

## 接続関係に基づいた物語文の構造解析

荻澤 義昭 乾 伸雄 小谷 善行 西村 恕彦  
(東京農工大学 工学部 電子情報工学科)

本研究では、物語文の解析を接続関係解析について述べる。物語文においては、登場人物の行動や、場面の説明などが連続的に述べられている。それゆえに、物語文の理解には、これらがどのように連続しているかということを理解する必要がある。本研究では、物語文の接続関係解析を、表記的な情報を用いて行った。まず、物語中のそれぞれの文が持つ機能を文タイプとして、表現した。この文タイプは、行動、説明、心情、情景の四種類である。そして、この文タイプと接続する文の主語のつながりなどの情報を利用して、物語文の接続関係を解析する。この解析のために、実際の物語を対象に調査を行い、それぞれの接続関係決定に有効な情報を抽出した。この結果を計算機上で実現することで、その妥当性を検証する。

## Analyzing Stories based on Relation between Two Sentences

Yoshiaki OGISAWA, Nobuo INUI, Yoshiyuki KOTANI, Hirohiko NISIMURA  
(Tokyo University of Agric. and Tech., Dept. of Computer Science)

In this paper, we discuss the relations between two sentences and the whole structure of stories. As a story sequentially expresses acts of a character or descriptions of scenes, to recognize the relations is necessary to understand the structure of the story. Therefore, we analyze the relation between two sentences using literal informations. First, we classify the sentences of stories into four types by the function of the sentence: act, description, feeling, scene. Using these types of sentences and results of surface analysis, for examples connections of words such as subject, we construct the rules which extract the relations between two sentences. To do this analysis, we find effective informations to decide rules by investigations of stories. We verify our method by means of implementing the system and applying it to several stories.

### 1 はじめに

談話処理の理解では、それぞれの文の内容の理解とともに、それらの文どうしによって作られている構造を理解する必要がある。この文章構造を理解する手法には、論説文を対象として、書き手の主張という観点からそれぞれの文を分類し、それらのつながりから接続関係を与え、構造化を行っているものやまた、表層に現れている接続詞などの修辭的な関

係に着目し、文章の構造をツリーで表現することで分析を行っているものもある。[4][5][6]

これらの論説文に対し、物語文においては同様の手法を適用するのは難しいとされる。これは、物語文は、論説文などに比べ、文章のつながりに論理性がないからである [2]。

それゆえ、物語文の構造解析では、場面を想定したスクリプトなどの手法などが数多くあり、これに

より、物語における事態の連鎖を理解する [11]。しかし、これは、文章の深層情報や、場面に応じたスクリプトを用意する必要が生じる。

本稿では、物語文中における表記的な情報から、物語文における接続関係を解析し、物語文の構造解析を行う。論説文は、文章に論理的なつながりがあるのに対し、物語文では、登場人物の行動、心情、情景の描写などが述べられていると考えられる。そこで、物語のそれぞれの文が持つ機能を、行動、説明、心情、情景の四つのタイプに分類し、これらの分類を接続関係解析の手がかりとした。この文のタイプのつながりとそれぞれの接続関係が持つと考えられる表記的な特徴の規則化を行った。そして、それらの規則を計算機上で実現し、妥当性を考察する。

## 2 接続関係と文タイプ

物語においては、論説文などのように、論理的な構造を持つものではなく、登場人物の行動の連続を表現した部分や、場面を説明した部分などが組み合わさっていると考えられる。

それゆえ、表記的な情報だけを用いて接続関係を決定することを考えた場合、これらの物語文としての特徴を表現し難い。そこで、これらの特徴を表す最小単位として、物語のそれぞれの文を位置づける。つまり、それぞれの文が表している内容と接続関係を関連づけるために、それぞれの文を、説明、行動、心情、情景の四つの文タイプに分類する。

