連体形イ・ナ形容詞に先行する格助詞句の係りに関するルールの抽出

菊池 浩三[†] 伊東 幸宏^{††}

係り解釈を正確に行うことは、自然言語解析における重要な課題である。近年大量のコーパスが利用できるようになり、統計的確率をベースに係りを決定する動きがある。本論文ではイ・ナ形容詞を含む連体修飾の係り関係の明確化に取り組んだ。係りを決定するために、大量の技術文と新聞文より約 4400 の対象文を抽出し解析した。「名詞 1 + 形容詞 + 名詞 2」の構造を持つ形容詞を含む連体修飾を各語間の関係として捉え、名詞 1 の係り先が形容詞か否かを調べた。その結果、係りに関与する 7 つの効果的ルールを検出することができた。また、それらは以下の 3 パターンに分類できた。

- 形容詞とは独立に前後の名詞により係りが決定できるもの。
- 形容詞と前後の名詞の関係により係りが決定できるもの。
- 形容詞単独の特性により係りが決まるもの。

3 パターンに分類された、7 つのルールを評価文に適用した結果、形容詞を含む連体修飾では、約 97% 精度で係りを決定することができた。

キーワード: 係り受け解析, 形容詞, 統計的解析

Extracting Dependency rules on dependency of postpositional phrases and I/NA-adjectives in adnominal forms

KOZO KIKUCHI[†] and YUKIHIRO ITOH^{††}

Improving the accuracy of dependency analysis in long Japanese sentences is a big problem in natural language analysis. Recently, there is a tendency to determine the dependency on the basis of a statistical probability by using the results of the analysis of a large amount of corpus. In this paper we research the dependency of adnominals which contain an I-adjective or NA-adjective using about 4400 sentences extracted from a large amount of technological documents and newspapers. To determine the dependency, we analyze the target sentences which include the adjective in the form of “<noun-1> + <adjective> + <noun-2>” considering the relationship of each word. As a result, we find seven effective dependency rules which are classified into the following three patterns:

- determined only from preceding or following nouns.
- determined from the relation between preceding or following nouns and adjective.
- determined from the characteristic of adjective itself.

We applied these seven rules to the set of evaluation sentences to determine the dependency of adnominals which include an adjective, and obtained an accuracy of about 97%.

KeyWords: *Dependency Analysis, adjective, statistical analysis*

1 はじめに

現在までに様々な構文解析手法が提案され、短い文に関しては比較的高い精度で解析が行えるようになってきた。しかし、長文の解析率は、依然として低く、これを解決することが自然言語処理の課題となっている。文が長くなると、語、句、あるいは節に係る(修飾する)可能性のある文中の要素が増え、文中の語句間の係り受け関係の可能な組み合わせの数が級数的に増大する。このため、係り受け関係の決定を一つでも間違えると、誤りが次々と増幅されてしまう。特に、文中に現れる用言の数が増すと、格助詞句、接続助詞句、副詞節などの係り先候補数が増し、解析の困難さを押し上げることになる。このため、長文の解析精度の向上には、係り解釈の精度向上が必須である。

長文中に現れる複数の用言が接続詞、接続助詞、連用形用言を介して係り受け関係を持つことが比較的多いことから、この観点での長文解析精度の向上をねらった研究はすでにいくつか報告されている。一方、長文中に複数の用言が出現するもう一つのケースに連体修飾節がある。この場合も、その用言に先行する格助詞句などが連体修飾中の用言に係るか、あるいはそれを飛び越えてさらに先の用言に係るかに曖昧性があり、解析精度向上を阻害する要因となりうる。

そこで、本論文では、大量のコーパスを利用して定量的に連体修飾中に出現する用言(形容詞)への格助詞句の係りに関する判定基準がルール化可能かどうか検証する。これは、連体修飾を構成する形容詞の場合、格助詞句を受ける文体(英語の叙述用法の形容詞に相当し、所有格関係節構造をなして連体修飾節を構成する文体:「色が赤い車を…」など)と格助詞句を受けない文体(英語の限定用法の形容詞に相当し、形容詞単独で連体修飾を構成する文体:「私が赤い車を…」など)が存在するため、係りの判定に用言の必須格という概念が利用できず、格充足の原則のみでは解析が困難であるからである。このような構文を採り上げて係り受け決定ルールを抽出しようという研究として、田中、荻野の研究がある(田中、荻野(1980), 荻野(1987))。この研究の概要については後述(5章)するが、意味情報に依存するところが比較的大きい。しかし、我々は、できる限り意味情報に依存しない形でルールを抽出した方が取り扱いが容易になると考えている。

現在の機械翻訳システムは、入力となる原文(今回の対象は日本語文)をより普遍的な情報空間である概念構造にマッピングし、意味トランスファ機構を使い翻訳文を生成する方式を採っている。この時、原文を表層情報、意味情報、文法規則を活用して解析し、解析精度の向上を図ってきた。しかし一般的な係り解釈は、格文法(case-grammar)に則り、以下のような原則で行っている。

- 後ろから順に格充足の処理に基づき解釈する(格充足の原理)
- 係りが交差しない(非交差の原理)

† 静岡大学大学院電子科学研究科, Graduate School of Electronic Science and Technology, Shizuoka University
†† 静岡大学情報学部, Faculty of Information, Shizuoka University

- 近くの文節にかける (卑近則)

しかし、この原則のみでは、係りの組み合わせの少ない単文には有効であるが、用言が複数存在する長文において、係りを正確に規定することが困難であった。そのため、係りの場合の数の削減や、係りを特定するケースの洗い出しがなされた。この作業において、文の構造の大域的決定が必要となるため、詳細な意味情報ではなく、文字列情報である表層情報に着目してその特徴分析がなされてきた。例えば、「～なり、かつ…」や「するが、…」などの句点を伴う連用中止、接続詞、接続助詞等の表層的形態に着目し、その特徴から重文を複数の独立した文に分割することにより、係りの組み合わせ数を減らし、単文に近い状態にすることで翻訳品質を向上させようとする試みがある (信国 1990)。また、従属節の述語の相関関係に着目し、それらを、同時、原因・中止、独立の3つのタイプに分け、句点の有無を考慮して従属節の係り関係を特定する研究がある (白井 1994)。また、並列句に関して、並列構造を構成する文字列の類似性をダイナミックプログラミング技法を用いて検出し並列関係を明確にすることで係りの場合の数を減らす試みもなされている (黒橋 1992, 黒橋 1994)。これらの試みで文の骨格構造の解釈手段が得られるようになってきている。

これらの研究からも分かる通り、係りの場合の数を減らすために、まだ活用されていない情報の活用が課題となっている。例えば、表層情報にはまだまだ活用されていない情報が多数残されており、高度な意味解析を追求するのみではなく、処理の簡便性の面からも未検出の表層情報を活用する試みが続けられている。表層情報はより広い範囲の相互関係を直感的に把握できるため、文の構造解析にはより効果があると考えられるからである。

