982002　名詞の意味を介した複合的連体節の構造解析について

計算機工学講座 池原研究室 浅野 純平

1　はじめに

日本語長文の構文解析の精度を向上させるためには述語間の関係の曖昧さを解消することが必要である。

述語間の関係のうち、問題となるのは連用節から連用節へ、連用節から連体節への係り受けである。このうち前者については、南の従属節分類 [2] を読点、動詞の種類などによって拡張した方法 [3] が提案され、述語間の係り受け解析の精度も飛躍的に向上したと言える。しかし、後者については、連体節が様々な形式の用言を包含することができるため、表層情報による規則では十分な効果が得られていなかった。

ところで、係り受け関係にある連用節と連体節では、格要素や底の名詞が動作主や対象として両者の用言と関係を持っていると考えられる。したがって、連用節と連体節の各々の動詞と格要素や底の名詞の意味の組合わせによって、係り受け関係を判定できる可能性がある。

そこで、本論文では用言の用法と要素となる名詞の意味に着目し、連用節から連体節への係り受け解析手法を提案する。

2　従属節の基本分類と問題の構造

ここで、連用節を含む連体節を複合連体節と呼ぶこととし、複合連体節を構成する可能性のある連用節と連体節のペアについて考える (図 1)。

一般に、連体節はさまざまな形式の用言を包含することができ、また、単独の用言は比較的自由につなぐことができるので、述部のみに着目した規則だけで係り受けの関係を判断することは難しい。

たとえば、図 1 の場合、「追跡し、」も「向けて」も B 類+読点にあたり、「測る」「開かれた」はどちらも B 類に分類される*。この場合、B 類+読点に分類される従属節は B 類の節に係ることはできず、従来の方法では正しく係り受け関係を判定できない。

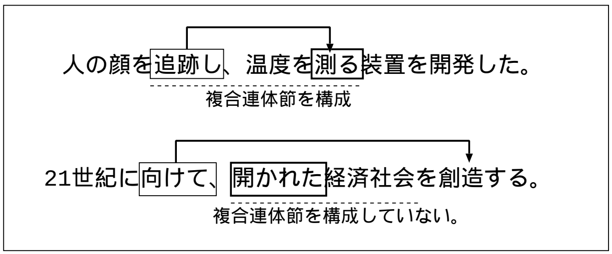

図 1: 対象とする連用節と連体節のペア

3　複合連体節の解析

ところで、連体節は、連体形の用言を末尾とし底の名詞を修飾する節で、係り受け関係にある連用節と連体節では、格要素や底の名詞が動作主や対象として両方の用言を繋いでいると考えることができる。したがって、それぞれの用言の要素となる格要素や底の名詞の意味や用法の組合わせによって、係り受けを判断できる可能性がある。

たとえば図 2 では、底の名詞「装置」が「追跡し、」「測る」のガ格としての役割を果たしていることから、係り受け関係にあると判断できる。また、この場合、「測る」のヲ格「温度」は、「追跡する」対象である「顔」の「温度」である相対関係であり、間接的に「追跡する」と「測る」を関係付けていると考えられる。

そこで、結合価パターンとのマッチングで取得した格要素と底の名詞との関係、用言の用法による係り受け規則を作成する。

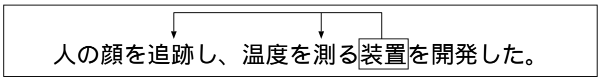

図 2: 要素と用言の関係

4　係り受け規則

4.1　結合価パターンと単語意味属性

本論文では、従属節を構成する格要素を判定し、用言の用法を決定するために日本語語彙大系 [1] に掲載されている「構文意味辞書」の結合価パターンを使用する。

結合価パターンは用言と格要素 (単語意味属性+助詞) の意味的関係を記述したもの (表 1) で、単語意味属性とは「単語の意味的用法を整理、体系化したもの」である。この単語意味属性体系は最大 12 段の木構造を構成しており、上位の意味属性の性質を下位の意味属性の性質に伝搬・継承できるという利点がある。

表 1: 結合価パターンの例

用言	結合価パターン　（）内は名詞意味属性
追跡する （身体動作）	（主体，乗り物）が （主体，動物，乗り物）を追跡する
追跡する （思考動作）	（主体，機械）が（情報，乗り物）を追跡する
測る （思考動作）	（主体，機械）が（名詞）を計る

4.2　結合価パターンの取得

まず、各々の用言に対して、格要素と底の名詞、また用言の用法を判断するために、結合価パターンのマッチングを行なう。これは、対象とする用言の標準表記をキーとして、辞書から結合価パターンの候補を読み出し、用例側の助詞と名詞の対応関係を照合することによって行う。この場合、誤った格要素は採用しないように、以下の制限を加える。

1 ハ格、モ格など、動詞を越えて係る可能性の高い格を除いて、着目する用言の直前までの格要素を候補とする。

2 重なる格があれば用言に近い方を候補とする。

こうして選択した格要素によって、結合価パターンを選択し、底の名詞が格要素として用言と関係を持てるかどうかを判定する。連体節述語の場合、底の名詞が格要素として動詞と関係を持てるのならば内の関係、持てないのであれば外の関係と仮定して、底名詞と用言の関係を含めたパターンを作成する。

