

新聞記事日本文における修飾句と被修飾複合語の関係

4 A E - 1 0

白井 諭*1 十河則子*2 木村淳子*2 横尾昭男*1 内野 一*1 池原 悟*3

*1NTTコミュニケーション科学研究所 *2NTTアドバンステクノロジー(株) *3鳥取大学工学部

1 はじめに

日本語の長い文には従属節や名詞句が数多く含まれるため、これらの成分の相互関係を正しく解析することが必要である。従属節に関しては、文法的関係に意味的な働きを加味した分類により相互関係が高い精度で決定される[白井 95]。名詞句に関しては、並列関係の解析はマーカとなる表現や単語の連なり類似性に着目するのが効果的であり[黒橋 92][山岡 95]、同格関係の認定は統語的意味的特徴に着目することにより高精度で同定可能である[横尾 90]。名詞の意味的な関係についても定式化が試みられている[島津 85][冨浦 95]。

日本語の構文解析では、処理の基本単位を「文節」とする場合が多いが、そうすると[首藤 86]で分類されている並列関係の一部には適用できないという問題がある。例えば、新聞記事には「日本、米国市場」のような並列表現がしばしば現れる。これは勿論「日本市場、米国市場」と解するべきであり、後側の文節が複合語である場合に発生する現象である。

本稿では、このような部分的な関係に着目して新聞記事 965 文に含まれる名詞句を分析した結果を示すとともに、構成要素の関係の解析方法を検討する。

2 新聞記事における名詞句の分布

初めに、新聞記事にどのような名詞句が現れるかについて調査した。ここでは、単一名詞または複合名詞を主要素とする文節(名詞文節)が連体・並列・同格の関係で結びついたもの、および、これらに埋め込み文修飾が付いたものを名詞句とした。連体には連体詞を含め、並列には助詞等による呼応や句接続詞を含めた。埋め込み修飾されている場合、埋め込み文の述語文節は名詞句に含めたが、格要素は「通

信速度が毎秒二千四百ビットの | 高速サービスへの | 対応を」のように一体化している場合を除き、それぞれを別の名詞句として集計した。

表 1 に新聞記事 965 文(日経産業新聞・情報欄・リード)に現れた名詞句の長さ(文節数)と修飾文節の機能別の出現件数を示す。表から、2 項関係のものが半数以上になることがわかる。また、名詞句が長くなると相対的に並列の含まれる割合が増している。

表 1 名詞句の長さ(文節数)と構成

文節数	1	2	3	4	5	6	7	8	9	10~18	合計
件数	2048	1227	399	132	38	21	6	7	4	5	3887
(比率)	52.7	31.6	10.3	3.4	1.0	0.5	0.2	0.2	0.1	0.1	100%
名詞句中の関係	連体	582	525	243	82	64	18	21	16	16	1567
	並列	73	77	82	42	29	15	20	11	21	370
	同格	66	29	17	7	2					121
	埋込	499	153	49	18	8	3	3	3	10	746
	格要素	7	14	5	3	2		5	2	3	41

3 名詞句の部分関係

名詞句を構成する文節数が 2 以上となると係り受け構造に多義を生じ、その解消には文法的関係だけでなく意味的關係が必要となる。しかし、2 文節から構成されている名詞句でも、後側文節の自立語部分が複合名詞である場合に問題が生じることがある。図 1 に前側文節が後側文節の自立語部分全体を修飾していない例を示す。

①部分連体

- ・構造 / 流体 / 解析 / の | 分野 / 向け / に
- ・自治体 / への | 納入 / 実績 / が
- ・日米 / 合弁 / の | ソフト / 会社 / ロゴ / ジャパン / は

②部分並列

- ・設計 / 事務 / 所 / や | 建設 / 会社 / 向け / に
- ・情報 / 事業 / 、 | 海外 / 技術 / 両 / 本部 / 内 / に
- ・CAE / 、 | AI / ソフト / 製作 / などに

③部分埋込

- ・(会場が | 数カ所に) 分散している | 各種 / イベント / 向け / で
- ・急激な | 円高 / 状況 / にも
- ・(…へ | 自動的に) 電話 / 連絡 / する | 非常 / 通報 / 装置 / ヘルプ / メイト / を

④部分格

- ・(××工場は) 昨年 / 秋 / に | 完成 / 予定 / だった / が

図 1 2 文節間の部分関係の例

このような部分関係は、日英機械翻訳では次のような問題がある。例えば、「日本、米国市場」の場合、「the Japan and U. S. markets」のように訳される

Relationships between Modifiers and Compound Words in Japanese Newspaper Articles

Satoshi SHIRAI*1, Noriko SOGO*2, Junko KIMURA*2, Akio YOKOO*1, Hajime UCHINO*1 and Satoru IKEHARA*3

*1NTT Communication Science Laboratories, *2NTT Advanced Technology Corporation and *3Faculty of Engineering, Tottori University

