

国立国語研究所学術情報リポジトリ

On Japanese Predicate Nouns With Appositive Clauses

メタデータ	言語: jpn 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): 作成者: 今田, 水穂 メールアドレス: 所属:
URL	https://doi.org/10.15084/00001508

外の関係の連体修飾節を伴う名詞述語について

今田 水穂 (文部科学省)

On Japanese Predicate Nouns With Appositive Clauses

Mizuho Imada (Ministry of Education, Culture, Sports, Science and Technology)

要旨

『現代日本語書き言葉均衡コーパス』述語項構造データを使って、人魚構文の抽出と、人魚構文になりやすい名詞の評価を行った。まず人魚構文を成立させる構造的要件を検討し、人魚構文を判別するための条件を設定した。次にコーパス中の名詞を語彙素ごとに集計し、対数尤度比統計量によって外の関係の連体修飾節の取りやすさと人魚構文へのなりやすさを評価した。また、nwjc2vec のベクトル表現を用いてそれらの名詞をクラスタリングし、類似語ごとに配列して可視化した。最後に個々の事例について調べ、抽出に失敗する事例の構造的特徴や、人魚構文になりやすい名詞の修飾節の特徴を検討した。

1. はじめに

本発表は、コーパスを用いた名詞述語文研究の下位タスクとして、人魚構文⁽¹⁾の分析を行う。人魚構文とは、次のような種類の文をいう。

(1) 人魚姫は泡になる運命だ。

この構文は一見すると「人魚姫」を主語、「運命」を述語とする名詞述語文のように見えるが、「運命だ」の部分が文法化して機能表現になり、「泡になる」の部分が主節化した動詞述語文である可能性が指摘されている。



図1 人魚構文の構造

この文法化の問題が、人魚構文が研究対象とされる主要な動機付けの1つと考えられるが、本発表は文法化を主要な問題としない。本発表は、コーパスから人魚構文を自動抽出し、どのような名詞が人魚構文に生起するかを分析することを研究の目的とする。この目的のために、本発表では人魚構文を図1左のような名詞述語文として扱うが、これは単に分析の手続き上、その方が都合がよいからである。

⁽¹⁾ 「人魚構文」という名称は角田(2011)によるものであり、文の前半が「人魚姫は泡になる」のような動詞述語文、後半が「運命だ」のような名詞述語文であることに由来する。「文末名詞文」と呼ばれることもある(新屋1989)。

本発表で使用するコーパスは『現代日本語書き言葉均衡コーパス』(BCCWJ)(Maekawa et al. 2014)である。BCCWJには多様な言語学的情報が付与されているが、本発表では主に形態論情報、係り受け情報、述語項構造情報を使用する。人魚構文は述語名詞が連体修飾節を持つが、その多くは外の関係の連体修飾節である。(1)の例で言うと、連体修飾節の述語「なる」は被修飾語「運命」との間に項関係を持たず、主節の主語「人魚姫」との間に項関係を持つ。こうした人魚構文に特有の構造を指定するために、述語項構造情報が役立つ。

本発表では、まず人魚構文を成立させる構造的要件について検討する。次に、その要件を満たす構造をコーパスから抽出する方法を使用するデータの特性に照らして検討し、分析対象を外の関係の連体修飾節を伴う名詞に限定することを説明する。次に、コーパス中の全ての名詞を語彙素ごとに集計し、対数尤度比統計量を用いて、外の関係の連体修飾節を取りやすい名詞と、人魚構文になりやすい名詞を評価する。また、それらの名詞をクラスタリングによって類似語ごとにグループ化し、分析しやすいように配列して可視化する。最後に個々の事例の分析に進み、人魚構文の判別に失敗する原因となる構造を検討し、また外の下位分類と人魚構文の関係を検討する。

