

上位語・下位語を含む連体修飾表現の言語的分析

安藤まや⁺ 関根聡[‡]
⁺慶應義塾大学 [‡]ニューヨーク大学

0. はじめに

我々は[1]で、名詞の下位語を自動的に抽出することを目的として、「トマトなどの野菜」のように上位語と下位語を含む連体修飾表現を大量の新聞記事コーパスを対象に収集した。テキストから自動的に上位下位概念の抽出が出来れば、シソーラスの自動構築の他、自然言語処理の応用においても役に立つと考えられる[7]。前述の報告では、「N1などのN2」のほかにも「N1などN2」、「N1というN2」、「N1のようなN2」、「N1に似たN2」、「N1以外のN2」、「N1と呼ばれるN2」の合計7つの表現が信頼度高く、かつ量も豊富に上位下位概念の抽出が出来ることがわかった。文脈に関係なく「N1(下位語)はN2(上位語)である」という文が成立する場合を正解とし、手作業で評価を行なった結果、「動物」「乗り物」など下位に一般名詞を多く含むものについて、ほぼ6割以上の正解率で下位概念が抽出された。その一方で、「人間」など一部の名詞については、定型表現ごとに正解率に差が見られた。本稿では、「などの」「という」「のような」および、類似した表現である「といった」の4つの連体修飾表現について、N1、N2に現れる名詞に着目し、それぞれの用法の違いを32年分の新聞記事を対象に分析、考察する。特にN2については「動物」「魚」「犬」など「下位語に普通名詞を多く含む名詞」の場合(以降「グループ1」と呼ぶ)と「人間」「女性」「学者」といった「下位語に固有名詞を多く含む名詞」(以降「グループ2」と呼ぶ)の場合を中心に、N1に現れる名詞が、用法によってどのような違いがあるのかを調査し、比較する。分析の結果「といった」などの「グループ」と、他の2つの表現はそれぞれ違った特徴が見られたため、それぞれについて個別に考察した後、全体を比較してまとめとする。

1. 「N1といったN2」と「N1などのN2」

「N1といったN2」は「運が良ければカモシカやウサギ、リスといった動物も見られるという」のように、N2の下位語にあたる名詞を提示することによって例示を表すことが多く、連体修飾表現において使用される[2]。例示は集合に属するものを個別に列挙するものであるため、N1に現れる名詞は、下位語になる。

「など」は「あの商店街には八百屋やお茶屋などがある」のように、多くのものの中から何か例をあげるときに使われる副助詞であり、連体修飾表現でなくとも使用することができるが、「N1などのN2」のように「の」を伴うことで、連体修飾表現のかたちでも使用することが可能である。

表1に32年分の新聞記事中に「N1といったN2」「N1などのN2」を含む文が何文あったのか(表中の頻度)、そのうちN1がN2の下位概念の単語となっていた文が何文あったのか(上下関係)、そのうち並列表現(本稿では、「,」「と」「や」によってN1の部分で名詞句が並列に表現されているものを並列表現と呼ぶ)を伴った文の割合(並列表現)、N1に固有名詞が出現した割合(固有名詞)を示す。並列表現と固有名詞の調査は、ランダムに取り出したサンプル30文について行なった。これらの頻度について、N2が「グループ1」の場合、「グループ2」の場合に分け、その特徴を述べる。また、最後に「などの」と「といった」の違いについて述べる。

<N2が「グループ1」の場合>

- ・ 「シカ、イノシシなどの動物の骨」のようにN1(シカ、イノシシ)がN2(動物)の例示となっている場合が非常に多く並列表現を伴う割合が高い
- ・ N1は「アブラハヤなどの魚」のように馴染みの薄い単語はあまり見られず、多くは「ネコ」「ウサギ」など非常に身近な単語である

<N2が「グループ2」の場合>

- ・ 上記の「一般名詞を下位語に持つ名詞」に比べ例示の出現頻度が低い
- ・ 「OL」「主婦」などの普通名詞と「レーニン」「ギルダール」などの固有名詞が出現しているが、固有名詞はす

べて並列表現とともに現れている。

<「などの」と「といった」の違い>

元々「といった」が使われていた次の文において「といった」を「などの」に置き換えてみる。「トドやカモ、カモメ（といった・などの）動物に囲まれ、潮の香りが心地よい生活ではあったが、不便さもまたひとしおだった」では、「などの」に置き換えるとトド、カモ、カモメの他に幾種かの動物がいたような印象を受ける。「といった」の方が「などの」に比べ、「他の例ではなく、この例でなくてはならない」意味合いが強く、いくらか限定して例をあげる傾向があるため、使用場面が制限される。そのために、「といった」の頻度が「などの」に比べ少なくなっているものと考えられる。