### 2.1 接続関係

文と文の間の関係を定める接続関係の設定には、さまざまな種類が挙げられている。ただし、それらの差というものは、どれだけ関係を細分化するかということに他ならない。そこで、本研究で設定する接続関係は、文献 [1] や文献 [6] を参考にして大まかに設定し、予備調査を行い選択した。以上のようにして、設定した接続関係は表 1 のとおりである。

表 1 設定した関係名

	関係名	説明
展開型	展開	前文の内容を後文で展開させる
	結果	前文の結果が後文で述べられている
逆接型	逆接	前文の内容と反対のことが述べられている
累加型	累加	前文の内容に後文の内容をつけ加える
	並列	前文とが並列的に述べられている
同格型	例示	前文の内容を後文で例示したりする
	同格	前文の内容を言い換えている
	反復	前文の内容を繰り返されている
根拠型	根拠	前文の内容に後文で補足をする。
対比型	対比	前文と後文が対比されている。
転換型	転換	話題や場面を転じる。
呼応型	呼応	前文と後文が呼応している。

### 2.2 文タイプ

物語文は、ある連続した事態が表されたものと考えることができる。例えば、ある状況の推移や事件の結果、それにより生じる登場人物の行動などが羅列されている。これらのことを理解するための情報は、それぞれの文が担っていると考えられる。

そこで、このような観点から、物語を構成している文を分類し、文タイプの情報として、用いる。この文タイプには、次の四つの種類を設定した。

**説明:** その場面の状況や状態などを記述した文

**行動:** 登場人物の行動を表した文

**心情:** 登場人物の気持ちを表した文

**情景:** 情景を描写した文

これらの分類にも、表記情報を用いる。この文タイプの決定は、基本的に次の部分に着目する。

**主部:** 主格にある語の意味属性に従う。

**述部:** 述部が動作を表しているかどうか。それが、「思う」などの心理動詞かどうか。

**副詞:** 文中に、量を表す副詞「よく」や、「毎朝」などの語が使われているかどうか。

例えば、「太郎はよくお酒を飲んだ。」では、

主部の「太郎」が人の属性を持ち、述部が「飲む」という動作を表すが、副詞「よく」が使われているために、この文は習慣的なことを表していると判断する。

従って、文タイプは説明に分類する。

## 3 事前調査

それぞれの接続関係が持つ表記的な特徴や、接続関係と文タイプとの間の特徴を規則化するために、調査を行った。

### 3.1 調査

調査は小学校国語の教科書 [12][13][14] の物語文 10 編 260 文について行った。ただし、本研究では、会話部は扱っていないので、会話や対話が少ない部分を抽出した。その結果、239 の関係が調査対象である。

得られた結果を文タイプごとに分類したものを表 2 に示す。なお、表中には、関係名、文タイプの頭文字だけを示す。

表に記載してない文タイプと関係は、

文タイプでは、心情-情景 (展開) 情景-心情 (結果) が 1 関係ずつ現れ、行動-情景と情景-行動は現れていない。また、関係では、反復が行動-行動で、呼応が心情-心情で一つずつ現れている。

表 2 調査結果

前-後	展	結	逆	累	並	例	同	根	対	転
説-説	18	4	5	12	4	3	6	7	1	4
説-行	8	2	4	17	-	-	-	-	-	3
説-心	3	-	5	6	-	-	-	1	-	-
説-情	-	-	-	2	-	-	-	-	-	-
行-説	2	9	1	7	-	-	1	2	2	2
行-行	3	10	4	20	2	2	1	-	2	4
行-心	-	3	1	-	-	-	-	-	-	1
心-説	-	-	2	6	-	-	-	4	-	1
心-行	3	1	2	-	-	-	-	-	-	2
心-心	-	-	-	2	1	1	3	-	2	1
情-説	-	-	-	1	-	-	1	-	-	1
情-情	1	-	1	2	-	-	2	-	-	-
合計	39	30	25	75	7	6	13	14	7	19

### 3.2 観察された特徴

この調査から、次のような特徴が観察された。

#### 3.2.1 展開語を多く含む関係

展開語とは、接続詞を始めとする、明示的に関係を表す表記情報である。接続詞の他にも、“～だからです。”といった表現や、“ある晩～”といった表現も展開語に含まれると考えられる。

