

形容詞を含む連体修飾の係り受け特性

菊池浩三 伊東幸宏
静岡大学

日本語の長文の解釈において、用言の振る舞いの明確化は、文の構造を把握する上で重要である。本論文では連体修飾中の用言の振る舞いについて、特にイ形容詞とナ形容詞を対象を絞って検討した。技術文中のイ・ナ形容詞と先行する「が」格・「の」格の係り方について大量の技術文より該当する連体修飾を含む文を抽出し、その特性を分析した。各イ・ナ形容詞を単語毎に、係りパターンを現す4つのTYPEに分類し、「が」格・「の」格に先行する名詞と連体修飾される名詞の特徴で補正を加えれば、係り関係を約96%の精度で正しく解釈できることが分った。

Dependency analysis of adnominals including adjective and adjective verb

Kikuchi Kozo
The Graduate School of
electronic Science and Technology
Shizuoka University

Itoh Yukihiko
Faculty of information
Shizuoka University

In the interpretation of a Japanese long sentence, the clarification of the behavior of the declinable word is important for understanding the structure of the sentence. In this paper, we examine characteristics of adnominals including the declinable word, especially adjective and adjective verb, by extracting from large amount of technological sentences. we obtain some profitable characteristics applicable to machine translation.

1. はじめに

機械翻訳の日本語解析において長文中の係り処理の精度向上は大きな課題である。重文・複文に対する用言の係り先の解析は、文の構造を並立関係に着目して大局的に解析した研究（例えば文献(1)(2)）や表層情報を有効に活用し従属節の係り先を解析した研究がある(文献(3))。また、連体修飾句の英文への訳し方についても研究がなされている（文献(4)）。しかし、用言の係り先を検討するためにも個々の用言が支配する節を正確に切り出せることが必要である。一方、連体修飾中に用言が含まれる場合、それに先行する格助詞がどこに係るかが決定できないケースが少なくない。そこで、本稿では、イ形容詞・ナ形容詞を含む連体修飾に関して、先行する「が」格・「の」格の形容詞への係り関係を局所情報と表層情報からどの程度まで解析できるかを調べた。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り、局所的情報に基づいて解析したのは、それが大局的解析に語毎の特性を組み込むことで補正・補完する可能性を持ち得るからである。

10万文を超える技術文のコーパスから形容詞を含む連体修飾（出現頻度が6件以上）を1645文抽出し、分析した。その結果、形容詞を4つのTYPEに分類し、前後の名詞に分析を加えれば約96%の精度で節を切り出せることが分かった。

2. 分析方法

抽出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3) 拡大・縮小しても字形のゆがみが少ないアウトラインフォントを採用した。

例1は、「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。すなわち「経験(名詞1)」は「少ない(イ形容詞)」に係ると解釈される。一方、例2では「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。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、形容詞の種類と、名詞1、名詞2の性質から分類できるか検証した。分析は、大量の技術文から例文を抽出し、その係り関係を人手で調べることにより行った。例3には「が」格の場合の例を示す。

3. 分類

手始めに、「が」格、「の」格それぞれが形容詞に係るか否かという基準を考える。この基準に基づくと、以下の4タイプを想定することができる。

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。

抽出した例文（同じ形容詞が6件以上検出できたものが対象）を事例の出現頻度に従ってこの4つのタイプにマップすると以下ようになる。ただし、慣用句は一語とし認識されるべきものであるので統計からはずした。例えば、「きめの細かい」のような語は「細かい」の係りの判定からはずしている。

(出現頻度6件以上のイ形容詞一覧)

分類1 (14語): 多い、大きい、少ない、高い、小さい、強い、無い、長い、激しい、低い、等しい、深い、短い、難しい

- 分類2 (1語) : 広い
 分類3 (2語) : 良い、悪い
 分類4 (4語) : 新しい、詳しい、正しい、若い

イ形容詞	「が」が形容詞に		「の」が形容詞に		
	係る	係らない	係る	係らない	
分類1	155	9	230	38	全体 641件
分類2	3	2	1	10	網掛け数=解釈可能数
分類3	3	4	59	5	575件
分類4	5	17	3	97	OK率 89.7%

