

ニュース文のスタイルに関する基礎的調査

金城 由美子* 熊野 正 西脇 正通 柏岡 秀紀 田中 英輝

A T R 音声言語通信研究所

1 はじめに

ニュース文はその語彙構成や、文の長さ、複雑さから書きことば的な特徴に注目されることが多い。本稿では、このような特徴をニュースコーパスによって検証するとともに、他のドメインのコーパスとの比較を通して、ニュース文が文末表現や指示表現に話しことば的な特徴も持つことや、ニュースに特徴的なモダリティ表現、特に引用に関する形式が発達していることなどを示す。同時に引用表現が、ニュース文の長さ、複雑さの要因となっていることを明らかにする。またニュース文の談話構造についても検討を行う。

2 語彙

A T R 音声言語通信研究所では、講演、ニュースなど独話の同時通訳研究の一環として、放送ニュースの日英機械翻訳システムの研究・開発を行なっている。研究に際しては、放送のために作成された日本語・英語原稿がデータベースとして蓄積されたニュースコーパス¹を利用している。コーパスの概要を表1に示す。

	日本語	英語
記事数	317,205 記事	73,802 記事
対訳記事	41,183 記事	41,982 記事
記事の長さ	5.6 文/記事	7.6 文/記事

(1995年3月～2000年4月)

表1: 放送ニュース汎用原稿コーパス

ニュースコーパスに含まれているのは、放送でアナウンサーが使う原稿の元となる汎用原稿と呼ばれる記事原稿であり[1]、一般に語彙調査の対象とされる書き起こしデータとは性質の異なるものである。

他のコーパスとの比較を行なうため、書きことばとして新聞、話しことばとして旅行会話のコーパスを用い、すべてのコーパスを茶筌によって形態素解析を行なった。ニュースは、放送ニュースコーパス中の1995年3月から2000年4月までの287,315記事(延べ14,879,546語/異なり141,704語)、新聞は日本経済新聞CD-ROM 1990年～1995年版の915,469記事(延べ27,6914,532

語/異なり482,038語)、旅行会話としてATRバイリンガル旅行会話コーパス[2] 618対話の日本語部分(延べ256,219語/異なり5,254語)を用いた。

2.1 品詞分布

各コーパスの品詞の分布を表2に示す²。新聞の品詞分布の特徴は名詞比率が非常に高いのに対して、旅行会話の品詞分布は名詞比率がそれほど高くなく、他のコーパスに比べて助動詞の比率が高い。ニュースの品詞分布は新聞に近い。

	新聞	ニュース	旅行会話
名詞	44.00	40.45	27.69
助詞	25.29	27.61	21.66
動詞	10.08	12.34	10.79
記号	12.64	9.42	14.81
助動詞	4.52	6.90	15.30
副詞	0.78	0.68	2.30
形容詞	0.74	0.63	0.92
その他	1.95	1.97	6.53

表2: 品詞(大分類)の分布 (%)

また細分された品詞に注目しても、書きことばに多いサ変名詞、話しことばに特徴的に現れる終助詞、接続詞、感動詞の分布は、表3に示すように新聞と同様の傾向を示している。

	新聞	ニュース	旅行会話
サ変名詞	8.77	7.25	3.90
終助詞	0.03	0.04	1.02
接続詞	0.29	0.31	1.31
感動詞	0.01	0.08	2.20

表3: ドメインに特徴的な品詞の分布 (%)

品詞の分布は、これまでの研究で指摘されてきた[3]のようにニュースの書きことば的特徴を示しているといえる。

*E-mail: yumiko.kinjo@slt.atr.co.jp

¹研究利用のためNHKから提供を受けたもの。

²大規模な語彙で、機能語を含めた品詞構成を比較するため延べ語数による分布を示した。

2.2 高頻度語から見た特徴

品詞分布では書きことば的な特徴を見せるニュースだが、個々の語彙に目を向けると、助動詞、指示表現、時間表現において、書きことばとは異なる話しことば的な特徴を示す。

まず、旅行会話と同様に丁寧さを表す助動詞の「です、ます」が非常に高頻度で現れる。丁寧さは聞き手めあてのモダリティとされるもので、聞き手の存在を前提とする話しことば的な要素である。

次に指示表現に注目すると、ニュース、新聞ともに「この」を始めとするコ系の指示表現が高頻度で現れるが、ニュースの方が多種のコ系指示表現が高頻度語として出現する。コ系の指示表現は直示的、ソ系の指示表現は文脈指示的な要素とされるが、新聞は「その」がより高頻度で出現し、「同社」のような表現に現れる接頭詞「同」が高頻度語として現れるなど、高い文脈依存性を示す。それに対して、ニュースは直示的なコ系指示表現への依存度が高い。