	N2	動物	魚	犬	人間	女性	学者
などの	頻度	419	225	8	31	90	43
	上下関係	417	214	5	4	20	2
	並列表現	63%	60%	40%	100%	50%	50%
	固有名詞	-	-	-	50%	5%	50%
といった	頻度	47	26	1	71	35	4
	上下関係	42	24	1	1	15	3
	並列表現	66%	64%	100%	0%	64%	100%
	固有名詞	-	-	-	-	7%	100%

表1:「などの」「といった」の分析結果

2.「N1というN2」

「N1というN2」は『田中という人』のように『N1という名前をもつN2』の意や、『教師という職業』のように『N1が属する範疇を表す』用法である[2]。いずれにせよ、N1が何者であるのかをN2で表し、上記で説明した「N1などのN2」「N1といったN2」のように、N2の例をあげるものとは異なった表現であると言える。そのため、並列表現を伴う例が見られないのが特徴である。

また、以下で説明するように「という」は、一般名詞の場合には身近でない名詞が多く、固有名詞との共通点は、それが何者であるかの判断が難しいことであり、そのような場合にN1を説明するために「という」が利用される傾向があることがわかる。

<N2が「グループ1」の場合>

- ・「動物」「魚」は先の「N1などのN2」「N1といったN2」に比べ頻度が減っている。「などの」「といった」の方が一般名詞を例にあげることが多いため、多用される傾向があると推測される。
- ・N1に現れる名詞は「などの」「といった」とは逆に「パラという動物」のような耳慣れない単語が、多く含まれる。特に「魚」では、N1だけでは魚であると筆者が判断できない単語が40%ほど存在した。
- ・「犬」は「タローという犬」のような固有名詞が68%を占めている。「動物」としてではなく、名前の付いた「ペット」としての側面が際立っている。「猫」も同様の傾向がある。「犬」や「猫」は一般名詞と固有名詞の両方を下位語に持つ名詞と考えられるが、その特徴が「などの、といった」と「という」の表現によって綺麗に分割されていることがわかる。

<N2が「グループ2」の場合>

- ・「N1などのN2」「N1といったN2」に比べ、頻度が格段にあがっており、そのうち、N2とN1が上下関係にある表現の数も多い。
- ・N1に固有名詞が含まれている確率は50%を超え、並列表現は登場しない。
- ・「人間」については、固有名詞の他に「彼女」「私」といったような代名詞や「自分」といった名詞が43%含まれており、このような特定名詞を除くとかなりの確率で固有名詞を抽出することが可能である。

	N2	動物	魚	犬	人間	女性	学者
という	頻度	87	110	35	449	381	37
	上下関係	77	97	28	384	286	34
	並列表現	-	-	-	-	-	-
	固有名詞	-	-	68%	50%	63%	100%

表2.「という」の分析結果

3.「N1のようなN2」

連体修飾表現「N1のようなN2」を調査した結果、大きく分けて、例示、属性、比喩の三種の用法が見られた。例示は「マナガツオやサワラのような魚の切り身を漬ける」といったものである。例示であるため、N2がN1を包摂している上に、N1に現れる名詞は代表として示されており、他にもN1に現れることができる名詞が存在するような印象を受ける。次に、「彼のような(すごい)人には会ったことがない」のように、N1の属性を問題にする場合がある[2]。これは、限定用法で、「彼」の部分は例示と異なり「彼」以外では意味がなくなる。この場合、N2はN1を包摂しているが、N1に現れる他の名詞を想定することはできない。また、「あのおじさんは僕にとってお兄さんのような(存在の)人」といった比喩を表すものもある。この場合、N2の示すもの(おじさん)はN1(お兄さん)ではない。

表3に分析結果を示す。「人間」「女性」についての「例示」「属性」「比喩」の分類は、ランダムに取り出したサンプル30に対して行なっている。

			動物	魚	犬	人間	女性	学者
のような	頻度		63	35	12	232	105	12
		上下関係	56	15	2	162	77	7
	例示	頻度	7	5	-	1/30	4/30	2
		並列表現	-	80%	-	100%	25%	100%
		固有名詞	-	-	-	100%	50%	100%
	属性	頻度	5	1	-	18/30	17/30	2
		並列表現	-	-	-	-	6%	-
		固有名詞	-	-	-	-	20%	100%
	比喩	頻度	43	9	-	6/30	3/30	1
		並列表現	-	-	-	-	-	-
		固有名詞	-	-	-	50%	-	100%
	判断不可	頻度	1	-	2	5/30	4/30	2
		並列表現	-	-	-	-	-	-
		固有名詞	-	-	100%	-	50%	100%