表1 イ形容詞の係りの数

(出現頻度6件以上のナ形容詞) 連体詞もナ形容詞として分類した。接尾辞は一まとめにした。

- 分類1 (4語) : 可能な、容易な、不可能な、不要な
 分類2 (4語) : 豊富な、有効な、必要な、簡単な
 分類3 (0語) : 該当なし
 分類4 (21語) : 新たな、大幅な、主な、完全な、最適な、様々な、主要な、詳細な、重要な、正確な、適切な、適当な、特殊な、特別な、複雑な、
 ~規模な、~速な、~大な、~的な、大きな、小さな

ナ形容詞	「が」が形容詞に		「の」が形容詞に		
	係る	係らない	係る	係らない	
分類1	93	0	14	14	全体 1004件
分類2	53	5	6	80	網掛け数=解釈可能数
分類3	0	0	0	0	960件
分類4	10	78	9	642	OK率 95.6%

表2 ナ形容詞の係りの数

この分類のみに依存して解析を行っても、表に示したようにイ形容詞で89.7%、ナ形容詞で95.6%の精度で係りの解釈ができる。このように、この4分類は係りの決定の標準ルールを設定するには有効ではあるが、更に解析率を上げるためにはより詳細なルールの設定が必要である。

4. 詳細な分析

これらの係り関係を、前後の名詞(名詞1、名詞2)で分析すると以下の3つのルールが係り関係を規定する上で効果的であることが分かった。

ルール1: 名詞1が、以下のものである場合は、名詞1は名詞2に係る

- 序数詞 (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
- 連体詞 (この、その、・・・) 例) この複雑な部分のベクトル化の促進
- まとめる語 (等の、他の) 例) ランダムエントリ、XIF等の豊富なエントリ手法
- 時の表現 (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) CEBの中の不要な分割を回復する

ルール2: 複合助詞は形容詞には係らない

- での、からの、への等 例) 中低速回転域での高いトルク～。

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ず名詞2に係ると考えるのが自然と思われる。

ルール3：名詞2が形式名詞の場合は、名詞1は形容詞に係る

場合、時、こと、方、順、側等 例) 計画と実績の差異が大きい場合の原因を知る

形式名詞は、意味的には希薄な語である。そして直前の用言を終止形的に働かせ係りを閉じさせる働きをする。そのため、形式名詞が具体的意味合いを持たない時は「の」「が」格は形容詞に係るとするのが自然と思われる。

これらの3つのルールを付加して分類した形容詞の係り関係の詳細分析を表3～表8に記載する。但し、イ形容詞の分類2は「広い」一語のみであるので、表1で代替することとした。この表から分かるように最終的には、以下のように、高い精度で係り関係を特定できた。

イ形容詞	全体	641件	解釈可能数	608件(94.9%)
ナ形容詞	全体	1004件	解釈可能数	976件(97.2%)

5. 更なる精度向上に向けて

ここで得られた精度は、個々の語の特性を加味することによりさらに改善される。例えば、「特別な」というナ形容詞は形式名詞「場合」を普通名詞化し、「高い」というイ形容詞は「～率」、「～性」と強い結び付きを持つという個別の付加条件を付けるというものである。この処理は語と語の関係に依存するので、単語間の細かいヒューリスティック処理となる。また、形式名詞が具体的意味を持つ場合があり、具体的意味を持つ確率から、形式名詞を2つのタイプ(タイプ1:場合、時、こと等、タイプ2:方、順、側等)に分けるというものである。例えば、『長さは「x」と「y」の長い方の長さがとられる。』のように、タイプ2には具体的なものを表す可能性が強く、意味論的補正を加えて係りの強弱のルールを追加するというものである。

