目 次

表 目 次

図 目 次

概 要

機械翻訳における問題点の一つに、言語表現の構造の曖昧性の問題がある。日本語では特に名詞句の構造が曖昧である。従来の文法的情報のみによる構文解析では、名詞句の構造を一意に決定することが困難であった。

そこで本研究では、形容詞＋「並列型名詞句」(形容詞＋名詞＋並列助詞＋名詞) における、形容詞が抽象度の高い名詞を修飾するという点に着目した形容詞の係り先決定法を提案する。

具体的には、「形容詞＋名詞Ａ＋並列助詞＋名詞Ｂ」の、名詞Ａと名詞Ｂの抽象度を一般名詞意味属性を用いて計算し、抽象度の大小により形容詞の係り先を決定する手法である。

本手法を検討するため、新聞記事１年分よりの 293 件に適用したところ、88.4 ％の正解率を得、本手法の有効性が示された。

1 はじめに

機械翻訳における問題点の一つに、言語表現の構造の曖昧性の問題がある。日本語では特に名詞句の構造が曖昧である。このため名詞句の研究は多く行われてきた。例えば、助詞「の」が結ぶ名詞の意味関係の解析 [?]、フレーム構造を用いた名詞句「ＡのＢ」の意味解析 [?]、名詞句「ＮＰのＮＰ」の意味構造 [?] などがある。しかしこれらは、体言のみを含んだ名詞句を扱っており、用言を含むものは扱っていない。また、文法的情報のみによる構文解析では、名詞句の構造を一意に決定することが困難であった。

そこで本研究では、形容詞＋「並列型名詞句」(形容詞＋名詞＋助詞＋名詞)の名詞間の抽象度の差に着目した形容詞の係り先決定法を提案する。本手法を新聞記事 1 年分より収集した 293 件に適用し、本手法の有効性を検討する。

本論文では以下、第２章に名詞句とその解析について、第３章に着目点について、第４章に形容詞の係り先決定法について、第５章に実験について、第６章に考察について、第７章に結論を述べる。

2　名詞句とその解析

形容詞＋名詞句において最も基本形となる「形容詞＋名詞Ａ＋助詞＋名詞Ｂ」を次の２つのタイプに大別する。

2.1　名詞句の分類

(1)「の型名詞句」

の型名詞句とは、「形容詞＋名詞Ａ＋の＋名詞Ｂ」の形をとり、形容詞と名詞間の係り受けと、名詞Ａと名詞Ｂ間との両者において係り受け関係が存在する。また、形容詞は名詞Ａに係る場合と、名詞Ｂに係る場合との２つの曖昧性を持つ。以下に例を示す。

広い門の下(広い→門、門→下)
長い北国の夜(長い→夜、北国→夜)　(1)

例において、「広い」は「門(名詞Ａ)」に、「長い」は「夜(名詞Ｂ)」に係り、形容詞と名詞間の係り受けが存在する。さらに、「門」が「下」に、「北国」が「夜」に係り、名詞間の係り受けが存在する。

(2)「並列型名詞句」

並列型名詞句とは、「形容詞＋名詞Ａ＋並列助詞＋名詞Ｂ」の形をとり、名詞Ａと名詞Ｂが並列に連なっている名詞句である。以下に例を示す。

珍しい犬と猫
古い記憶やイマジネーション　(2)

例において、名詞「犬、猫」「記憶、イマジネーション」には係り受け関係はなく、並列に連なっている。並列助詞には、助詞「の」以外で形容詞の係り先に曖昧性を持つ名詞句を形成する全ての助詞が含まれるが、この様な並列助詞には、「と」「や」「か」「とか」「やら」「だの」がある。「とか」「やら」「だの」については「や」と完全に置き換えが可能である。本研究では比較的多くの抽出が可能であった「と」「や」について扱っている。

2.2 「形容詞＋並列型名詞句」の曖昧性

「形容詞＋名詞Ａ＋と/や＋名詞Ｂ」には次の２つの場合の曖昧性がある。

(1) 形容詞が名詞Ａのみに係る場合（以下「Ａ係り」とする）
以下に例を示す。

白い 壁 と フローリング(白い→壁)
貧しい 人 や 高齢者(貧しい→人) (3)