表3.「のような」の分析結果

<N2が「グループ1」の場合>

- ・ 「のようなN2」の出現頻度に対して、N2とN1が上下関係にある場合の頻度が、他の表現に比べて低い
- ・ 「動物」のように下位にさまざまな種類のものを含む名詞は、N1に下位語が含まれることが多く、特に比喩表現にて下位語が示されることが多い。「生物」「道具」なども同様の傾向がある。
- ・ 「魚」のように、下位にある名詞と特徴が似ている名詞はN1に下位語が含まれることが少ない。「鳥」も同様の傾向がある。

<N2が「グループ2」の場合>

- ・ 固有名詞や個人を表す単語は例としてあげられにくく、属性にフォーカスが当たった表現で取り上げられることが多い。
- ・ N1には固有名詞以外に「私」などの代名詞が多く、このように個人を示したものが約80%存在した

- ・ カテゴリ「人間」以外でグループ2に含まれる名詞では、「会社」では固有名詞、代名詞が多く見られたが、「川」「島」などでは、が低くなった。これは、先にあげた「魚」と同じように、下位にある名詞との特徴が非常に似ているため頻度がさがったものと考えられる

4. 比較

「といった」「などの」「という」「のような」の違いを表4に示す。

用法	N1などのN2	N1 といったN2	N1 というN2	N1 のような N2		
	例示	例示（「などの」に比べ限定的）	範疇	例示	属性	比喻
並列表現						
N1=身近な名詞、N2=グループ1						
N1=固有名詞、N2=グループ2						

表4. 比較

- ・ 「などの」「といった」「のような」が例示を示す場合、並列表現が多く、それ以外ではほとんどみられない。これは例示のみが「あるものの中からいくつかを示す」という機能を持っているためだと考えられる。
- ・ 例示表現であっても、表現によって N1 が他のものに変えられるかどうかの程度が異なる。「などの」では N1 に他の名詞に置き換えても差し障りのない身近な単語が多く、「といった」では時に差し障りがある。「のような」では N1 に現れる属性が重要情報となる。「寺田寅彦や柳田国男、南方熊楠のような学者の本もいいでしょう」のように、他の学者、特に無名の学者などにはどうして置き換えられないことが多い。「のような」は例示であっても N1 を何らかの基準で限定していることが多いため N1 を他のものには置き換え難くなり、また固有名詞が例としてあげられることが多くなるものと考えられる。
- ・ 比喻や属性にフォーカスをあてた「のような」や、「という」と共に現れる N1 は、置き換えると文意が変わる。したがって、並列表現と共に使われることがほとんどなくなるものと考えられる。
- ・ N1 が身近な名詞で且つ N2 がグループ1の場合、「などの」「といった」と共に現れることが多く、N1 が固有名詞で且つ N2 がグループ2の場合、「のような」「という」と共に現れることが多くなる。

5. まとめ

上下概念関係をあらわす「などの」「という」「のような」「といった」の4つの連体修飾表現の特徴を分析、考察した。例示、範疇、属性、比喻の4つの用法が見られ、上位の概念が固有名詞を下位に持ちやすいかどうかによって、頻度などに明らかな違いが見られた。

参考文献

- [1] 安藤まや・関根聡・石崎俊, 定型表現を利用した新聞記事からの下位概念単語の自動抽出, 情報処理学会研究報告, 2003-NL-157, pp77-82, 2003.
- [2] 益岡隆志・田窪行則, 基礎日本語文法 改訂版, くろしお出版, 1992.
- [3] 宮島達夫・仁田義雄編, 日本語類義表現の文法(下), くろしお出版, 1995.
- [4] 安田芳子, 連体修飾形式「ような」における<例示>の意味の表れ, 日本語教育 92号, pp.177-187, 1997.
- [5] 藤田保幸, 研究ノート「～トイウ」と「～トイッタ」, 河内国文 11, pp.8-14, 1986.
- [6] 田添文博・椎野努・榎井文人・河合敦夫, “名詞AのようなB”表現の比喻性判定モデル, 自然言語処理 Vol.10 No.2, pp.43-58, 2003.
- [7] 緒方典裕, 型理論に基づいた特定領域テキストからの動的な Taxonomy Mereology 構成, 情報処理学会研究報告, 98-NL-127, pp.133-140, 1998.