最後に時間表現における特徴だが、ニュースには「きょう」を始めとして「今後」「今年」のような現在を基準とした相対的時間表現が高頻度語として現れるのに対し、新聞では相対的時間表現より「三月」のような絶対的時間表現が高頻度語として現れるという傾向の違いが見られる。

ニュースの指示表現と時間表現は直示的であるという点で共通しており、この点も話しことば的特徴として捉えることができる。

3 構文

ニュースの構文の特徴として、モダリティ表現と構造について取り上げる。また、ニュースの構造に複雑さをもたらす要因として、名詞修飾節と引用表現について考察を行なう。

3.1 ニュースのモダリティ

日本語のモダリティを大別すると、事態に対する話し手の判断や態度を表すモダリティと、聞き手に対する態度や働きかけを表すモダリティに分類される。「既に起きた事実を中立的に伝える」という役割を持つ放送ニュースの性質から予想される通り、ニュースにはあまり多くのモダリティ表現は現れない。また話しことばや、書きことばと異なる用法を持つモダリティ表現も見うけられる。

事態に対するモダリティでは「はずだ」などの当為を表すもの、「だろう、はずだ」などの概言を表すものなどがあるが、これらの表現がニュース文に現れるこ

とはほとんどない³。数少ない主観的判断を表すモダリティとしては「動詞連用形＋そうです」が使われる。

(1) 今後、各社の顧客獲得競争が本格化しそうです。通常、「動詞連用形＋そうです」は、ニュースと異なり、動的事態の生起が予想される外的事態が存在する場合に使われるのが一般的用法である [4]。

(2) この空模様ではどうも雨になりそうです。

一方、概言の中でも、他者から得た知識をそのまま伝える伝聞には、ニュースに特徴的な「ということです」「としています」といった表現が頻出する。

(3) 移植チームによりますと患者の呼吸の状態は人工呼吸器を外す前と変わり無いということです。

聞き手に対するモダリティでは、2.2 節で述べたように丁寧さを表す助動詞「です、ます」は現れるが、聞き手に直接働きかける命令や、問いかけなどのモダリティが現れることはほとんどない。

3.2 複雑な構造を持つ文

ニュース文が長く、複雑な構造を持つことは、過去の研究においても指摘されている [5],[6],[7]。各コーパスの1文の長さを表4に示す。

ニュース	新聞	旅行会話
78.3	49.6	21.1

表 4: 1 文の長さ (字)

ニュース文の長さの要因と構文の特徴を調べるために、各コーパスからそれぞれ 10,000 文を日本語解析システム⁴により係り受け解析を行い、従属節となりうる要素について調べた。主要素とその頻度を表5に示す。

	ニュース	新聞	旅行会話
動詞連用句	6,291	14,869	3,845
動詞連体句	12,167	5,795	439
と格を伴う句	5,283	1,641	469

表 5: 従属節を構成する句

動詞連用句はいわゆる動詞の連用中止や、接続助詞「が」を伴う述語句で、並列節を構成する要素となる。新聞に多い述語句だが、旅行会話にも次の例のように数多く現れる。

(4) たぶん大丈夫だと思いますが、空き室状況を調整しますので、少々お待ちください。

³後述する発言の被引用部などを除く。

⁴NHK で開発されたもの。

動詞連体句は、動詞の終止形が名詞句に接続する形式を持つ述語句で、名詞修飾節を構成する要素となる。と格を伴う句は述語の終止形にと格が続く述語句で、補足節の一種である引用節を構成する要素となる。

名詞修飾節と引用節がニュースに特徴的であり、文の長さや複雑さに深く関係する要素であると考えられるので、続いて名詞修飾節と引用節を含む直接引用について考察を行う。

3.3 名詞修飾節

名詞修飾節は、補足語修飾節、内容節などに分けられ⁵、さらに補足語修飾節は限定的修飾を行うものと非限定的修飾を行うものに分類される。

- (5) 近くを通りかかった子どもたち (限定的修飾)
- (6) 日本を訪れているトルコのジェム外相
(非限定的修飾)
- (7) 激しい雨が降る恐れ (内容節)

各コーパスに現れる名詞修飾節の特徴を知るため、それぞれ 100 文を日本語解析し、その結果に修正を施し、名詞修飾節と認められるものを人手で分類した。名詞修飾節のタイプと頻度を表 6 に示す。