(2) 形容詞が名詞Ａ、名詞Ｂ両方に係る場合（以下「Ａ＆Ｂ係り」とする）
以下に例を示す。

激しい 空爆 と 砲撃（激しい→空爆＆砲撃）
古い 記憶 や イマジネーション（古い→記憶＆イマジネーション） (4)

実際にＡ係り、Ａ＆Ｂ係りがどの様な割合で存在するのか調べるために、新聞記事１年分より抽出した「形容詞＋名詞Ａ＋と／や＋名詞Ｂ」の係り先の割合を次に示す。

表 1: 形容詞の係り先割合（単位：個数）

	Ａ係り	Ａ＆Ｂ係り	合計
と	17	148	165
や	23	105	128
合計	40	253	293

（Ａ係り：名詞Ａに係る, Ａ＆Ｂ係り：名詞Ａ、Ｂ共に係る）

「形容詞＋名詞Ａ＋と／や＋名詞Ｂ」の純粋な形での抽出は全293件にとどまるが、形容詞と同じ働きをする形容動詞の存在、過去の活用形、並列助詞の多様さを考慮すると、さらに多数存在すると思われる。

本研究の目的は、「形容詞＋並列型名詞句」(形容詞＋名詞Ａ＋と／や＋名詞Ｂ）において形容詞の係り先の曖昧性を判定する手法を提案することである。

3　着目点

新聞記事等から抽出した並列型名詞句の用例分析を行った結果より、「 形容詞 + 名詞Ａ + と / や + 名詞Ｂ 」における形容詞の係り先決定法を提案するために、以下の仮説を作った。

(1) 共起の有無

Ａ係りとなるのは形容詞と名詞Ｂとの意味的結合力が極めて弱いときのみであり、「 形容詞 + 並列型名詞句 」には「 形容詞 + 名詞Ｂ 」が共起しない場合がある。以下に例を示す。

強い 意志 と 誠実さ　(5)

「 強い + 意志 」は共起するが、「 強い + 誠実さ 」は共起しない。

(2) 抽象度の差

並列型名詞句は形容詞を利用し、名詞Ａと名詞Ｂの抽象度の差を小さくする性質があり、名詞Ａに比べて名詞Ｂの意味がより抽象的と見られる場合は、Ａ＆Ｂ係りとなる傾向が強い。これは、形容詞の限定用法の性質を反映したものと考えることができる。以下に例を示す。

危ない 家具 (名詞Ａ)や 品物 (名詞Ｂ)　(6)

例において、形容詞「 危ない 」は「 家具 」のみでなく「 品物 」にも係るが、「 品物 」は「 家具 」よりも抽象度が高いことが指摘できる。

(3) 「 こと 」「 もの 」

抽象名詞である「 こと 」「 もの 」は、日本語の中で最も抽象度の高い名詞であり、他のどんな名詞よりも情報含有量が小さいために、常に形容詞の修飾を必要とし、形容詞との強い結合力を得る。以下に例を示す。

悪い こと や 職業 (Ａ係り)　(7)

抽象名詞「 こと 」は「 職業 」に比べ非常に情報含有量が少ないため、形容詞の修飾を必要としている。そのため抽象名詞「 こと 」は「 職業 」よりも形容詞との強い結合力が強い。

(4) 意味距離

Ａ＆Ｂ係りは、Ａ係りに比べ意味の似通った名詞Ａと名詞Ｂを並べる傾向が強く、Ａ係りとＡ＆Ｂ係りでは名詞間の意味距離に差がある。以下に例を示す。

長い 歴史 と 必然性（Ａ係り）
重い テント と 寝袋（Ａ＆Ｂ係り） (8)

Ａ係りでは「歴史、必然性」と名詞Ａと名詞Ｂの意味距離が大きいのに比べ、Ａ＆Ｂ係りでは「テント、寝袋」と意味距離が小さい。

4　係り先決定法

前述の着目点に基づいて、「形容詞＋名詞Ａ＋と／や＋名詞Ｂ」において、次の係り先決定法を提案する。

(1) 共起の有無人手によって「形容詞＋名詞Ｂ」が共起しないと判断した場合は、Ａ係りとする。

(2) 抽象度名詞Ｂの抽象度が名詞Ａの抽象度以上ある場合はＡ＆Ｂ係りとする。
抽象度の計算法については以下に示す。

【抽象度の計算法】

本研究では名詞の抽象度を一般名詞意味属性 [?] を用いて計算する。
まず名詞Ａ、名詞Ｂに名詞意味属性を付与する。
名詞意味属性は性質上、上位ノードを辿っていくことで、名詞の情報を得ることができる。よって、意味属性の深さの浅い名詞の方が上位ノード数が少なく、得られる情報も少ないので、より抽象度が高いと考えられる。

