日本語名詞述語文の類型と主語の意味分類について
京都大学テキストコーパスと分類語彙表を用いた調査・分析

今　田　水　穂

1．はじめに

本研究では，京都大学テキストコーパスを対象として，日本語名詞述語文の意味特徴の調査と分析を行う。特に, 主語名詞句の意味分類に注目する。まず, 京都大学コーパスに付加されている形態論情報，構文情報，格関係情報などを手掛かりにして，主語名詞句および述語名詞句の抽出を行う（第 2 節)。次に, 分類語彙表データベースを用いて，抽出した名詞句に意味分類情報を付与し, 基本的な計量データを提示する（第 3 節)。最後に,「生産物」を主語とする名詞述語文を例にとって，個別の文に対する分析を行い，名詞述語文の意味構造の類型化と記述の方法, および設定すべき課題について検討を行う (第 4 節)。名詞述語文の類型的な多様性，体系性を分析，記述するためには，大量の言語資料に基づく実例主義的研究が不可欠であるが，機械処理を併用することによって，大規模な実例調査を効率的に実施することができる。

2．データの抽出

2.1.　京都大学コーパス

京都大学コーパスは，毎日新聞 1995 年度版のテキストデータの一部，約 4 万文に対して，機械処理によって形態論・構文情報を付与し，人手によって修正を施したものである（以下，syn コーパス)。また，そのうちの 5000 文に対しては，格関係，照応・省略関係，共参照の情報が付与されている（以下, rel コーパス)。本研究では機械処理の便宜上，これらのコーパスを XML 形式のファイルに変換したものを使用した。解析には Ruby と MSXML を用いた。

2.2.　判定詞の抽出

京都大学コーパスの品詞体系は益岡・田窪（1989）に基づいており，コピュ

ラの「だ」「です」等には助動詞ではなく判定詞という品詞が割り当てられている。本調査では，述部に判定詞を含む文を名詞述語文と見なし，調査対象とした。名詞述語文は「だ」などのコピュラを伴わない場合もあるが，そのような事例については抽出の難しさのため調査対象から除外した。京都大学コーパス全文を対象として判定詞の全事例を抽出した結果，syn コーパスからは 9010 例，rel コーパスからは 992 例の判定詞の事例が得られた（図 1）。

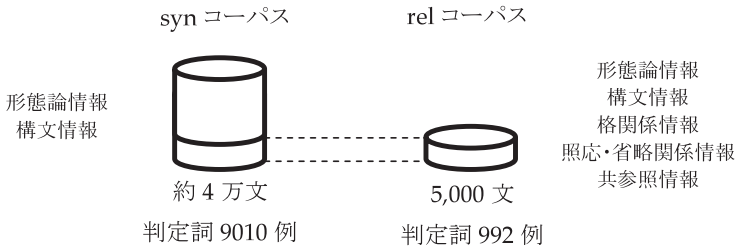

図 1　京都大学テキストコーパスにおける判定詞

2.3.　主語名詞句の抽出

2.3.1.　構文情報に基づく抽出

コーパスから抽出した名詞述語文に対して，二通りの方法で主語名詞句の抽出を行った。第一の方法は構文情報（係り受け情報）に基づく主語名詞句の抽出である。syn コーパスについては利用可能な情報が形態論情報と構文情報のみなので，基本的には構文情報に基づいて主語名詞句を抽出することになる。具体的には，図 2 のような手続きによって主語名詞句の抽出を行った。

この方法による名詞述語文の抽出では，syn コーパスの判定詞 9010 例のうち 5131 例（56.95%）について主語名詞句が抽出された。また，syn コーパスの一部である rel コーパスについては，判定詞 992 例のうち 562 例（56.65%）について主語名詞句が抽出された（図 3）。

どちらのコーパスにおいても，主語名詞句の抽出数は全体の 57% 程度であり，次節で述べる格関係情報に基づく抽出と比べると数が少ない。これは，構文情報に基づく方法では，述語と直接的に係り受け関係にある要素の中からしか主語名詞句を抽出することができない，という事情による。例えば，主語が省略されている文などにおいては，構文情報に基づく方法で主語名詞句を抽出

条件1　述語にかかり、文節末が副助詞(「は」など)または格助詞の「が」である文節を主語名詞句と見なす。ただし、副助詞の直前が格助詞の場合(「には」など)、動詞テ形の場合(「とっては」など)は除外する。

条件2　条件1を満たす文節が複数ある場合は、述語に最も近い位置にあるものを主語名詞句と見なす(複主語構文については取り扱わない)。

条件3　条件1を満たす文節が一つもない場合は、述語句が連体修飾的であるならば、係り先の文節(被連体修飾句)を主語名詞句と見なす。

図 2　主語名詞句の抽出

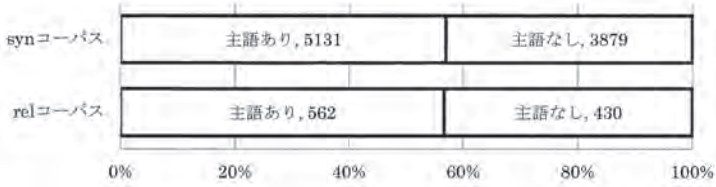

図 3　構文情報に基づく主語名詞句の抽出

することはできない。より多くの主語名詞句を抽出するためには，ゼロ代名詞の判別や照応解析など，より高度の解析処理を行う必要がある。

2.3.2.　格関係情報に基づく抽出

第二の方法は，格関係情報に基づく主語名詞句の抽出である。rel コーパスには構文情報などの他に格関係情報などの情報が付加されているため，この情報を用いて主語名詞句を抽出することができる。具体的には，判定詞を含む句に付与されている格関係情報のうち，「ガ」格項として指定されている要素を

主語名詞句と見なし，要素が複数指定されている場合には，便宜的に，最も後方にある要素を主語名詞句とした。この結果，rel コーパスの判定詞 992 例のうち，901 例（90.83%）について主語名詞句を抽出した（図 4）。

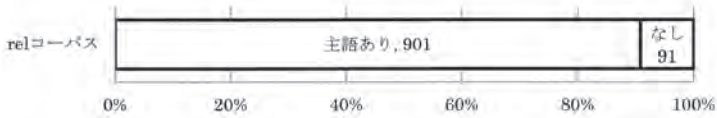

図 4　格関係情報に基づく主語名詞句の抽出

構文情報に基づく抽出（図 2）と格関係情報に基づく抽出（図 3）を比べると，前者の主語抽出数が全体の 57% 程度に留まるのに対して，後者の主語抽出数は全体の 90% に及ぶ。これは，前者の方法では述語と係り受け関係にある要素しか主語名詞句として抽出することができないのに対して，後者の方法では，格関係情報として述語と係り受け関係にある以外の要素も「ガ」格項として指定されており，より広い範囲から主語名詞句を抽出することができるためである。格関係情報に基づいて抽出された主語名詞句 901 例の中には，次の種類のものが含まれる。