一方上記のような、ルールベースの機械翻訳システムに対して、人間の翻訳作業の内省から、よりこなれた翻訳のためには用例ベースを活用することが有効であるという考え方がある。この考え方に基づいて、事例に基づく機械翻訳 (Example Based Machine Translation : EBM T) なる機構の研究も進んでいる (Nagao 1984, Sumita 1995)。このアプローチが注目を集めたのは、既存の伝統的ルールベースの翻訳処理は、ルールの抽出そのものが難しいことに加え、ルールが詳細化されてゆくに従ってルール間の相互干渉がおこり、ルールの修正を一貫した体系を保持しつつ行うことが困難になってきていることによる。また、コンピュータの性能の向上と記憶装置の大規模化に伴い、大規模コーパスという形で大量の文書が電子媒体に蓄積されるようになったという背景も無視できない。

大規模コーパスの出現により、各種情報を自動抽出する試みもなされるようになってきている (新納 1995)。さらに、クラス n-gram モデルを利用した確率的形態素解析器の精度向上 (森 1998) の取り組みや、係りの特性解析に統計的処理を利用する試み (藤尾 1997, 今一 1998) がなされており、分析のためにより大規模なデータを活用し、統計的な処理の中で自然言語を扱うようになってきた。

本論文では、大量のコーパスを利用し、表層情報の更なる活用の可能性を目指し、形容詞を

含む連体修飾における係り決定の規則を形容詞とその前後の名詞に焦点をあて、これらの3項関係という非常に局所的な情報を使ってどの程度まで解釈が可能か分析するものである。また統計的情報も形容詞それ自身の持つ属性として係り解釈に利用することも試みる。基本的にはルールベースの解析方式ではあるが、ルールの確かさについても、確率として表現することを試みた。

本論文の第2章では、対象とする構文を具体的に定義する。第3章では、係りの決定には品詞分類ではなく語ごとの振る舞いの分析のみならず何らかのルールが必要なことを説明する。第4章では、詳細な事例分析により得られた形容詞の係りを規定する7つのルールと、形容詞自身の持つ係りの特性について説明し、それらの適用結果について述べる。第5章では、同一構文を扱う先行研究との比較を行い、類似点と相違点について明確化する。第6章では、提案したルールの機械翻訳システムへの適用可能性及び派生効果について述べる。

2 分析対象文の形式

本論文では、用言としてイ形容詞・ナ形容詞を含む連体修飾について、先行する「が」格・「の」格の形容詞への係り関係を、局所的な表層情報とできるだけ少ないシソーラス情報から解析を行う。この2つの格を調べたのは、これらが文の構造に大きな影響を持つ主格になりえるからである。

分析したのは、次の連体修飾形を含む文である。

名詞1 + < * の | * が > + [副詞] + < イ形容詞 | ナ形容詞 > + 名詞2

この形式に該当する連体修飾の例を以下に示す。

例1) 経験の少ない利用者

例2) 運転条件の広い範囲

例1は、「経験(名詞1)」は「少ない(イ形容詞)」に係ると解釈される。これは「経験」に付属する「の」格が「が」格に格シフトを起こし、「経験が少ない」と解釈されるためである。

例2は、「運転条件(名詞1)」は「範囲(名詞2)」に係り、「広い(イ形容詞)」には係らないという例である。

このように、品詞としては同じイ形容詞であっても係り関係から見ると違いが生じる。このような違いをどのように解釈すべきかを以下で明確にしていこう。

3 予備的調査

詳細な分析を行うに前に、イ形容詞とナ形容詞について統計的にみてどのような係り特性があるか概観する。

表 1 形容詞の係りの分析

分類	対象語	「が」格が		「の」格が	
		係る	係らない	係る	係らない
イ形容詞	高い	77	9	266	27
	新しい	1	26	1	140
ナ形容詞	簡単な	7	5	0	54
	主要な	0	4	0	22

3.1 品詞と係りの関係

イ形容詞とナ形容詞について、係りの傾向をみるために、各々2語について分析する。分析する形容詞は「高い」、「新しい」、「困難な」、「主要な」である。分析結果を表1に示す。

表1からも分かる通り、イ形容詞の「高い」には、先行する「が」「の」格が高い確率で係るが、「新しい」は全く逆の傾向を示す。また、ナ形容詞についても語によって係りの振る舞いが全く異なる。すなわち、係りの解析では、イ形容詞・ナ形容詞という品詞の単位ではなく、語ごとに係りの規則を決定しなければならないことが分かる。さらに、「高い」の例のように、先行する「が」格が「高い」に係る場合もあるが係らない場合もある。そのため、各語の係りを決定するためには、語ごとの統計的処理のみではなく、係り判定する何らかのロジックが必要となる。以下でこれらについて説明する。

3.2 分析の対象とした語とコーパス

分析では、大規模なコーパスから分析対象文の形式を持つ例文を抜き出し、形容詞ごとに先行する「が」格や「の」格が形容詞に係るか否かを調べることとなる。そのために、分析に先立って、係りの正解集を整備する必要がある。我々は、以下に示すコーパスから、対象とした語の分析対象文を抽出し、それらに対して人手で係りの正解集を作成した。

分析に利用したコーパスは、富士通で利用している10万文を超える技術文(マニュアル文)と毎日新聞95年度版1月の記事(約8Mバイト)、10月の記事(約8Mバイト)である。そして、そこで得られた結果を毎日新聞95年度版12月の記事(約8Mバイト)で評価した。

分析対象とした形容詞は、技術文から抽出し、同じ形容詞が6件以上抽出できたものとした。対象語は50語となった。なお分析用コーパスから抽出した分析対象文は、イ形容詞2029文、ナ形容詞2418文の計4447文である。表2に対象となった語を示す。連体詞と分類される語(「大きな」等)も対象に含めた。その理由は、連体詞であれば先行する「の」格は係らないはずであるが、実際には「の」格に係る場合もあったためナ形容詞扱いとしたからである。なお、これらに対し、一般に想定される最も単純な品詞分類での係りのルール

- イ形容詞には前の名詞に係る。
- ナ形容詞には前の名詞が係らない。

表 2 分析対象語

品詞	分析対象語
イ形容詞 (21 語)	若い, 詳しい, 新しい, 正しい, よい, 悪い, 良い, 広い, 強い, 高い, 低い, 小さい, 少ない, 大きい, 長い, 短い, 深い, 多い, 難しい, 激しい, 無い
ナ形容詞 (29 語)	～規模な, ～速な, ～大な, ～的な, さまざまな, 完全な, 最適な, 主な, 主要な, 重要な, 小さな, 詳細な, 新たな, 正確な, 大きな, 大幅な, 適切な, 特殊な, 特別な, 複雑な, 簡単な, 困難な, 必要な, 豊富な, 有効な, 可能な, 不可能な, 不要な, 容易な

