

文章構造化のための文の接続関係の解析

斎藤葉子¹ 柴田昌宏¹ 福本淳一¹¹

¹ (株) 沖テクノシステムズ ラボラトリ ¹¹ 沖電気工業 (株)

1 はじめに

現在、我々は論説文等の文章の構造化を目的として、新聞社説記事中の各文の接続関係の解析を行っている[1]。一般に、文の接続関係は接続詞等の語句により明示的に決定されるが、新聞社説記事において接続詞を含む文は少なく、その他の要因により文の接続関係が決まると考えられる。これまでに我々は、文末表現による文タイプの分類、主題提示情報を用いることで文章の構造化を行ってきた[1]。

本稿では、前文と後文の結び付きを決定する要因として、後文に現れる指示語や助詞の使われ方、同語反復の3つに注目し、これらの情報が文の接続関係にどうかかわっているかについて解析を行った。

2 文の接続関係の分類

文の接続関係を分類したものととして、永野[2]や市川[3]によるものがある。永野は接続関係を9個に、また、市川は大分類として8個、さらに小分類として31個に分類している。

我々はこれらをもとに、図1のような2文間の接続関係を設定した。そして、この関係に基づき昭和62年8月～10月の朝日新聞朝刊社説記事40編を用いて、2文間の接続関係について1173件のデータを抽出した。このとき、隣接した2文間だけではなく、離れた2文間に対して接続関係を認めた場合にも、データに含めた。また、接続関係の選択は人手により行った。

背景	累加	対立	根拠	逆接
前提	序列	転換	制約	比較
結果	追加	反復	補充	限定
目的	並列	例示	連係	呼応

図1 文の接続関係

3 文の接続関係の調査

2文間(前に現れる文を前文、後ろに現れる文を後文とする)の接続関係を決定する要因としては、さまざまなものが考えられるが、今回は、特に、後文に現れる助詞「も」や指示語、同語反復と文の接続関係について注目した。調査対象である接続関係のデータのうち、接続詞等の機能語によって接続関係が決定されているものは除外した。これは、これらの語句が接続関係の決定に影響を与えているためである。

これらのデータについて、以下のように調査を行った。

1. 後文に現れる助詞「も」

後文に助詞「も」が現れるものについて文の接続関係との対応を分類したが、特徴は出なかった。そこで、文の接続関係に影響を与える助詞「も」を抜き出すために、助詞「も」を含む文節情報の違いと、2文間に影響を与えるものとそうでないものとの対応を調査した。

表1 文節情報と2文間への影響 (数値: データ数)

表層情報	影響あり	影響なし
(形式名詞)(に)も	16	3
(名詞)も	42	36
(名詞)にも	13	15
(名詞)でも	2	30
(動詞)でも	2	26
についても	2	8
とも	2	8
よりも	0	6

表1より「(形式名詞)(に)も」という表現をとる場合、2文間に影響を与えることが多い。また、文節情報だけでは区別できない「(名詞)も」「(名詞)にも」を分類する方法として、後文の構文情報(単文であるか、重文であるか)に着目して2文間に影響を与えるものとそうでないものとの対応を調査した。

表2 構文情報と2文間への影響 (数値: データ数)

構文情報	影響あり	影響なし
単文	43	14
重文(前部に「も」)	7	14
重文(後部に「も」)	5	23

表2より単文である場合、2文間に影響を与えることが多い。

これらの分析より、(1)後文に「(形式名詞)(に)も」を持つ(2)後文に助詞「にも」または「も」を持ち、単文である場合について文の接続関係との対応を調査した。

2. 後文に現れる指示語

指示語については、後文において指示語の現れる位置を文の先頭か、そうでないかに分けて調査した。さらに詳細な分類として、「これ・それ」「この・その」「こう・そう」という指示語の種類による違いと文の接続関係との対応を調査した。

3. 同語反復

同語反復については、前文と後文の結び付きがもっとも強いと思われる(複合)名詞が完全に一致する(完全一致)場合と、(複合)名詞の一部が一致する(部分一致)場合とに分類した[4]。部分一致については、さらに詳細に分類を行った。

- A. 上位-下位(上下)(例)教育-教育水準
- B. 部分-部分(部分)(例)石油危機-石油供給
- C. 下位-上位(下上)(例)税制改革-税制

これらの分類による違いと文の接続関係との対応を調査した。

Analysis of Cohesion for Text Structure

¹Yoko SAITO, Masahiro SHIBATA

¹¹Jun-ichi FUKUMOTO

¹Oki Technosystems Laboratory, Inc

¹¹Oki Electric Industry Co., Ltd.