（1）　a. 述語と係り受け関係にある要素（901 例中 513 例）
　　例）［ 帆走は 科学である。 ］
　　b. 述語と同一文中だが，係り受け関係にない要素　（901 例中 112 例）
　　例）［ 韓国の 米は ［ 日本と 同じ 短粒種で ］ コシヒカリに 劣らず うまい。 ］
　　c. 同一文中にない要素 1（先行文脈中の要素）　（901 例中 197 例）
　　例）［メードの 収入は 月三千七百香港ドル。］　［マニラの標準的労働者の 約三倍だ。 ］
　　d. 同一文中にない要素 2（「不特定：状況」など）　（901 例中 74 例）
　　例）［ロシア側の苦戦が続いている 模様だ。 ］

構文情報に基づく主語名詞句の抽出では，（1a）のものについては抽出できる可能性があるが，（1b-d）のものについては原理的に抽出不可能である。このため，構文情報に基づく主語名詞句の抽出数よりも，格関係情報に基づく主

語名詞句の抽出数の方が，はるかに数が多い。

2.3.3.　抽出精度の検証

前節までに述べたように，構文情報に基づく主語名詞句の抽出は，格関係情報に基づく主語名詞句の抽出と比べて，抽出数がかなり少ない。しかしながら，格関係情報は京都大学コーパスの一部（rel コーパス）にしか付与されていないため，それ以外の部分については構文情報に基づき主語名詞句を推定する他ない。

そこで，構文情報に基づく抽出結果を「推定データ」，格関係情報に基づく抽出結果を「正解データ」と見なして，前者の方法による主語名詞句抽出の精度を検証した。検証には，両方の方法で抽出を行った rel コーパスを用いた。まず，二つの方法による主語抽出の有無を単純に交差集計したものを表 1 に示す。

表 1　推定データと正解データの交差集計（1）

推定データ＼正解データ	主語あり	主語なし	計
主語あり	547　(a)	15　(b)	**562**
主語なし	354　(c)	76　(d)	**430**
計	**901**	**91**	**992**

このうち（b）と（c）については，推定データと正解データが一致していないため，不正解であると見なすことができる。また（d）については，推定データと正解データがともに主語なしと判定しているので正解であると見なすことができる。(a）については，実際に同一の要素を主語名詞句と判定しているかどうかについて，さらに確認する必要がある。同一の要素が主語名詞句として判定されているか否かを区別して再集計したものを表 2 に示す。

表 2　推定データと正解データの交差集計（2）

推定＼正解	主語あり		主語なし	計
	一致	不一致		
主語あり	465　(a1)	82　(a2)	15　(b)	**562**
主語なし	-	354　(c)	76　(d)	**430**
計	**464**	**437**	**91**	**992**

このうち，（a1）を正解，（a2）を不正解の事例と見なすと，（a-d）までを集計した推定データの正解精度は表 3 の通りである。括弧内の数字は行の総計に対する比率である。

表 3　推定データの正解精度

推定データ	正解	不正解	計
主語あり	465 (82.7%)	97 (17.3%)	**562 (100.0%)**
主語なし	76 (17.7%)	354 (82.3%)	**430 (100.0%)**
計	**541 (54.5%)**	**451 (45.5%)**	**992 (100.0%)**

全体としては，推定に正解した事例は全事例の 54.5% であり，高い数値であるとは言えない。特に，正解データで主語ありとされていながら，推定データでは主語なし（主語が抽出できなかった）事例が 992 例中 354 例と，かなりの数に上る。これは 2.3.1 節で述べたように，構文情報に基づく抽出方法では主語名詞句の一部のものしか抽出することができないという事実を反映したものであると考えることができる。

一方で，推定データで主語ありとされた事例にのみ注目すると，562 例中 465 例（82.7%）と，8 割を超える精度で主語の推定に成功している。syn コーパスにおいても同程度の確率で構文情報に基づく主語名詞句抽出に成功しているものと仮定すると，syn コーパスにおいて主語ありとされた事例 5131 例（図 3）のうち，4245 例程度については主語の推定に成功しているものと考えられる。

2.4.　述語名詞句の抽出

判定詞と同一文節中にあり，判定詞より先行する部分を述語名詞句と見なして述語名詞句の抽出を行った。この結果，syn コーパスの判定詞 9010 例のうち，9007 例について述語名詞句が抽出された。述語名詞句が抽出されなかったのは，「だとしたら」の「だ」が判定詞としてタグ付けされていた場合などである。構文情報に基づく推定で主語ありと判定された事例 5131 例については，全事例について述語名詞句が抽出された。従って，syn コーパスに含まれる判定詞 9010 例のうち 5131 例について，主語名詞句と述語名詞句の対を推定，抽出することができたということになる（主語名詞句の推定正解率は，前節で述べた通り 82.7% 程度と考えられる）。

3．意味分類情報の付与と分析

3.1.　意味分類情報の付与

前節までに抽出した名詞述語文の主語名詞句および述語名詞句のデータについて，分類語彙表データベースを用いて意味分類情報を付与した。分類語彙表では，類（体，用，相，他），部門（関係，主体，活動，生産物，自然），中項目，分類項目という階層で分類が施されているが，ここでは部門を意味分類として用いた。多義語については語義の推定等は行わず，最も若い分類番号を付与した。この結果, syn コーパスの主語名詞句 5131 例中の 4829 例（94.11%），述語名詞句 9007 例中の 8547 例（94.89%）に意味分類情報を付与した。内訳は表 4 の通りである。「その他」は「その他の類」に属する事例を，「不明」は意味分類情報が付与されなかった事例を表す。

表 4　意味分類情報の付与

意味分類	主語		述語	
関係	3062	(59.68%)	5108	(56.71%)
主体	640	(12.47%)	881	(9.78%)
活動	957	(18.65%)	2196	(24.38%)
生産物	99	(1.93%)	151	(1.68%)
自然	67	(1.31%)	168	(1.87%)
その他	4	(0.08%)	43	(0.48%)
不明	302	(5.89%)	460	(5.11%)
計	**5131**	**(100.00%)**	**9007**	**(100.00%)**

