主題の補完のための文の分類*

横山　晶一　　三文字　陽子
（山形大学工学部）

1．はじめに[1]

主題、焦点は、日本語の文章を構成する基本要素であり、これらを正確に抽出することによって、談話解析のさまざまな場面に役立つ。しかしながら、日本語の文章においては、主題が省略されているものも多い。人間は、それらの文章を読んだときでも、直感的に主題が必要かどうかを判断し、必要な場合には無意識のうちにそれらの主題を補って文章を理解している。

機械処理で主題、焦点を抽出した場合には、これらの省略を補うことは非常に難しい。これまでは、漠然と前文の主題や焦点を次の文の主題として implicit に取り上げたり、筆者を主題にするなどの方法が取られてきた。しかしながら、機械処理の観点から、実際の文章に対して主題をどのように補完すればよいのかについては、余り調査が行われてこなかった。

本稿では、実際の新聞コラムを人手で調査し、主題が欠けている場合にどのような主題を補うべきかについて、文章を分類した。その結果、主題の補完に関する文章の分類をある程度行うことができたので報告する。

２．主題、焦点とその抽出システム

主題、焦点を次のように定義する[1]。

・主題:その文中で話題となっている要素であり、前述された既知の情報

・焦点：その文で新しく導入された情報

我々はすでに、この基準に従った、主題・焦点抽出システムを構築した。当初は、「完全な構文解析を前提とする」という厳しい条件があったが、その条件を形態素解析レベルまで緩和して、主題・焦点を抽出するシステムを構築した[2]。

具体的には、次のような手順で主題・焦点を抽出する（主な手順のみ述べる）。

(1) 述語の形から、文を名詞文、動詞文、形容詞文の３種類に分類する
　動詞文：彼は談話を発表した。
　形容詞文：財政は苦しい。
　名詞文：単身赴任が多く見られるのは日本だけだ。

(2) 主題・焦点を次のように抽出する
　(a) 「は」があるときは、主題を「は」の名詞とし、焦点を名詞文では述語名詞、動詞文、形容詞文では必須格を抽出する
　(b) 主格の「が」が存在するときは、焦点は主格名詞。ただし、「が格」と述語の間に動詞が存在したときには、この「が格」は修飾句内のものとみて焦点にはしない

(3) 次のように主題の補完を行う
　(a) 名詞文：前文が名詞文なら、前文の述語名詞を補完する。動詞文なら焦点、あるいは前文を名詞化して補完する
　(b) 動詞文、形容詞文：動詞、形容詞とマッチするものを前文、前々文とさかのぼって調べ、補完する

(4) 主題・焦点の前が、動詞、助詞「の」の場合、修飾部として抽出する

この手順で抽出した文例を以下に一つ示す。この文では、下線部が主題、二重下線部が焦点である。

(文 1) <u>国連高等難民弁務官事務所</u>はコンピュータによる初の<u>子供検索システム</u>をスタートさせる。
主題：国連高等難民弁務官事務所

＊ Classification of Sentences to Complement Themes by YOKOYAMA Shoichi and SAMMONJI Yoko (Yamagata University)

焦点：子供検索システム
焦点修飾＝コンピュータによる初の

上記のように抽出された主題・焦点を用いて、キーワード抽出[3,4]や、重要文抽出[5]が行われ、その有効性を実証してきている。

しかしながら、上記(3)のような主題補完方法は、実際の文章を詳細に調査した結果として採用された手法ではないため、実際の文章でどのような主題の補完をすべきか、あるいは、また、主題の補完がすべての文章で必要かどうかといった点について、これまで調査は行われなかった。

３．資料

天声人語[6]の 2005 年度分のうち 113 回の記事をデータとして用いた。ここに現れた文の数は 2357 文である。これらの文を、システム[2]の出力と比較対照しながら、人手で分類した[7]。その大雑把な内訳は表１の通りである。

表１ 天声人語の文の分類

主題のある文		964
主題の補完の必要がない文		797
	主格の「が」が存在	461
	主格の「が」なし「も」あり	112
	その他	224
補完の必要がある文		497
不明な文		99
	疑問文	9
	不明	90
合計		2357