2. 人魚構文の構造的要件

人魚構文を名詞述語文と見立てた場合、この構文は次の特徴を持つ。

- (2) a. 主語と述語が同一関係でも包摂関係でもない。
- b. 述語が連体修飾節を伴う。
- c. 文末の「名詞 + コピュラ」を削除しても文として成立する。

特に述語が連体修飾節を伴うこと(2b)、その連体修飾節を主節に格上げし得ること(2c)が人魚構文を成立させる基本要件であり、この2つの条件を満たして初めて文法化、主節化という現象が成立し得ると考えられる。また、この2つの条件を満たすものであっても、(2a)の条件を満たさないものは通常、人魚構文と見なされない。例えば「猫はやさしい動物だ」は「猫はやさしい」と言い換えられるが、これを人魚構文に含めて考えることは普通はないようである。文法化という観点から見ても「猫は気性のやさしい動物だ」のようにガ・ノ交替が成り立つことから、連体修飾節の主節化はしていないように見える。

述語名詞は連体修飾節を持つが、この連体修飾節は多くの場合、外の関係の連体修飾節である。すなわち、連体修飾節の述語は、被修飾語との間に項関係を持たない。一方で、連体修飾節の述語は、主節の主語と項関係を持つ。(3)の例で言うと、連体修飾節の「助ける」は被修飾語「予定」とは項関係を持たないが、主節主語の「人魚姫」(3a)または「王子」(3b)と項関係を持つ。ただし、(3c)のような主節が主語を持たない構造の場合はその限りではない。

- (3) a. [s 人魚姫₁ は [s' ϕ_1 王子を助ける] 予定だ]
- b. [s 王子₁ は [s' 人魚姫が ϕ_1 助ける] 予定だ]
- c. [s [s' 人魚姫が王子を助ける] 予定だ]

この構造は(2c)の条件を満たすための下位条件と考えることができる。「予定」が「助ける」

の項だとしたら、「予定」を削除したときに「助ける」の項が足りなくなってしまう。「人魚姫」が「助ける」の項でないとしたら、「予定だ」を削除したときに「人魚姫は」と対応する述語が無くなってしまう。特に前者の条件を考える限り、連体修飾節が外の関係でなければ人魚構文は成立しなさそうに見える。

ところが、従来人魚構文として報告されてきた事例の中には、連体修飾節が内の関係であるものも存在する。(4)は角田(2011)が挙げている例である。

- (4) a. 明夫は明るい表情だ
b. あの力士は立派な体格だ

これらの文では「表情が明るい」「体格が立派だ」のように連体修飾節述語と被修飾語の間に項関係が成立している。一方で、被修飾語を削除して「明夫は明るい」「あの力士は立派だ」と言い換えても文としては成立する。このような文が成立するのは「明夫は表情が明るい」のようなハ・ガ構文の第1項が主節主語に、第2項が主節述語(被修飾語)になる場合であり、かつ第2項を削除してもほぼ同等の意味で解釈できる場合に限る。この構文も(2)の条件を満たしている点で人魚構文の可能性を持つものであるが、外の関係の人魚構文とは構造が異なる。まとめると、(2c)は以下の下位条件に分割することができる。

- (5) $(2c) \leftarrow (a \wedge (b \vee c)) \vee (d \wedge e)$
- a. 主節述語が連体修飾節述語の項ではない。(外の関係)
b. 主節主語が連体修飾節述語の項である。
c. 主節主語がない。
d. 主節述語が連体修飾節述語のガ項である。(内の関係)
e. 主節主語が連体修飾節述語のハ項である。

3. コーパスにおける人魚構文

コーパスから、人魚構文を自動抽出することを考える。まず(2a)の条件について、BCCWJは意味情報を含む各種の言語学的情報が付与されているが、(2a)の情報を直接取得することは難しそうである。(2b)は形態論情報と係り受け情報から容易に取得することができる。(2c)は前節で述べたように(5)の下位条件に分割して考えることができるが、この内(5a, b, d)については述語項構造情報から取得することができる。一方、(5c, e)については、述語項構造情報を用いても悉皆的に取得することは難しそうである。そこで、ここで(5a)∧(5b)の場合に限って抽出を試みることにする。この条件を満たすとき、文は次のような構造を持つ。

