名詞の意味を介した複合的連体節の構造解析

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

日本語では従属節などの述語句が連結され長文となる傾向がある。そのため、日本語長文の構文解析の精度を向上させるためには従属節間の関係の曖昧さを解消することが必要である。

従属節は、主節の他に、連用節、連体節、引用節があり、こ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　複合的連体節

2.1　従属節の基本分類

従属節は主節の他に、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。引用節、連体節の係り先は、述語の活用形や係られる側の動詞のタイプなどによって比較的容易に決定することができる。また、連用節の係り先の問題のうち、連用節から連用節へ、連用節から引用節への係り受けは、従来の方法で比較的高い精度を得ることができるので、問題となるのは連用節から連体節への係り受け関係の判断となる。

そこで、図1のように、連用節を含む連体節を複合的連体節と呼ぶこととし、複合的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。

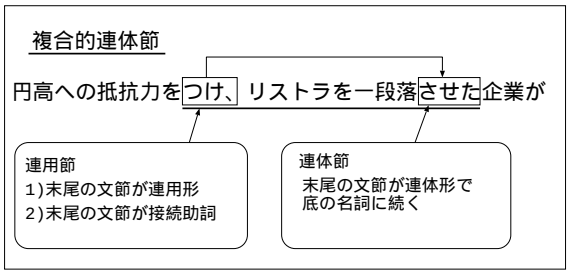

図 1: 複合的連体節の定義

2.2　問題の構造

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2の場合、従来の分類法では「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ないので、どちらの場合も連用節は連体節を越えることになり、従来の方法では正しく係り受け関係を判定できない。

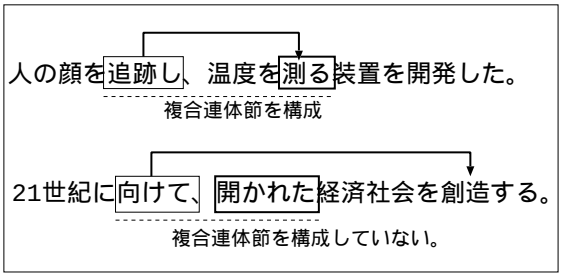

図 2: 対象とする連用節と連体節のペア

*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。

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の場合、格要素や底の名詞が動作主や対象などとして両者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 3 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。

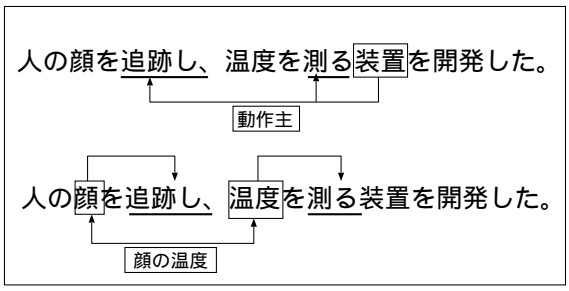

図 3: 要素と用言の関係

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。また、用言と格要素の組によって規定される用言意味属性も最大 4 段の木構造を構成しており、名詞意味属性と同様の性質を持つ。

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 (身体動作)	(主体, 機械) が (名詞) を測る

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と、用言の結合価パターンから、用言の用法と格要素の絞り込みを行う。この場合、用例の格が適合する結合価パターンの名詞意味属性の深さと、適合した格の数をスコアとした。

「人の顔を追跡する」という文であれば、「追跡する」には「身体動作」と、「思考動作」のふたつの用法がある。この場合「顔」の持つ意味属性のうち動物の一部としての「顔」という属性が「身体動作」の意味の「追跡する」のヲ格を取ることができる「動物」の下位属性にあたるので、「追跡する」の用法は「身体動作」であり、「顔を」はこの「追跡する」の格要素であると判断される (表 2)。

表 2: 結合価パターン選択

	ガ格		ヲ格		用言
用例	(なし)		人の顔		追跡する
結合価パターン	主体 乗り物	×	主体 動物 乗り物	○	追跡する (身体動作)
	主体 機械	×	情報 乗り物	×	追跡する (思考動作)

また、連体節の場合では、用法と格要素を選択したのち、選択された結合価パターンの中で空いている格に対して底の名詞を当てはめることで、底の名詞が格要素として扱うことができる内の関係と、そのときの格助詞を判断する。

こうして、各用言に対し、格要素、底の名詞との関係を含めたパターンを作成する。

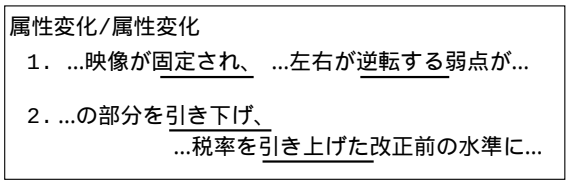

図 4: 意味属性による規則 (属性変化の場合)

4.3 用言意味属性,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。具体的には、各ペアに対して用言意味属性に加え、連用節、連体節の格要素と底の名詞の意味属性の組と「係る/係らない」によりテーブルを作成、「係る/係らない」を分けることができるような属性に着目する。その際、規則に使用する属性の組を絞るために決定木を使用した [4]。

