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

平成12年2月28日

目 次

概要

日本語では従属節などの述語句がつなげられ長文となる傾向がある。そのため、日本語長文の構造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。

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、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[5]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の組合わせに着目し、従来の方法では解析の困難だった連用節から連体節への係り受け規則を作成した。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、効果を検証した結果、正解率80.7%となり、従来法に比べ全体で10%余りの改善の可能性があることが分かった。

1　はじめに

自然言語の構文解析の方法として、従来から多くの研究が行われ、様々な方法が提案されている。それらの中で、比較的自由な語順を取る日本語の構文解析の方法としては、省略などに強い係り受け解析の方法が適していると考えられる。

日本語では、長い名詞句を含む名詞句の並列、従属節などの述語句が連結されることによって長文になる傾向がある。したがって、構文解析において解析精度の低下をさせる大きな原因として、並列を伴う名詞句、及び名詞句間の関係の曖昧さと述語間の解析の曖昧さが問題となってきた。

並列を伴う名詞句については、文節間の類似度をマトリクスを利用して比較し抽出する方法が長尾らによって提案され[4]、解析精度の向上が得られる見込みとなった。

また、述語間の関係は、従属節の種類によっていくつかに分けることができる。従属節は、主節の他に、連用節、連体節、引用節に分けることができこのうち係り先の曖昧さが特に問題となるのは連用節から連用節へ、または引用節、連用節から連体節への係り受けである。

前者については、南の従属節分類[2, 3]を読点、動詞の種類などによって拡張した方法[5]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、述語同士は比較的自由に繋ぐことができ、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、述部のみに着目した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。また、連体節の場合、従属節の要素は要素数、種類のどちらもが多く、いずれの節内の要素も述語によって左右される。

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

2.1 従属節の基本分類と性質

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。

本論文で言う連用節とは以下にあげる範疇のものとする。

1 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」、「追跡し、」など
2 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」、「知らないので」、「引き下げて」など

ただし、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(放す)など

また、連体節は以下にあげる範疇のものとする。

1 末尾が連体形で終わり、底の名詞を修飾する
　「温度を測る(装置)」、「開かれた(経済社会)」、など

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、連用節から引用節、連用節から連用節への係り受けは、従来の方法で比較的高い精度を望むことができる。したがって、連用節から連体節への係り受けの解析が問題となる。

本論文では、図1のような連用節を含む連体節を複合連体節と呼ぶこととし、図2に示される「追跡し、」「測る」、「向けて、」「開かれた」のように複合連体節を構成する可能性のある述語と連体節のペアを対象とする。

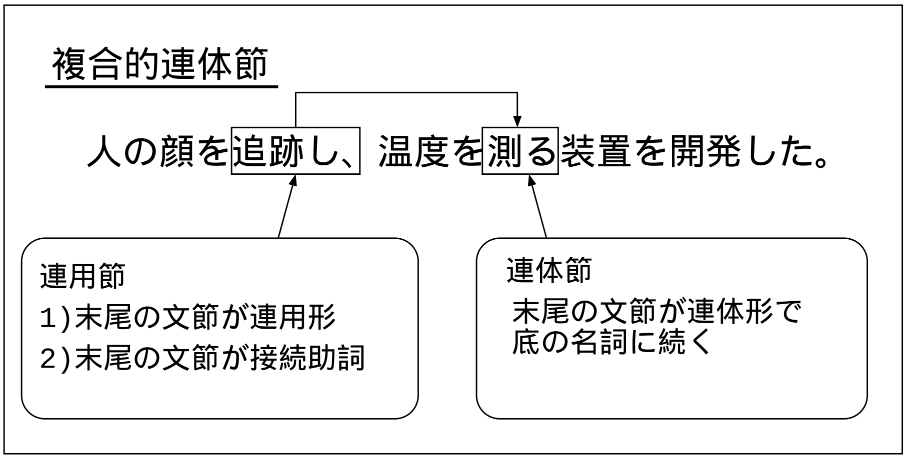

図1: 複合的連体節の定義

2.2 従来法の問題点 (誤り分析)

