ICOT Technical Memorandom: TM-1099

TM-1099

日本語文章の構造化解析

福本　淳一、安原　宏（沖）

September, 1991

© 1991, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03)3456-3191～5
Telex ICOT J32964

Institute for New Generation Computer Technology

日本語文章の構造化解析

福本 淳一　　安原　宏

沖電気工業(株)総合システム研究所

概要

論説文は、ある事柄についての書き手の考えや意見など書き手の主張を述べることを目的とした文章であり、このような文章においては、書き手の主張を読み手である読者に伝えるために論旨が展開されている。この論旨の展開構造が、文章の構造であり、書き手が文章において主張したいことであると考えられる。本稿では、このような文章の構造化のため、新聞社説記事を用いた文脈現象の解析を行った。その結果を基に文章構造のモデルについて考察した。

Japanese Text Structure Analysis

Jun-ichi Fukumoto, Hiroshi Yasuhara

Systems Laboratory, Oki Electric Industry Co., Ltd.
11-22, Shibaura 4-chome, Minato-ku, Tokyo 108, Japan

Abstract

Japanese texts such as newspaper editorials are written to show writer's opinion and thinking to readers. I call such opinion "Writer's intention". Such kinds of texts are composed centering around the Writer's Insistence. In this paper, we analyzed context phenomena in Japanese newspaper editorials and stored context data from these texts. We discussed text structure model based on this analysis.

1　はじめに

論説文は、ある事柄についての書き手の考えや意見などの書き手の主張を述べることを目的とした文章であり、このような文章においては、書き手の主張を読み手である読者に伝えるために論旨が展開されている。この論旨の展開構造が文章の構造であり、書き手が文章において主張したいことであると考えられる。現在までに筆者らは、論説文として新聞の社説記事をとりあげ、書き手の主張という観点から文章の構造化についての分析を行ってきた[1]。構造化のため、まず、文章中の各文を書き手の主張に基づいてタイプ分類を行い、各文のタイプ情報と主題提示情報、および、文章中に表れる接続詞等の機能語の情報を用いることで文の連接関係を求め、これにより文章を木構造の形式に構造化を行ってきた。文章全体は、話題の転換部分でいくつかの部分構造の分割し、各部文構造においては、その中で中心となる主張文を根とする木構造となっている。

本稿では、この文章構造の詳細化のため、新聞社説記事を用いて文章中に表れる文脈現象として照応現象と文の連接関係についてのデータの抽出を行い、そのデータの解析から文章構造のモデル化のための考察を行った。

以下では、2章で文脈現象のデータとして抽出を行った照応現象データ及び文お連接関係データについて述べた後、3章で文章構造のモデル化について述べる。

2　文間関係の解析

文間の関係を決定する要因としては、名詞語句等の照応関係、指示語による関係、一貫性関係、接続詞等の語句による関係等がある[4][6][7]。ここでは、新聞社説記事を調査対象とし、社説文章中に表れる照応現象、文の連接関係についてのデータの収集を行い、その解析を行った。

2．1　照応現象の調査

照応現象とは、コンテキスト内のある要素をその前後にある文の中において、別の形式で表現する現象のことを言う。あるいは、名詞などの語句が、その前後のコンテキスト内で言及されているモノ・コトを指示参照する現象であるとも言える[2]。我々は、文章中の関連性のある語句の関係が文間の関係の決定にどのような影響を与えているかの調査を目的をして照応現象を調査を行ってきた[2]。そこでは、文章中で関係があると思われるもの語句を広く収集し、これをデータとして蓄積した。照応現象データは、照応語とそれが指示参照するものである先行詞とこれらの間の照応関係について、図1に記入フォーマットを用いて収集した。