それぞれの意味分類にどのような語が含まれるかを，出現頻度順に上位 5 位まで表 5 に示す。括弧内の数字は出現頻度である。

全体の傾向としては，「関係」や「活動」など抽象的な概念を表す語，および主体を表す語の比率が高い。特に，「関係」を表す語の比率が高く，主語，述語いずれの場合も過半数を占める。これは「の」や「こと」などの形式名詞など，出現頻度の高い機能的な表現が「関係」に分類されるためであると考えられる。一方，「生産物」や「自然」など具体的な事物を表す語の数は少なく，主語，述語いずれの場合も 1 ～ 2% 程度である。また，主語，述語ともおおよその語彙の分布は類似しているが, 主語の方が主体を表す語彙の比率が大きく，述語の方が活動を表す語彙の比率が大きいなど，若干の違いが見られる。

表 5 各意味分類に含まれる語彙（出現頻度上位 5 位まで）

意味分類	主語	述語
関係	の（1023），こと（210），それ（114），これ（98），さん（43）	もの（599），こと（383），から（352），だけ（276），ため（84）
主体	氏（31），日本（23），私（15），大統領（13），首相（13）	国（38），機関（16），道（15），日本（13），人間（11）
活動	問題（49），選（14），会議（12），制度（10），発言（9）	問題(113)，予定(58)，課題(50)，方針（38），見通し（37）
生産物	官邸（5），車（3），兵器（3），スキー場（3），建物（3）	商品（4），車（4），定番（3），ロケット（3），産物（3）
自然	地震(4)，ビタミン(3)，光景(2)，死（2），列島（2）	沖（11），地震（6），物質（5），反映（4），いるか（4）
その他	それこそ（3），もしも（1）	だから（17），ん（8），なぜ（8），どうして（3），もちろん（2）

3.2. 主語の意味分類と出現形式

次に，主語の意味分類と出現形式（主語が「は」で表示されるか，「が」で表示されるか，被連体修飾語として現れるか，など）の相関について調べた。結果を表 6 に示す。括弧内の数字は行の総計に対する比率である。

表 6 主語の意味分類と出現形式

意味分類	は	が	も	その他の副助詞	被連体修飾語句	計
関係	1785 (58.30%)	811 (26.49%)	209 (6.83%)	54 (1.76%)	203 (6.63%)	**3062** **(100.00%)**
主体	325 (50.78%)	115 (17.97%)	43 (6.72%)	16 (2.50%)	141 (22.03%)	**640** **(100.00%)**
活動	591 (61.76%)	200 (20.90%)	58 (6.06%)	28 (2.93%)	80 (8.36%)	**957** **(100.00%)**
生産物	65 (65.66%)	15 (15.15%)	6 (6.06%)	2 (2.02%)	11 (11.11%)	**99** **(100.00%)**
自然	36 (53.73%)	18 (26.87%)	4 (5.97%)	3 (4.48%)	6 (8.96%)	**67** **(100.00%)**
その他	1 (25.00%)	3 (75.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**4** **(100.00%)**
不明	155 (51.32%)	63 (20.86%)	13 (4.30%)	9 (2.98%)	62 (20.53%)	**302** **(100.00%)**
計	**2958** **(57.65%)**	**1225** **(23.87%)**	**333** **(6.49%)**	**112** **(2.18%)**	**503** **(9.80%)**	**5131** **(100.00%)**

全体の傾向としては「その他」を除く全ての意味分類で「は」が過半数を占めるが，主語の意味分類によって出現形式にかなりの偏りがあることが分かる。「は」と「が」の使用比率で見ると，「生産物」は「は」の，「関係」「自然」は「が」の使用比率が他の意味分類より高く，「活動」はその中間である（図 5）。「主体」は「は」と「が」の比率だけを見ると「活動」と同程度だが，他の意味分類と比べて被連体修飾語句として現れる比率が高い（図 6）。

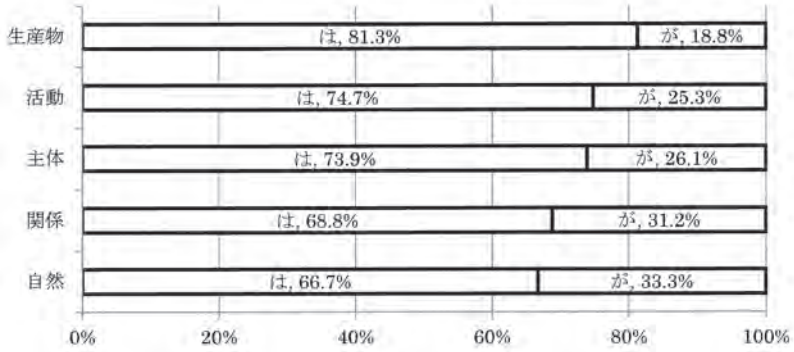

図 5　「は」と「が」の比率

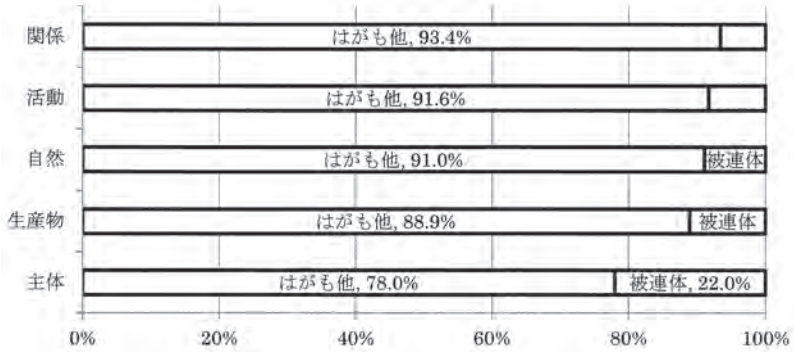

図 6　被連体修飾語句の比率

このうち「関係」が「が」で生起する比率が高いのは，前節の表 5 で示した通り，「関係」に分類される語彙の中に形式名詞の「の」が含まれており，これが「～のが～だ」の形式で多数用いられていることが原因の一つであると考えられる。

「の」は「関係」に分類される主語 3062 例のうち 1023 例を占めるが，その出現形式の内訳は「のは」644 例,「のが」310 例,「のも」69 例であった。表 7 に,「関係」における「は」と「が」の比率を,「の」と「の」以外に分けて示す。

表 7 「関係」を表す主語における「は」と「が」の比率

	は		が		計	
「の」	644	(67.5%)	310	(32.5%)	**954**	**(100.0%)**
「の」以外	1141	(69.5%)	501	(30.5%)	**1642**	**(100.0%)**
計	**1785**	**(68.8%)**	**811**	**(31.2%)**	**2596**	**(100.0%)**

表 7 を見ると,「のが」の比率の高さが,「関係」全体における「が」の比率を押し上げていることが分かる。ただし「の」以外の比率を見ても,「生産物」「活動」「主体」などよりは「関係」の方が「が」で現れる比率が高いようである（図 5 と比較されたい)。