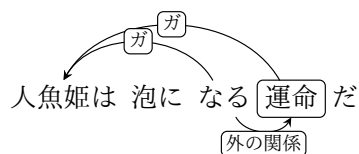


図2 述語項構造と人魚構文

実際の検索処理では、条件を満たすような名詞を検索する処理を行う。具体的には、次の条件を満たす語 N を検索する。

- (6) a. N は名詞または代名詞である。
- b. N は連体修飾節 C を持つ。(=2b)
- c. N は C の述語 V の項ではない。(=5a)
- d. N は述語である。
- e. N と V が共有する項がある。(=5b)

(6) の各条件は $a \leftarrow b \leftarrow c$ 、および $(b \wedge d) \leftarrow e$ の依存関係を持ち、上位の条件が偽の場合は下位の条件も偽となる。最終的に、 $c \wedge e$ を満たすものが検索対象の語である。

これらの条件に基づき、コーパスから条件を満たす語の検索を行った。使用したコーパスは BCCWJ コアデータで、サブコーパス (レジスタ) は特に区別せず全て検索対象とした。語の単位は長単位とし、総語数は 0000 長単位だった。形態論情報は BCCWJ-DVD-1.1 版収録の長単位タブ区切りデータを、係り受け情報と述語項構造情報は BCCWJ-DepParaPAS-3.3.0 を使用した。検索処理は独自に作成した Ruby スクリプトで行い、集計処理は R で行った。(6) の各条件を満たす語の頻度を以下に示す。

表 1 コーパスにおける人魚構文

品詞	(6a) 名詞類	(6b) 連体修飾	(6c) 外の関係	(6d) 名詞述語	(6e) ゼロ照応	(6c, e) 人魚構文
代名詞	9948	161	44	327	0	0
名詞 固有名詞	25414	1745	494	583	9	4
名詞 助動詞語幹	151	5	2	0	0	0
名詞 数詞	22344	617	282	1229	17	11
名詞 普通名詞	223260	36841	20297	12382	645	602

4. 対数尤度比統計量による評価

ある名詞が人魚構文になりやすいか否かを評価することを考える。といっても、人魚構文になりにくい名詞は非常に多く存在すると考えられるので、まず外の関係の連体修飾節を取りやすい名詞に限定した上で、その中で人魚構文になりやすい名詞とになりにくい名詞を調べることにする。外の関係と人魚構文は、それぞれ (6c) と (6c, e) の条件で判定し、各条件の満たしやすさは対数尤度比統計量によって評価する。

$$G = 2 \sum_i O_i \log O_i / E_i$$

O_i は頻度の実測値、 E_i は期待値である。これは語 N と条件 C の分割表を用いて次のように計算することができる。 $O_1 O_4 - O_2 O_3 < 0$ の場合は、 G に -1 を乗じる補正を加えた。 O_5 は

コーパス中の全ての名詞・代名詞の頻度の和である。

	C	¬C	計
N	O_1	O_2	O_6
¬N	O_3	O_4	O_7
計	O_8	O_9	O_5

$$G = 2 \left(\sum_{i=1}^5 O_i \log O_i - \sum_{i=6}^9 O_i \log O_i \right)$$

(6c) の G 値が 50 以上だった 59 種の名詞について、(6c) と (6c, e) の G 値を表 2 に示す。先行研究 (新屋 1989, 角田 2015) で挙げられている名詞に印を付した。