	ニュース (109)	新聞 (62)	旅行会話 (1)
補足語修飾節			
(限定的)	52	32	0
(非限定的)	10	15	0
内容節	47	15	1

表 6: 名詞修飾節のうわけ

名詞修飾節のほぼ半数が、限定的修飾を行う点ではニュース、新聞とも共通しているが、新聞は非限定修飾が多く、ニュースは内容節が非常に多いことがわかる。また同じ内容節でも、ニュースに現れるものは、半分近くが「という」「との」などの接続形式と共に出現している点も新聞と大きく異なる点であった。

- (8) 予算案の採決を阻止していくという方針

ニュースに内容節が多いことは、ニュースの頻出パターンに「恐れ、動き、意見、認識」など内容節をとる名詞が多く含まれるという点からも裏づけられる [8]。

トイウ内容節は、被修飾名詞より内容節の方に情報上の重点があり、トノ内容節は引用に関係するとの指摘がある [9]。発言・思考を表す名詞はどちらの内容節にも共通して被修飾名詞として現れるという点で、両

者とも引用に関係している。さらにトノ内容節も内容節に情報上の重点を置く機能があると考え、ニュースにおいて接続形式を伴う内容節をとる被修飾名詞は、引用内容に情報の重点を置くための表現として機能していると考えられる。つまり、そのような被修飾名詞は、間接引用の形式として機能語化しているといえる。機能語化という見方は、トイウ内容節では被修飾名詞が意味的には副次的要素として働き、範疇化の機能を持つという指摘 [9] とも一致する。

また内容節が接続表現を伴う場合は、「だろう」などのモダリティを含む表現を取ることが可能である。

- (9) a. 決定が延期される (* だろう) 予測

- b. 決定が延期されるだろうという/との予測

このことは、内容節を含む表現がニュースの構造をより長く、複雑なものにしている可能性を示唆している。

3.4 直接引用

次にニュースに特徴的な引用節を含む表現である直接引用について調査を行なった。直接引用であることが明らかな「述べる」「話す」を引用の述語⁶として含む文を取りだし、被引用部の特徴について調べた。

3.1 節で述べたように、ニュースの本文にモダリティ表現はほとんど現れない。誰かの発言内容を紹介する直接引用の場合は、被引用部に事態に対するモダリティ、聞き手に対するモダリティの両方が現れると予想される。しかし、実際の被引用部には、本文とは異なり事態に対するモダリティは現れるものの、聞き手に対するモダリティが現れることは少ない。

また、被引用部には本文に現れる丁寧さを表す「です、ます」が基本的には現れないという特徴もある。つまり直接引用の表現であっても、発言通りの引用がなされているわけではなく、間接引用的な特徴を備えているといえる。

被引用部に聞き手に対するモダリティが現れるかどうかは、引用の述語によって異なる傾向を見せる。引用の述語が「述べる」の文は 70,925 文 (ニュース文の約 4.5%) あるが、「です、ます」を被引用部に含む文は約 1.5% しかないのに対して、「話す」の文 42,384 文 (ニュース文の約 2.7%) では約 34.0% の文に「です、ます」が現れるなど、大きく異なる。

また「述べる」は主に政治と国際、「話す」は経済ニュースを中心として政治と国際以外のニュースに現れるなど、著しい分野の偏りを見せる。これらの事実、ニュースの分野によるスタイルの違いが存在する

⁵他に相対名詞修飾節があるが、今回は調査の対象としない。

⁶引用の述語としては「いう」が最も高頻度であったが、間接引用との区別が難しいため調査の対象としなかった。

ことを示唆しているが、これについては今後の検討課題としたい。

4 談話構造

ニュース文の談話構造についての先行研究として、ニュース原稿の調査を行なったものがある [10]。この論文では、記事の冒頭に現れるリード文とよばれるニュースのあらましを述べる部分と本文との関係を調べ、ニュースの基本構造が「リード文+背景+詳細+展望・付加」という形であると報告し、ニュースの大半 (77.8%) は表 7 に示すように 4 つのニュースの型にあてはまるとしている。

構成	件数
リード文 + 詳細	47 件 (13.0%)
リード文+背景+詳細	98 件 (27.1%)
リード文 + 詳細+展望・付加	66 件 (18.3%)
リード文+背景+詳細+展望・付加	70 件 (19.4%)

表 7: ニュース文の構成

また、同論文では記事全体の要約となるリード文の重要性が指摘され、リード文が本文に比べて長いこと、リード文に現れた単語が同じ順序で本文に再度現れることが報告されている。