南の分類を拡張した従来の方法を用いた係り受け解析の方法の誤りを見てみると、表1のように連体節には係らないとして誤る場合が多く、誤り全体のうちの半数近くを占めている。

南の分類の場合、連体節が含むことができるのは名詞句の一部になることができるB類であるが、実際の文ではB類の出現数が多いうえ、分類に意味的な判断も必要なため、それだけでは判断が困難である。そのため白井らは南の分類を生かし、計算機処理を可能にするために再分類した [5]。白井らの分類によると一般名詞を底とする連体節はB類に属し、A類、A類+読点、B類以外に分類される連用節は係らないとなる。

しかし、読点は通常遠い係り先をさす場合に使われるが、連体節の場合、述語の並列や因果関係などをさすなど、述語を対比させて並べることもあり、近い係り受けでも読点が使われる場合も多い。また、接続助詞「 ~のに 」が、C類に分類される「 ~が 」「 ~けど 」で言い換えのできるような場合で使われることもある。そのことから、従来の述部の表層情報による分類では充分な精度を得られていなかった。

たとえば、図2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。

表 1: 従来法での誤り

分類	従来法	
	正	誤
係る (63)	25(14%)	38(22%)
係らない(108)	90(52%)	18(11%)
合計 (171)	115(67%)	56(32%)

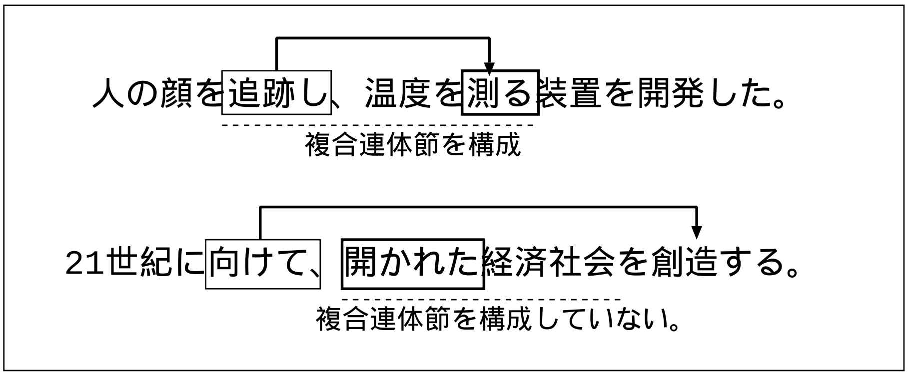

図 2: 問題の構造

---

[*] 南 [2] によって提案された分類を細分類したもの [5] 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

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の場合、底の名詞「装置」が「追跡し、」と「測る」の両者の用言のガ格としての役割を果たしている。したがって「追跡し、」は、「測る」に係り、共に底の名詞である「装置」を修飾しており、係り受け関係にあると言える。

また、この場合、「測る」対象である「温度」は、「追跡する」対象の「顔」の「温度」であると考えられ、これも間接的に「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。

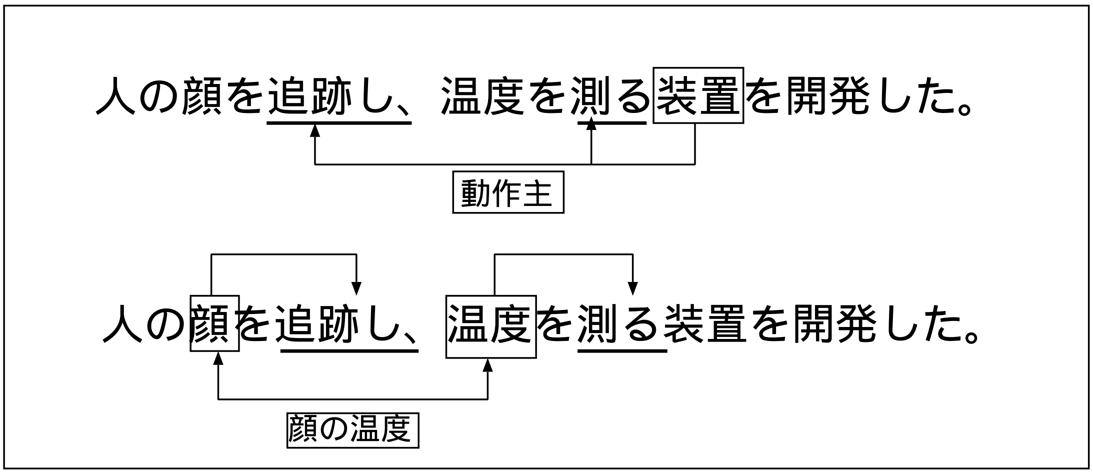

図 3: 要素と用言の関係

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 2) 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 (図 4)、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 (図 5)。

