

## 日本語文章推敲支援における係り受けを誤解される文の抽出

小野 貴博 菅沼 明 谷口 倫一郎  
九州大学

誤解される文は書き手の意図が読み手に正しく伝わらないため、文章を推敲する際に書き手に提示して推敲を促す必要がある。本研究では、文節の係り受けが原因で読み手が誤解してしまう文の抽出法の構築を目指している。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、誤解される文とみなす。しかし、一般に数多くの係り受け候補が得られるため、その中から読み手の解釈として不自然な係り受けだけを振り落とす必要がある。そのために本研究では、文節のランク付けによる制約や名詞句における係り受けの絞込みを導入した。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。その結果について述べる。

### Extrantion of the sentences whose modification relation is misunderstood for a writing tool

Takahiro Ono Akira Suganuma Rin-ichiro Taniguchi  
Kyushu University

Since the sentences misunderstood by readers seem to make a communication gap between the author and them, it is necessary to inform him of these sentences in his document when he polishes it. In this paper, we describe a method extracting the sentences misunderstood by readers because of their modification relation. Our method considers that the sentence is misunderstood if it has a number of its modification relations. The basic rule discovers lots of solutions of modification relation of a sentence. To find natural modification relations for readers, we have brought in the restriction about the predicate phrase and the refinement of the modification relations in noun phrase. We have applied, furthermore, our method to a writing tool and done an experiment to evaluate it. We describe our experimental results.

#### 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ある。よって、計算機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思われる。

図1に我々が考える推敲支援の概要を示す。書き手が文章を計算機に渡すと計算機が解析する。計算機は文章内の書き直すべき文を指摘し、推敲に役立つ情報を提示する。こうして計算機

が書き手に情報を与えることで書き手の推敲作業を手伝う。ただし、計算機は推敲に有用な情報を書き手に提示するだけで、実際にどのように書き直すかは書き手に委ねられる。

推敲対象となる、書き直すべき文の一つに、複数の係り受けをもつ文が挙げられる。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。

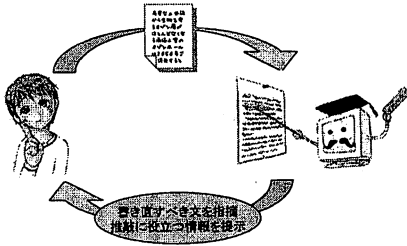


図 1: 計算機による推敲支援

係り受けは文中の各文節間の係り受け関係で構成される。各文節は後方にある文節のうち、文節の主辞や機能語、非交差条件などの制約を満たす一文節に係ることができる。しかし、これらの制約を満たす文節が複数存在する場合、係り先文節を一意に定めることができず、複数の係り受けが生じる。こうして求められる複数の係り受けの中には、読み手の解釈として不自然なものも含まれている。これは、読み手が文の構造や単語の意味なども用いて読解を行っているためである。よって、複数の係り受けを持つ文全てを誤解される文とみなすと、誤解されない文もその中に含まれてしまう。

そこで、従属句の構造に着目し、その構造に基づいた係り受け制約を導入する。また、一部の名詞句に対する係り受けの絞込み方法を提案する。これらによって、文が取りうる係り受けのうち不自然でないものを求めることができる。本論文ではこれらの手法を用いた誤解される文の抽出について述べる。

## 2. 従属句の階層的認識構造による係り受け制約

### 2.1 日本語文の階層的認識構造

日本語文には物事を客観的に表現する部分と、それに対する書き手の心情を表現する部分が存在する。前者を「叙述」、後者を「陳述」と呼び、その分類に関しての研究が行われてきた。これらの分類は主に従属句に重点を置いている。よって、従属句に注目することで、日本語文の構造を分析することができると考えられる。

南山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、3段階の分類を行ってい

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。

「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。

「風邪を引いたので」「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。

「手をつないで」「かきまぜながら」

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。

これらの分類は要素の違いによって分けられたものである。文中において各段階が現れる順番は決まっておらず、上位段階の直後に下位段階が現れることもあり、その逆もある。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。