この表で分かるように、全体の 40.9%に当たる 964 文に主題が存在する。主題の補完の必要がない文も、全体の 33.8%くらいある。このうちの主なものは、以下の例文のように、主格の「が」が存在するものである。

(文 2) 11 月下旬、米将兵らが現れた。

この場合には、「米将兵ら」が焦点になるが、特に主題を必要としない。

上の表の「主格の「が」は存在しないが「も」がある」という意味は、次のような文である。

(文 3) 同じ課題に終戦直後の出版界も直面した。

この文の場合は、「終戦直後の出版界」を焦点として採用してよい。その他は、いろいろなタイプがある。

(文 4) 幹部に外部の人を入れる。市民や議員から職員に要望があれば、文書に残して情報公開の対象にする。組合との不透明な協議をやめる。それらが改革案の内容だ。

この文では、上記の点線部が「内容」の具体例を示している。この場合には主題を補完しない。

(文 5) 「６連覇、いきたいね。もっと強くなりたい」。
(文 6) 朱に交われば、赤くなる。
(文 7) 大久保利通、乃木希典、犬飼毅、…

(文 5)のような引用文のみから成る文も主題を補完しない。さらに、条件文(文 6)、名詞のみの文(文 7)、感嘆文なども主題を補完しない。

また、不明な文とは、次のような文である。

(文 8) ２０００年ミレニアム記念の北米練習船レースでは、１等航海士として乗り込んだ海王丸が、世界の強豪を抑えて１位に輝いている。なぜ、この時代に帆船で実習をするのだろうか。
(文 9) あまりに傍若無人ではないのか。こんなオヤジの小言を口にしたら、同僚が教えてくれた。彼女たちは「カスタネット娘」とか「カンカン女」と呼ばれているのだ、と。

(文 8)では、下線点線部は疑問文になっているが、補完の必要性の有無は不明である。また、(文 9)でも、下線点線部は引用に近い文であるが、現在のところ、主題を補うべきかどうかは分からない。

以下では、全体の 21.1%ほどであるが、主題の補完の必要な文について詳細を述べる。

４．主題の補完の必要がある文

表１の 497 文をさらに細かく分類すると、表２のような内訳になる。以下ではこの内訳に沿って、詳細を述べる。

表 2　主題の補完の必要な文の内訳

前文の主題	171
前文の焦点	94
筆者	94
読者	3
一般者	10
前文の全体	39
前出の複数文全体	7
前文以前の文の主題	34
前文以前の文の焦点	32
主題でも焦点でもない成分	11
補完の必要な文の小計	497

４．１．前文の主題が次の文の主題となる文

(文 10) 浜口は財政を引き締めるとともに、軍縮と国際協調の外交を進めた。(Φ=浜口)そこを右翼に襲われた。

最も典型的な文である。最初の文が主題を有する文で、次の文が主題の欠けた文の場合、前文の主題（下線部）が、次の文の主題となる。

４．２．前文の焦点が次の文の主題となる文

(文 11) 日本の気象学者らが世界の大河の流量の予測をまとめた。（Φ= 気象学者ら）地球の温暖化による降水量の変動などを計算した。

これも比較的典型的な例である。

４．３．主題として筆者を補う文

これには大きく分けて２種類考えられる。一つは、次の文の例のように、文章の冒頭に来る文、即ち第１文である。

(文 12) （Φ= 筆者）切手の位置に子犬をあしらった、来年のはがきの１枚を手に取る。（Φ= 筆者）裏返す。

もう一つは、述語が「～ほしい」、「～たい」のような希望を表す場合である。

(文 13) （Φ= 筆者）民主党にそんな危機感をもって今後の論戦に臨んでもらいたい。

いずれの場合も、主題として「筆者」を補うことができる。

４．４．主題として読者を補う文

(文 14) （Φ= 読者）AED をご存知ですか。（Φ= 読者）自動体外式除細動器といっても、ますますわかりにくいかもしれない。

上記のような「か」で終わる問いかけをする文では、「筆者」ではなく、「読者」を補った方がよい。

４．５．主題として一般者を補う文

前項の文と似ているが、同じ問いかけでも、問題提起をする文の場合には、「読者」よりも「一般者」を補う方がよい場合がある。

(文 15) 昨年７月の新潟、福島、福井３県の豪雨でも、多くのお年寄りがなくなった。（Φ= 一般者）こうした支援の必要な人たちをどう助けるのか。

４．６．前文全体

(文 16) 「天気がよかろうと、悪かろうと、ドナウ河の流れは同じ。ただ定めなき人間のみが、地上をさまよい歩くのです」。（Φ= 前文全体）１９世紀のルーマニアの国民詩人といわれるエミネスクの詩の一節だ。

