

## トピック・マーカを利用した談話構造の表示の試み

中川奈津子<sup>†</sup> 横森大輔<sup>†</sup> 浅尾仁彦<sup>‡</sup><sup>†</sup> 京都大学大学院 人間・環境学研究科<sup>‡</sup> 京都大学大学院 文学研究科

## 1 はじめに

本発表<sup>\*1</sup>は、トピック (文の中でそれについて語られるところのもの: [5]) をマークする助詞やトピックを導入する構文などを用いて、日本語話し言葉の構造を表示する表記法と手順の提案を目的とする。

## 2 節の情報構造

この節では、談話の構造を判定するのに必要な情報構造の諸概念を概観する。

まず、トピック・マーカについて見る。先行研究においてすでに、トピックをマークする助詞には少なくとも、ハ、トイウノハ (とその変種、ツテ、ツテイウノハなど)、モがあることが分かっている ([11])。<sup>\*2</sup>

次に、トピックを導入する構文について見る。場所格を取るタイプの自動詞 (e.g. *be, live, arrive*) や分裂文はトピックを導入する傾向がある ([2, 9, 11])。

最後に、節の情報構造の種類とそれによる助詞の違いについて見る。日本語では、(2) のように節の情報構造に少なくとも以下の 4 種類がある ([9, p.121])<sup>\*3</sup>

(1) - トピック-コメント構造 (述語フォーカス<sup>\*4</sup>)

- a. (子どもたちは次に何をしたの?) 子どもたちは学校に行っただよ。

- 同定構造 (項フォーカス)

- b. (誰が学校に行ったの?) 子どもたちが学校に行っただよ。

- イベント報告構造 (文フォーカス)

- c. (何が起こったの?) 子どもたちが学校に行っただよ!

- 前提構造 (フォーカスなし)

- d. (その日の朝、太郎は忙しかった。) 子どもたちが学校に行っただよと、掃除をして、パーティーの買い物に行

かなければならなかったのだ。

(1a) では、「子どもたち」はハでマークされている。(1b) と (1c) では、「子どもたち」はガでマークされている。最後に、(1d) でも、「子どもたち」はガでマークされている。ここからも、日本語のトピックはハでマークされ、その時に節全体はトピック-コメント構造となることがわかる (1a)。また、主語だけがフォーカスの場合と文全体がフォーカスの場合には、主語はガでマークされる (1b-c)。最後に、前提構造文も主語はガでマークされる (1d)。

本発表ではこれを利用し、談話の構造を表示することを試みる。

## 3 先行研究と本発表の課題

## 3.1 先行研究

本節では談話構造を表示する理論として修辞構造理論と相互行為の観点からの分析を概観する。

修辞構造理論 (Rhetorical Structure Theory, RST: [10]) では、談話内の 2 つの文の間に見られる 23 種類の関係が規定されており、各関係を認定する際の基準が明示されている。

相互行為の観点からの分析 (e.g. [12]) では、進行中の談話に対して聞き手が抱くであろう「疑問」という観点を取り入れ、話題導入表現の機能は談話の目的を表す (2) のような形式の疑問を導入することであると考えられている:

- (2) 話し手は X の Z は W かを伝えようとしている

このとき X はトピックに、Z は X に伴うフレーム知識を総括するある種の抽象名詞、W は疑問詞に相当する。(2) の X, Y, Z の値が満たされたのち、話題導入表現によって設定された疑問への回答となる「評価」が現れ、その周辺にある文は疑問・回答の「背景文」とされる。

## 3.2 本発表の課題

本発表では、先行研究とは異なり、2 にあげられた道具立てに加え、名詞句の言及回数 (持続距離: 後述) を使い、機能主義的な方法を用いて談話の構造を表示することを試みる。<sup>\*5</sup>

例えば (3) のように 3 つの節が連なっている場合、(3a) 「バス」と (3b) 「φ」は同じ指示対象が言及されているため、単純に指示対象が同じものを追跡する素朴な理論でも捉えやすい。ところが (3c) の場合は、「バス」と指示対象が同じでないため

<sup>\*1</sup> 第一発表者連絡先: nakagawanatuko@gmail.com

<sup>\*2</sup> ハに関する古典的な研究は以下を参照: [8, 7]

<sup>\*3</sup> 原文は英語で、(i) のように助詞ではなく強勢の違いにより現れている。SMALL CAPITAL で表示された部分は強勢を表す。

(i) a. (What did the children do next?) The children went to SCHOOL.  
b. (Who went to school?) The CHILDREN went to school.  
c. (What happened?) The CHILDREN went to SCHOOL!  
d. (John was very busy that morning.) After the children went to school, he had to clean the house and go shopping for the party.