それ以外のものについてはまだ分析が十分でないが，参考として,「生産物」＋「は」,「自然」＋「が」，および「主体」が被連体修飾語句として表れる事例を以下に示しておく。

(2) 「生産物」＋「は」
  a. 二十一世紀の新幹線はどんなものなのか。
  b. カレーは三百円だが，サラダは三百十円する。
  c. ともに使われた凶器は，けん銃だった。
(3) 「自然」＋「が」
  a. ビタミンが大切なものであることはだれもが知っている。
  b. 大小２頭のゴリラが主役で，盗まれたバナナを取り戻す旅に出る設定。
  c. ペンギンの羽毛が黒と白なのに対し，ゴイサギは黒とグレー。
(4) 被連体修飾語句の「主体」
  a. 松が取れてからしばらくして，小学三年生だった二男が，ポツンと言った。
  b. 前政権の大統領だったデクラーク現副大統領を強く非難した。
  c. 世界一の地震国である日本も見習うべきだろう。

同じ具象物を表す語でも，「生産物」と「自然」で出現形式に違いが観察されるという点は興味深い。しかしながら，どちらの意味分類も収集した用例の数が少ないため，本調査の結果だけでは明らかに差があると結論することはできない。「主体」については「生産物」や「自然」と比べると用例数が多く，少なくとも新聞テキストに関する限り，他の意味分類よりも被連体修飾語句の位置に生起する比率が高くなる傾向があると考えられる。

3.3.　主語の意味分類と述語の意味分類

次に，主語の意味分類と述語の意味分類の相関について調べた。

表 8　主語の意味分類と述語の意味分類

主語＼述語	関係	主体	活動	生産物	自然	その他	不明	総計
関係	1802 (58.85%)	225 (7.35%)	731 (23.87%)	43 (1.40%)	61 (1.99%)	17 (0.56%)	183 (5.98%)	**3062** **(100%)**
主体	273 (42.66%)	219 (34.22%)	102 (15.94%)	5 (0.78%)	8 (1.25%)	0 (0.00%)	33 (5.16%)	**640** **(100%)**
活動	617 (64.47%)	35 (3.66%)	253 (26.44%)	7 (0.73%)	8 (0.84%)	4 (0.42%)	33 (3.45%)	**957** **(100%)**
生産物	67 (67.68%)	3 (3.03%)	15 (15.15%)	10 (10.10%)	1 (1.01%)	1 (1.01%)	2 (2.02%)	**99** **(100%)**
自然	37 (55.22%)	5 (7.46%)	11 (16.42%)	1 (1.49%)	10 (14.93%)	0 (0.00%)	3 (4.48%)	**67** **(100%)**
その他	1 (25.00%)	0 (0.00%)	3 (75.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	**4** **(100%)**
不明	146 (48.34%)	47 (15.56%)	67 (22.19%)	9 (2.98%)	8 (2.65%)	3 (0.99%)	22 (7.28%)	**302** **(100%)**
総計	**2943** **(57.36%)**	**534** **(10.41%)**	**1182** **(23.04%)**	**75** **(1.46%)**	**96** **(1.87%)**	**25** **(0.49%)**	**276** **(5.38%)**	**5131** **(100%)**

表 8 の各行を全体の平均（一番下の総計の行）と比較すると，「主体」「生産物」「自然」を表す主語は，それぞれ「主体」「生産物」「自然」を表す述語を取る比率が高く，「活動」も同様の傾向がある程度見られる。これらの意味分類については，主語と述語の意味範疇がある程度一致する傾向が見られると言える。以下に例を示す。

(5)　主体 ― 主体

a. 【私たち】は今まで，決して，【おしどり夫婦】ではなかったですね。

b.【犯人】は外務審議官付の【女性秘書】だった。

c. ところが，もとを正せば【日本】は農業本位の【国】であった。

(6) 活動 ― 活動

a.【朗読】は女優の【仕事】である。

b.【放映】は五日の【予定】だが, ＣＢＳが四日, 発言要旨を公表した。

c. この地域主義の【克服】も今後の【課題】だろう。

(7) 生産物 ― 生産物

a.【発泡酒】は 【アルコール飲料】だが, ・・・

b. 問題になるのはその【薬】が 【抗がん剤】だった場合。

c. 阪神間を結ぶ【道路】は, 生活用品や復旧資源を運ぶ貴重な【幹線】だからだ。

(8) 自然 ― 自然

a.【豆】は静岡独特の【ゆで落花生】だ。

b.【東海地震】は海溝型の【地震】だが, ・・・

c.【アスベスト】は，肺がんの原因にもなる危険な【物質】だが, ・・・

一方,「関係」を表す主語が「関係」を表す述語を取る比率は全体の平均と比べてそれほど高くない。この理由を詳しく調べるために,「関係」を表す主語の中でも特に出現頻度の高いものを区別して，述語との共起関係を再集計した。結果は以下の通りである。

表 9 「関係」を表す主語と，述語の意味分類

主語＼述語	関係	主体	活動	生産物	自然	その他	不明	総計
の	595 (58.16%)	82 (8.02%)	253 (24.73%)	12 (1.17%)	11 (1.08%)	11 (1.08%)	59 (5.77%)	**1023** **(100%)**
こと	130 (61.90%)	14 (6.67%)	59 (28.10%)	2 (0.95%)	2 (0.95%)	0 (0.00%)	3 (1.43%)	**210** **(100%)**
それ	60 (52.63%)	7 (6.14%)	38 (33.33%)	1 (0.88%)	2 (1.75%)	1 (0.88%)	5 (4.39%)	**114** **(100%)**
これ	49 (50.00%)	1 (1.02%)	43 (43.88%)	1 (1.02%)	2 (2.04%)	0 (0.00%)	2 (2.04%)	**98** **(100%)**
その他	968 (59.86%)	121 (7.48%)	338 (20.90%)	27 (1.67%)	44 (2.72%)	5 (0.31%)	114 (7.05%)	**1617** **(100%)**
総計	**1802** **(58.85%)**	**225** **(7.35%)**	**731** **(23.87%)**	**43** **(1.40%)**	**61** **(1.99%)**	**17** **(0.56%)**	**183** **(5.98%)**	**3062** **(100%)**

表 9 を見ると，「の」「それ」「これ」などが「関係」を表す述語を取る比率が低いことが分かる。「の」は極端に比率が低いわけではないが，「関係」を表す主語 3062 例中 1023 例を占める。「それ」「これ」は「の」ほど出現頻度が高いわけではないが，「関係」を表す述語を取る率が顕著に低い。「の」は「〜のは〜だ」「〜のが〜だ」など分裂文の形式で用いられる語であり，「それ」「これ」は人や場所以外の様々な事物に対する指示詞として用いられる語である。そのため，いずれも述語に様々な意味範疇の語を取るのだと考えられる。

