助動詞に着目した従属節係り受け関係の解析について

浅野 純平　池原 悟　村上 仁一
鳥取大学大学院工学研究科
{asano,ikehara,murakami}@ike.tottori-u.ac.jp

1 はじめに

日本語の構文解析において述語間の係り受け関係の曖昧さは長文解析の精度を低下させる大きな原因となっている。日本語では述語の連結により長文になる傾向があるから、用言の相互関係を知ることが構文解析全体の精度の向上につながると期待できる。用言の相互関係に関して、南は従属節を3分類し、分類間の制約に関する性質を明らかにした[1],[2]が、分類には意味的分析を必要とするなど、構文解析に直接利用することは困難であった。これに対して、白井らは表層的な情報から計算機で処理しやすいよう長単位で従属節末尾を再分類するとともに、読点、述語の種類による係り受けの規則等を加味して、述語句を細分類する方法を提案した[3]。この方法によって、係り受け解析の精度は飛躍的に向上した。しかし、述節に含まれる助動詞の役割が必ずしもに十分考慮されていない。

従属節間の係り受け関係には話し手の認識から表現までの過程が階層として現れることが指摘されている。ここで話し手の認識が直接的に現れるのは助動詞であるため、その有無と用法は従属節間の階層構造を把握する上で有効と期待される。日本語文は主体的表現が客体的表現を包み込む入れ子構造を持つ。助動詞は用言に接続して対象への話し手の判断を示す機能を持つことから、述部で用いられた助動詞によって、述語間の階層構造がより明確になると期待できる。また、春野らは決定木を使った係り受け解析において、文節語尾の形、助動詞を含む文節タイプ、句読点の有無、他8種の属性に関する分析を行った[4]。その結果、係り受け関係にあるうち前方の文節のタイプの効果が高く、従属節の係り受け関係を決定するうえも重要な役割を果たすのは文節語尾と助動詞である可能性が高いと指摘されている。

そこで本論文では、従来の日英機械翻訳システムALT-J/E方式で係り先を決定したのち、助動詞の有無によって係り先を修正することで、助動詞の効果を検証する。

2 従属節の基本分類

(1) 対象範囲

日本語が長文になる場合に問題となるのが、主語や目的語などの一文になるための要素を持つことができ、比較的大きなまとまりになることができる述語句である。そこで、本論文で言う従属節とは以下にあげる範疇のものとする。

- 末尾の文節が連用形で終っているもの
  「自然に放し」、「制限することができ」、「避けられない情勢になってきたことに加え」、など
- 末尾の文節が接続助詞で終っているもの
  「繰り返されるが」、「吸い出して」、「知らないので」、など
- 末尾の文節が形式名詞のうち「際」、「ため」、「ところ」などで終っているもの
  「高まった際」、「進んでいたため」、「顕微鏡で調べたところ」、など

ただし、末尾が連用形、接続助詞で終わっていても後続の要素とのつながりが強く、分解することが妥当でない(間に要素が入り得ない)と思われる場合、ひとつの要素と考え従属節の末尾と考えないとする。

やらなく(てもよかった)、見えなく(なる)、
押し(つぶされたり)、など

また、用言の連用形または「用言＋テ、デ」の形であっても他の成分を伴わないで、意味、用法が状態副詞に類するものは、副詞としての役割が強いことから状態副詞として従属節としては扱わないとする。

大量に(放し)、手早く(塗る)、いばって(話す)、とんで(帰る)、など

(2) 対象とする従属節

言語処理の立場から節を分類すると、主節のほか、連用節、連体節、引用節の3種類の従属節に分けられる。このうち主節は、他の節の係り先にはなるが、

係り元にはならない。連体節は述語の活用形などの文法的性質によってほぼ係り先が決定される。また、引用節は引用の助詞等を伴い、また、係られる側の動詞のタイプが限定されるなど、形態的に係り先がほぼ決定できる。したがって、述語間の係り受け解析で問題となるのは、係り元が連用節である場合、すなわち、連用節から引用節、連体節、連用節への3つの係り受け関係となる。そこで、引用節の係り先は特定の用言に絞られ、連体節の係り先は埋め込み文の問題として、本論文では除外した。