これらの展開語を数多く含んでいる関係は、接続関係の決定がその展開語に依存している率が高いと考えられる。これらの展開語を含んでいる率が高かった関係を次の表 3 に示す。

表 3 展開語を含む割合が多かった関係

	結果	逆説	根拠	転換
関係の数	16	23	11	10
割合	(50%)	(92%)	(78%)	(52.8%)

#### 3.2.2 接続する文タイプに制限がある関係

物語は論説文に比べ、展開方法に制限が少ないと考えられる。この制限が少ない事が、物語の構造化を難しくしている一つの要因である。次の関係では、接続する文タイプに制限があった。

- 前文と後文の文タイプが同一である関係
  - ・ 並列、例示、( 対比 )
- 後文の文タイプが説明である関係
  - ・ 根拠

前文と後文の文タイプが同一である関係では、後文で、前文の内容と類似したことを述べたり、例を挙げたりするので、文中の単語は異なっても、文タイプまでは、変わらないと判断できる。同様に、「根拠」

では、前文の理由付けが後文で行われるために、説明の形になると考えられる。

また、文の構成までみてみると、並列の関係では、7 関係中 5 の関係で、述部の動詞が、同一、あるいは、文献 [9] での同一の分類項目に含まれる。

#### 3.2.3 展開と累加の特徴

先に挙げたような点では、数多く現れている展開と累加の特徴がない。そこで、「展開」の特徴として、「語のつながり」が挙げられるために、「展開」と「累加」での語のつながりを調べた。その結果を表 4 に示す。

表 4 「展開」と「累加」における語のつながり

	同一の主語である	類似した語がある
展開	17.4%	61.5%
累加	41.3%	9.3%

「同一の主語」は、それぞれの文で出現している主語が同じであった割合を示し「類似した語」というのは、後文の主語、または、それに類似した語が前文にあった割合である。

なお、“類似している”という基準は、同一の語、または、文献 [9] で同一の分類項目、または、共通の分類項目までのそれぞれの距離のうち、最小が 1 であるものとした。

これらの観察された関係の特徴をふまえた上で、次節でそれぞれの関係の規則化を行う。

## 4 接続関係解析のための規則化

### 4.1 規則化を行うための情報

接続関係の規則化では、何をその情報として取り扱うかが問題となる。文献 [7] では、次のような情報を挙げている。

- ・ 接続表現 ( 展開語 )
- ・ 文の種類 ( 平叙文、疑問文、命令文 )
- ・ 文の意味内容の型 ( 命題、事象、状態、… )
- ・ 主題要素間の関係
- ・ 格要素間の関係
- ・ 文の意味内容間の関係

本研究でも、ほぼこれらと同様の情報を取り扱う。ただし、文の種類については、考慮にいれず、文の意味内容の型については、文タイプを用いる。さらに、次についても接続関係決定の情報として扱う。

- ・ 述部に使われている動詞の類似性
- ・ 述部表現の類似性

また、接続表現は、次の語で表される。

- 接続表現 ( 接続詞など )
- 関係表現 ( “～だからです” など )
- 指示詞
- 特定の単語 ( “ある晩” など )

これらは、接続関係を決定するのに、非常に有効である。ただし、指示詞の扱いについては、照応問題解決の範囲でもあるために、取り扱わない。この指示詞を除いた接続表現を、本研究では、展開語と呼ぶ。

## 4.2 文タイプの規則化

ここで先に述べた文タイプの規則を詳しく述べる。主語には、「人称」、「情景」、「その他」を設定した。「人称」は、固有名詞と「父」などの人の属性を持つと考えられる語に割り当て、「情景」は、対象が物語であるから、街の情景や、自然の風景を記述できる語を抽出した。