<sup>\*4</sup> フォーカスとは、トピックに関して述べられるところのものである (cf. [9])。

<sup>\*5</sup> したがって本発表は、談話トピックを表示しようとする機能的な試み (e.g. [1]) の発展と捉えることができる。

「運転手さん」がバスの下位トピックとでも言うべきものとしてはたらいっていることが捉えにくい。「運転手さん」はトピックをマークするハが使われており、ここではバスが最低だったことの一部として「運転手さんは酔っ払っていた」が捉えられるべきである。

- (3) a. 昨日乗ったバスは最低だった。  
b.  $\phi$  汚かったし、  
c. 運転手さんは酔っ払っていた。

また、(4)における「私」をトピックと考えることも、情報構造の理論に混乱を招く。先行研究 ([9]) では、「私」がバスに乗ることの理由として (4) が発話された場合に「私」はトピックであるとされていたが、日本語では、「私の車」はガでマークされており、「壊れたのは私の車だ」(同定構造)あるいは「何が起ったか」と私の車が壊れた」(イベント報告構造)のパラフレーズに相当するような、「私の車」がフォーカス(の一部)となる情報構造を持つと考えられる(cf. (1b-c))。「私」をトピックだと考えると、フォーカスの一部としてトピックがあるような場合を想定することになり、情報構造の理論が混乱する恐れがある。この場合、「(私の)車が」は文フォーカスの一部であってトピックではなく、「私」は文を越えた上位トピックであると考えた方がよい。

- (4) [いつも車に乗っている「私」がバスに乗るのを不思議そうに見つめる人に対して]  
(私の)車が壊れたんだ。

本発表では、(3)の「運転手さん」を「バス」の上位トピックの下位トピックとして、(4)の「私」を節内のトピックではなくより上位のトピックとして、適切に表示できるような方法を提案する。

## 4 調査

### 4.1 調査手順

コーパスとして「日本語話し言葉コーパス」(CSJ)の独話データの1つS00F0014を用いた。このデータを著者の1人がイントネーション・ユニット(Intonation Unit, IU: [4, 6])と言われる聴覚印象に基づく発話の単位に分割し、747個のIUと974個の名詞を得た。各々のIUに含まれる名詞の指示距離(Referential Distance)と持続距離(Persistence)を[5]の方法に基づき計測した。<sup>46</sup> 指示距離とは、ある名詞に関して、その指示対象が言及されたIUとそれが直前に言及されたIUとの間の数によって近似される。最大値は恣意的に100とした。以前に全く言及がない場合は100となる。持続距離とは、ある名詞句に関して、その指示対象があるIUで言及された後に言及された回数で近似される。最大値はこれも恣意的に200とした。なお、述語の項である場合に限り、ゼロ代名詞の存在を認定した。その場合、述語と同じIUに存在するものとして計測した。述語的名詞句(e.g. 「AはBだ」におけるB)は指示対象を持たない(e.g. [3])ため、それ自身の指示距離と持続距

離は計測していない。

[9]に基づく4種類の情報構造を持つ節(cf. 2)を、(1)を例にとって、表記法を(5)のようにする:

- (5) a. トピック-コメント構造  
    ● 子どもたち:  
    = 学校に行った  
        スクールバス:  
    = 出発しそうだった  
    = 走った  
b. 同定構造  
    ● xが学校に行った:  
    = x=子どもたち  
c. イベント報告構造  
    @ 子どもたちが学校に行った  
d. 前提構造(表示しない)

談話構造の認定手順は以下である: まず、場所格を取るタイプの自動詞(e.g. *be, live, arrive*)や分裂文はその後の談話でトピックとなる指示対象を導入する<sup>47</sup> 傾向があるという一般化([2, 9, 11])を利用し、導入されたトピックを選び出す。これは(5c)の形式で表記できることが多い(cf. [9])。このとき指示距離が100で持続距離も大きいことも参考にする。ゼロ代名詞、ハ、トイウノハ、モなどでマークされた名詞はトピック-コメント構造を持つので、(5a)のように、トピックにはコロ付で、コメントには行頭に“=”を表記する。設定されたトピックの言及が後まで続くのに新たにトピック・マーカでマークされた名詞が言及された場合は、下位トピックとして行頭を下げて表示する(cf. (5a)の「スクールバス」)。<sup>48</sup> その場合コメントは、対応するトピックと同じ階層に書くことにする(e.g. (5a)の「出発しそうだった」)。同定構造の表記法は(5c)のように表記することにするが、今回のデータでは同定構造は見られなかった。前提構造は表示が困難なため、表示しない。