2.1 従来の従属節分類

南は、従属節に現れる助動詞や助詞の種類の違いが、客体的表現から主体的表現に至るまでのいくつかの段階を明らかにするための言語上の特徴を表していると考え、従属節を次の3種類に分類した [1],[2]。白井ら [3] は南の考えを生かしながら、計算機処理を可能にするために語尾表現を長い単位でとりだして同様に3つに再分類した。

- A類 :「 ~ながら 」「 ~とともに 」等、同時の表現
- B類 :「 ~て 」「 ~ため 」等、「 原因 」「 中止 」の表現
- C類 :「 ~が (逆接)」、独立の表現

A類、B類、C類の順で優先度が高いとし、以下の規則にしたがって係り受け関係を決定している。

- 優先度の低い従属節は優先度の高い従属節に係る。
- 優先度の高い従属節は優先度の低い従属節には係らず、それを越えてより後ろの従属節、もしくは文末に係る。

その他に、読点、連用形の中止性、述語の状態性と動作性を用いて詳細な優先度設定を行っている。

読点の有無 : 同類同士の従属節の間では、読点の付与された従属節の方が、読点の付与されていない従属節よりも優先度が高い、つまり、従属節の優先度の大小関係は、A < A+読点 < B < B+読点 < C < C+読点、となる。

連用節の中止性 : B類、B類 + 読点、同士の従属節では、表現の意味的な流れの中止性の強弱により、以下の2種類に分類でき、中止性の強い従属節は中止性の弱い従属節よりも優先度が低い。

- 中止性の弱いもの : 用言連用形、「 ~て 」、「 ~ため 」など7種類
- 中止性の強いもの :「 名詞 + で 」、「 ~ており 」など4種類

述語の状態性と動作性 : B類、B類 + 読点、同士の従属節は、動作性の強い順に、他動詞性、自動詞性、形容詞性、名詞性の4種類に分類でき、動作性が強いほど優先度が高い。

引用節と連体節の扱い : 連用節から引用節への係り受けにおいては、「 ~すると (発表する)」などの引用節の優先度は「 C類+読点 」に準じ、「 ~するよう (依頼する)」などの引用相当節述語の優先度は「 B類+読点 」に準ずる。一方、連用節から連体節述語への係り受けにおいては、形式名詞に係る連体節述語の優先度は「 B類+読点 」に準じ、その他の通常の連体節述語の優先度はB類に準ずる。

以上の方法は日英機械翻訳システム ALT-J/E によって実現されている。

2.2 従属節の係り受けの定義

通常述語間の係り受け解析では、ひとつの要素の係り先はひとつとされているが、並列構造を含む文では、ひとつの要素が並列構造のそれぞれに対して係ることもありうる。

本論文では最も直接的に関係のある従属節同士を係り受け関係にあるとした。また、並列関係にあるものは、ALT-J/E と同様に並列関係にある節同士を係り受け関係にあるとし、並列構造以前の節が並列構造全体に係る場合並列構造の最初の節に係る、並列構造の係り先は最後の要素の係り先として表す。(図 1)

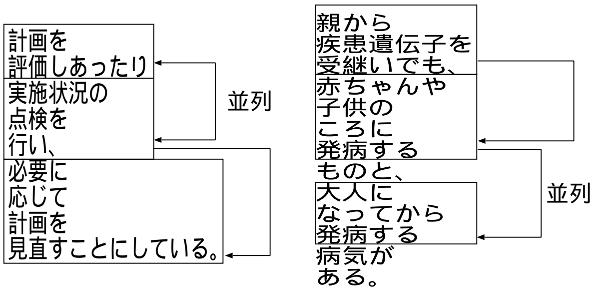


図 1: 並列構造での係り受け関係の定義

表 1: 係り先候補数と従属節の個数（括弧内は全体からの割合）

係り先候補数	1	2	3	4	5	6	7	8	9	10
従属節数	16(6%)	61(24%)	57(23%)	46(18%)	31(12%)	22(8%)	9(3%)	2(1%)	1(-)	1(-)