しかしながら，これら出現頻度の高い語を除いた「その他」の比率を見てみると，「関係」を表す述語を取る比率は 59.86% とやや高くなるが，顕著に高いとは言い難い（表 8 の総計と比較されたい）。従って，「の」「それ」「これ」などを除いてもなお，「関係」を表す主語は「関係」を表す述語と共起する比率が，それほど高くないのだと言うことができる。これについては事例をさらに詳しく分析する必要があるが，本稿ではこれ以上は立ち入らない。

3.4.　まとめ

分類語彙表データベースを用いて，名詞述語文の主語名詞句，述語名詞句に意味分類情報を付与した結果，主語名詞句の 94.11%，述語名詞句の 94.89% に意味分類情報を付与することができた。主語名詞句の意味分類と出現形式の関係を調べたところ，「主体」を表す主語は他の類と比べて被連体修飾語句の位置に生起する比率が高いなど，主語の意味分類に応じて出現形式に違いがあることが分かった。また，主語名詞句の意味分類と述語名詞句の意味分類との関係を調べたところ，「関係」を表す主語は形式名詞の「の」を多く含むなどの理由から述語の意味分類に顕著な偏りは見られなかったが，それ以外の類については主語と述語の意味分類が一致する比率が高くなる傾向が見られた。

4．個別事例の分析

4.1.「生産物」を主語とする名詞述語文

前節では，主語名詞句の意味分類と，その出現形式や述語名詞句の意味分類との間に一定の共起傾向があることを観察した。このような計量的調査は全体の傾向を見るためには有用だが，個々の事例を詳細に分析するためには十分ではない。名詞述語文の文法的構造の詳細を明らかにするためには，個々の事例を詳細に分析していくことが不可欠である。そのような個別的な調査，分析を

進める際にも，名詞句の抽出や意味分類情報の付与といった機械的手続きは役立つ。コーパスに基づく事例分析においては，まず収集した事例を分類，整理することが基本的な手続きの一つとなるが，機械処理によって，この手続きをある程度自動化することができる。例えば，特定の意味分類の語を主語とする文のみを機械的に抽出して，それらを集中的に分析するといった手続きが可能となる。

ここでは，「生産物」を主語とする名詞述語文を取り上げて，文の意味構造の詳細な分析を行う。分析は，文の意味構造と情報構造を独立した構造として扱う立場（今田 2009）から行うものとし，文の意味構造のみを問題とする。名詞述語文の意味構造については，名詞句の意味の観点から分析するもの（西山 2003 など）と，主語と述語の意味関係の観点から分析するもの（高橋 1984 など）があり，どちらの観点も必要であるが，本稿は後者の観点における意味構造の類型化を試みるものとする。以下では「生産物」を主語とする名詞述語文を，述語との意味関係に基づき，「範疇叙述型」「属性叙述型」「外延叙述型」「その他」の四つの類型に分類して記述する。分析と分類が未完了であるため，事例数の集計等は行っていない。

4.2. 範疇叙述型の文

名詞述語文の主要な類型の一つは，主語名詞句が表す事物の帰属する範疇を述語名詞句が述べるものである。(9-13) は述語名詞句が範疇的概念を表す名詞述語文の例であるが，(9) や (10) のように作品や飲料といった類の下位種を表す語が現れる場合や，(11) や (12) のように比較的上位の類概念を表す語が現れる場合，さらには (13) のようにごく抽象的なレベルの範疇しか表さない形式名詞が現れる場合などがある [1]。

(9) トルコの【「汗と絨毯」】は，トルコ東南部の町，ウルファで昔ながらの方法で絨毯を作る職人たちを追った【ドキュメンタリー】だ。

(10) 【発泡酒】は【アルコール飲料】だが，麦芽使用率が六五％以下のため，日本の酒税法の分類ではビールと見なされず，通常のビールより低い税率が適用される。

(11) お正月の獅子舞は，我々を守ってくれる神の「権現」であり，村境に張った【「しめ縄」】は，悪霊の侵入を防ぐ【道具】だった。

(12) ここ一，二年のうちに朝の食卓に登場してきた【オレンジジュース】

も新しい市場を開拓した【商品】である。

(13)【ワゴン車】はレンタル会社から借り出された【もの】で，借用書類からの指紋採取が進んでいた。

範疇的概念の中には，事物の博物学的分類を表すものもあれば，そうではないものもある。例えば (9) の「ドキュメンタリー」や (10) の「アルコール飲料」は博物学的分類の一種であるが, 次の「問題」「メッセージ」「温床」「幹線」「シンボル」などは，それぞれ「核の傘」「ラップ」「商業地」「道路」「石油」などの博物学的分類を示しているわけではない。これらの述語も範疇的概念を表すものではあるが，事物の博物学的分類を表しているというよりは，何らかの目的や機能を共有する事物や概念のグループを表している。

(14)【核の傘】は日米安保体制の根幹にかかわる【問題】だ。

(15)【ラップ】は若者への【メッセージ】だ。

(16) その半面，バブルの【温床】だった【商業地】は，大都市圏で二ケタの大幅マイナスとなった。

(17) 阪神間を結ぶ【道路】は，生活用品や復旧資材を運ぶ貴重な【幹線】だからだ。

(18) ところが, 同委員長は「【石油】はわが国の主権の【シンボル】である」として，真っ向から対立。

範疇を述べる文には，範疇情報自体が文の主要な伝達内容であるものもあるが，範疇情報自体よりも，それを修飾する部分の方が文の主要な伝達内容であるものも多い。(9) や (10) では，作品や商品がどのようなカテゴリーに分類されるかということが問題となっており，「ドキュメンタリー」や「アルコール飲料」といった範疇情報自体が文の主要な伝達内容となっている。一方，(11-13) では「道具」「商品」「もの」は文の主要な伝達内容ではなく，「どのような」にあたる部分の情報が文の主要な伝達情報となる。

「どのような」にあたる部分の情報には様々な種類のものがある。いくつか例を挙げると，(11) の「悪霊の侵入を防ぐ」や (19) の「国内向けの」は，事物の目的を表す。