以下に詳細な分類表を示す。

イ形容詞 分類1	「が」が形容詞に係る				「が」が形容詞に係らない				「の」が形容詞に係る				「の」が形容詞に係らない			
	RO	R1	R2	R3	RO	R1	R2	R3	RO	R1	R2	R3	RO	R1	R2	R3
	多い	<u>8</u>	0	0	<u>28</u>					<u>16</u>	0	0	<u>0</u>			
大きい	<u>6</u>	0	0	<u>13</u>	3	0	0	0	<u>18</u>	0	0	<u>7</u>				
少ない	<u>6</u>	0	0	<u>14</u>					<u>15</u>	0	0	<u>1</u>				
高い	<u>5</u>	0	0	<u>11</u>	1	<u>1</u>	0	0	<u>107</u>	0	0	<u>7</u>	7	<u>2</u>	<u>3</u>	0
小さい	<u>7</u>	0	0	<u>12</u>	0	0	0	1	<u>5</u>	0	0	<u>2</u>	2	<u>2</u>	<u>0</u>	0
強い	<u>2</u>	0	0	<u>4</u>	1	0	0	0	<u>6</u>	0	0	<u>2</u>	3	0	<u>2</u>	0
無い	<u>0</u>	0	0	<u>8</u>					<u>12</u>	0	0	<u>1</u>				
長い	<u>2</u>	0	0	<u>5</u>					<u>3</u>	0	0	<u>2</u>	0	<u>1</u>	<u>0</u>	2
激しい	<u>2</u>	0	0	<u>0</u>					<u>4</u>	0	0	<u>0</u>	0	<u>1</u>	<u>1</u>	0
低い	<u>3</u>	0	0	<u>2</u>					<u>14</u>	0	0	<u>0</u>	1	<u>1</u>	<u>0</u>	0
等しい	<u>1</u>	0	0	<u>9</u>	1	0	0	0	<u>1</u>	0	0	<u>0</u>				
深い									<u>3</u>	0	0	<u>0</u>	2	<u>2</u>	<u>1</u>	0
短い	<u>1</u>	0	0	<u>2</u>					<u>2</u>	0	0	<u>2</u>	1	<u>2</u>	<u>1</u>	0
難しい	<u>2</u>	0	0	<u>2</u>	1	0	0	0					1	<u>1</u>	<u>0</u>	0
合計1	<u>45</u>	0	0	<u>110</u>	7	<u>1</u>	0	1	<u>206</u>	0	0	<u>24</u>	16	<u>12</u>	<u>8</u>	2
合計2				<u>155</u>				9				<u>230</u>				<u>38</u>

R0:ルール1～ルール3以外の一般ケース
R1:ルール1に該当するもの
R2:ルール2に該当するもの
R3:ルール3に該当するもの
表中空白は抽出例なしを示す
(以下の表は全てこの形式)

下線付き数字が解釈可能数
斜体の数字が解釈の誤りの数

合計1がルール適用後の合計
合計2は4分類のみ合計

表3 イ形容詞(「が」「の」格は形容詞に係るケース)

イ形容詞 分類3	「が」が形容詞に								「の」が形容詞に								
	係る				係らない				係る				係らない				
	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	
良い/よい	0	0	0	1	3	0	0	0	46	0	0	6	2	0	1	0	
悪い	0	0	0	2	1	0	0	0	4	0	0	3	2	0	0	0	
合計1	0	0	0	3	4	0	0	0	50	0	0	9	4	0	1	0	
合計2	3								4				59				5

表4 イ形容詞(「が」格は形容詞に係らないが、「の」格は形容詞に係る)

イ形容詞 分類4	「が」が形容詞に								「の」が形容詞に								
	係る				係らない				係る				係らない				
	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	
新しい					10	0	0	0	0	0	0	3	37	23	11	0	
詳しい													13	0	4	0	
正しい	0	0	0	5	3	0	0	2					3	0	1	0	
若い					2	0	0	0					2	3	0	0	
合計1	0	0	0	5	15	0	0	2	0	0	0	3	55	26	16	0	
合計2	5								17				3				97