(3)「こと」「もの」名詞Ａが「こと」「もの」である場合はＡ係り、
名詞Ｂが「こと」「もの」である場合はＡ＆Ｂ係りとする。

(4) 意味距離名詞Ａと名詞Ｂとの意味距離が閾値より大きい場合にＡ係り、
小さい場合はＡ＆Ｂ係りとする。意味距離の計算法については以下に示す。

【意味距離の計算法】

まず名詞Ａ、名詞Ｂに意味属性を付与する。名詞Ａの意味属性の深さを a、名詞Ｂの意味属性の深さを b、共通の親ノードの深さを c とする。意味距離 d は次の式で計算する。

$$\text{意味距離}\ d = 1 - \frac{2 \times c}{a + b} \tag{9}$$

そして、Ａ係りとＡ＆Ｂ係りの意味距離の閾値を求める。

【一般名詞意味属性体系】

一般名詞意味属性体系とは、概念化された対象と単語との対応関係を木構造（約 2,700 ノード、12 段）の形式に整理分類したものである。

図 1: 一般名詞意味属性体系（上位部位）

5 実験

提案したルールを評価するために次の手順で実験する。

5.1 実験方法

(1) 新聞記事１年分より抽出した「形容詞＋名詞Ａ＋と／や＋名詞Ｂ」(293 件) に対して提案したルールを適用する。
(2) カバー率、適合率を以下の式で計算し、提案した形容詞の係り先決定法の有効性を評価する。それぞれの計算式を以下に示す。

$$\text{カバー率} = \frac{\text{ルール適用数}}{\text{全実験データ数}} \tag{5}$$

$$\text{適合率} = \frac{\text{係り先正解数}}{\text{ルール適用数}} \tag{6}$$

(3) 最後に適合率の高いルールのみを順番に適用し、ルールを総合した正解率を求める。

5.2 実験結果

(1) 新聞記事 1 年分より収集したデータ 293 件に、提案したルールを個別に適用した実験結果を以下に示す。

表 2: ルールごとの評価 (単位：%)

	カバー率	適合率
1; 共起の有無	5.1	100
2; 抽象度の差	61.8	94.5
3;「こと」「もの」	0.34	100
4; 意味距離	-	-

ルール３「こと」「もの」は新聞記事 293 件に対して、１件のみのカバーであった。またルール４の意味距離の差については閾値が得られず、ルールが成立しなかった。

(2) ルール１、ルール２のみを適合率の高いルール１、２の順に適用し、カバーできなかったものはＡ＆Ｂ係りとした結果を次に示す。

表 3: 総合評価（単位：％）

	カバー率	正解率
総合評価	100	88.4

これは全てをＡ＆Ｂに係るとする場合の 83.3 ％から、約 5 ％の精度向上であることが示された。

6 考察

6.1 提案したルールについての考察

提案したルールに対しての考察を次に述べる。

(1) ルール２：抽象度の差

失敗例を以下に示す。

若い サラリーマン や 主婦 (7)

例において、形容詞「若い」は単に名詞を修飾し情報を加えるだけでなく、「人間」のカテゴリを分類する形容詞である。従ってカテゴリとして見た場合に、「サラリーマン」と「主婦」は全く同等であり、抽象度の差だけでは係り先を決定できないと考えられる。

(2) ルール３：「こと」「もの」

ルール３は、抽象名詞の「こと」「もの」に対してのみでなく、上位概念語にも適用できると考えられる。以下に例を示す。

若い 人 や 知識人（Ａ係り）
大きな 木 と 人間（Ａ係り） (8)

例において、上位概念語である「木」や「人間」が名詞Ｂに比べ非常に含有情報量が少ないために、形容詞の修飾を独占し、形容詞との強固な結合力を持つために、Ａ係りになると考えられる。