表 2: 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	用言意味属性
追跡する	(主体, 乗り物) が (主体, 動物, 乗り物) を追跡する	身体動作
追跡する	(主体, 機械) が (情報, 乗り物) を追跡する	思考動作
測る	(主体, 機械) が (名詞) を計る	思考動作

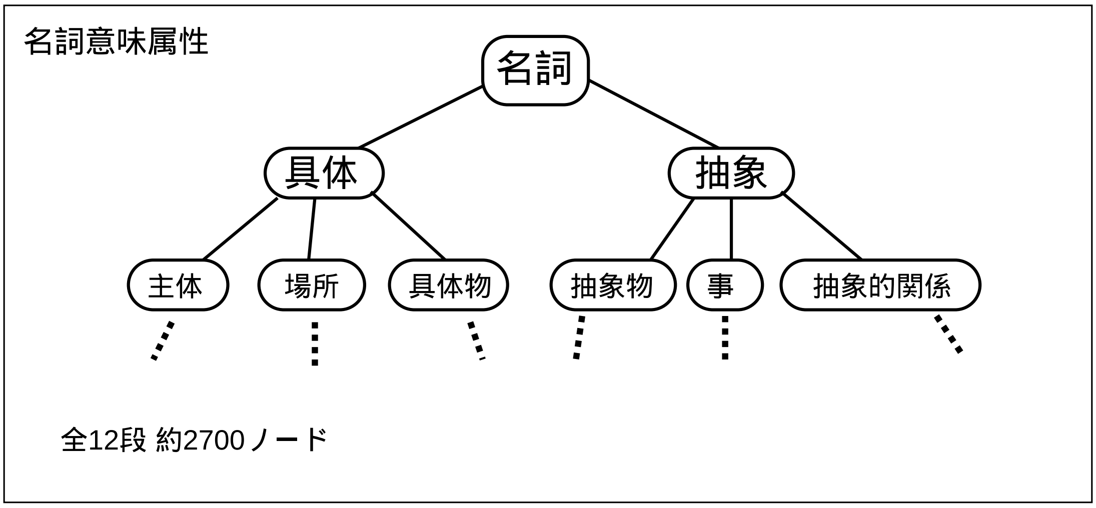

図 4: 名詞意味属性

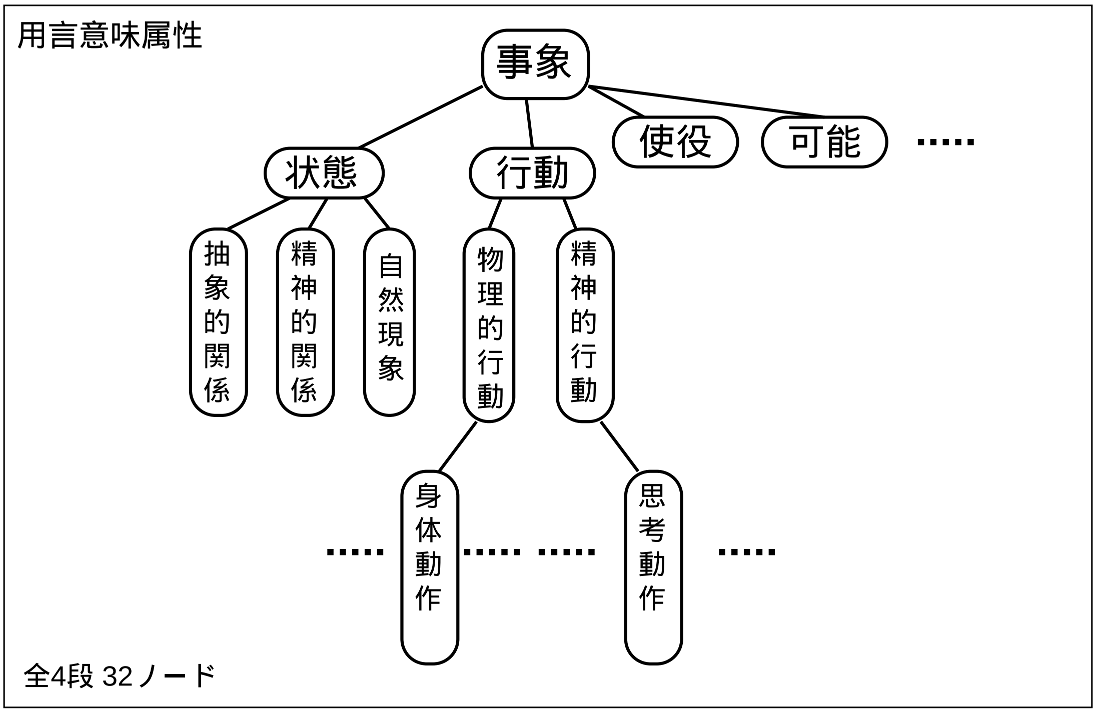

図 5: 用言意味属性

4.2 結合価パターンの取得

まず、各々の用言に対して、どの格要素と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ために、以下の制限を加える。

1 着目する用言の直前までの格要素とする。
×遠足に 行ったけど、○雨が 降ったので...

2 ただし、ハ、モ、など述語を越えて係る可能性が大きい (京都大学のコーパスで調査の結果、直近の述語に係る確率が 90%未満) ものは、用言を隔てていても格要素の候補として挙げるとする
○僕らは 、×遠足に 行ったけど、○雨が 降ったので...

3 重なる格があれば用言に近い方を採用する。
僕らの ×自由を 、僕らの ○青春を 大げさに言うのならば...

このような制限を加え、候補となる用例中の格要素と結合価パターンとのマッチングを行い、用法と用言の取る格要素を判断する。この際、結合価パターンにおける、格要素が適合した格の名詞意味属性の深さをスコアとすることで、より具体的な名詞を

取ることのできるパターンを選択することになり、最尤なパターンを判断することができる。