述部は、動詞を動作、状態動詞と心理動詞の二つに分類する。これは、IPAL 動詞辞書 [10] の意味素性に基づく。テンスやアスペクトなどの情報も、有効と考えられるが、物語では表現技法の一つとして、意識的に用いられることがある。そのため、本研究では、テンス、アスペクト情報は取り入れていない。ただし、文タイプが情景に分類される文は、語尾がテイル形であるものが、8 文中 7 だったので、情景の判断にはこれを利用する。

「よく」といった、習慣的なことを表す語は、文獻 [9] から抽出して、利用する。

これらをまとめると次の表 5 のようになる。

表 5 文タイプの決定規則

主語	動詞	テイル	副詞	文タイプ
人称	動作	-	×	行動
人称	動作	-	○	説明
人称	心理	-	×	心情
人称	心理	-	○	説明
人称	×	-	-	説明
その他	-	-	-	説明
情景	動作	×	-	説明
情景	動作	○	-	情景
情景	×	-	-	情景

## 4.3 関係の規則化

事前調査からの特徴と、4.2 節で挙げた情報を加味して、接続関係の規則化を行った。

それぞれの接続関係には、先に挙げたすべての情報が使われるわけではない。それぞれの関係に有用であると考えられるものを使用して規則化を行った。適用した規則の大まかな内訳を表 6 に示す。

表 6 それぞれの関係に適用される規則

制限	A	B	C	D	E	F	G
展開	-	○	○	-	-	-	○
結果	-	○	○	-	○	-	-
逆接	-	○	○	-	-	-	-
累加	-	○	○	-	-	-	○
並列	○	○	○	-	○	○	○
例示	○	○	○	-	○	○	○
同格	○	○	○	-	○	○	-
反復	○	○	○	-	○	○	-
根拠	○	○	○	○	○	○	-
対比	-	○	-	○	○	-	○
転換	-	-	○	○	○	○	-
呼応	○	-	-	○	○	○	-

- A: 文タイプの制限
- B: 主語どうしの類似に関する規則
- C: 主語の出現に関する規則
- D: 後文の主語と前文中の関係に関する規則
- E: 述部の類似性に関する規則
- F: 格要素間の関係に関する規則
- G: 助詞の出現に関する規則

これらは、適用される規則を大まかに分けたもので、関係内の文タイプのつながりごとに変わるものもある。例えば、展開 ( 説明 - 行動 ) と 展開 ( 説明 - 説明 ) では、異なる主語の関係に関する規則が用いられる。

一方、展開語が含まれているかどうかは、すべての関係でチェックを行う。それぞれの関係には、それに対応する展開語が設定してある。

逆接 しかし、けれども、ところが、…

転換 ところで、ある晩、…

結果 すると、～からだ、～のためだ、…

…

### A: 文タイプのつながりに制限があるもの

関係の中には、文タイプのつながりに制限を設けたものもある。一例を挙げると、

- 前文と後文の文タイプが等しい
  - ・ 並列、例示、反復
- 後文の文タイプが「説明」である
  - ・ 根拠

### B,C,D: 前文と後文の主語の関係

表記的な情報で接続関係を決定する場合には、主語のつながりは重要である。本研究では、主語が省略されている場合は、その補完は行わない。そこで、この省略をふまえた規則が必要となる。