図 4 の例は属性変化をあらわす動詞同士の場合で、1. の場合は「左右」が「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。これに対して、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、ここでは「属性変化」という用言意味属性と「左右」の持つ「抽象的関係」という意味によって、「係る/係らない」を分けることができる。このようにして、用言の意味と格要素、底の名詞の意味属性によって、全 26 個のルールを作成した。ルールの一部を表 3 に示す。

表 3: 作成したルール (一部)

判定結果	条　　件
係る	rear_yimi_3 = 属性変化,rear_*_2=抽象的関係 (連体節述語の用言意味属性が「属性変化」で 格要素に「抽象的関係」に属する名詞がある場合)
	case_parent = 両方 (格要素に親子関係がある場合)
	rear_yimi_3 = 結果 (連体節述語の用言意味属性が「結果」である場合)
	rear_* = 0,rear_yimi_1 = 行動 (連体節述語に格要素が無く、用言意味属性が 「行動」である場合)
係らない	rear_yimi_3 = 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)
	fore_ga_2 = 具体物,fore_yimi_3 = 因果関係, (連用節述語の用言意味属性が「因果関係」で 連用節述語のガ格に「具体物」に属する名詞が 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、図 5 に示す場合を並列構造とし、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。

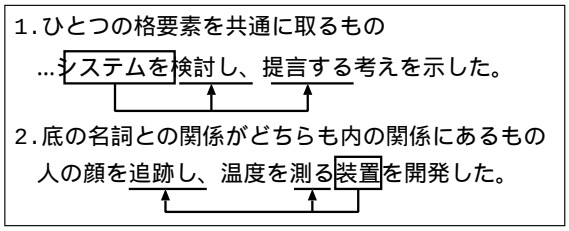

図 5: 並列となる場合

5　実験結果

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)

規則を適用した結果、係る場合を正しく係ると判定する場合が 25 件から 34 件に、係らない場合を正

表 4: 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%)

しく係らないと判定する場合が 90 件から 104 件に増え、全体での正解率は 80.7%となった。また、従来の手法で正しく解析できるが、本手法で誤る例は、係る場合の判定において 8 件、係らない場合の判定において 3 件で合計 11 件であった。解析例を図 6 に示す。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%余りを改善できる可能性がある (表 4)。

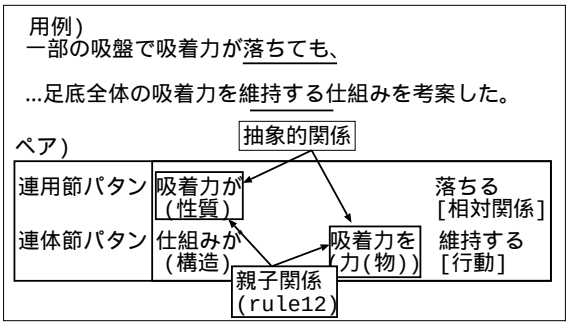

図 6: 解析例 (係る場合)

6 本手法の問題点と課題

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による誤り]

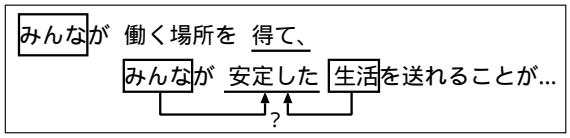

図 7: 結合価パターン選択誤り

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こと、格要素選択の誤りなどが原因として挙げられる。例えば、図 7 の場合、「安定する」の要素として格要素の「みんな」と底の名詞の「生活」が考えられる。実際は「生活が安定する」であるが、ここだけに着目して「みんな」よりも「生活」を結合価パターンで選ぶことは難しい。今回、「みんなが安定する」のパターンを選択してしまうことから「得る」のガ格「みんなが」との共通性から係ると判断してしまうような場合があった。今回は結合価パターンの格における名詞意味属性の深さでスコア付けを行ったが、このような場合深さによるスコアに違いが出ない場合も少なくない。ただし、一文全体の係り受けのスコアによって解決できるものもあると思われる。

[**2.** 形式名詞による埋め込み文での誤り]

また、図 8 のように形式名詞を使った格要素「 ~ ものを探り、」などを持つ場合、「問う」の格要素で

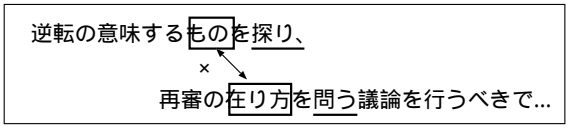

図 8: 形式名詞による埋め込みの場合

ある「在り方」と「問う」の格要素の間の親子関係などを正しく判断できない。ここでは「在り方」は「再審の在り方」で、「逆転の意味するもの」との関係は薄いが、今回の方法では「在り方」は「もの」の下位属性として扱われてしまい、誤る結果となる。

[**3.** その他の問題]

また、これらの問題の他に、今後の課題として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、用言の並列関係の判断にも効果があると思われる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。また、今回は 171 例という少ない用例での規則作成であったので、規則の増補も大きな課題と言える。

7 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] Quinlan,J.:AI によるデータ解析, トッパン (1995)