表 3 単純ルール適用時の正解率

対象語	係り正解数	総件数	正解率
イ形容詞	1454	2029	71.7%
ナ形容詞	2065	2418	85.4%

を適用した場合の、係り解釈精度を表 3 にまとめる。表中の「係り正解数」は上記ルール適用時に正解集と一致したものの数を示し、「総件数」は分析対象文の数である。「対象語」は、分類単位としてのイ形容詞とナ形容詞である。表 3 に示す通り、このような単純な処理では、実用的に十分な係り解釈の精度は得られない。

4 事例の詳細分析と 7 つのルール

語ごとに係りを分析すると、顕著に係りに特性のある語 (例えば「詳しい」) も見受けられるが、係りにばらつきが見られた。そのため、精度の高い係りの決定には、語ごとの係りの統計的情報のみではなく、他の要因も分析対象に含める必要があることが分かった。従来は、格充足の考えに基づき主述語から順に解析を進め、係りのばらつきも文の解析の中で解決されるとするものが一般的である。しかし、ここでは局所情報の更なる利用の観点から、名詞 2 直後の係助詞「は」の存在を加味しつつ、前後の名詞 (名詞 1 または、名詞 2) と対象となる形容詞の相互関係から分析を加えた。

これらの関係を意識しつつ分析対象文を詳細に分析した結果、7 つのルールが係りを決める大きな役割を果たしていることが分かった。7 つのルールで係りの全てが網羅できるわけではないが、7 つのルールが係りの揺らぎの主要な部分を占めるので、7 つのルールを適用した後の係りの統計的頻度は、形容詞が持つ単独の係り特性と考え、それをその形容詞が持つ係りの Default 属性とした。

4.1 詳細ルール

分析対象文の詳細な分析により得られた7つのルールについて説明する。これらの7つのルールは、以下の3つにカテゴリに分類されるので説明もこのカテゴリ順に行う。

- 前後の名詞の特性のみで係りが決定され、形容詞とは無依存に係りが決まるもの。
- 前後の名詞と形容詞との関係で係りが決まるもの。
- 形容詞の特性で係りが決まるもの。

なお、慣用句や共起関係の強い語は助詞いかに関わらず係りが決まるので、これらから出る揺らぎをなくすために、これらの語は辞書に登録されているという前提をおいた。

前処理：共起関係の強い語の登録

形容詞は慣用句になる語が多いことや慣用句とまではいえないものの共起関係の強い語がある。これらは、係り解析にあたっても、一語として処理する方が望ましい。これには以下のようなものがある。これらは分析対象文中に複数回出現したものから抽出した。

長い目, 長い間, 息の長い
広い範囲, 幅の広い
年齢の若い, 長さの長い
周期の長い, 周期の短い
見通しの／が良い 等

例えば、「年齢の若い」は「若い」と同じ意味であるが、意味合いをより明確にするための強調であり、言葉の結合力は非常に強くなると解釈できる。そのためこれらの語は事前に単語として登録されたものとして解析する。

このような慣用句的な語の処置を済ませたものに対してルールが適用される。以下、各々のルールを前述の3つのカテゴリに従って説明し、検討事例に対し、各ルールに基づいて係りを判断し正解が得られる割合を示す。なお、各ルールは以下に取りあげる順(ルール番号の小さい順)に適用するものとして評価を行っている。

4.1.1 中間の形容詞に無依存のルール

中間の形容詞には依存せず、名詞1と名詞2の特性から係りが決まるものがある。それらについて以下に説明する。

ルール 1 : 名詞 1 が下記のものである場合は、名詞 1 は形容詞には係らない。

数詞 (一つの, 二つの ..)	例) もう一つの大きな特徴
単位 (キロの, 個の, %の ..)	例) 前年度比 19% の高い伸び
まとめる語 (他の, 一連の)	例) 繊維その他の複雑な成形材
時詞 (日常の, 将来の ..)	例) 将来の望ましい研究方向
位置 (外の, 中の ...)	例) C E B の中の不要な分割を回復する
区切り記号がある場合	例) 新党が, とともに低い評価を下している

上記ルール適用時の精度を表 4 に示す。表 4 の「総件数」は分析対象文 (4447 文) の中で、このルールに該当する文の数であり、「係り正解数」はそれらにこのルールを適用した時に係りが正解集と一致した文の数である。

表 4 名詞 1 の特性で係りの決まるもの

対象語	係り正解数	総件数	正解率
イ形容詞	78	78	100%
ナ形容詞	275	275	100%

ルール 2 : 複合助詞 (での, からの, への等) 及び「等の, 等が」は形容詞には係らない。

一般に複合助詞は、構成する複数の助詞の特性を引き継ぐので、「の」以外の助詞が形容詞に係らないものであれば、その複合助詞は形容詞に係らないと考えるのが自然と思われる。「等の」はここで挙げた複合助詞とは異なる品詞構成を持つが、ここでは複合助詞に相当すると考え、このルールに集約した。このルール適用時の精度を表 5 に示す。

表 5 複合助詞の係りの精度

対象語	係り正解数	総件数	正解率
イ形容詞	69	71	97.2%
ナ形容詞	187	187	100%

例) 中低速回転域での高いトルク～。

「中低速回転域」は複合助詞「での」を伴うので直後の形容詞には係らない。係りの解釈を誤る例は、次の 2 例であり、量の多寡を示す語である。

例 1) 親からが一番少ないおとし玉

例 2) デジタル情報のハードコピーには、にじみなどの少ない合成紙が不可欠だ

ルール 3 : 形式名詞は係りに強い影響を持つ。

ルール 3-1 : 名詞 2 が形式名詞の場合, 名詞 1 は形容詞に係る。

形式名詞は, 直前の用言を終止的に働かせ, 係りを閉じさせる働きをする。そのため, 名詞 1 に続く「の」「が」格は形容詞に係るとするのが自然と思われる。

形式名詞は文献(名柄 1987)より抽出し, 副助詞と思われるものを除いたものについて検証し, 以下のものを対象とした。

くせ, ぐらい, せい, そう, だけ, ため, はず, 方(ほう・かた), まま, みたい,
故(ゆえ), 由(よし), 訳(わけ), 間(ま, あいだ), 上, 内, 分, 場合,
こと, 時