(3) ルール４：意味距離の差

名詞Ａと名詞Ｂ間の意味距離に関して、Ａ係りとＡ＆Ｂ係りにおいて期待した特徴が見られない理由として形容詞の観点が考えられる。

形容詞の観点とは、例えば名詞「猫」と「ネズミ」では、形容詞の観点「大きさ」においては、名詞間の意味距離が大きいと考えられるが、観点「速度」においては意味距離が小さいと考えられる。従って形容詞の観点により、名詞間の意味距離に多大な影響を及ぼすと考えられ、観点の考慮なしの単なる意味距離だけでは、Ａ係りとＡ＆Ｂ係りの間に明確な閾値を得られず、係り先を決定できなかったと考えられる。

6.2 名詞間のバランス

「形容詞＋並列型名詞句」（形容詞＋名詞Ａ＋並列助詞＋名詞Ｂ）において、並列助詞は名詞Ａと名詞Ｂとの間に「バランス」をとる傾向がある。この傾向は助詞「や」において、さらに顕著に現れている。Ａ係り、Ａ＆Ｂ係りにおける名詞間のバランスのとり方について以下に詳しく述べる。

(1) Ａ係り

(i)「形容詞＋名詞Ａ」と「名詞Ｂ」で抽象度のバランスを保つ。以下に例を示す。

悪いこと や 職業 (9)

名詞「こと」は「職業」に比べ抽象度が高いが、形容詞「悪い」を独占することによって、その抽象度を下げ、「悪い＋こと」対「職業」で抽象度のバランスをとっている。

(ii)「形容詞＋名詞Ａ」と「名詞Ｂ」で「観点」についてのバランスを保つ。以下に例を示す。

大きい ネズミ と 猫（Ａ係り） (10)

観点「大きさ」において、「ネズミ」は「猫」よりも小さいが、形容詞「大きい」が「ネズミ」に係ることによって、「大きい＋ネズミ」対「猫」で、観点「大きさ」においてバランスをとっていると考えられる。

(2) Ａ＆Ｂ係り

Ａ＆Ｂ係りには、名詞Ａと名詞Ｂに意味の似た名詞を並べ、「名詞Ａ＋名詞Ｂ」はひとまとまりとして扱われる傾向がある。ひとまとまりとなった「名詞Ａ＋名詞Ｂ」は、名詞Ａに続け名詞Ｂを置くことで例示したり、名詞Ａと名詞Ｂを対比させたりする。以下に例を示す。

大きな ホテル や 別荘（例示）
小さい 電子 と 陽電子（対比） (11)

例において、同一カテゴリ「宿泊施設」に属する「ホテル」に続き「別荘」が例示されている。また「電子」と「陽電子」は対比関係にある。「ホテル＋別荘」、「電子＋陽電子」はそれぞれ、例示、対比の関係において強く結び付いている。形容詞が強く結合した「名詞Ａ＋名詞Ｂ」に係るため、Ａ＆Ｂ係りになると考えられる。

6.3 形容詞の分類

「形容詞＋名詞Ａ＋並列助詞＋名詞Ｂ」においては、形容詞の観点が名詞Ａと名詞Ｂ間との意味距離に影響するなど、形容詞の分類が重要である。形容詞は量的観点を持つ形容詞と、質的観点を持つ形容詞に大別することができると考えられる。分類について以下に示す。

(1) 量的観点を持つ形容詞

量的観点を持つ形容詞とは、主観を程度によって表現できる形容詞である。量的観点を持つ形容詞はさらに次のように分類できる。

(i) 空間的量の大小

「大きい」「あらい」など

(ii) その他の数量
「多い」「少ない」など
(iii) 速度
「はや(速・早)い」「遅い」など
(iv) 新旧
「新しい」「古い」など
(v) 強弱
「強い」「弱い」など
(vi) 難易
「難しい」「易しい」など
(vii) 味
「おいしい」「まずい」など
(viii) 音
「大きい」「小さい」など
(ix) 温度
「暑い」「暖かい」「涼しい」「寒い」など
(x) 美醜
「美しい」「醜い」など
(xi) におい
「香ばしい」「くさい」など

(2) 形容詞の質的観点

質的観点とは、主観を程度による相対的な表現でなく、絶対的な主観によって表現する観点である。質的観点をさらに、話し手の主観を反映する場合と、名詞の概念に付随する場合とに分類する。