### 4.2 分析例

例として(6)をあげる。トピックとそれをマークする助詞はゴシック体で表示されており、トピックを導入していると見なされる節には下線が引かれている。“//”は統語的に認定された文末位置を示している。右端に記されている数字は、ゴシック体で示された名詞の指示距離(左)と持続距離(右)である。1行が1つのIUに対応する。

- (6) a. それで  
b. あの: 夜に  
c. あの 食事をこうしてましたら  
d. n=窓のところと言うかベランダに何か

<sup>47</sup> 「その後の談話でトピックとなる指示対象を導入する」を省略して、「トピックを導入する」と言うことにする。

<sup>48</sup> (5a)の表示は、たとえば(ii)のような談話を想定している。

- (ii) a. 子どもたちは学校に行った。  
b. スクールバスは出発しそうだったので、  
c. 走った。

<sup>46</sup> ただし、[5]の方法ではIUの数ではなく節の数を数える。

- e. 小動物<sub>i</sub>がこうちょこちょこと来たんですね// (100, 5)
- f. で最初こう
- g. あの そちらの外の方から  $\phi_i$  覗いたもんですから (2, 4)
- h. 私は最初
- i.  $\phi_i$  リスかなと思ったんです// (2, 3)
- j. で
- k. t=さっところ
- l. あれと思って行ったら
- m. さっと  $\phi_i$  逃げちゃいまして (4, 2)
- n. で
- o. あの: 主人がいや  $\phi_i$  リスだ (2, 1)
- p. って言うんですけども
- q. いやリスにしてはちょっと尻尾の感じが  $\phi_i$  違うなど (2, 0)
- r. 思ってたんですね//
- s. で
- t. あの:
- u. 次の日ゴルフ場に行きましたら
- v. ゴルフ場の その ブッシュと言うか
- w. ま あの 小さいこう下草みたいのが生えてるところで
- x. あの 昨日見たのと同じようn
- y. な感じのする動物<sub>j</sub>が (100, 5)
- z. さささ:って動いたんです//
- aa. で
- ab. それ<sub>j</sub>は何だろうと思って後で聞いたら (3, 4)
- ac. あの  $\phi_j$  マングース (1, 3)
- ad. らしいんですね//
- ae. でマングース<sub>k</sub> っていうのは あの (3, 10)
- af. ま あたしはあの フェレット<sub>l</sub> って最近 あのイタチの仲間 (100, 1)
- ag. 結構  $\phi_l$  ペットにしている人がいますけど (1, 0)
- ah. あ
- ai.  $\phi_j$  イタチかなと思ったら (6, 2)
- aj.  $\phi_j$  マングースで (1, 1)
- ak. でマングース<sub>k</sub> も あの イタチ (1, 7)
- al. とか ま きつとそういう あの
- am. テンの仲間だと思うんですが
- an. あの
- ao.  $\phi_j$  マングース (5, 0)
- ap. だそうです//

(6) の談話の構造を表示すると (7) のようになる:

(7) @ 小動物<sub>i</sub> が来た//

小動物<sub>i</sub>:  
= 覗いた  
・ 私:  
= リスかなと思った//  
= 逃げた  
= 主人がリスだと言う

= リスにしては尻尾の感じが違う//

@ 昨日見たのと同じような感じのする動物<sub>j</sub> が動いた//

同じような感じのする動物<sub>j</sub>:

= マングースらしい//  
・ マングース:  
・ フェレット:  
= イタチの仲間  
= ペットにしている人がある  
= マングース  
・ マングース:  
= イタチとかテンの仲間だと思う  
= マングースだそう//

場所格をとる自動詞で文フォーカスのものは新しいトピックを導入していると考えられ ([9]), その後トピックが自動的に設定されると考える。このことは、(6e, y) の「小動物」、「昨日見たのと同じような感じのする動物」の指示距離が 100、持続距離が両方 5 と、高いことから支持される。この指示対象が言及され続ける限り、他の指示対象がトピックとして言及された場合には (トピック・マーカにより言及されても) 下位トピックとして記されていく。この例において、異なるトピックに移行する時には文末候補が現れるという傾向が観察される。これは [2] など指摘されたように、節がいくつも集まってきた文というかたまりにつき 1 つの大きなトピックを表しているという仮説を支持する。

2 つ目の例は (8) である:

- (8) a. で
- b. あのホノルルというのは (100, 2)
- c. ま あの う
- d. 運転免許が四ドルで取れるとかっていうの聞いたことあるんですけども
- e. ま そういうようなところなんです
- f. ま: あの 右折するにも
- g. 左折車線をこう変更するにも
- h. ウィンカーも出さない (100, 0)
- i. 止まるにも何にも出さないようなところで
- j. ちょっと その
- k. 車も込んでますし運転は (12, 0), (100, 2)
- l. なかなか あの:
- m. n=慣れない人には (100, 1)
- n. 難しいと言うか 危ないところがあるんですが
- o. ま
- p. マウイ島もそうですけれどもハワイ島なんていうのはとにかく あの (100, 0), (97, 6)
- q. 道がちょっと あの 町を外れれば
- r. まっすぐで
- s. こう n=何にもないような感じなんですね//
- t. ですから
- u. あの: わりと ま 初心者と言うか
- v. その

- w. 外国での運転に慣れない人でも (100, 0)  
 x. あの: ハワイ島とかの あの  
 y. レンタカーを借りての運転は (100, 0)  
 z. まわりとし易いんじゃないかなという風に  
 aa. 思います//

(8) の分析例は (9) である:

- (9) • (ハワイ島:)  
 ホノルル:  
 = 運転免許が 4 ドルで取れる  
 = そういふようなところ  
 ・ ウィンカー:  
 = 出さない  
 ・ 何にも:  
 = 出さない  
 = ようなところ  
 ・ 車:  
 = 込んでいる  
 ・ 運転:  
 ・ 慣れない人:  
 = 難しい  
 = 危ない  
 マウイ島:  
 = そう  
 • ハワイ島:  
 = 道がちょつと町を外れればまっすぐ  
 何にも:  
 = ないような感じ//  
 外国での運転に慣れない人:  
 ・ レンタカーを借りての運転:  
 = わりとし易いと思う//

(9) の冒頭の「(ハワイ島:)」というのは、「ハワイ島」がこれ以前にトピックとして言及されていたことを示す。

## 5 まとめ

本発表では、トピック・マーカ、トピックを導入する構文、持続距離などを用いて談話の構造を表示する表記法と手順の提案を試みた。本発表と先行研究 ([10, 12]) の分析との関連が明らかでないで、それを明らかにすることは今後の課題とする。また、背景的なことを述べる節と疑問の回答となる節 ([12]) の区別が本発表ではできていない。これも今後の課題とする。また本発表では、場所格を取るタイプの自動詞 (e.g. *be*, *live*, *arrive*) がトピックを導入する傾向にあることを利用したが、同じタイプの自動詞が背景的なことを述べる節で言及されると、必ずしもトピックを導入しない。したがって、どのような文のタイプがトピックを導入するのか、言語学的な研究を行うことが必要である。

## 参考文献

- [1] G. Brown and G. Yule. *Discourse Analysis*. Cambridge

University Press, Cambridge, 1983.

- [2] Patricia Clancy. Referential choice in English and Japanese narrative discourse. In Wallace Chafe, editor, *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, Vol. 3 of *Advances in Discourse Processes*, pp. 127–202. Ablex, New Jersey, 1980.
- [3] John W. Du Bois. Beyond definiteness: the trace of identity in discourse. In Wallace L. Chafe, editor, *The Pear Stories: Cognitive Cultural, and Linguistic Aspects of Narrative Production*, pp. 203–274. Ablex, New Jersey, 1980.
- [4] John W. Du Bois, Stephan Schuetze-Coburn, Susanna Cumming, and Danae Paolino. Outline of discourse transcription. In Edwards and Lampert, editors, *Talking Data: Transcription and Coding in Discourse Research*, pp. 45–89. Lawrence Erlbaum Associates, Hove/London, 1993.
- [5] Talmy Givón. *Topic Continuity in Discourse*. John Benjamins, Amsterdam/Philadelphia, 1983.
- [6] 岩崎勝一. 音声言語コミュニケーション研究のための分析単位: イントネーション単位 (IU). 人工知能学会誌, Vol. 23, No. 1, pp. 105–112, 2008.
- [7] Susumu Kuno. *The Structure of the Japanese Language*. MIT Press, Cambridge/London, 1973.
- [8] Sige-Yuki Kuroda. *Generative Grammatical Studies in the Japanese Language*. Garland Publishing, New York/London, 1979.
- [9] Knud Lambrecht. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, 1994.
- [10] W. C. Mann and Sandra A. Thompson. Rhetorical structure theory: toward a functional theory of text organization. *Text*, Vol. 8, No. 3, pp. 243–281, 1988.
- [11] Natsuko Nakagawa, Daisuke Yokomori, and Yoshihiko Asao. The relationships between phonological size and the role of intonation units in Japanese discourse. In *2nd UK Post Graduate Conference in Cognitive Linguistics at Brighton*, 2008.
- [12] 高梨克也・竹内和広・森本郁代・仲本康一郎・井佐原均. 談話を語る/聞く動機とエピソード構造. 日本語用論学会第6回大会, pp. 76–79, 2003.