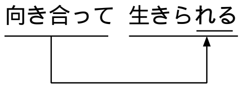
3　助動詞による規則

3.1　助動詞による規則

助動詞は、話し手の意思や判断を直接表現するもので、南の分類と従属節の包含関係は、従属節の認識の段階と包含関係を表しており、認識の段階が進んでいる従属節は認識の段階が進んでいない従属節を含むことができない (係ることができない) と言うことができる。したがって助動詞の有無によって同一分類の中でも表現の段階の違いを判断できる可能性がある。そこで、以下のヒューリスティックスを導入する。

(1) 従属節は助動詞を持たない同類の従属節を飛び越えず、それに係る。

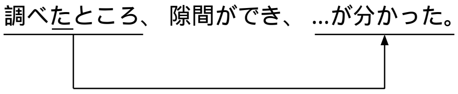
例文 1：



例文 1 の場合、直前の助動詞を持たない「向き合って」は、後方に助動詞を持つ「生きられる」に含まれる (係る) とする。

(2) 助動詞を持つ従属節は、助動詞を持たない同類の従属節を飛び越え、それ以降の従属節または文末に係る。

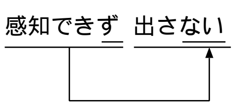
例文 2：



例文 2 の場合、過去の助動詞を含む「調べたところ、」は、直後の語幹連用形のみである「でき、」に含まれることができず、それ以降 (ここでは文末) の「...が分かった。」に係る。

(3) 助動詞を持った従属節同士であれば、係ることができる。

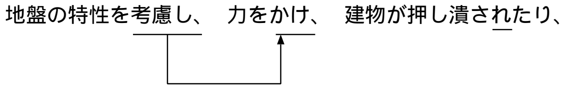
例文 3：



例文 3 の場合、「感知できず」、「出さない」のどちらも助動詞を含んでいるから、どちらも同程度の表現の段階にあるとし、「感知できず」は「出さない」に含まれるとする。

(4) ALT-J/E の決定した係り先は 1) ～ 3) の条件を満たさない限り飛び越えず、それに係る。

例文 4：



ALT-J/E の結果では、「考慮し、」は「(力を) かけ、」に係るとなっている。助動詞の効果を検証するために、助動詞による係らない条件に適合しない限り、「考慮し、」は「(力を) かけ、」に係るものとする。

これらのヒューリスティックスによって、以下の効果が期待できる。

- 計算機で処理する場合に意味による判断が必要で分類を正しく付与できないもののうち、長単位で取られたキーワードから漏れてしまった従属節に対して、正しい分類を行うことができる可能性がある。
- 同一分類内であっても、認識の段階の差をより明確にできる可能性がある。

4　実験結果と考察

4.1　実験の方法

95 年毎日新聞の経済欄、科学欄より、述部を 6 個以上含み形態素解析の失敗の無いものを合計 107 文（それぞれ 46 文、61 文）を抽出した。抽出した文に含まれる連用節 246 個（それぞれ 116 個、130 個）を対象に係り受け解析における助動詞の効果を机上実験により検証する。

対象とする文における連用節と係り先候補数との関係を表 1 に示す。これより、ランダムに係り先を決定した場合の誤り率は約 65%である。

今回、分類は人手によって行ったが、計算機による判断の難しいと思われる従属節は B 類とすること

とした。つまり、実際は引用節であってもALT-J/Eの結果で、B類の従属節が越えている個所は、「C類+読点」とせず、B類が越えている場合はB類、「B類+読点」が越えている場合は「B類+読点」として扱った。また、同一分類で助動詞により係るかどうかの判断できない場合(両方の従属節に助動詞のない場合)は係らないとした。