ただし, 「ことである」, 「ことだ…」等の判定詞を伴い助動詞になるもの(益岡 1989)は対象から除外した。

表 6 名詞 2 が形式名詞の場合の係りの精度

対象語	係り正解数	総件数	正解率
イ形容詞	327	331	98.8%
ナ形容詞	80	83	96.4%

例) 計画と実績の差異が大きい場合の原因を知る。

例では, 形式名詞「場合」が名詞 2 の位置にあるので, 「差異」は「大きい」に係る。係りの解釈を誤る例は, 次のようなものである。

- イ形容詞では, 「こと」の扱いで正しくない場合がある。

例) 自分が非常に悪いことをしたのではないかと自分を責め, ~。

- ナ形容詞では, 「主な(連体詞に分類する場合もある)」, 「特別な」の場合であり, 「特別な」は, 形式名詞を普通名詞的に働かさせるためと思われる。

例) これは変数内誤差問題の特別な場合である。

ルール 3-2 : 名詞 1 が形式名詞の場合, 名詞 1 は形容詞に係らない。

ただし, ここで言う形式名詞はルール 3-1 で示した形式名詞から「こと」「方」を除いたものである。

表 7 名詞 1 が形式名詞の場合の係りの精度

対象語	係り正解数	総件数	正解率
イ形容詞	17	17	100%
ナ形容詞	101	102	99.0%

例) 配偶者を失った時の深い絶望感が、～

例にある形式名詞「時」は「深い」には係らない。一方、係りの解釈を誤る例は、次のようなものがある。

例) 薬を売り込むための謝礼的な意味で支払うケースも多い。

ルール 4 : 「が」格は名詞 2 以降にある「は」格を越えて係ることはない。

これは、形容詞連体修飾とは関係なく、係り解釈の一般的な原則である。係助詞「は」は文末に係る機能を備えているが、格助詞「が」は近くの用言に係る確率が高い。係りの非交差の原則からもっともらしい原則である。但し、「に関しては」は挿入的に用いられるので適用外とする。このルール適用時の精度を表 8 に示す。

表 8 係助詞「は」を含む場合の係りの精度

対象語	係り正解数	総件数	正解率
イ形容詞	2	2	100%
ナ形容詞	3	3	100%

例) 同様に評判が悪いズボンずり下げには「せめてパンツが見えない程度に」の
声が、

例の「評判」に続く「が」は係りにおいて「には」を飛び越さないので、「悪い」に係ることになる。

4.1.2 中間の形容詞に依存するルール

名詞 1、名詞 2 の属性と間に入る形容詞の関係により係りが決まる場合がある。それらについて以下に説明する。

ルール 5 : 動きや相互関係に関連する形容詞はサ変名詞や転成名詞と結び付く。

形容詞の中には、サ変名詞や連用形が名詞に転化した転成名詞と強く結び付く語が存在する。このような形容詞には以下のようなものがある。

量を表す語	→ 深い, 大幅な
評価を表す語	→ 可能な, 不可能な, 不要な, 必要な
難易を表す語	→ 難しい, 困難な, 容易な, 複雑な, 簡単な
行為の状態を示す語	→ 激しい

このルール適用時の精度を表9に示す。なお、名詞1、名詞2がともにサ変名詞か転成名詞の場合は本ルールは適用しない。

表9 サ変名詞・転成名詞と形容詞の係り解釈精度

対象語	係り正解数	総件数	正解率
深い	22	22	100%
大幅な	22	22	100%
可能／不可能な	92	92	100%
不要な	5	5	100%
必要な	63	66	95.5%
難しい	9	9	100%
困難な	19	19	100%
容易な	15	15	100%
複雑な	8	8	100%
簡単な	39	42	92.9%
激しい	50	50	100%
強い	10	16	62.5%
適切な	6	7	85.7%

ここでは相関の在りそうな「強い」と「適切な」をも含めて評価した。しかし、この2語はルール化するには精度が低いのでルール適用語とはしない。

具体例を「難しい」、「激しい」を使った用例で説明する。

例1) 査察の難しい生物兵器や化学兵器でさえ全面禁止が実現した。

例2) 分不相応の難しい金融取引に走った理由はここにある。

例1では、「査察(サ変)の」は「難しい」に係る。一方、例2では「難しい」のは「金融取引(サ変)」であるため、「分不相応の」は「難しい」に係らない。

例3) 各勢力の激しい戦いが繰り広げられている。例4) 運賃引き下げ競争の激しい米国で勝ち抜くためには、…

例3では、「各勢力」は「激しい」には係らない。「激しい」は「戦い(転成名詞)」に結び付くためである。その結果「各勢力」は「激しい」には係らない。

例4では「激しい」は直前の「競争(サ変)」との結び付きが強く「競争(サ変)」が「激しい」に係る。

なお、今回は分析対象としなかった語「速い」等の動作を表す語にも同様の傾向が見られる。

表 10 「人」等との結びつきの弱い形容詞の係り解釈精度

対象語	係り正解数	総件数	正解率
大きい	2	2	100%
小さい	4	4	100%
広い	1	1	100%
高い	60	60	100%
低い	14	14	100%
長い	22	22	100%
短い	1	1	100%
深い	7	7	100%
強い	56	56	100%
よい・良い	19	19	100%
悪い	25	25	100%
必要な	8	8	100%
有効な	3	3	100%
印象的な	3	3	100%
難しい	5	5	100%
困難な	2	2	100%
複雑な	2	2	100%
激しい	3	3	100%

ルール 6 : 「人」, 「組織」, 「国」などの属性を持つ名詞と特定の形容詞には係りに相関がある。

ルール 6-1 : 「人」, 「組織」, 「国」などの属性を持つ名詞と結び付かない形容詞。

形容詞には, 「人」, 「組織」, 「国」そのものの特徴ではなくそれらの部分が持つ特徴を表す語や, それらの行為の程度を表す語がある。そのような語は, 「人」等に直接結び付かず, 条件を明確化する名詞に結び付くことになる。このような語には以下のものがある。

空間的量を表す語 → 大きい, 小さい, 広い, 高い, 低い, 長い, 短い, 深い
 強弱を示す語 → 強い
 評価を表す語 → よい・良い, 悪い, 必要な, 有効な, 印象的な
 難易を表す語 → 難しい, 困難な, 複雑な
 行為の状態を表す語 → 激しい