表 2: 人魚構文へのなりやすさ

語彙素	頻度	(6c) 外の関係		(6c, e) 人魚構文		新屋 (1989)	角田 (2015)
		頻度	G 値	頻度	G 値		
物	1887	816	1834.17	71	279.78		○
予定	116	80	276.36	31	246.80	○	○
方針	97	60	187.59	18	128.18	○	○
感じ	173	94	260.81	11	53.67	○	○
タイプ	44	21	51.42	7	47.39	○	○
見込み	24	17	60.14	5	36.78		○
積り	105	82	318.01	7	34.77	○	○
見通し	43	25	73.80	5	30.51		○
奴	88	31	55.23	6	30.07		
事	5214	3663	13446.35	33	27.65	○	○
考え	113	45	91.73	6	27.08	○	○
状態	101	54	147.49	5	21.87	○	○
場所	166	47	64.14	4	12.00		
一方	175	80	187.93	4	11.61		○
形	201	69	119.43	4	10.60	○	○
疑い	94	41	91.83	3	10.57		○
様子	69	29	62.52	2	6.68	○	○
通り	146	44	65.07	2	3.99		
時代	177	59	98.67	2	3.35		
所	653	293	677.85	4	3.10	○	○
状況	192	75	149.85	2	3.09		○
方向	66	26	52.37	1	2.17		○
恐れ	68	45	149.63	1	2.11		
仕組み	74	41	115.76	1	1.97	○	○
訳	403	334	1375.72	2	1.04	○	○

表 2: 人魚構文へのなりやすさ

語彙素	頻度	(6c) 外の関係		(6c, e) 人魚構文		新屋 (1989)	角田 (2015)
		頻度	G 値	頻度	G 値		
思い	146	42	58.52	1	0.92	○	○
点	150	56	106.50	1	0.88		
方法	161	71	160.87	1	0.79		
姿	176	48	62.30	1	0.68		
筈	235	220	1031.86	1	0.36		○
日	334	72	65.60	1	0.09		
必要	343	185	510.39	1	0.07		
途端	26	20	76.41	0	-0.11		
事等	29	28	136.45	0	-0.13		
瞬間	34	23	78.01	0	-0.15		
会	48	22	51.77	0	-0.21		
場面	51	27	73.04	0	-0.22		
傾向	59	35	105.27	0	-0.26		○
度	60	40	133.89	0	-0.26		
以上	62	25	51.62	0	-0.27		
分	63	26	55.00	0	-0.28		
ケース	79	45	130.38	0	-0.35		
時期	104	38	70.55	0	-0.46		
限り	105	95	427.74	0	-0.46		
機会	114	52	121.84	0	-0.50		
為	1472	696	1706.05	2	-0.55		
際	133	60	139.06	0	-0.58		
事件	138	43	66.29	0	-0.61		
頃	221	83	158.95	0	-0.97	○	
可能性	224	163	592.18	0	-0.98		
理由	225	91	188.69	0	-0.99		
結果	229	56	62.26	0	-1.01		
俣	337	143	311.92	0	-1.48		
内	427	103	112.30	0	-1.88		
気	450	133	192.24	0	-1.98	○	○
場合	653	418	1354.52	0	-2.87		
他	664	124	87.21	0	-2.92		
後	707	147	126.15	0	-3.11	○	○
時	1171	718	2245.72	0	-5.16		

6. 抽出の失敗

実際には人魚構文ではないが人魚構文と判定されたものを第1種過誤(偽陽性)、実際には人魚構文だが人魚構文と判定されなかったものを第2種過誤(真陰性)とする。第1種過誤は、内の関係と外の関係の判別に失敗したものが多い。連体修飾節の述語と被修飾語の間に項関係がありながら、実際には被修飾語ではなく先行文脈中の語が項としてアノテーションされている場合などにこの過誤が起きる。図4の左が期待する構造、右が過誤の原因となった実際のアノテーションである。「もの」「こと」「ところ」などの語は、形式名詞ではない実質名詞としての用法の場合であっても、連体修飾節述語の項と認識されず、この過誤に該当するものが比較的多いようである。また、表2で人魚構文の条件(6c, e)を満たす頻度が1以上だった語のうち、先行研究で言及のないものが10語あったが、そのうち「奴」「場所」「時代」「点」「方法」「姿」「日」の7語は全てこの過誤に該当するものだった。