(19) これについて, 冷蔵庫の査察を拒否した工場は, その冷蔵庫にある【製

品】は国内向けの【もの】だったからだと弁明している。

(12) の「新しい市場を開拓した」や (13) の「レンタル会社から借り出された」は，事物の来歴や過去の履歴を表す。類例としては次のようなものがある。

(20) 写真で首に巻いている【マフラー】は，インタビューの直前に歌手の荻野目洋子さんにプレゼントされた【もの】である。
(21) アウシュビッツで公開されている【ガス室】は解放後に共産党政権がでっち上げた【もの】で，ナチス指導者はユダヤ人絶滅を計画したことなどなかった，と指摘している。
(22) 男は陸軍士官学校卒の現役中尉で，【銃】は同士官学校から盗み出した【もの】だった。
(23) 同副首相は七日夜のロシア・テレビとのインタビューで，現在ドゥダエフ政権部隊が使用している大量の【兵器】は，一九九一年にロシア軍がチェチェンを撤退した後に同政権に引き渡した【もの】だと言明。
(24) 一カ月にわたるロシア軍の進攻に激しく抵抗しているドゥダエフ･チェチェン政権の【兵器】が旧ソ連軍の残した【もの】で，ドゥダエフ大統領との兵器引き渡し交渉に，グラチョフ・ロシア国防相があたっていたことがロシアで問題化している。
(25)【バッジ】は，実行グループの一人が落とした【もの】であることも裏付けられている。

(26) の「ドライバーの柄に千枚通しのようなものをつけた」は, 事物の構成, あるいは製造方法を表している。(27-29) の「大使の」「太閤秀吉の」「自社の」は，事物の所有者や生産者を表す。

(26)【凶器】はドライバーの柄に千枚通しのようなものをつけた【もの】だった。
(27)【公邸】は大使の【もの】ではない。国民の財産です。
(28) 太閤秀吉の【城】であった【伏見城】は，家康の持ち城となっていたからである。
(29) 製薬メーカー側は「分からない」「うちの製品ではない」と回答を避けるケースもあるが，【抗がん剤】が自社の【製品】だった場合「分から

ない」では不自然。

こうした様々な属性は，Pustejovsky のクオリア構造を想起させる。これは本来，語の意味の記述のために考案されたものであるが，事物の構成，分類，目的, 生成といった諸属性を統合的に扱うことができる。例えば (11) では,「道具」は事物の分類に関する属性（formal qualia）を表しているが,「悪霊の侵入を防ぐ」は事物の目的に関する属性（telic qualia）を表しており，述語名詞句全体は実質的に（30）のようなクオリア構造を持った事物を表す記述句として機能しているように見える[2]。このうち，文の主要な叙述内容となっているのは，事物の分類ではなく目的である。

（30）「悪霊の侵入を防ぐ道具」 = $\begin{bmatrix} const & \\ formal & instrument(x) \\ telic & prevent(x, the\ evil\ spilit's\ invasion) \\ agentive & \end{bmatrix}$

Pustejovsky の分類が様々な属性の全てを記述するために十分であるかどうかは分からないが，いずれにせよ，範疇を述べる文は実際には事物の分類をするだけではなく，他の様々な種類の属性を述べるために用いられるという点に注意する必要がある。そうした属性を何らかの方法で整理し，記述することが重要である。

4.3.　属性叙述型の文

前節では述語名詞の主要部が範疇的概念を表す語である場合を見たが，次のような事例では述語は主語が表す事物の帰属する範疇を表しているわけではなく，味，色，図柄，材質，数量，価格などの属性的概念を表している。

（31）皮が薄い分, たっぷり詰まった【つぶしあん】は抑えた【甘さ】で, さっぱりした仕上がり。
（32）【偽造投票用紙】は本物と同じ薄い【緑色】だが，紙がやや薄めで朱色で印刷された文字や公印が不鮮明。
（33）衣装一つをとっても黒繻子に松竹梅の模様に肩から袖にかけて金糸よりのしめ飾り，その下が火焔太鼓と桜の絵柄，前に垂れた【帯】は鯉の滝昇りの【図】である。
（34）また，宴会客のたばこの火の不始末の疑いも浮上したが，その後の調

べで【ステージ】はたばこの火などでは燃えにくい【材質】であることも分かった。

(35)【タンク】は【グラスファイバー製】で上部に直径五十センチのネジ込み式のふたがある。

(36) 日本の海運会社が運航している【外航貨物船】はおよそ【二千隻】である。

(37) 日本電子機械工業会が 11 日に発表した昨年 11 月のカラーテレビ・VTR 統計によると, 国内出荷台数は, カラーテレビが前年同月比 7·8% 増の 81 万 8000 台, 【VTR】は 12・0% 増の【51 万 7000 台】で, ともに 4 カ月連続して前年実績を上回った。

(38) 建設省のキャッチフレーズになった【「住宅」】は, 九五年度予算, 九六年度要求予算ともに, 一七・六%と同じ【シェア】だ。

(39)【カレー】は【三百円】だが, サラダは三百十円する。

(40)【実験器具】は【実費】で, 参加費は二万五千―三万円。

(41)「免許行政下で【一物一価】だった【ビール】が, 一物多価の時代に突入した」節目の年だった。

名詞述語文によって叙述される属性の種類は多様であり, それらをどのように整理するかは課題である。しかし, これらの属性の少なくとも一部は, クオリア構造のような形式で記述することができる。例えば,「甘さ」や「色」は対象の constituitive qualia の一種である。また上の例には含まれていないが,「ソニー製」や「子供用」のような属性は, それぞれ agentive qualia や telic qualia の一種と考えることができる。

また, 構文的特徴についてもう少し詳しく区別すると, 属性を表す述語名詞句の主要部は, 属性の種類を表す場合と, 値を表す場合がある。例えば (31) では,「甘さ」は属性の種類を表しており, それを修飾する「抑えた」の部分が, 属性の値が低め (抑えめ) であることを示している。コンピュータ·データベースなどで用いられる EAV モデルというデータモデルで記述すると, これらの文の意味構造は表 10 のように記述することができ, 属性または値のいずれかが述語名詞句の主要部になっているということになる。

実体, 属性, 値の三項の現れ方には, いくつかのパターンがある。代表的なものとしては, 表 11 のような構文が考えられる。E+A/V 型の構文においては, 主語は実体ではなく属性にシフトしており, 文全体は後述する外延叙述型の構文 (いわゆる役割 - 値文) にシフトしている。

表 10　EAV モデルを用いた属性名詞述語文の意味構造

実体（Entity）	属性（Attribute）	値（Value）
つぶしあん	甘さ	抑えた
偽造投票用紙	色	緑色
帯	図柄	（鯉の滝登りの）図
ステージ	材質	燃えにくい
タンク	材質	グラスファイバー製
外航貨物船	数量	二千隻
VTR	数量（国内出荷台数）	51 万 7000 台
住宅	数量（シェア）	同じ
カレー	価格	三百円
実験器具	価格	実費
ビール	価格	一物一価