前文と後文の主語が現れている場合は、その類似性を調べる。

ここで、類似性は次の三段階に分類される。

- 同一の語が使われている。
- 表記は異なるが、類似性が高い。

- 類似性が低い。

ここで、類似性が高いというのは、文献 [9] を用いて行い、共通の親までの距離が短いほど、類似性が高いとした。これは、他の規則でも同様である。

例えば、

規則 1a: 二つの文の主語が出現している場合、それらの語の類似性は高い。

規則 1b: 二つの文の主語が出現している場合、それらの語の類似性は低い。

これらが適応する関係は、

規則 1a 対比, 並列, 反復, 結果 (心情-行動) など  
規則 1b 展開, 根拠, 例示, 結果 (行動-説明) など

また、一方の主語が省略されている場合がある。例えば、結果 (行動-説明) では次のような規則がある。

規則 2a: 展開語を含む場合は、二つの文で省略される

規則 2b: 展開語を含まない場合は、前文だけ省略される

また、展開では、

規則 4: 後文の主語が、前文に出現している。

逆に、累加では、

規則 5: 後文の主語が、前文に出現していない。

などの調査結果からの規則がある。

E: 述部に類似性がある関係

述部に特徴が見られる関係がある。例えば、

規則 6: 述部の動詞が類似している

この規則を適応するのは、例示 (行動-行動など)、同格 (心情-心情)、並列、対比である。それ以外の関係に対しては、類似性は見られない。

F: 格要素間の類似性

それぞれの格要素が類似するもの、文中に存在する格が類似するものがある。並列では、ある内容を並べて記述するために、この規則を適応する。

G: 助詞の出現

関係によっては、出現する助詞に特徴がある。例を挙げると、

規則 7a: 前文、後文で助詞「は」を伴う要素がある

規則 7b: 後文で、助詞「も」を伴う要素がある

規則 7a は対比に、規則 7b は並列に適応される。

関係には、このような規則がそれぞれいくつか用意されている。そして、その適応した規則の数 (適応度) で関係を決定する。逆接のように、ほぼ展開語で決まる関係は、展開語への依存度を高くした。展開語が含まれていない場合は、それぞれの関係における規則の重要度を考えて設定した。

## 5 物語文の構造解析システム

4 節に基づいた接続関係解析を中心としたシステムの試作を作成し、実験を行った。入力データは、形態素解析されたタグ付きの文とし、入力データから、その文の文タイプを決定し、得られた文タイプと表記情報を用いて、接続関係を決定する。

それぞれの規則の適応具合により、その適応度が四段階に分類される。すべての関係に、適応度を調べ、最も高い関係が選択される。

### 5.1 実験

実験は、調査した結果がどれほど規則として判定されているかを調べるために、調査で使った物語を解析した。システムは単文処理なので、元データは一部改変した。ただし、省略語の補完などは行っていない。

まず、文タイプに関しての結果を表 7 に示す。

表 7 文タイプ分類結果

文タイプ	元データ数	一致した数	一致率 (%)
説明	124	94	75.8
行動	89	84	94.4
心情	35	25	71.4
情景	8	4	50.0
合計	256	207	80.6

全体での一致率は約 80 % である。

次に、接続関係の結果を示す。表 8 は、調査のために与えた接続関係と規則から判定された接続関係について示した。逆接の関係については、ほぼ展開語に依存するために表には示していない。

表 8 に示すように、本システムでは、同格の関係を決定することができなかった。これは、文の内容推論を必要とすることが多いためであると考えられる。それには、さらに文の内容を制約する規則が必要である。

さらに、表 7 で示したように、本システムでの文タイプ判定率は、約 80 % 程度である。そこで、表 8 の結果と、文タイプの判定がどのような関係であるかを表 9 に示す。

表 8 接続関係の結果

関係	与えた数	一致した数	一致率
展開	40	13	32.5 %
結果	39	27	69.2 %
累加	70	47	67.1 %
並列	5	3	60.0 %
例示	5	2	40.0 %
同格	9	0	0 %
反復	1	1	100 %
根拠	12	7	58.3 %
対比	9	7	77.8 %
転換	20	8	40.0 %
呼応	1	1	100 %
合計	211	114	54.0 %

表 9 接続関係と文タイプの関係

関係	関係一致		関係不一致	
	文タイプ一致	文タイプ不一致	文タイプ一致	文タイプ不一致
展開	13	0	13	14
結果	25	2	9	3
累加	42	5	14	9
並列	3	0	1	1
例示	2	0	2	1
同格	0	0	3	6
反復	1	0	0	0
根拠	6	1	3	2
対比	6	1	1	1
転換	5	3	6	6
呼応	1	0	0	0