以上のような構造がニュースコーパスに認められるかを検証するために、ニュース文の出現順ごとの 1 文の長さとして出現単語の頻度について調査を行なった。出現順と 1 文の長さを表 8 に示す。この結果が示すように、ニュースの第 1 文は他の文に比べて明らかに長く、第 3 文以降は単調に減少していく。

全体	第 1 文	第 2 文	第 3 文	第 4 文	第 5 文
78.3	86.9	82.4	86.2	81.2	76.7
	第 6 文	第 7 文	第 8 文	第 9 文	第 10 文～
	71.4	66.5	61.1	55.5	42.9

表 8: 出現順と 1 文の長さ (字)

出現順ごとの品詞の分布についてみると、第 1 文と第 2 文において、名詞、特に固有名詞の頻度が高く、第 3 文以降では減少する。高頻度語では、指示表現の「この」は第 1 文のみ出現頻度が極端に低く (64 位)、他の文ではコーパス全体での出現頻度 (28 位) に近い出現頻度を保っている。また、ニュース中の語句や背景情報を説明する時に使われる表現「～ものです」に含まれる形式名詞「もの」は、第 2 文に高頻度で出現 (28 位) し、第 1 文での頻度は低く (148 位)、第 3 文以降はコーパス全体での出現頻度 (52 位) に近い出現頻度を推移している。

これらの事実から、ニュース文の基本構造をニュースコーパスに適用することは妥当だと思われる。先行

研究におけるリード文、背景、詳細、展望・付加の区分は、文中の事態の時系列関係に基づいて人が判断を行なったものであるため、これらを機械的に推定する方法について検討を行なっていく必要がある。

5 おわりに

本稿では、ニュースコーパスの調査に基づいて放送ニュースの言語的特徴を語彙、構文、談話構造、それぞれのレベルについて調査を行ない、ニュースは書きことば的な特徴と話しことば的な特徴を合わせ持つと同時に、ニュースに特徴的な表現、内容節による引用表現などがあることを示した。またニュースの談話構造についても明らかにした。

今後の研究課題として、引用表現の詳細、談話構造の推定方法などがある。またニュースの言語的特徴を翻訳のための知識としてどのように生かすかもこれらの課題である。

謝辞

本稿をまとめるにあたり貴重な助言をいただいた城西国際大学の石野博史先生と NHK 放送技術研究所の江原暉将氏に感謝いたします。

参考文献

- [1] 熊野正, 田中英輝, 金淵培, 浦谷則好. 日英ニュース原稿の対訳コーパス化に関する基礎調査. 言語処理学会第 2 回年次大会発表論文集, pp. 41-44. 言語処理学会, 1996.
- [2] 竹沢寿幸, 白井諭, 大山芳史. バイリンガル旅行会話コーパスに見られる話し言葉の特徴分析. 情報処理学会研究報告, Vol. 2001, No. 9, pp. 137-144, 2001.
- [3] 石井正彦. 番組ジャンルの特徴語とジャンル間の関係. テレビ放送の語彙調査 III, 国立国語研究所報告 115, pp. 76-106. 大日本図書株式会社, 1999.
- [4] 益岡隆志, 田窪行則. 基礎日本語文法. くろしお出版, 東京, 1989.
- [5] 菅野謙, 石野博史. ニュース文章の分析方法について (1). 文研月報, 12 月号, pp. 33-47, 1968.
- [6] 井上和子. 日本語の談話構造. 放送文化基金 (編), 放送とことば—放送文化基金・研究報告会から—, pp. 64-78. 放送文化基金, 1984.
- [7] 西脇正通, 浦谷則好, 畑田のぶ子. テレビニュース原稿の特徴分析. 言語処理学会第 5 回年次大会発表論文集, pp. 169-172. 言語処理学会, 1999.
- [8] 西脇正通, 金城由美子, 柏岡秀紀, 田中英輝. 放送ニュース文における頻出パターン調査. 言語処理学会第 7 回年次大会発表論文集, pp. P1-13. 言語処理学会, 2001.
- [9] 益岡隆志. 名詞修飾節の接続形式—内容節を中心に—. 田窪行則 (編), 日本語の名詞修飾表現, pp. 5-27. くろしお出版, 東京, 1994.
- [10] 川村よし子, 金庭久美子. TV ニュースの基本構造の分析. 東京国際大学論叢 商学部編, Vol. 58, pp. 109-121, 9 1998.