表 11　EAV モデルに基づく構文類型

構文	例
E/V+A	［$_{E}$ 東京タワー］は［$_{V}$ 333m］の［$_{A}$ 高さ］だ。
E+A/V	［$_{E}$ 東京タワー］の［$_{V}$ 高さ］は［$_{A}$ 333m］だ。
E/A/V	［$_{E}$ 東京タワー］は［$_{V}$ 高さ］が［$_{A}$ 333m］だ。
E/A	［$_{E}$ 東京タワー］は［$_{A}$ 333m］だ。

属性と値は，一般的には後述する役割と値の関係で捉えることができるが，表 10 ではそうではないものも便宜的に EAV モデルで記述してある。例えば，「緑色」のような範疇的概念や「二千隻」のような数量的概念は“色”や“数量”の値と見なすことができるが，「抑えた」「燃えにくい」のような性質的概念は“甘み”や“素材”の値そのものとは解釈しにくい。また，実体と属性の関係についても，実体と属性というよりは全体と部分の関係に相当するものが現れる場合があり得る。例えば次の例では，「タンク」「ふた」「グラスファイバー製」の三項は，実体，属性，値の関係にあるというよりは，全体，部分，部分の属性（値）とでもいうべき関係にある。表 11 では E+A/V 型構文は外延叙述型になったが，表 12 の E+A/V 構文は外延叙述型ではなく，属性叙述型である。

こうした多様性について，さらに詳しい分析と体系化が必要であるが，一般的にいえば，範疇叙述型の文が主語（対象）と述語（範疇）の二項関係であるのに対して，属性叙述型の文は EAV モデルで構造化されるような三項関係で記述することができる。E 項（実体）と V 項（数量や性質など）は，基本的

表 12　全体 - 部分関係と EAV モデル

EAV	例
E/V+A	［$_{E}$ タンク］は［$_{V}$ グラスファイバー製］の［$_{A}$ ふた］だ。
E+A/V	［$_{E}$ タンク］の［$_{A}$ ふた］は［$_{V}$ グラスファイバー製］だ。
E/A/V	［$_{E}$ タンク］は［$_{A}$ ふた］が［$_{V}$ グラスファイバー製］だ。

には独立した意味論的要素であり，A 項が両者を結び付ける働きをする。A 項が明示的に示されない場合には，E 項と V 項の関係はしばしば不明瞭になり，属性叙述型の構文はウナギ文に接近する。例えば（37）は，「国内出荷台数」が明示的に示されていなければ，「51 万 7000 台」が出荷台数を表すのか廃棄台数を表すのかは解釈次第であり，ウナギ文的になる [3]。

4.4.　外延叙述型の文

名詞述語文のある種のものは，主語に内包的概念を表す語を取って，述語でその外延を叙述する。これは従来の研究では倒置指定文や役割 - 値文といった名称で知られる文類型に相当する。以下に例を示す。役割を表す名詞には，(42) -(44) の「主力ロケット」「凶器」「商品」のように字義通りの意味で用いられるものもあれば，(45) -(48) の「秘密兵器」「落とし穴」「旗」「引き金」のようにやや比喩的な意味で用いられるものもあるが，ここでは問題としない。

(42) 宇宙研の【主力ロケット】が今回の【M3S2】で，宇宙開発事業団は「きく 6 号」を打ち上げた H2 ロケットだ。
(43) ともに使われた【凶器】は，【けん銃】だった。
(44) 【商品】は【一年物スーパー定期】で，10 万円につき一回の抽選権がある。
(45) コークの造るアメリカキューブの【秘密兵器】は【何】なのか。
(46) 第二の【落とし穴】は，私たちの神経はエネルギーをブドウ糖に頼っており，B1 の欠乏が神経に重大な障害をもたらす【こと】です。
(47) 掲げる【旗】は【民主主義】だ。
(48) 新方式の【引き金】は「学校名や序列を安上がりの選抜の手段として利用し，有力大学出身者以外を寄せ付けない大手企業の採用システムは，受験競争の過熱に重大な一役を買ってきた」との【中教審中間報告】だったという。

前節でも述べたことだが，役割 - 値文のある種のものは EAV モデルを用いた三項関係で記述することができる。例えば（42）は，「宇宙研」「主力ロケット」「M3S2」の三項で構成される E+A/V 型の構文と見なすことができる。この文は「宇宙研は M3S2 の主力ロケットだ」のような E/V+A 型の構文に言い換えることはできないが，「宇宙研は M3S2 だ」のような E/V 型のウナギ文にすることはできる。実際，（42）の後半の文には A 項が含まれておらず，「宇宙開発事業団」と「H2 ロケット」の二項によるウナギ文になっている。

外延を述べる文の中には，特定少数の要素を値として持つような役割概念ではなく，不特定多数の要素を外延として持ち得るような範疇的概念が主語の位置に現れる事例もある。次の例では，「陽性の植物」や「二階建ての洋風の建物」は範疇的概念を表しており，述語にはその外延が例示，列挙されている。（50）の場合はたまたま外延が二つしかないが，「二階建ての洋風の建物」という名詞句自体は役割というより範疇的概念を表すと考える方が自然であろう。

（49）陽性の【食物】は地面の下で育つ野菜，海草，大豆，【小魚など】で，陰性の食物は夏に取れる野菜や果物，ビールなど。

（50）そう，二階建てで洋風の【建物】は広小路の名古屋郵便局と【ここだけ】でした。

従って外延叙述型の文は，必ずしも主語が役割を表すとは限らないのであるが，どのような場合にこのような構文が可能になるかについては現段階では十分に明らかではない。上記の例では，外延として複数の要素が列挙されていること，例示であることを示す「など」や要素の限定を示す「だけ」のような要素が現れていることなどを特徴として挙げることができるが，一般的な規則の発見のためには，より多くの事例を収集し，分析する必要がある。

4.5.　その他の文

名詞述語文の主語と述語の関係は，帰属関係，対象と属性の関係，内包と外延の関係などの一般的な意味論的関係に即して解釈されるが，そのような関係が認められない場合には，両者の関係は様々な語用論的原則に従って解釈される。そのような事例の代表的なものは，ウナギ文という名称でよく知られている。どのような文をウナギ文と認定するかは難しい問題であるが，主語と述語の関係を一般的な意味論的関係に即して類型化することができない名詞述語文

の事例としては,次のような文を挙げることができる。(51) では「中継」は「桑田」の出演方法を,(52) では「音楽」は「CD」の収録内容を,(53) では「ロスタイム」は「時計」の進み具合を表している。

(51) アルバム大賞になった【桑田】も,公演先の横浜からの【中継】だった。
(52)【CD】は【音楽ばかり】ではありません。
(53)【時計】は既に【ロスタイム】だった。