また、用言が受け身、使役になっている場合については、用言の標準表記(終止形)である場合に取る格助詞と異なる格がある。このような場合は用例の格助詞を標準表記に合わせ、変換し用言と格要素の候補の対を作成する。今回は、以下の変換を行った(表3,4)。この変換については今回の用例での場合であり、他の用例での場合この限りではないと思われる。格助詞の変換についても課題が残っている。

表 3: 格助詞の変換 (使役)

		格助詞	
		変換前	変換後
使役	ニ格あり	ヲ ニ	ヲ ガ
	ニ格なし	ヲ	ガ

表 4: 格助詞の変換 (受け身)

		格助詞	
		変換前	変換後
受け身	ニ格あり	ヲ ニ ガ	ヲ ガ ニ
	ニ格なし	ヲ ヲ	ガ ヲ

例えば、「温度を測る装置」であれば、「測る」の標準表記は「計る」となり、結合価パターンには3つの用法が登録されている。用例の格要素のうち、「温度」が最初のパターンのヲ格「名詞」に当てはまり、「名詞」は深さ1のノードであるので、スコアは1となる。また、他のパターンに対しては当てはまる格がなく、この「温度を測る装置」の用法は「身体動作」である(1)の用法が選択される(表5)。

こうして選択した結合価パターンに対し、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。

前述の「温度を測る装置」であれば、先ほど選択された「測る」のパターンに対して、空いている格であるガ格に対し、「装置」を入れて「装置が温度を測る」と言える用法であるかどうかを判断する。この場合、「装置」にふられている「機械」として

表 5: 結合価パターン選択

	ガ格		ヲ格		用言
用例	（なし）		温度を		測る
結合価パターン	主体 機械	×	名詞	○	測る（1） （身体動作）
	人	×	生理(その他)	×	測る（2） （身体動作）
	主体	×	思想 意向	×	測る（3） （事象）