4.2 規則の検証

解析例を図2に示す。この場合、「考慮し、」はALT-J/Eの結果にしたがって「(力を)かけ、」に係るとし、「(力を)かけ、」は、正解が「(応力計算を)する。」であるのに直後の同じ「B類+読点」であって助動詞「れる」のついた「押しつぶされたり、」にかけるとしてしまう。また、「押し潰されたり」は、以後の「倒壊したり」、「しないかという」を越えて文末にかけてしまう(ALT-J/Eも同様)。「倒壊したり」は、直後の同じB類であり助動詞「ない」を含む「しないかという」にかける。(ALT-J/Eは文末にかけてしまう。)

規則を適用した結果、正解が誤りになる個所が20、誤りが正解になる個所が42あった。したがって、総合的には22個所についての誤りを修正する結果となり、ヒューリスティック適用後のALT-J/Eでの誤り率が48%だったことを考えると9%の向上となった。(表2)決定木による属性の導入の仕方と、試験文に違いがあるので完全に公正な比較にはならないが、決定木における前文節タイプ、後文節タイプの効果がそれぞれ9.34%、0.52%であった[4]ことを考えると、述語間の係り受け関係の決定における文節タイプの効果の多くは助動詞と接続助詞によるものではないかと考えられる。

規則を適用した結果、誤りになる個所は、語幹連用形や接続助詞「て」などから、「せる(使役)」、「れる(受け身)」に係る場合がいくつか見られた。この場合、使役と受け身の助動詞はA類でも現われることができるため、判断の段階の評価が助動詞の有無だけでは判断できなかったと考えられ、前文節と後文節での要素のつながりを検討する必要があると考えられる。

表2: 実験結果

	ALT-J/Eの結果	規則適用後
正解	127(52%)	149(61%)
誤り	119(48%)	97(39%)
合計	246(100%)	246(100%)

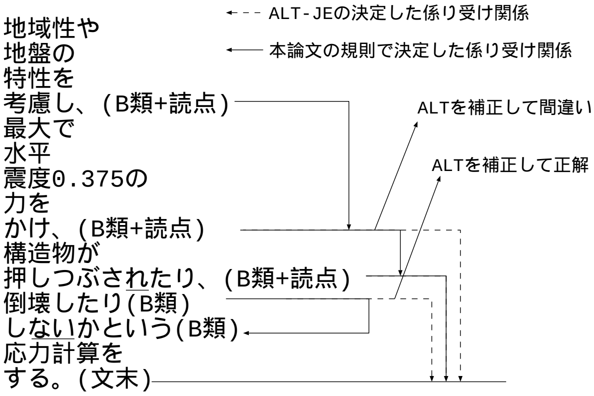


図2: 解析例1

5 おわりに

本論文では、従来のALT-J/E方式によって係り先を決定したのち、助動詞の有無によって係り先を修正することで、係り受け解析における助動詞の効果を確認した。その結果、助動詞によるヒューリスティックスによって、毎日新聞95年分記事(科学欄より61文、経済欄より46文、合計107文、総従属節数246個)を対象にした机上実験で、都合22個所(全体に対する割合9%)が改善され、助動詞の有無によって構文解析の精度向上の可能性があるという結果が得られた。

参考文献

[1] 南 不二男: 現代日本語文法の輪郭, 大修館書店(1993)

[2] 南 不二男: 現代日本語の構造, 大修館書店(1974)

[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,No10,pp.2353-2361(1995)

[4] Haruno,M.,Shirai,S.and Ooyama,Y.: Using Decision Trees to Construct a Practical Parser, COLING-ACL '98 Vol.1,pp.505-511(1998)