このルール適用時の精度を表 10 に示す。なお, 名詞 1, 名詞 2 とも当該属性を持つ場合は本ルールは適用しない。

具体例 (「長い」) で説明する。

例 1) 被災者が長い列を作ったため, …

例 2) 滞米生活が長い彼には, …

例 1 の「被災者」は「人」の属性を持つため「長い」には係らない。一方例 2 の「彼」は「長い」に結びつかず「長い」は名詞 1 の「滞米生活」に係ることとなる。

なお、今回分析の対象としていない「弱い」等の「強い」に対する反意語も同等の傾向を持つ。

ルール 6-2：名詞 2 が「人」の属性を持つ場合、その語と強く結び付く形容詞。

人の様態を直接表す語は、名詞 2 が「人」の属性を持つ時にその語と結びつき、名詞 1 は形容詞に係らない。この様な語には以下のものがある。

人の様態を表す語 → 若い

このルール適用時の精度を表 11 に示す。

表 11 「人」等との結びつきの強い形容詞の係り解釈精度

対象語	係り正解数	総件数	正解率
若い	30	30	100%

例) 副艦長はエリート大学卒でよき家庭人の若い黒人。

ルール 7：接尾語と結び付きの強い形容詞は接尾語のある語に結び付く。

これは、ものの性質を表す語はその性質の程度を示す語との結び付きが強くなるためである。この様な語には以下のものがある。

少ない	→ 性, 色
強い	→ 性, 色, 力
よい・良い	→ 性, 率
高い・低い・小さい・大きい・大きな	→ 性, 率, 度

なお、名詞 1、名詞 2 とともに上記接尾語を伴う場合は、本ルールは適用しない。このルール適用時の精度を表 12 に示す。

表 12 接尾語との結びつきの強い語の係り解釈精度

対象語	係り正解数	総件数	正解率
上記形容詞	245	246	99.6%

以上の共起及び 7 つのルールを分析対象文 (4447 文) に適用した結果を表 13 と表 14 に示す。この表で「出現数」は分析対象文中の各形容詞が用いられている文の中で、「が」格 (「の」

表 13 イ形容詞の出現頻度

対象語	「が」格が						「の」格が					
	係る			係らない			係る			係らない		
	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外
よい	7	4	3	2	1	1	82	37	45	4	2	2
悪い	15	13	2	4	2	2	47	31	16	7	6	1
良い	7	4	3	2	1	1	51	27	24	4	3	1
強い	54	37	17	8	6	2	62	40	22	66	58	8
短い	4	4	0	1	1	0	7	6	1	3	2	1
高い	77	52	25	9	8	1	266	171	95	27	25	2
低い	28	17	11	1	1	0	40	19	21	7	6	1
少ない	68	35	33	1	0	1	65	12	53	0	0	0
多い	165	91	74	0	0	0	108	8	100	0	0	0
大きい	45	27	18	3	0	3	55	16	39	0	0	0
小さい	20	12	8	1	1	0	12	3	9	7	6	1
無い	11	9	2	0	0	0	15	0	15	0	0	0
深い	12	11	1	6	5	1	22	12	10	21	16	5
広い	9	5	4	4	3	1	7	4	3	27	19	8
難しい	17	14	3	5	5	0	2	2	0	2	2	0
激しい	11	10	1	8	8	0	17	17	0	32	32	0
長い	12	12	0	14	12	2	22	16	6	26	22	4
詳しい	0	0	0	0	0	0	0	0	0	26	4	22
新しい	1	1	0	26	2	24	1	0	1	140	31	109
正しい	5	5	0	10	1	9	0	0	0	6	4	2
若い	2	1	1	10	5	5	3	3	0	55	45	10

格) が形容詞に係る (係らない) 事例の総数を示す。そのうち、これまでに説明した共起及び 7 つのルール of のいずれかによって、係る / 係らないが決定される事例の数を「共起 ルール」に示す。残りの事例 (事例数を「適用外」欄に示す) は、これらのルール適用範囲外の事例であり、係りの決定のためには別途基準が必要である。これについては次節で述べる。

表 14 ナ形容詞の出現頻度

対象語	「が」格が						「の」格が					
	係る			係らない			係る			係らない		
	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外	出現数	共起 ルール	適用外
不可能	12	12	0	0	0	0	0	0	0	0	0	0
不要な	7	6	1	0	0	0	4	1	3	6	5	1
容易な	16	16	0	0	0	0	3	3	0	3	3	0
困難な	34	34	0	0	0	0	1	0	1	5	5	0
必要な	102	92	10	6	6	0	10	9	1	5	2	3
可能な	89	81	8	0	0	0	7	7	0	4	3	1
簡単な	7	5	2	5	5	0	0	0	0	54	45	9
有効な	9	7	2	5	3	2	3	0	3	22	4	18
豊富な	3	0	3	0	0	0	3	0	3	11	2	9
～規模	0	0	0	3	1	2	0	0	0	17	8	9
～速	0	0	0	2	0	2	0	0	0	40	1	39
～大	2	2	0	9	0	9	0	0	0	58	25	33
～的	8	6	2	107	21	86	9	1	8	734	203	531
様々な	1	0	1	6	1	5	0	0	0	46	18	28
完全な	0	0	0	1	0	1	0	0	0	18	4	14
最適な	0	0	0	2	0	2	0	0	0	12	3	9
主な	0	0	0	36	13	23	0	0	0	163	36	127
主要な	0	0	0	4	1	3	0	0	0	22	9	13
重要な	4	3	1	20	2	18	0	0	0	69	26	43
小さな	0	0	0	8	3	5	2	2	0	54	28	26
大きな	0	0	0	116	9	107	12	5	7	154	52	102
詳細な	0	0	0	1	0	1	0	0	0	16	2	14
新たな	0	0	0	16	1	15	0	0	0	73	34	39
正確な	0	0	0	0	0	0	0	0	0	22	1	21
適切な	0	0	0	4	1	3	0	0	0	11	5	6
特殊な	0	0	0	1	0	1	0	0	0	21	10	11
特別な	0	0	0	3	1	2	0	0	0	15	6	9
大幅な	0	0	0	4	4	0	0	0	0	29	24	5
複雑な	5	5	0	4	3	1	0	0	0	18	10	8

4.1.3 係りの Default 属性

以上の7つのルールを適用した後の係りの特性は、係りの揺らぎを取り去ったものとなるので、各形容詞が持つ係りの特性であると考えられる。それでこれを各形容詞が持つ「係りの Default 属性」と呼ぶことにする。係りの Default 属性は、形容詞に先行する「が」格、「の」格が形容詞に係ってよいか否かを決定するために用いられる。よって、表13、表14の「適用外」の事例において「が」格、「の」格が形容詞に係る傾向が強いのか、係らない傾向が強いのかによって係りを決定することとした。具体的には、「が」格、「の」格の各々の欄で、「係る」「係らない」の各々の「適用外」事例数を比較し、単純に多い方をその語の係りの Default 属性とすることにした。例えば、「よい」は「が」格も「の」格も共に係るものと考えたこととした。このようにして決定した形容詞の係りの Default 属性を以下の分類に従って分類する。

分類1： 「が」「の」格は形容詞に係る。

分類2： 「が」格は形容詞に係るが、「の」格は形容詞に係らない。

分類3： 「が」格は形容詞に係らないが、「の」格は形容詞に係る。

分類4： 「が」「の」格は形容詞に係らない。

この分類に従ってマップしたものを表15に示す。

表 15 イ・ナ形容詞の係りの Default 属性

分類	分類された対象語
分類1	よい, 悪い, 良い, 強い, 短い, 高い, 低い, 少ない, 多い, 大きい, 小さい, 無い, 深い
	不可能な, 不要な, 容易な, 困難な
分類2	広い, 難しい, 激しい
	必要な, 可能な, 簡単な, 有効な, 豊富な
分類3	長い,
	ナ形容詞は, 該当するものなし
分類4	詳しい, 新しい, 正しい, 若い
	～規模な, ～速な, ～大な, ～的な, さまざまな, 完全な, 最適な, 主な, 主要な, 重要な, 小さな, 詳細な, 新たな, 正確な, 大きな, 大幅な, 適切な, 特殊な, 特別な, 複雑な