「正しい」はルール3が有効に働く

表5 イ形容詞(「が」「の」格は形容詞に係らないケース)

ナ形容詞 分類1	「が」が形容詞に								「の」が形容詞に								
	係る				係らない				係る				係らない				
	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	
可能な	52	0	0	12					7	0	0	0	1	3	0	0	
容易な	10	0	0	5					3	0	0	0	2	0	2	0	
不可能な	2	0	0	4													
不要な	5	0	0	3					4	0	0	0	1	5	0	0	
合計1	69	0	0	24					14	0	0	0	4	8	2	0	
合計2	93												14				14

「の」格に対してルール1～ルール3を適用しないと特徴が見出せなくなる

表6 ナ形容詞(「が」「の」格は形容詞に係る)

ナ形容詞 分類2	「が」が形容詞に								「の」が形容詞に								
	係る				係らない				係る				係らない				
	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	
豊富な	1	0	0	0					1	0	0	0	4	3	0	0	
有効な	3	0	0	5	2	0	0	0	3	0	0	0	17	1	0	0	
必要な	23	0	0	15	1	0	0	0	1	0	0	1	4	1	1	0	
簡単な	3	0	0	3	1	1	0	0					44	5	0	0	
合計1	30	0	0	23	4	1	0	0	5	0	0	1	69	10	1	0	
合計2	53								5				6				80

表7 イ形容詞(「が」格は形容詞に係るが、「の」格は形容詞に係らないケース)

ナ形容詞 分類4	「が」が形容詞に								「の」が形容詞に							
	係る				係らない				係る				係らない			
	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3	R0	R1	R2	R3
新たな					1	0	0	0					2	0	5	0
大幅な					2	0	0	0					11	0	1	0
主な					7	1	0	0					70	9	1	0
完全な													9	3	0	0
最適な					2	0	0	0					10	2	1	0
様々な					1	0	0	0					6	0	0	0
主要な					2	0	0	0					7	6	0	0
詳細な					1	0	0	0					8	2	2	0
重要な	1	0	0	4	11	0	0	0					14	15	2	0
正確な													12	1	2	0
適切な	0	0	0	0	1	0	0	0					6	1	0	0
適当な	0	0	0	1	3	0	0	0					0	7	0	0
特殊な					1	0	0	0					8	9	0	0
特別な					2	0	0	0					5	2	0	3
複雑な	2	0	0	0	4	0	0	0					2	6	1	0
～規模な													2	3	1	0
～速な													19	1	2	0
～大な													3	8	1	0
～的な	0	0	0	2	19	0	0	0					303	0	0	0
大きな					11	1	5	0	8	0	0	0	23	7	1	0
小さな					3	0	0	0	1	0	0	0	8	9	0	0
合計1	3	0	0	7	71	2	5	0	9	0	0	0	528	91	20	3
合計2				10				78					9			642

連体詞の解釈には追加補正が必要

表8 イ形容詞(「が」「の」格は形容詞に係らないケース)

6. まとめ

本稿ではイ・ナ形容詞を含む連体修飾について、局所的な表層情報から分析し、精度良く節を特定できることを示した。これからは、これらの分析結果を新たな大量の文に適用し、その正当性を検証すると共に、特定の単語や形式名詞について更なる特徴抽出を行い解釈精度の向上を計って行く予定である。また数多くの必須格や自由格を持つ動詞を含む連体修飾の統計的な特徴抽出についてもこれからの課題である。本稿の分析は、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、この分類が効果的であると判断している。

参考文献

- 1) 黒橋、長尾：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、Vol.33 No.8 1022-1031 (1992)
- 2) 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 No.1 35-57 (1994)
- 3) 白井、横尾、池原、木村、小見：日本語従属節の依存構造に着目した係り受け解析、自然言語処理 102-9(1994)
- 4) 北村、荻野：日英翻訳における連体修飾句の訳し分け 自然言語処理 75-10(1990)
- 5) 益岡隆志、田窪行則(1989) 基礎日本語文法 くろしお出版