このように、正しく接続関係が決定されたときは、文タイプの正しく決定されていると考えることができる。また、文タイプの決定に失敗している、関係が正しく決定されているときは、文タイプによる制限が緩やかか、あるいは、展開語によって決定されているとみることが出来る。

これらの表 8,9 から文タイプの判定が、接続関係の一致率に影響を及ぼしていることがわかる。

以上のように、本システムでは、表記情報による文タイプと主語のつながりなどの表記的な情報を用いることで、約 50 % の接続関係を正しく解析した。

## 5.2 考察

本研究では、用いる情報は、文タイプ情報とその他の表記情報である。この表記情報には、接続関係を直接的に示す展開語も含まれている。展開語が含まれている場合、文タイプなどの他の情報によらず、関係が決定する事が多い。本実験での展開語の有無と接続関係の一致率は表 10 のとおりである。

表 10 展開語と接続関係と文タイプの関係

関係	展開語あり		展開語なし	
	文タイプ一致	文タイプ不一致	文タイプ一致	文タイプ不一致
一致	54	9	66	15
不一致	5	5	50	47

この表より、展開語が存在しなくても、文タイプの情報が正しければ、約 57 % の割合で正しい接続関係が決定できることになる。それゆえに、物語文解析における接続関係の決定のために、本研究で設定した規則が有用であると考えられる。

また、設定した規則は、対比や並列など、表記的な特徴を持つ関係などにおいては、他の関係の規則より制限がきつくなっている。そのために、これらの関係を満たす文の解析を行うと複数の関係が挙げられることがある。それに対して、出現頻度が高いと考えられる累加などの関係においては、複数の関係が挙げられることが少ない。本実験においても、累加の関係では、正しく判定された 47 関係中 40 の関係 (約 85 %) が一意に決定されている。これを落とすことなく、複数候補を落とす規則を構築していくことが精度を上げるためには必要である。

## 6 おわりに

本稿では、物語文解析における接続関係の決定について述べた。表記情報を利用して、文が持つ機能を、四種類の文タイプに分類する。さらに、これらの情報と文の主語のつながりなどの表記情報を用いて、接続関係を規則化した。この規則を用いて実験を行い、表記情報だけで、約 50 % の接続関係が正しく決定できた。

## 参考文献

- [1] 永野賢：文章論総説，朝倉書店，1986。
- [2] 所一哉：日本語 思考のレトリック，匠出版，1986。
- [3] 長尾真：自然言語処理，岩波講座ソフトウェア科学 15，岩波書店，1996。
- [4] 小野頼司，浮田輝彦，天野真家：文脈構造の分析，情報処理学会自然言語処理研究会研究報告，NL70-2，pp.1-8，1989。
- [5] 福本淳一：筆者の主張に基づく日本語文章の構造化，情報処理学会自然言語処理研究会研究報告，NL78-15，pp.113-119，1990。
- [6] 福本淳一，安原宏：日本語文章の構造化解析，情報処理学会自然言語処理研究会研究報告，NL85-1，pp.81-88，1991。
- [7] 桃内佳雄：文章における接続関係の解析のための基礎的考察，情報処理学会自然言語処理研究会研究報告，NL78-13，pp.97-104，1990。
- [8] 小方孝，掘浩一，大須賀節雄，物語のための技法と戦略に基づく物語の概念構造生成の基本的フレームワーク，人工知能学会誌，Vol.11 No 1，pp.148-159，1996。
- [9] 国立国語研究所：分類語彙表，秀英出版，1964。
- [10] 情報処理振興事業協会技術センター：計算機用日本語基本動詞辞書 IPAL(Basic Verbs) - 解説編 -，情報処理振興事業協会，1987。
- [11] J. グリーン：言語理解，長町三生監修，認知科学研究会訳，認知心理学講座 4，海文堂，1992。
- [12] 小学校国語 4 上，日本書籍株式会社，1985。
- [13] 小学校国語 4 下，日本書籍株式会社，1985。
- [14] 小学校国語 5 上，日本書籍株式会社，1986。