形容詞「詳しい」の「が」格のように、そのケースが存在しないものに対しては、「の」格での出現数の多い方を「が」格の場合にも適用し、係りを想定することで分類した。また7つのルールを適用することで「適用外」が存在しなくなる場合もあるが、これはルール適用前の出現数の多い方をその語の持つ係りの Default 属性とした。係りの Default 属性で特徴的なことは、イ形容詞にも「詳しい」「新しい」「若い」等のように先行する「が」「の」格を取らない語がある一方で、ナ形容詞にも先行する「が」「の」格を取る語があるということである。

4.2 分析文での係り解釈精度

以上の7つのルールと係りの Default 属性を分析文に対して適用した結果を表16にまとめる。この表は分析対象文に対して得られた係り解釈の最終結果である。

表 16 詳細分析を適用した結果

対象語	係り正解数	総件数	正解率
イ形容詞	1981	2029	97.6%
ナ形容詞	2378	2418	98.3%

4.3 評価

上記の共起, 7つのルール, 及び係りの Default 属性を毎日新聞 95年度版 12月の記事(約8Mバイト)に適用し評価した。その結果をイ形容詞については表17に, ナ形容詞については表18に示す。例えば, イ形容詞の「よい」は, 対象文形式の連体修飾が評価文中に20文存在し, 本方式では全ての係りを正しく解釈することができるというものである。

表 17 イ形容詞の評価文での評価結果

対象語	「が」格が		「の」格が		出現数	係り誤り数	正解率
	係る	係らない	係る	係らない			
よい	0	1	19	0	20	0	100%
悪い	2	1	22	2	27	1	96.3%
良い	0	1	15	3	19	1	94.7%
強い	34	5	38	47	124	11	91.1%
短い	5	0	3	1	9	0	100%
高い	54	9	79	23	165	5	97.0%
低い	9	1	12	8	30	0	100%
少ない	13	0	25	0	38	0	100%
多い	45	1	47	0	93	1	98.9%
大きい	18	0	13	1	32	1	96.9%
小さい	7	0	4	4	15	2	86.7%
無い	1	0	1	0	2	0	100%
深い	2	1	7	4	14	1	92.9%
広い	1	0	6	6	13	1	92.3%
難しい	21	2	2	0	25	0	100%
激しい	3	4	8	19	34	3	91.2%
長い	4	4	12	8	28	1	96.4%
詳しい	0	2	0	5	7	0	100%
新しい	0	13	0	30	43	0	100%
正しい	0	3	0	5	8	0	100%
若い	0	8	0	37	45	0	100%

表 18 ナ形容詞の評価文での評価結果

対象語	「が」格が		「の」格が		出現数	係り誤り数	正解率
	係る	係らない	係る	係らない			
不可能	2	0	0	0	2	0	100%
不要な	1	0	0	0	1	0	100%
容易な	2	0	0	0	2	0	100%
困難な	6	0	0	0	6	0	100%
必要な	43	1	2	0	46	0	100%
可能な	7	0	1	0	8	0	100%
簡単な	2	0	0	1	3	0	100%
有効な	0	0	0	5	5	0	100%
豊富な	0	0	0	2	2	0	100%
～規模	0	0	0	3	3	0	100%
～速	0	0	0	6	6	0	100%
～大	0	3	0	22	25	0	100%
～的	6	48	4	223	281	5	98.2%
様々	0	6	0	17	23	0	100%
完全な	0	1	0	0	1	0	100%
最適な	0	0	0	0	0	0	—
主な	0	5	0	33	38	0	100%
主要な	0	3	0	2	5	0	100%
重要な	0	8	0	16	24	0	100%
小さな	0	1	0	30	31	0	100%
大きな	0	33	5	53	91	1	98.9%
詳細な	0	0	0	5	5	0	100%
新たな	1	4	0	29	34	1	97.1%
正確な	0	0	0	3	3	0	100%
大幅な	0	2	0	13	15	0	100%
適切な	0	2	0	0	2	0	100%
特殊な	0	0	0	2	2	0	100%
特別な	0	0	0	3	3	0	100%
複雑な	0	1	1	6	8	0	100%

上記の結果を集計すると表 19 の値が得られた。

表 19 詳細分析を適用した結果(まとめ)

対象語	係り正解数	総件数	正解率
イ形容詞	763	791	96.5%
ナ形容詞	668	675	99.0%
計	1431	1466	97.6%

以上より分かる通り、形容詞連体修飾に対して高い精度で係りを特定できた。但し、一部の語(「強い」等)は更なる改善が必要と思われる、ルール強化の余地を残したものとなっている。以下の誤り例に示すように、「強い」は「作用」に係るべきであるが、本方式では、製剤を受けてしまう。「強い」とサ変名詞との結合強度についてさらに詳細な分析が必要と思われる。

誤り例) 副作用発生率は製剤の強い作用で十個以上採卵したケースで高まり、十五個以上では五人に一人に上った。

5 先行研究との比較

以上の分析により、形容詞に先行する「が」「の」格が形容詞に係るか否かについて、実用的な精度で決定できることが分かった。我々が分析対象とした「名詞1 + “の／が” + 形容詞／形容動詞 + 名詞2」という構文の係りに関しては先行研究があり(田中, 荻野 1980), 形容詞の係りは以下の基本原則で決まるとしている。

- (1) 「名詞1ノ名詞2」がなりたてば、形容詞あるいは形容動詞は名詞2を修飾し得る。
- (2) 「名詞2ノ名詞1」がなりたてば、形容詞あるいは形容動詞は名詞1を修飾し得る。
- (3) 「名詞1ノ名詞2」と「名詞2ノ名詞1」の両方がなりたてば、形容詞との意味的結合度により名詞1に係るか、名詞2に係るかが決まる。

これらの基本原則は、人にとって非常に理解し易いものである。しかし、対象となる文を上記3つのいずれに分類するかについての具体的方法は明確でなかった。その明確化は、格助詞が「の」の場合について別の論文で明らかにされた(荻野 1987)。その方法は、「名詞1」と「名詞2」の相互関係を以下の7種類に分け、形容詞無依存に決定しようとするものである。

- (1) 全体部分関係:「全体」表す名詞と「部分」を表す名詞がある場合、形容詞は「部分」に係る。
- (2) 上下関係:「上位」表す名詞と「下位」を表す名詞がある場合、形容詞は「下位」に係る。
- (3) 属性関係:「親」表す名詞と「属性名」を表す名詞がある場合、形容詞は「属性名」に係る。