また、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。

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。図2に文の構造と係り受けの例を示す。

これにより、係り受け関係に以下のような制約を与えることができる。

- 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
- 自身より上位の段階を越えて係ることはない

図2の文では、「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

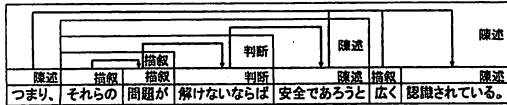


図 2: 階層構造と係り受け

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。

## 2.2 文節のランクと係り受け制約

福本らは各段階を係り受けのランクで表現し、同様の制約を導入した係り受け解析を行っている<sup>2)</sup>。このランク付けを参考に、本研究では、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。各文節のランクは表1の通りである。

Aが下位、C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で付けられる<sup>3)</sup>。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ランクの文節に読点が付いたものにはA'、Bランクの文節に読点が付いたものにはB'、Cランクの文節に読点が付いたものにはC'というランクを与える。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、このランクはAを最下位とし、A'、B、B'、C、C'の順に上位にあるものとする。

このランクを基に、前節で示したものと同様の制約を導入する。

- 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
- 自身より上位ランクの文節を越えて係ることはない

## 2.3 文読解モデルによる解析

前に述べた制約を用いて、文の係り受けを求める。係り受け解析は以下の手順で行う。解析対象文はCaboCha<sup>4)</sup>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。

- (1) 先頭文節を係り候補にする。
- (2) 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。
- (3) 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2へ。
- (4) 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3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2へ。
- (5) 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。
- (6) 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4へ。
- (7) 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4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3、6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. 名詞句に対する絞込み

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。以下はそのような文の例である。

例文 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「動きをを求める」の二つに分けることができる。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、

表 1: 文節のランク

A	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	実的順接条件句 (ので ため)		
C	事	実的逆接条件句 (のに けれど)	接	続詞	用	言終止形

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。

この例文のように、従属句から変化した名詞句は複数の係り受けを持つが、読み手は一意に解釈する。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。よって、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。

以下、従属句の推測法と解析への導入について述べる。

### 3.1 名詞句化の変化規則

名詞句から元の従属句を推測するために、従属句が名詞句化する際の変化規則②について述べる。

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。「～する」という形の動詞は「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」
- **ニ格** 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 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。

今回は、用言を体言化したものの中で形態素解析によって求められるサ変接続名詞を手がかりにする。

(1) **格助詞+「の」+サ変接続名詞** 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** 論理プロセッサ上でのキャッシュの共有  
**従属句** 論理プロセッサ上でキャッシュを共有する

「論理プロセッサ上での」は「キャッシュの」と「共有」のどちらにも係ることができる。しかし、元の従属句では「論理プロセッサ上で」であり、「キャッシュを」に係ることができない。よって、「共有」にのみ係ると判断できる。

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、係り先候補に用言が体言化した形であるサ変接続名詞があれば、その文節に優先的に係ると考えられ、他の文節に係り先とする係り受け解を取り除く。

(2) **「の」+サ変接続名詞+「の」+サ変接続名詞** 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** 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、元の従属句では「選択する」に対し「接続(する)先を」に係る。よって、従属句における「エンドホスト同士を」は「選択する」に係ることができず、「エンドホスト同士の」は「接続先の」に係ると判断できる。

「の」単独の場合、ほとんどが格助詞「ガ」「ヲ」の連体修飾化であり、これらは一つの文節に対し一つしか係らないという係り受け関係の制約を持つ。よって、第1文節は第2文節に係ると考えられ、他の文節に係り先とする係り受け解を取り除く。

#### 4. 抽出実験

これまで述べた手法を用いて誤解される文の抽出実験を行った。実験対象に用いた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無作為に抽出した1000文である。階層構造による従属句の制約、名詞句化の変化規則に基づく名詞句の絞込みを導入することで生じる抽出数の変化を調べ、比較する。表2に実験結果を示す。

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。従属句の制約、名詞句の絞込みを導入することで、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。

以下に5つの文を示す。これらは誤解されると感じず抽出されない文、誤解されると感じず抽出される文、誤解されると感じず従属句の制約により抽出されない文、誤解されると感じず名詞句の絞込みにより抽出されない文、そして誤解されると感じないが抽出される文の例である。また、それらの文に対する解説も加える。