の意味が「測る」のガ格の「機械」に当てはまるので、「装置が温度を測る」というパターンとなる(表 6)。

表 6: 底の名詞の挿入

	ガ格		ヲ格		用言
用例	（装置）		温度を		測る
結合価パターン	主体 機械	○	名詞	-	測る（1） （身体動作）

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。

手順としてはまず、用言意味属性、各格要素の意味属性、底の名詞の意味属性によりテーブルを作成し、「係る」と「係らない」を分けられるような属性に着目し、規則を生成する。また、その際、規則に使用する属性の組を絞るために、決定木[7]を利用した。

図6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7に示す。

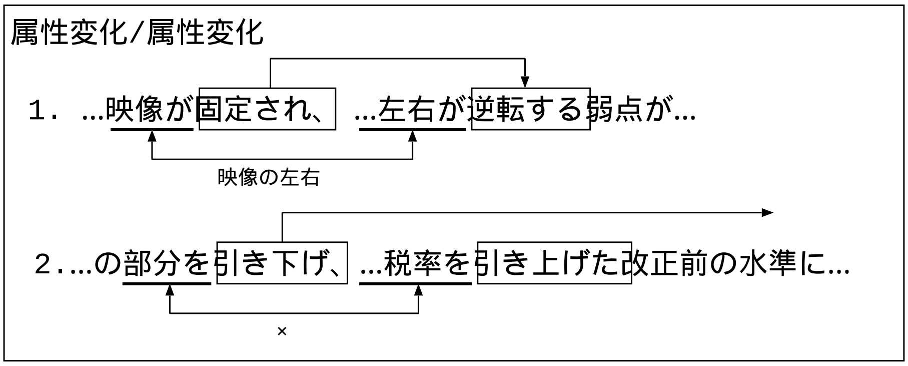

図 6: 用言意味属性によるルール, 属性変化の場合

表 7: 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

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、名詞意味属性と用言意味属性の規則に加え、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。

本論文では、図7のような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。

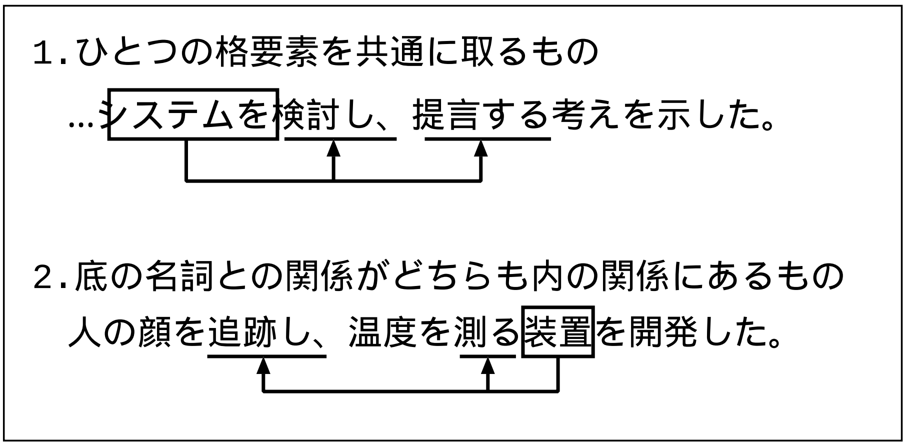

図 7: 並列関係にあるとするもの

図8の「追跡し、測る装置」の場合では、前文節の「追跡する」の取得した格要素と用言の組において、ガ格に相当する要素が欠ている。この空いている格に後文節である「測る」の要素である「温度」または「装置」が前文節のガ格として入ることのできる場合、その要素が「追跡する」と「測る」に共通の要素とできるものとする。この場合であれば、底の名詞である「装置」を「追跡する」のガ格として扱い、「装置が追跡する」と言える用法かどうかを結合価パターンの名詞意味属性によって判定することで共通の要素であるかを判定する。