(i) 話し手の主観を反映する質的観点
「ふさわしい」など、名詞の質を詳細に修飾する形容詞である。
これらは、名詞の概念に影響されることなく、状況によって様々な名詞が共起し、話し手の主観を状況に応じて反映する形容詞である。
(ii) 色に関する質的観点
「青い」「白い」など、名詞の色を修飾する形容詞である。色に関する表現は、話し手の主観を反映することがなく、名詞固有の概念に付随している。
そのため係り先を判定するには、個々の名詞の情報が必要となる。

6.4 量的観点と係り先

形容詞の係り先は、名詞A、名詞Bのいずれが形容詞の観点を持つかによって決定すると考えられる。以下に例を示す。

太い ロウソク と マッチ
若い 女性 と 主婦 (12)

例において、観点「太さ」において「ロウソク」は太く、「マッチ」は太くない。よって形容詞と同じ観点を持つ名詞Ａにのみ形容詞は係り、Ａ係りとなる。
同様に、観点「若さ」において、社会通念上「主婦」よりも「女性」の方が若いと考えられ、形容詞「若い」は「女性」に係ると考えられる。

7　結論

「形容詞＋並列型名詞句」(形容詞＋名詞Ａ＋と／や＋名詞Ｂ）に対して、

(1)「形容詞＋名詞Ｂ」の共起の有無
(2) 抽象度の差
(3)「こと」「もの」
(4) 意味距離の差

以上４つの点に着目し、次の形容詞の係り先決定法を提案した。

(1) 共起の有無
　「形容詞＋名詞Ｂ」が共起しない場合はＡ係りとする。
(2) 抽象度
　名詞Ｂの抽象度が名詞Ａの抽象度より高い場合、Ａ＆Ｂ係りとする。
(3)「こと」「もの」
　名詞Ａが「こと」「もの」である場合はＡ係り、
　名詞Ｂが「こと」「もの」である場合はＡ＆Ｂ係りとする。
(4) 意味距離の差
　名詞Ａと名詞Ｂとの意味距離が閾値よりも大きい場合にはＡ係り、
　小さい場合にはＡ＆Ｂ係りとする。

以上のルールを新聞記事１年分より抽出された 293 件に適用したところ、次の結果を得た。

表 4: ルールごとの評価（単位：％）

	カバー率	適合率
1; 共起の有無	5.1	100
2; 抽象度の差	61.8	94.5
3;「こと」「もの」	0.34	100

また適合率の高いルール１、２の順で適用したところ、次の結果を得た。

表 5: 総合評価（単位：％）

	カバー率	正解率
総合評価	100	88.4

これは全てをＡ＆Ｂ係りとする場合より、約５％の精度向上を得た。また、今後の課題としては、次の点が考えられる。

(1) 形容詞の分類

考察で述べた形容詞の分類に加え、次の視点による分類が必要と考えられる。

(i) 抽象度

例えば、形容詞「膨大な」は、後述の名詞に「数、量」などの抽象度の高い上位概念語が来るが、形容詞の種類によって後述の名詞の抽象度に影響がある。

(ii) 共起の難易

例えば、形容詞「美しい」は一般的に、「男性」よりも「女性」の修飾に用いられ、またあまりに幼い「少女」に対しては用いられにくい。この様に形容詞の中には、修飾しやすい名詞や、修飾しにくい名詞が社会通念上あらかじめ決まっている場合がある。

(2) 名詞の概念の考慮

名詞には、その本来の意味に加えて、人間が日常生活などで得る様々な情報をも内包している。それらの様々な概念を名詞に付与することが必要と考えられる。次に例を示す。

若い 先生 や 生徒
白い 壁 や フローリング (13)

人間の通念において「先生は一般的に生徒よりも年長」であり、「フローリングは一般的に白くない」。これらの通念は人が形容詞の係り先を決定する際にも多分に関係しており、考慮する必要があると考えられる。

参考文献

[1] 島津:助詞「の」が結ぶ意味関係の解析、計量言語学会１５巻第７号、pp.247-266(1986).

[2] 横山：フレーム構造を用いた名詞句「ＡのＢ」の意味解析、電子情報通信学会技術研究報告書、NLC97-3,pp.17-24,(1997)

[3] 日高：名詞句「NP の NP」の意味構造、情報処理学会論文誌、Vol.36,No.6,pp.1441-1448.(1995)

[4] 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)