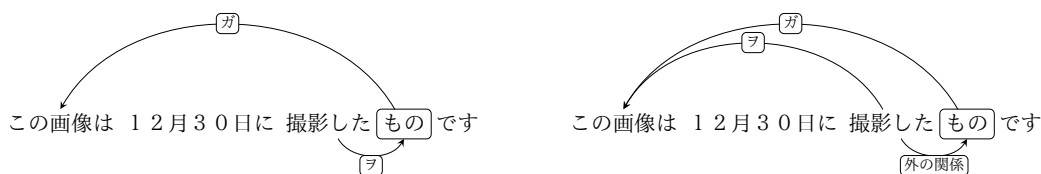


図4 第1種過誤 [00019_A_OY04_00112]

第2種過誤は、名詞に述語タグが付与されていなかったため判別に失敗したものが多い。「もの」「こと」「ところ」「つもり」「まま」「わけ」「はず」「とおり」などの形式名詞は、述語としてアノテーションされておらず、この過誤に該当するものが多いようである。なお、先行研究で人魚構文として言及されているにもかかわらず、表2で条件(6c, e)を満たす頻度が0だった語が4語あったが、これらはこの過誤に該当するわけではなく、単に人魚構文の構造に該当する例がなかった。



図5 第2種過誤 [00009_A_PM12_00011]

いずれの過誤においても、形式名詞や意味的な抽象度の高い実質名詞について、述語タグや項タグが付与されない傾向があることが、過誤の一因となっているものと考えられる。ただし、そもそも述語項構造データは必ずしも統語構造を配慮してアノテーションする方針で作られているわけではないので、抽象語より実質語に優先的にタグが付与されていたとしても、一概にアノテーションの誤りとみなすことはできない。本研究の目的の上では期待する検索処理ができなかったため、今後、検索条件の見直しを行うか、または追加の情報をアノテーションするなどの方策を検討する必要がある。

7. 外の関係の分類と人魚構文

外の関係の連体修飾節には内容補充修飾節、付随名詞修飾節、相対名詞修飾節があるとされる。

- (7) 田中君が結婚するという噂 (内容補充修飾節)
- (8) 関東大震災が起こった翌年 (相対名詞修飾節)
- (9) ジュースを買ったおつり (付随名詞修飾節) (日本語記述文法研究会 2008:51)

人魚構文になりやすい名詞は、内容補充修飾節を伴うものが多い。内容補充修飾節は、一定の内容を持つ名詞に対して、その内容を補充する。被修飾名詞には、様々な様相の事態や命題的情報を表すもの(「状態」「予定」「考え」など)、事態や命題的情報を表象する言語表現や知覚様相(「話」「光景」など)、命題的情報によって特徴付けられる属性を表すもの(「タイプ」「仕組み」など)などがある。

- (10) 甲板は傾いていて、かなり意識してバランスをとらないと立ってられない状態であった。 [00023_A_PB39_00009]
- (11) 一行は十九日に軍のヘリで下山する予定だったが、 [00163_C_PN1e_00011]
- (12) 農水省は、生産者団体を支援していく考えだ。 [00069_B_PN4c_00005]
- (13) 夫の両親は亭主関白の舅を姑が常に顔色を伺っていて、自分の本心を口にしないタイプ。 [00053_A_OC10_00005]

「結果」は内容補充修飾節を伴う場合と付随名詞修飾節を伴う場合があるが、何れにせよ今回の調査で人魚構文に該当するものは検出されなかった。しかし「だ」「です」などの助動詞だけでなく「となる」などの変化表現もコンピュータに含める考え方を取ると次のような例があり、この場合の修飾節は内容補充修飾節である。