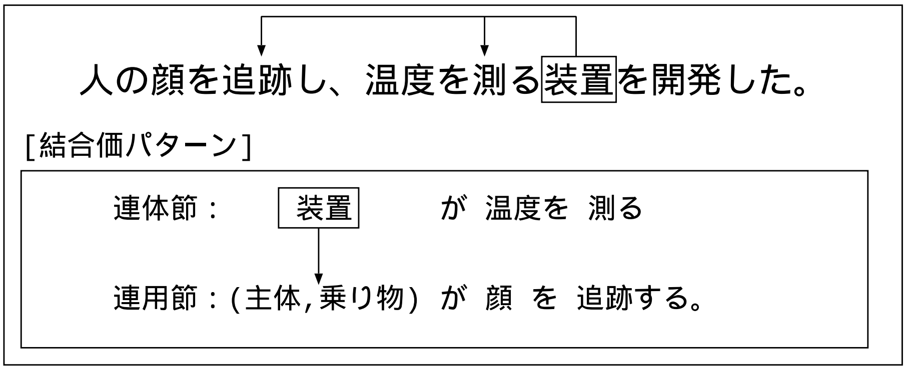

図 8: 共通要素のマッチング

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)

表 8: 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%)

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、本手法で解析を誤る例は、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 9 に示す。この場合、連用節「落ちても」の要素としては「吸着力が」が、また、連体節「維持する」の要素としては、「仕組みが」「吸着力を」が候補として挙がり、「落ちる」「維持する」の用言意味属性はそれぞれ「相対関係」「行動」が選ばれる。また各格要素の名詞意味属性は、「落ちる」の要素とされた「吸着力」が「性質」としての「吸着力」、「維持する」の要素とされた「仕組み」「吸着力」は、それぞれ「構造」「力 (物)」として格を埋めている。このうち、どちらの節にも存在する「吸着力」という名詞の「性質」「力 (物)」という意味属性は親子関係にあり、優先順位が 12 番にあたる「親子関係」の規則に適合し、「係る」と判断される。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 8)。

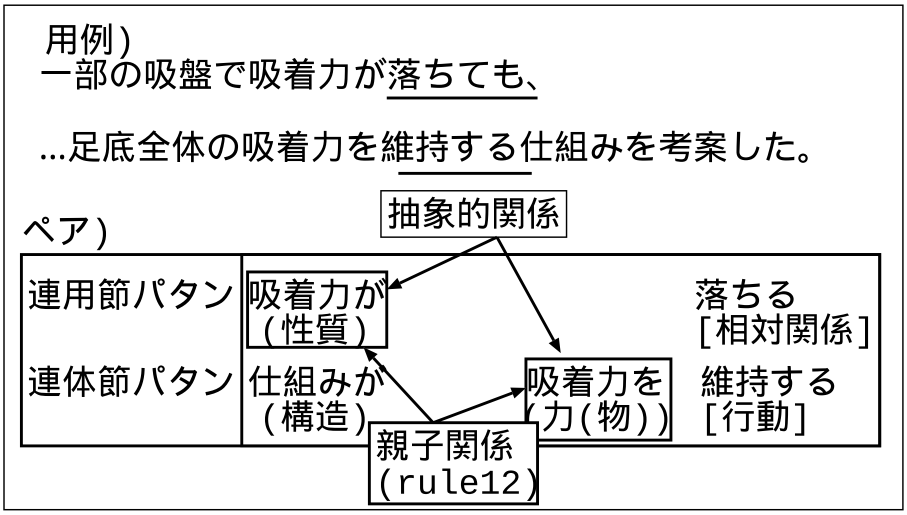

図 9: 解析例 (係る場合)

6　考察

6.1　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]

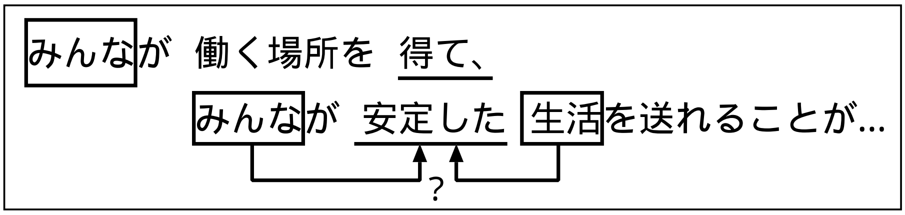

図 10: 結合価パターン選択誤り

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10の場合、「安定する」の要素として考えられるのは、格要素の「みんなが」と底の名詞の「生活」である。実際は「みんなが ~ 生活を送る」、「生活が安定する」ということを意味しているが、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。ここで「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適合する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大きな違いが出ない場合も少なくない。ただし、このような格要素選択、結合価パターン選択誤りは、一文全体の係り受けのパターンをスコア付けした場合、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]