- (4) 格支配関係：「格成分」表す名詞と「述語成分」を表す名詞がある場合、形容詞は「述語成分」に係る。
- (5) 限定関係：「限定成分」表す名詞と「親」を表す名詞がある場合、形容詞は「親」に係る。
- (6) 評価関係：規定なし。
- (7) 所有関係：規定なし。

なお、最後の2つの分類(6. 評価関係, 7. 所有関係)については、分析の対象とした用例数が少ないこともあり、係りの決定法は明確にされず、「形容詞」と名詞の関係を先に分析した方が良い場合があることを指摘している。

この研究では「名詞1」と「名詞2」の関係を中心に分析が行われているため、名詞の厳密な属性分類を必要とし、実際のシステムへの組み込みは大変難しいものと思われる。また「名詞1」と「名詞2」の関係が明確に規定できない場合の具体的な処理については明確にはふられていない。我々の分析方法は「名詞1」または「名詞2」と「形容詞」の関係から係りを決めようとするものであり、係りの解析方式は全く異なる。

先行研究での名詞間の7つの関係と、我々が抽出した7つのルールにどのような関係があるかを以下で分析する。先行研究では、「数詞」、「単位」、「形式名詞」、「が」格、「複合助詞」等は対象としていないので、ルール2, ルール3, ルール4は比較の対象にはならない。関係するルールは、ルール1の一部, ルール5, ルール6, ルール7である。各ルールと先行研究との関係は以下の通りである。

<ルール1>は形容詞無依存のルールであり、時詞や位置に関するルールである。これは「5. 限定関係」と解釈できる。すなわち形容詞は「親」に係ることになる。言い換えれば、「限定成分=名詞1」は形容詞には係らないことになる。

<ルール5>は形容詞依存のルールであり、サ変名詞や転成名詞を示すルールである。これは「4. 格支配関係」と解釈できる。この結果、形容詞は「述語成分」に係る(結び付く)ことになる。しかし、格支配関係が「で」格の場合、形容詞が「述語成分」に係らないことがある。その例を示す。

例1) 地球の温暖化の詳細な経緯。→ある経緯 で 温暖化した。

例2) 日米関係の新たな役割。→ある役割 で 関係する。

例3) レーザを利用するがん治療の新しい方法。→ある方法 で がん治療する。

我々の方式では、ルール5は形容詞依存ルールであるため、形容詞がサ変名詞と関係を持たないものはルール適用対象外となり、係りの解釈は、係りのDefault属性を適用することとなる。

<ルール6>は形容詞依存のルールであり、人等との関係である。これは先行研究の「1. 全体部分関係」、「3. 属性関係」と関連を持つ。すなわち、形容詞は「部分」、「属性名」等に係

ることになり、人には結び付かないことになる。ここで、我々のルールが形容詞依存となっているのは、“「人」、「組織」、「国」そのものの特徴ではなくそれらの部分が持つ特徴を表す形容詞や、それらの行為の程度を表す形容詞”を基準に形容詞を選択したからである。これは「少ない」等の量を示す形容詞等をルール適用除外とするためである。なぜなら、「少ない」は量的な名詞であれば何にでも結び付くことできるからである。以下に人の属性を持つ名詞を含む「少ない」の例を2つ示す。

例1) 受験回数の少ない人

例2) 顔見知りの少ない仮設住宅

例1では、「少ない」は「受験回数」を受けるので人とは結び付かない。しかし例2では、「少ない」は「顔見知り」を受け人と結び付く。例1は先行研究では、「3. 属性関係」に分類されるが、例2は何らかの別の分類に入れられるものと思われる。このように先行研究とは、ルールに相関関連はあるが、係りの解析法が異なるため、両者のルールのカバー範囲は当然異なる。

<ルール7>は形容詞依存のルールであり、接尾語のルールである。接尾語はものの属性を表すから、これは「3. 属性関係」と解釈できる。その結果形容詞は「属性名」に係る(結び付く)ことになる。ただし、我々の扱う属性関係は非常に狭く、明らかにものの属性と判断できる接尾語のみであるため取り得る形容詞が限定されたものになっている。

このように、「の」格に限定し、普通名詞で、その相互関係が規定できる場合は、両者は強い相関を持つ。一方、先行論文では、名詞間の相互関係を利用していることに起因する以下の課題が述べられている。

- 1) 解釈精度は、名詞の意味項目を記述した辞書の質による。
- 2) 比喩表現のような文脈中で意味が確定されるものは、解釈が困難。
- 3) 「形容詞」との結びつきを先に解釈したほうが良い場合もある。
- 4) 意味分類体系のうち、特に属性関係については関係付けが不十分である。

このように、名詞間の関係が規定可能となる位まで属性を厳密に規定することは大変難しい。さらに、3)のような特殊ケースの導入を先行研究では必要とした。一方、我々の方法は、名詞と形容詞の関係を議論し、名詞については表層情報か簡単な意味属性で済ませている。これによって現段階での機械翻訳システム等への組み込みが容易になると思われる。

また、形容詞をキーとしたので大量のコーパスを使った分析が可能となった。このことにより、ルール化されなかった部分に対しても、係り解釈を統計的に扱うヒューリスティック(係りのDefault 属性)が導入可能となった。また、対象とする構文を「の」格に限定する必要がなくなり、「が」格も取り扱えるようになった。「が」格の場合には、形容詞は直前の「が」格の述部としない可能性があるため、「名詞1」と「名詞2」の関係のみでは係りが決まらない場合がある。「が」格への拡張を考えると「形容詞」と「名詞」の關係に着目する方が柔軟性を持っていると思われる。

6 実用可能性と省略問題の解釈

本論文で利用した方式は、単に係りの曖昧性を解消するのに有用であるだけでなく、ある種の省略の補完の問題にも適用が可能であるように思われる。本章では、提案した方式の機械翻訳システムへの適用可能性と、省略補完への応用可能性について考察する。

6.1 既存システムへの適用可能性

現在市販されている翻訳システムは、翻訳精度がかなり高く、実用的なものである。しかし、このような連体修飾の問題は体系的に対処されているわけではない。現在の翻訳システムは、格充足の原則に則り主述語から順に解釈する方式を採っているものが多い。しかし、この方式のみでは、述語省略や、複雑な文章等では係り解釈が大きく間違うことがある。実際、ある翻訳システムを用いて、「新しい」という語を含む事例の翻訳実験を行ったところ、十分な翻訳率をあげることができなかった。

本方式は、局所的な情報を活用しているので、既存の格充足の考え方に基づく構文解析法と背反するものでなく、格充足法を補完するものとして活用されるべきものである。そのような情報を活用する手段として、ほとんどの機械翻訳システムは語ごとに個別のルールを定義することが可能な枠組みを有している。従って、現在の翻訳システムの枠組みを変更することなく、ルールを付加していくことにより比較的容易に本方式を組み込むことが可能である。

6.2 述語省略を検出することの可能性

ナ形容詞では、名詞1がナ形容詞に係らない場合が多い。このような場合で、文が体言止めになっているケースでは、述語が省略されているものとして、これを補わなければ正しい係り受け構造が得られない。例えば、次の文を考えてみる。

例) 米政界では、レポートのイメージが重要な要素。

この文は、最後に「である」という主述語が省略されており、「が」格はその述語に係るものとして解釈されるべきである。例えば翻訳を行う場合でも、

動詞追加) 米政界では、レポートのイメージが重要な要素だ。

訳文例) In the US political world, reporter's image is an important element.

