を比較し、両者が一致しているかどうかを調べる。実験の結果を表4に示す。

読み手が考える誤解の原因と、順位付けによって第一位になる原因文節はほぼ一致している。ただし、一致しない文も2文存在する。この原因として、まず文頭の接続詞が挙げられる。この接続詞は、文節のランクではAランクに分類され、文末に係りやすい性質を持つ。よって、係り受け木では高さが低くなりやすく、高い順位が与えられる。一方、読み手はこうした接続詞を文と文の関係を表現する要素として捉えており、文単体の読解時にはさほど重要な要素と感じない。そのため、この接続詞が複数の係り先を持つ場合でも、その係り先の違いが解釈に与える影響は小さく、読み手は誤解される原因とは考えない。

6. おわりに

本稿では、文中に存在する誤解される文を抽出する手法を提案した。可能な係り受けをすべて見つけ、従属句の階層構造による制約や名詞句化規則を用いた絞込み、共起情報による絞込みなどを導入し、不自然な解釈になる係り受け解を取り除くことで、抽出精度を向上させた。さらに、誤解される原因となる文節を推定し、影響の大きさを順位付けして書き手に提示する手法を提案した。

本稿で提案する抽出手法では、誤解されると感じる文はほぼ抽出されるが、誤解されると感じ

ない文も多く抽出される。全体の抽出数に対し、誤抽出の割合が大きいと、書き直す必要がない文の中に書き直すべき文が埋もれてしまう。これでは書き手が書き直すかどうか判断する作業に大きな負担を与えることになる。よって、抽出精度をより向上させることが今後の課題である。

本研究では、誤解されるかどうかの判断を文単位で行っている。これは、文単位で誤解されない文を書くことができれば、文脈などの情報を用いて判断してもこの文は誤解されない文だと言えるからである。しかし、文単位では誤解される文の中には、文脈など、その文の外部で得た情報を用いることで一意に解釈できる文も存在する。このような文を誤解されない文と判断することができれば、より文章を意識した推敲を支援することができる。

参考文献

- [1] 菅沼 明, 山村 広臣, 牛島 和夫: “日本語文における名詞句の並列構造の推定およびその推敲支援への適用,” 情報処理学会論文誌, Vol. 38, No. 7, pp. 1296–1307, 1997.
- [2] 横林 博, 菅沼 明, 谷口 倫一郎, “係り受けの複雑さの指標に基づく文の書き換え候補の生成と推敲支援への応用,” 情報処理学会論文誌, Vol. 45, No. 5, pp. 1451–1459, 2004.
- [3] 南 不二男, “現代日本語の構造,” 大修館書店, 1974.
- [4] 福本 文代, 佐野 洋, 斎藤 葉子, 福本 淳一, “係り受けの強度に基づく依存文法一制限依存文法一,” 情報処理学会論文誌, Vol. 33, No. 10, pp. 1211–1223, 1992.
- [5] 白井 諭, 池原 悟, 横尾 昭男, 木村 淳子, “階層的認識構造に着目した日本語従属節間の係り受け解析の方法とその精度,” 情報処理学会論文誌, Vol. 36, No. 10, pp. 2353–2361, 1995.
- [6] 工藤 拓, 松本 裕治, “チャンキングの段階適用による係り受け解析,” 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842, 2002.
- [7] 黒橋 禎夫, 齋藤 由衣子, 坂口 昌子, “コーパス作成の作業基準,” 京都大学テキストコーパス, 1997.
- [8] 川崎 恵里子, “ことばの実験室,” プレーン出版, 2005.
- [9] 飯田 龍, 小町 守, 乾 健太郎, 松本 裕治, “NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション,” 情報処理学会研究報告, 2007–NL–177, pp. 71–78, 2007.