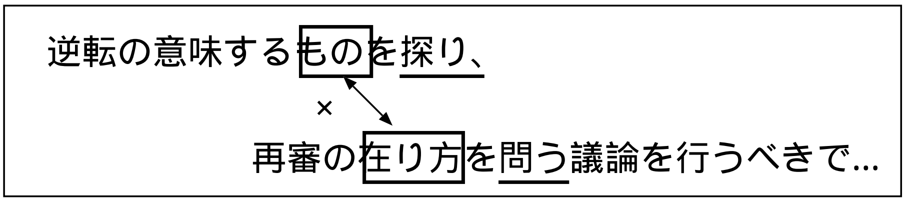

図 11: 形式名詞による埋め込みの場合

また、図12のような場合、「問う」の格要素である「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. 形式名詞による埋め込み文の場合 (その 2)]

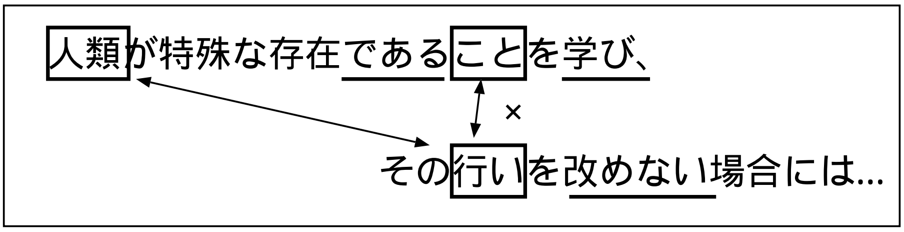

図 12: 形式名詞による埋め込みの場合

また、図12のように形式名詞を使った格要素「~ことを学び」などを持つ場合、「改める」の格要素である「行い」と「学ぶ」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「行い」は「特殊な存在である人類」の「行い」をあらわしているが、今回の方法では「こと」と「人類」、「存在」の関係を捕らえることは難しい。ここで、「場合」は、「学ばず、行いを改めない場合」であり、「こと」は「行い」の上位属性であるので、「行い」は、その「こと」の「行い」と判断してしまう。係り受けの判定としては係る結果となり正解としたが、今後このような名詞間の関係の問題にも取り組む必要がある。

[4. 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。

6.2　規則に対する考察

係り受け規則については連用節、連体節それぞれの要素の組合わせになると考えていたが今回作成した規則を見ると、用言意味属性や名詞意味属性が単独で決定されるものも多かった。これは、今回171例と少ない用例での規則作成であったこと、また、単独でも係られにくい用法、係りやすい用法などがあることが考えられる。例えば、連体節述語の用言意味属性が「相対関係」の場合、係らないと判定されるが、「相対関係」にあたる用法は「～に関連する」や「影響を与える」などで、親子関係や共通要素など強い名詞同士の繋がりがなければ係り先になることは少ないと考えられる。今回の結果からは用言意味属性に関する規則では、「因果関係」や「結果」なども係り受けに影響が強いことも分かる。

今後、用例を増やすことで、今回の少ない用例にのみ当てはまるような規則を削り、また、必要なルールを増補することにより、正解率の向上を望めると考えている。また、品詞体系の一部の変換が可能になれば、コーパス中の大量の用例により、今回作成した規則における属性の着眼点、組合わせを利用することで、決定木のみによる学習での規則作成も可能であると考える。

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] 南 不二男：現代日本語の構造, 大修館書店 (1974).

[4] 黒橋, 長尾: 並列構造の検出に基づく長い日本語文の構文解析, 自然言語処理, Vol.1,No1,pp.35-58(1994).

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

[6] 林 良彦：結合価構造に基づく日本語解析情報処理学会自然言語処理研究会,87-NL-62(1987).

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

参考資料