のように述語の補完を行わないと適切な翻訳文を導けない。

しかしながら、一般に、「が」格に係る主述語が省略されているか否かの判断は容易ではない。格充足の原理を用いると、末尾に近い述語である「重要な」に対し「が」格を充足させてしまうため、何らかの方法で「レポートのイメージ」が「重要な」の主格となり得ないと判断しなければならない。

本方式を用いると、この構文において「重要な」は先行する「が」格を受けないと決定できるため、「が」格の係りを完結させるために動詞を補う必然性を導くことができる。

一般に、この種の問題に対処するためには、単純に文法を適用すると係り先の候補が想定できる場合に、何らかの方法でそのような係りは実現しないということを積極的に主張する基準が必要になる。本方式は、限られた範囲ではあるが、深い意味に立ち入ることなく係りが生じないことを判断できるため、この種の問題にも比較的容易に適用することが可能である。

7 まとめ

本論文ではイ・ナ形容詞を含む連体修飾について、局所的な表層情報と若干のシソーラス情報から分析し、精度高く節を特定できることを示した。また、既存の機械翻訳システムにも容易に適用できる部分のあること、後続する主述語の省略に関しても省略が同定できる場合があることを付帯効果として示した。しかし分析の対象とした語は技術文での出現頻度から選択した形容詞であり、一般文での出現頻度を必ずしも反映しているものではない。今後は、一般文での出現頻度を考慮し、反意語や類似語に対して同様の分析をする予定である。また、検出した7つのルールはかなり幅広く形容詞群に適用できそうであり、その適用範囲についても詳細に分析する予定である。例えば、今回分析した語(深い, 広い, 強い)の反意語(浅い, 狭い, 弱い)や動作と関連する(速い, 早い)等の語について評価する予定である。また、抽出したルールの内、ルール5, 6, 7は、形容詞をクラス分けするための属性になりうるものと考えている。これ以外の係り決定ルールがあるかについても今後検討する予定である。

本論文の分析は、机上分析の段階であり機械翻訳システムとして実装したわけではないが、形容詞に関して本方式が効果的であると判断している。

謝辞

本研究にあたり、多くの支援をいただいた(株)富士通静岡エンジニアリング社長田口尚三氏に心より感謝いたします。また、コーパスの利用を許諾していただいた富士通(株)の機械翻訳グループと毎日新聞社に感謝いたします。

参考文献

- 藤尾正和, 松本裕治 (1997). “統計的手法を用いた係り受け解析.” 情報処理学会研究報告 自然言語処理, 117-12.
- 今一修, 松本裕治, 藤尾正和 (1998). “統計情報と文法制約を統合した統語解析手法.” 自然言語処理, 5 (3), 67-83.

- 菊池浩三, 伊東幸宏 (1996). “形容詞を含む連体修飾の係り受け特性.” 情報処理学会研究報告 自然言語処理, **96** (114), 116-7.
- Kikuchi, K. and Itoh, Y. (1997). “Dependency Analysis of Adnominals Including I-adjective and NA-adjective in Japanese.”. pp. 581-584.
- 黒橋禎夫, 長尾眞 (1992). “長い日本文における並列構造の推定.” 情報処理学会論文誌, **33** (8), 1022-1031.
- 黒橋禎夫, 長尾眞 (1994). “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, **1** (1), 35-57.
- 益岡隆志, 田窪行則 (1989). 基礎日本語文法. くろしお出版.
- 森信介, 長尾眞 (1998). “形態素クラスタリングによる形態素解析精度の向上.” 自然言語処理, **5** (2), 75-103.
- Nagao, M. (1984). “A Framework of a Mechanical Translation between Japanese and English by Analogy Principle.”. pp. 173-180 North-Holland.
- 名柄迪, 広田紀子, 中西家栄子 (1987). 形式名詞 (外国人のための日本語例文・問題シリーズ 2. 荒竹出版.
- 信国佳之 (1990). “自然言語における長文分割方式.” 第 39 回情報処理学会 全国大会論文集, 2F-1.
- 荻野孝野 (1987). 意味体系を用いた単語間結合の分析. 計量国語学と日本語処理. 秋山書店.
- 新納浩幸, 井佐原均 (1995). “片方向の共起性による述語型定型表現の自動抽出.” 自然言語処理, **2** (3), 73-86.
- 白井諭, 横尾昭男, 池原悟, 木村淳子, 小見佳恵 (1994). “日本語従属節の依存構造に着目した係り受け解析.” 情報処理学会研究報告 自然言語処理, 102-9.
- Sumita, E., Oi, K., Furuse, O., Iida, H., and Higuchi, T. (1995). “Example-Based Machine Translation using Associative Processors.” 自然言語処理, **2** (3), 27-47.
- 田中穂積, 荻野孝野 (1980). “形容詞もしくは形容動詞の修飾先の名詞を決める原則について.” 計量国語学, **12** (5), 191-203.

略歴

菊池 浩三 (学生会員): 1970 年大阪大学理学部物理学科卒業. 1972 年大阪大学大学院修士課程修了. 1973 年富士通株式会社入社. 1983 年 (株) 富士通静岡エンジニアリング出向. 1995 年静岡大学大学院電子科学研究科博士課程に社会人学生入学. 自然言語処理の研究に従事. 言語処理学会, 情報処理学会各会員.

伊東幸宏 (正会員): 1980 年早稲田大学理工学部電子通信学科卒業. 1987 年同大学院博士後期課程修了. 同年, 早稲田大学理工学部電子通信学科助手. 1990

年静岡大学工学部情報知識工学科助教授。現在静岡大学情報科学科助教授，工学博士。自然言語理解，知的教育システムなどに興味をもつ。言語処理学会，電子情報通信学会，情報処理学会，人工知能学会，日本認知学会，教育情報システム学会各会員。

(1998 年 8 月 25 日 受付)

(1998 年 11 月 4 日 再受付)

(1999 年 1 月 8 日 再々受付)

(1999 年 1 月 22 日 採録)