- (14) COP6は中断し、平成13年(2001年)7月に開催されるCOP6再開会合に議論を持ち越す結果となった。 [00011_A_OW6X_00014]
- (15) 白54、黒55の交換は却って白56・58の二間を苦しめる結果となった。 [00034_B_PM41_00060]
- (16) 西安事件後、張学良氏は蒋介石に同行して南京へ赴き、中国大陆と台湾で半世紀にわたる拘束と軟禁を受ける結果となった。 [00186_D_PN2d_00012]

人魚構文になりにくい名詞は、相対名詞修飾節や付随名詞修飾節を伴うものが多い。相対名詞修飾節は時間・空間などを表す相対名詞を修飾する節であり、「後」「内」などがこれに当たる。「頃」「機会」「際」「度」「途端」などの時間表現も、修飾節が表す事態を参照点として初めて時点や時間範囲を特定できるので、相対名詞に準ずるものと考えられる。「時」「時期」「時代」「瞬間」などは相対名詞とは言いにくく、他の分類を割り当てる必要があるかも知れないが、いずれにせよ今回の調査では、時間を表す名詞が人魚構文になる例はほとんど見られな

かった(ただし新屋(1989)や角田(2011)では「頃」「後」などの例が挙げられているので、人魚構文にならないというわけではないようである)。時間・空間表現以外では、「他」「以上」「限り」なども相対名詞の一種と考えられる。付随名詞修飾節は事態の結果、原因、属性、評価のような隣接物を被修飾語とするもので、「結果」「理由」「可能性」「必要」などが該当する。今回の調査では、これらを述語とする人魚構文の例はほとんど見られなかった。

8. まとめ

人魚構文を成立される構造的要件を検討し、特に述語名詞が外の関係の連体修飾節を伴う構造のものについて、BCCWJ 述語項構造タグ付きコーパスからの自動抽出を試みた。また、抽出結果を語彙素ごとに集計して対数尤度費統計量を計算し、外の関係の連体修飾節の取りやすさと、人魚構文へのなりやすさを評価し、クラスタリングによって類似語をグループ化した上で可視化した。抽出誤りの原因となる構造を分析し、形式名詞などに述語タグ、項タグが付与されていない場合に過誤が生じることを確認した。人魚構文へのなりやすさを外の関係の下位分類と照らして検討し、人魚構文になりやすい名詞は内容補充修飾節を伴うものが多く、なりにくい名詞は相対名詞修飾節や付随名詞修飾節を伴うものが多いことを確認した。

本研究は名詞述語文の分析を通じて名詞の項構造記述の形式化を推し進めるすることを目的としており、本発表はその部分問題として人魚構文を扱った。人魚構文の述語名詞はオントロジー的には命題 (Proposition)、情報媒体 (ContentBearingPhysical)、属性 (Attribute)、時間 (TimePosition) などがあり、また主語と述語の関係は is-a 関係ではなく、has-a 関係のものが多いなどの特徴がある。今後、コーパスから得られた事例の分析や意味情報のアノテーションなどを通じて、名詞の意味記述の形式化を進めていきたい。

謝 辞

本研究は JSPS 科研費 17H00009 の助成を受けたものです。

文 献

- 角田太作 (2011). 「人魚構文: 日本語学から一般言語学への貢献」 国立国語研究所論集, pp. 53–75.
- 新屋映子 (1989). 「文末名詞」について」 国語学, 159, pp. 1–14.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014). “Balanced Corpus of Contemporary Written Japanese.” *Language Resources and Evaluation*, 48, pp. 345–371.
- 浅原正幸・岡照晃 (2017). 「nwjc2vec: 『国語研日本語ウェブコーパス』に基づく単語の分散表現データ」 言語処理学会第 23 回年次大会発表論文集.
- 日本語記述文法研究会 (編) (2008). 『現代日本語文法 6: 複文』 くろしお出版, 東京.