**文1** 我々は歩行動作を再現するデジタルヒューマンを作成するにあたって、必要なデータを実験から得ている。

この文は誤解を生じない文である。解析でも係り受けが一意に求められるため、抽出対象にならない。

**文2** 修士研究では、タンパク質配列という一連の文字列から得られる数値情報を用いて、有用な特徴抽出を行うための手法について研究を行う。

この文では「タンパク質配列という一文字列」と「タンパク質配列という一数値情報」という

二通りの解釈ができ、誤解される文であるといえる。この文からは複数の係り受けが求められるため、誤解される文だと判断され抽出される。

**文3** 携帯型情報端末はユーザに常に持ち歩かれるので、誤動作しないことが重要である。

この文は誤解を生じないが、基本的な係り受け解析では「ユーザに」が「持ち歩かれるので」と「重要である」の二つの係り先を持ち、複数の係り受けが求められる。しかしランクによる制約を用いると係り先が「持ち歩かれるので」のみとなり、抽出されない。

**文4** 本演習では、空間データベースでの検索法に関する最新の文献を紹介する。

この文も文3と同様、誤解を生じないが複数の係り受けを持つ文である。ランクによる制約を用いても「空間データベースでの」には「検索法に関する」と「文献を」の二つの係り先が存在する。しかし、名詞句化の規則による絞込みで係り先が「検索法に関する」に決定され、係り受けが一つだけ求められるため、抽出対象にならない。

**文5** よりリアルな人間の動きを表現するために自然法則を活用する。

この文も誤解されると感じないが複数の係り受けを持つ文である。ランクによる制約、名詞句化の規則による絞込みを用いても係り受けを一意に決定できない。その結果、誤解される文と判断され抽出されてしまう。

誤解されると感じる文は全て抽出されるが、抽出された文における誤解されると感じる文の割合は44%であり、誤抽出文の割合が大きい。このような誤抽出の原因として次のようなことが挙げられる。

- 変化規則で絞り込めない名詞句  
名詞句化した従属句を探す手がかりとしてサ変接続名詞を用いたが、実際には動詞の体言化によって名詞化した従属句も存在する。これは形態素解析の結果からは判断が

表 2: 実験結果

	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	269	518

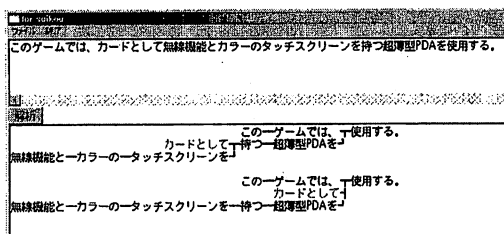


図 3: 解析結果の表示

できず、絞込みが行えないので、人が読む場合には一意に理解できる文でも複数の解が残る。また、従属句内の連用修飾要素が連体修飾化した文節を探索できずに絞り込めなかった文もある。

- 一般的と思われる知識によって誤解を生じない文  
読み手は各主辞間の関係を知識として持ち、読解に用いている。この知識には個人差があるが、一般的な語に関するものなどほぼ全ての人が共通して持っていると思われる知識も多くある。そのような知識を用いることで誤解を避けることができる文でも、本手法では誤解される文として抽出される。

## 5. おわりに

図 3 は文の係り受け解析を行った際の、解析結果を表示したウィンドウである。上のテキストボックスに文を入力し解析を実行すると、文が取りうる係り受け構造が下のウィンドウに示される。また、複数の係り先を持つ文節は他の文節と色を変えて表示している。これは、この文節が複数の係り先を持つことで文に複数の係り受けが生じることから、この文節を誤解され

る原因とみなしているためである。こうして、誤解の原因と解釈の構造を提示することで、書き手の推敲を支援する。

推敲支援の観点では、推敲対象文の抽出にとどまらず、推敲に有用な情報を提示することが望ましい。そのような情報の一つに、書き直すべき文の書き換え候補が挙げられる。取りうる解釈ごとに書き換え候補を示すことで、書き手はその候補を参考に文を書き直す。このような解釈ごとの書き換え候補をいかに生成し提示するかが今後の課題である。

## 参考文献

- [1] 南 不二男, “現代日本語の構造,” 大修館書店, 1974.
- [2] 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—制限依存文法—,” 情報処理学会論文誌, Vol. 33, No. 10, pp. 1211–1223, 1992.
- [3] 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.
- [4] 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
- [5] 寺村 秀夫, “日本語のシンタクスと意味 III,” くらしお出版, 1991.