上記の文のように、前文が引用文で、同じ段落にあり、しかも次の文が主題を欠く場合には、前文全体を主題として補完するとうまく行く。表２に記載した 39 文中、この種のものは 11 文あるが、これらはすべて補完が可能になる。

４．７．前出の複数文の全体

(文 17) 明治４年の初夏、岩倉具視ら高官が集まり、開国日本の服装はどうあるべきか激しく論じた。和服派は「衣服まで外国をまねるのは愚か」と訴えたが、洋服派が「外国との交際に欠かせない」と説き伏せた。（Φ= 点線部全体）世にいう「洋服大評定」である。

上の文章では、前の複数文が、次の文の主題になっている。

４．８．前文以前の文の主題

(文 18) 作家の池波正太郎さんは、並外れて準備のいい年賀状について随筆を残している。「年が明けると間もなく、来年の年賀状の絵を描き、印刷屋へ出す」（中略）（Φ=池波正太郎さん）夏から秋、そして師走にかけて、少しずつ宛名を書いてゆく。

この文章では、（中略）の部分も含めてずっと同じ主題が保たれている。

４．９．前文以前の文の焦点

(文 19) 日本で一番古いという仏像を見たのは、２年前の春だった。奈良県明日香村の飛鳥寺の本尊、釈迦如来の座像で、３メートルちかくある。７世紀の初め、女帝の推古天皇が仏師の鞍作鳥（くらつくりのとり）（止利）につくらせたという。（Φ= 座像）後年火災で焼けた。

このような文章の場合には、間に新しい焦点（推古天皇）が入っているので、意味解析をしないと補完は難しい。

４．１０．前出の主題でも焦点でもない成分

(文 20) 戦前の逓信相小泉又次郎氏の肉声を聞く機会が先日あった。（Φ=逓信相小泉又次郎氏）小泉首相のおじいさんである。

ここに分類された文章は、主題や焦点の修飾部や、副詞句（例：「民主主義について、…」という形の副詞句が、後で主題として現れる）など、さまざまなバリエーションがある。

５．まとめ

前節の表２や例文から分かるように、当初考えていた補完方法（２(3)）よりもはるかに複雑な補完が必要である。表２の「前文の主題」から「前出の複数文の全体」（４．１～４．７）に至る分類は、比較的容易にシステムに組み込める可能性があるので今後検討したい。それ以外のものは、解析の手法や組み込む必要性、可能性をさらに追求したいと考えている。

参考文献

[1] 吉田悦子・横山晶一：主題・焦点を用いた文脈解析の一手法、電子情報通信学会技術報告 NLC97- 29 (1997)

[2] 廣町潤・横山晶一・西原典孝：形態素解析を用いた主題・焦点抽出システム、情報処理学会第 65 回全国大会講演論文集 5B-3 (2002) pp.2-11,12

[3] 横山晶一・菅野崇：主題・焦点のスコアを用いたキーワードの抽出、言語処理学会第７回大会論文集(2001) pp.177-180

[4] 菅野崇・横山晶一・西原典孝：主題・焦点のグループ化によるキーワード抽出、言語処理学会第９回大会論文集(2003) pp.393-396

[5] 横山晶一・菅野崇・西原典孝：主題・焦点リンクを用いた重要文抽出システム、情報処理学会自然言語処理研究会資料 NL156-1 (2003)

[6] 天声人語、朝日新聞 http://asahi.com/paper/column.html

[7] 三文字陽子：主題の補完のための文の分類、山形大学工学部卒業論文 (2006)