なら部分関係を考えなくても済むが，“Japan and the market in the U. S.”のように訳されれば誤りとなる。常に前者のように訳出されるわけではないので、何らかの処理を導入する必要がある。

表1の名詞句のうち、図1のような部分関係がどのくらい含まれているかを表2に示す。次節以降では件数の多い部分連体関係の処理について検討する。

表2 部分関係の出現数

関係の種類	連体	並列	埋込	格
出現件数	1567	370	746	41
部分	116	22	65	8
(比率)	7.4%	5.9%	8.7%	19.5%

4 部分連体関係の分析

次に、部分連体関係の検出方法の確立を目的として、部分連体関係の分析を行なった。後側の文節に含まれる複合語の構成単語のうち、部分連体関係の原因と思われる要素に着目した分類を表3に示す。

「名詞+固有名詞」型の複合語では、名詞が直前要素の修飾を受け、固有名詞と同格となっていた。「用・向け・程度・サービス」などの語の前に部分連体が多く見られた。事例が少ないので現在は個別にルール化しているが、機能的な接尾語や範疇を表す語などの場合に部分関係が発生することが多いようである。

表3 部分連体関係の分類

着目要素	件数	例	他の語例
固有名詞	43	日本電気のパソコンPC9800シリーズ	—
接尾語	31	通信サービス関連のソフト用	型、向け
サ変名詞	22	自社ブランド製品の輸出拡大	対応、開始
一般名詞	17	将来像についての調査研究結果	実績、技術
連用形名詞	3	英和辞典の販売差し止め	引き上げ
合計	116		

5 部分連体関係の処理

部分関係の処理方法として、文節を分割して部分並列関係を解消することにより通常の並列名詞句に帰着させる方法が提案されている[白井93]。例えば、図2のようになる。

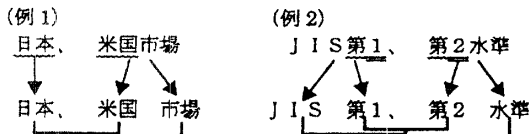


図2 文節分割による部分並列関係の解消

この方法を部分連体関係に適用すると図3のようになる。図3では、接尾語「用」を中心要素とする文節が生成される。しかし、接尾語のみを独立した文節とするのは極めて不自然である。また、接尾語を単独で訳出するのは可能でない場合も少なくない

ため、その観点からも問題である。

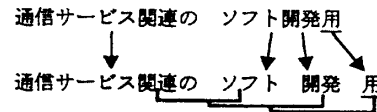


図3 部分連体関係の場合の問題点

一方、名詞の連続により構成される複合語を翻訳する際、核となる機能語に着目して構造化する方法(機能語支配型複合名詞翻訳処理)が提案されている[内野95]。しかし、図3の例のような場合、後方の複合語を単独で訳出するだけでなく、複合語の前の要素と一体的に処理することが必要で、名詞句処理と密接な連係を保つ必要がある。

そこで、本稿では部分関係にあずかる文節を疑似的な複合語として束ね、複合語解析の中で一体的に処理する方法を提案する。この方法によれば、結果的に複合名詞翻訳処理の中で直前の名詞句を考慮した処理を行なうことが可能になる。また、名詞句処理と複合語処理の間のインターフェース等を変更しなくても、処理に必要な情報が渡せるため、この点からも有利であると考えられる。

6 おわりに

本稿では、新聞記事に含まれる名詞句のうち、特に前側の文節が後側の文節に含まれる複合語の全体を連体修飾していない場合、すなわち、部分的な連体関係が生じている場合を検討し、処理の方法を提案した。現在、本提案に基づいて処理系を改良中であり、その結果については別の機会に報告する。また、本提案を突破口にして、[水谷93]で示されているような表現への適用を進めていく予定である。

参考文献

- [黒橋92] 黒橋,長尾:長い日本語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, pp.1022-1031
- [水谷93] 水谷:意味・構文を考へる九十例, 計量国語学, Vol.19, No.1, pp.1-14
- [島津85] 島津,内藤,野村:日本語文意味構造の分類 --名詞句構造を中心に--, 情報処理学会 自然言語処理研究会, 47-4, pp.25-32
- [白井93] 白井,横尾,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, 情報処理学会第47回全国大会, 3M-1, Vol.3, pp.97-98
- [白井95] 白井,池原,横尾,木村: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, No.10, pp.2353-2361
- [首藤86] 首藤,吉村,津田:日本語技術文における並列構造, 情報処理学会論文誌, Vol.27, No.2, pp.183-190
- [富浦95] 富浦,中村,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
- [内野95] 内野,横尾,池原:日英機械翻訳における複合名詞翻訳処理, 第1回言語処理学会年次大会, C2-3, pp.273-276
- [山岡95] 山岡:並置構造の分析と計算機による実験, 計量国語学, Vol.20, No.2, pp.55-79
- [横尾90] 横尾:同格名詞句の日英翻訳について, 情報処理学会第41回全国大会, 4S-2, Vol.3, pp.153-154