---

*南 [2] によって提案された分類を細分類したもの [3] で、A 類 $<$ A 類 + 読点 $<$ B 類 $<$ B 類 + 読点 $<$ C 類という係り受け優先度がある。

4.3　用言意味属性と名詞意味属性による規則

複合連体節を構成する可能性のある連用節と連体節のペアに対して、用言意味属性と格要素、底名詞の一般名詞意味属性によって係り受けルールを作成する。

属性変化/属性変化
１．…映像が固定され、 …左右が逆転する弱点が…
２.…の部分を引き下げ、…税率を　引き上げた改正前の水準に…

図 3: 用言意味属性によるルール、属性変化の場合

図 3 の例では、同じ属性変化をあらわす動詞同士でも、1. の場合は「逆転する」のガ格「左右」が「固定され、」のガ格、「映像」の「左右 (抽象的関係)」をあらわし、「映像を固定する→ (映像の) 左右が逆転する」という因果関係をあらわしているのに対し、2. の場合は部分と税率の間にそのような関係が仮定できない。

したがって、前者の場合、「映像が固定され、」という節は「左右が逆転したりする弱点」という名詞句の一部として判断でき、後者の「部分を引き下げ、」という節は「税率を引き上げた改正前の」という名詞句の一部ではないと考えられる。このようにして、述語と格要素、底の名詞の意味属性によって、ルールを作成した。ルールの一部を表 2 に示す。

表 2: 作成したルール (一部)

判定結果	条　　件
係る	case_parent = 両方 (格要素に親子関係がある場合)
係らない	rear_yimi = 相対関係 (連体節述語の用言意味属性が「相対関係」である場合)

4.4　共通要素による規則

また、連用節述語と連体節述語が並列関係にある場合、それぞれの動詞には共通の対象や動作主などになる格要素、または底の名詞が存在すると考えられる。したがって、係り元と係り先候補の選択された結合価パターンによって、それぞれの用言に対し同じ関係にある要素があるかどうかを調べることで、並列構造の判定をすることができる可能性がある。

そこで、本論文では、以下のものを並列構造とし、係り受け関係にあるとする。また、通常、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こととする。

1 ひとつの格要素を共通に取るもの
2 底の名詞との関係がどちらも内の関係にあるもの。

5　実験結果と考察

5.1　実験の方法

毎日新聞科学欄、経済欄 (総数 97,912 文) から述部を 6 つ以上含み形態素解析の失敗のないもの 107 文を取りだし、その中から連体節述語と複合連体節を構成する可能性のある連用節述語のペア (総数 171) に対して作成したルールを以下の手順で適用する。

1 結合価パターンのマッチにより格要素と用言の組を得る (4.2 参照)
2 用言意味属性と名詞意味属性による規則で係り受けを判定 (4.3 参照)
3 残りのペアに対して共通要素により並列関係によって係り受けを判定 (4.4 参照)

表 3: 実験結果

分類	従来法		提案手法	
	正	誤	正	誤
係る　(63)	25(14%)	38(22%)	34(19%)	29(16%)
係らない(108)	90(52%)	18(11%)	104(61%)	4(2%)
合計　(171)	115(67%)	56(32%)	138(80%)	33(19%)

5.2　規則の検証

規則を適用した結果、連体節での正解率は 80.7%となる。日英機械翻訳システム ALT-J/E の連体節における誤りが誤り全体の 35%だったことを考えると、10%程度を改善できる可能性がある (表 3)。本手法の問題点としては以下の点が挙げられる。

1 結合価パターン選択の誤り
2 「もの」､「こと」などでの誤り

結合価パターン選択を誤る場合、名詞に対して用法にあった意味属性がふられていない場合が多い。また、「もの」､「こと」によって「～することを学び、」となるような場合、格要素の名詞による親子関係が正しく取得できない。また、今後の課題としては、以下のものがあげられる。

1 省略要素の文外照応
2 外の関係の取得

省略要素の文外照応に関しては、既存の技術の応用が期待できるが、外の関係の取得は困難である。ただし、「変質したりする (と言う) 危険」のような場合に限って、「～と言う」で制限されるような同格の関係であることから関係を推測できる可能性もあると考えている。

6　おわりに

従来の方法で解析の難しかった連用節から連体節の構造解析を格要素の名詞、底の名詞と用言の関係に着目して解析する方法を提案した。この方法を、毎日新聞科学欄経済欄から 107 文より連体節述語と連用節述語のペアに対して適用し、効果を検証した。その結果、正解率 80.7%となり、従来の方法に比べて、10%余りの改善の可能性があることが分かった。今後、ルールの増補などによって、正解率の向上を見込むことも可能であると考えている。

参考文献

[1] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林：日本語語彙体系, 岩波書店 (1997).

[2] 南 不二男 ： 現代日本語文法の輪郭, 大修館書店 (1993).

[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] 浅野純平, 池原悟, 村上仁一：助動詞に着目した従属節係り受け関係の解析について言語処理学会第 5 回年次大会,pp76-79(1999).

[5] 浅野純平, 池原悟, 村上仁一：名詞の意味を介した複合連体節の構造解析について言語処理学会第 6 回年次大会 (予定).