ウナギ文に関する現在の研究課題としては,少なくとも二つのものが考えられる。一つは,ウナギ文の意味はいかにして解釈されるかという解釈論上の研究である。かつては,ウナギ文を省略文の一種と見なす変形文法的解釈が盛んであったが,近年ではメトニミーなどの考え方を用いた認知言語学的なアプローチも多い。もう一つは,どのような種類の文をウナギ文と認めるか,ウナギ文にどのような種類のものがあるか,という記述的研究である。既に述べたように,属性叙述型の名詞述語文の中にはウナギ文に接近すると見られる事例があり,両者の境界は必ずしも明確なものではない。また,ウナギ文と見られる事例を何らかの規則性に則って,体系化,下位類型化することができるか,という点も明らかではない。特に後者の記述的研究においては,より多くのウナギ文と見られる事例を収集し,分析することが必要であり,コーパス基盤の実例主義的研究が非常に重要性を帯びることになる。

今回の調査で収集した名詞述語文の事例の中には,形態・統語論的な理由から,主語と述語の関係が一般的な意味論的関係から逸脱したものも見られた。次の例では,名詞述語文は連体修飾構造の形で用いられているが,この文の主語を被連体修飾語の全体と見なすと,主語と述語の関係は帰属関係や内包・外延関係のような一般的な意味論的関係からは逸脱する。

(54) 同国の【主要産業】である【農業耕地】の三分の一が耕作不能になった,といわれる。
(55) 供述によると,遺体は自分の車で群馬県利根郡片品村花咲まで運び,関根容疑者経営の「アフリカケンネル」の【役員】だった【知人宅】の敷地で遺体を損壊,細かく切断したうえでドラム缶で燃やしたという。

これらの文においては，述語名詞の意味上の主語は被連体修飾語の全体ではない。(54) の「主要産業である」の主語は「農業耕地」ではなく「農業」であり，(55) の「役員だった」の主語は「知人宅」ではなく「知人」である。こうした事例は，いわゆるウナギ文とは区別して扱うべき事例であるが，形態・統語論的には興味深い事例であり，またコーパスに格関係情報や意味情報を付与する際にも一定の注意が必要であろう。

5．まとめ

京都大学テキストコーパスを用いて，名詞述語文の主語と述語の抽出，意味分類情報の付与，個別の事例の分析を行った。構文情報に基づく主語名詞句の推定では，主語名詞句の抽出率は全体の 57% 程度に留まり，また正解率はそのうちの 83% 程度である。特に抽出率については，本稿の方法で抽出可能な主語は，述語と直接係り受け関係にある要素にのみ限られ，より多くの主語を自動抽出しようとすれば照応解析などより高度な解析手法が必要である。意味分類情報については, 分類語彙表データベースを用いた情報の自動付与により，94 ～ 95% 程度の事例に意味分類情報を付与することができた。これらの一連の機械処理については，精度の向上等の課題が残されてはいるものの，最終的に人手による分析を行うことを前提とした調査においては，データの収集と整理の工程を効率化する上で，かなりの効果を期待することができる。

個別事例の分析については，実際の言語資料に現れる多様な事例をいかにして体系的に整理し，記述するかが問題となる。本稿では，特に文の意味構造に注目して範疇叙述型，属性叙述型，外延叙述型，その他という類型を設定し，クオリア構造や EAV データモデルを用いた意味構造の記述方法を検討した。しかし，このような類型化の妥当性については，まだ十分に検証できているとは言い難く，より多くの事例の分析を進める必要がある。また，こうした意味論的類型と主語や述語の意味分類との間にどのような相関があるのかについても，現時点では十分に明らかになっていない。予想としては，範疇叙述型は主語と述語の意味分類が一致する傾向があるが，属性叙述型は主語の意味分類によらず述語が「関係」に属するような抽象概念（色，形，量など）になる傾向があるなど，何らかの偏りが観察される可能性が考えられる。このような相関を分析するためにも，調査対象を広げてより包括的な分析と集計を進めることが不可欠である。また，今回の分析では，述語の主要部だけでなく修飾語句の

部分も，名詞述語文の意味構造を記述する上で重要な要素であることが再確認された。今後は主語と述語の抽出だけでなく，EAV モデルのような三項構造を念頭においた機械解析の方法についても検討する必要がある。

注

1 ここで言う範疇叙述型は，高橋（1984）の言う「種類づけ」におおよそ相当する。高橋（1984）は「種類づけ」を「類づけ」「種づけ」「別種類づけ」に分類しているが，(9, 10) は「種づけ」に，(11-13) は「類づけ」に相当するものと考えられる。また（14-18）は「別種類づけ」に相当する可能性があるが，高橋（1984）の「別種類づけ」は比喩的表現などを中心として扱っているので，ここでは異同を問題としない。

2 ここでは，単に多様な属性の束を記述するためにクオリア構造の記法を借用しているだけであり，句の意味の構成に関して必ずしも Pustejovsky（1995）の理論に忠実な方法を取っているわけではない。技術的詳細の検討と理論的整合性の検証については，今後の課題としたい。

3 数量名詞述語文とウナギ文の類似性については，西山（2003）の研究がある。

参考文献

荒木健治・桃内佳雄（1989）「名詞述語文における名詞句間の意味関係の解析と学習」『全国大会講演論文集』39(1)，pp.640-641．情報処理学会.

今田水穂（2009）「日本語名詞述語文の意味論的・機能論的分析」筑波大学博士論文.

影山太郎（2002）「語彙と文法」北原保雄［監修］斎藤倫明［編］『朝倉日本語講座 4 語彙・意味』朝倉書店，pp.170-190.

国広哲也（2002）「語義の構造」北原保雄［監修］斎藤倫明［編］『朝倉日本語講座 4 語彙・意味』朝倉書店，pp.153-169.

新屋映子（2009）「形容詞述語と名詞述語：その近くて遠い関係」『国文学解釈と鑑賞』74(7)，pp.30-40.

高橋太郎（1984）「名詞述語文における主語と述語の意味的な関係」『日本語学』3(12)，pp.18-39.

角田太作（1996）「体言締め文」鈴木 泰・角田太作［編］『日本語文法の諸問題：高橋太郎先生古希記念論文集』ひつじ書房，pp.139-161.

西山佑司（2003）『日本語名詞句の意味論と語用論』ひつじ書房.

益岡隆志・田窪行則（1989）『基礎日本語文法』くろしお出版.

Fauconnier, Jilles (1985) *Mental Spaces*. MIT Press.

Jackendoff, Ray (2002) *Foundations of Language*. Oxford University Press.

Pustejovsky, James (1995) *The Generative Lexicon*. MIT Press.

謝 辞

本研究は科学研究費補助金（課題番号 23720225「Ruby と MSXML による日本語名詞述語文の実例調査とコーパス分析ツールの構築」）の助成を受けたものである。