*本研究は、第5世代コンピュータプロジェクトの一環として
I-COTからの委託で行われたものである。

4 調査結果¹

1. 後文に現れる助詞「も」

表3に、(1) 後文に「(形式名詞)(に)も」を持つ(2) 後文に助詞「にも」または「も」を持ち、単文である場合について、文の接続関係ごとのデータ数を示す。「(形式名詞)(に)も」については累加・連係が合わせて47%、単文については累加・並列が合わせて49%であった。

表3 文の接続関係ごとのデータ数

接続関係	(形式名詞)(に)も	単文
累加	5	16
並列	2	12
追加	2	7
連係	4	1
転換	1	9
補充	1	7
その他	4	5

2. 後文に現れる指示語

表4に、後文に現れる指示語の位置による文の接続関係ごとの集計を示す。後文の先頭にきた場合には、累加が48%、連係が16%となった。

表4 後文に現れる指示語の位置と文の接続関係

接続関係	先頭	先頭以外
累加	61	32
連係	21	12
転換	13	27
背景	13	6
補充	7	10
その他	12	10

表5に、指示語が後文の先頭にくる場合について、指示語の種類による文の接続関係ごとの集計を示す。「これ・それ」「この・その」という表現では共に累加が43%で、「こう・そう」という表現では連係が42%となった。

表5 後文の先頭に現れる指示語の種類と文の接続関係

接続関係	これ・それ	この・その	こう・そう
累加	10	26	4
連係	5	6	8
背景	3	7	4
転換	1	9	1
前提	0	4	0
補充	3	4	0
その他	1	5	2

3. 同語反復

表6にあるように、完全一致する場合には、転換の割合が37%、累加が18%であった。部分一致では、「上位-下位」の場合には、転換の割合が28%、累加が36%であった。「下位-上位」の場合には、転換の割合が24%、累加が20%、連係が22%であった。

表6 同語反復と文の接続関係

接続関係	完全一致	部分一致		
		上下	部分	下上
転換	42	15	16	11
累加	21	19	15	9
補充	13	7	12	3
連係	13	4	7	10
その他	25	8	15	13

¹表の数値は、すべてデータ数である。

5 考察

以上の調査結果から、文の接続関係に影響を与える要因として以下のことが考えられる。

- (1) 表3より「(形式名詞)(に)も」という表現が後文にある場合、文の接続関係は累加あるいは連係が目立つ。
- (2) 表3より「も」「にも」という表現が後文にあり、その文が単文である場合、文の接続関係は累加・並列が多い。
- (3) 表4より後文の先頭に指示語がきた場合、累加が多い。また、連係の数も目立つ。
- (4) 表5より「これ」「それ」「この」「その」という指示語の場合、累加になりやすい。また「こう」「そう」という指示語の場合連係が目立つ。
- (5) 表6より同語反復において完全一致の場合、転換が多い。
(複合) 名詞のみの一致では、2文間にあまり強い結び付きはあらわれないといえる。

部分一致の「上位-下位」の場合、文の接続関係は累加と転換が目立つ。同語反復については転換が多いが、その後の調査より後文の直前で段落の切れ目になっていることが多かった(95%)。その情報を使って転換を除くと、部分一致の「上位-下位」は累加になりやすいといえる。

また、調査を行って行く中で次のような要因も得る事ができた。

- (1) 2文間に影響を与えない、いいわし的な表現としては、「どの〜も」「これまで」「それほど」「〜も〜ない」などが多い。
- (2) 「でも」「よりも」「もの」「(数詞)も」という表現では、強調する場面が多い。
- (3) 重文の後部に「も」という表現があった場合は、文中で物事を並べている場面が多い。

6 おわりに

本稿では、文の接続関係を決定する要因として、助詞「も」・指示語・同語反復の3つに注目し、接続関係データを用いて調査を行った。その結果、接続詞のほかに表層情報によって文の接続関係を決定する要因を得る事ができた。

今後の課題としては、複数の接続関係の候補が出たものについて、区別する必要がある。指示語と同語反復などそれぞれの情報が重なったときにどの情報が有効となるかを調べることも必要である。

今回は3点に止まって調査したが、「〜に加えて」「〜と対照的に」などの決まり文句による表層情報からも文の接続関係を決定することができると思われる。

【参考文献】

- [1] 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報告 78-15 (1990)。
- [2] 永野 賢：文章論総説、朝倉書店 (1986)。
- [3] 市川 孝：国語教育のための文章論概説、教育出版 (1978)。
- [4] 田中、柴田、福本：文章構造解析システムにおける同語反復解析処理、情報処理学会第43回全国大会 5H-1 (1991)。