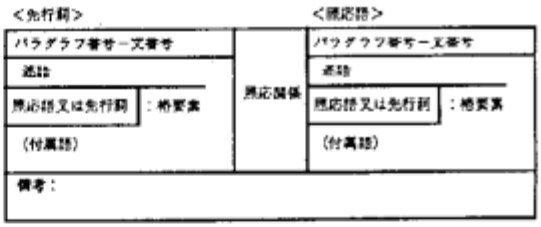

図1　照応現象データ調査フォーマット

調査対象としては、昭和62年 8月から10月までの朝日新聞社説記事70編を用いた。各項目の抽出基準は以下の通りである。

(a) 照応語

- 人称代名詞　「彼」など
- ゼロ代名詞　（明かに省略されていると判断できる主格のものについてのみ調査する）
- 指示代名詞　「これ」「それ」
- 指示連体詞　「この」「その」＋体言
- 指示連用詞　「こう」「そう」＋用言

・サ変動詞の語幹
・名詞、名詞句
・代名詞的表現　「両国」「同氏」など

(b) 先行詞

照応語によって指示参照されている語、句、文、パラグラフを選択した。また、照応語が文章中の内容を指示参照している場合は、その内容を含む文、パラグラフを選んだ。背景的な知識などにより推定されるものについては、対象外とした。

(c) 照応関係

・同一名詞による照応
・上位－下位関係
・部分－全体関係
・指示語による照応
・ゼロ代名詞による照応

(d) 照応の距離

照応関係の存在する文間の距離を測定する。前方照応の場合に正値をとる。

以上の基準に基づき3003件のデータの収集を行った。照応関係別に分類したものを表１に示す。

表１　照応関係分類表

関係名		件数
同一名詞		２１０８
上位－下位関係	上位－下位	７５
	下位－上位	８０
部分－全体関係	部分－全体	５３
	全体－部分	５８
指示語	前方照応	１０
	後方照応	４５３
ゼロ代名詞	前省略	１６６
	後省略	５
合計		３００８

２．２　文の連接関係の調査

文章構造の解析を行うため、照応現象データの収集に用いた同じ社説文章を用いて文章中の任意の文の間に認められる文間関係データを図２の記入フォーマットにしたがって収集した。

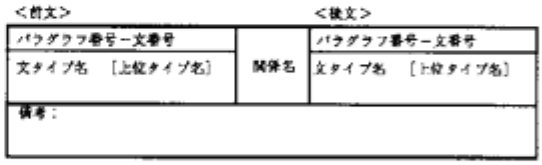

図２　文間関係データ調査フォーマット

調査は、新聞社説記事において文章中の任意の２文を選択し、その間にある設定された文間関係が存在する場合、その前の文を前文、後ろの文を後文とし、その文間に存在する関係を設定した。また、それぞれの文タイプ情報についても設定した。この調査で用いた文タイプ名、関係名は以下のとおりである。

(1) 文タイプ名

これまでに、書き手の主張という観点から文章中の文を主張文と叙述文に分類を行い、主張文を８つに細分類してきた。この主張文の分類に加え、叙述文についても１０に細分類を行ったものを文タイプ情報として用いた。以下にその意味と新聞社説記事中に現れた表現を示す。

○主張文

・問掛文…疑問表現を用いて書き手が読み手に対して問い掛けを行う
　のではあるまいか　　どう～のか
　恐れはないか　　どう～だろうか
　ありはしないか　　あるのだろうか
・断定文…書き手がある事柄について断定を行う
　である　　ことである
　のである　　明らかである

のではない　　のだ

・推量文…書き手が推測で物事を述べる

だろう　　ではあるまい

べきだろう　　とはいえまい

かもしれない　　やむをえまい

・要望文…書き手がある要望を述べる

～たい　　期待する

～したい　　てもらう

てほしい　　てもらいたい

・判断文…書き手がある判断を行う

と考える　　はずだ

はずはない　　といえる

とは言いがたい　ているようだ

・意見文…書き手の意見を述べる

てはならない　　必要がある

望ましい　　必要である

当然だ　　大切だ

・理由文…話題になっている事柄についての原因理由を述べる

からだ　　ているからである

わけだ　　わけではない

ためだ

・義務文…義務的な表現を述べる

なければならない　　べきだ

ねばならない　　べきである

○叙述文

・現在　　「る」で終わる形

・過去　　「た」で終わる形

・可能　　できない　　いえない

・伝聞　　と聞く　　という

・様態　　強い　　厳しい　　多い

・叙述　　名詞＋だ

・存在　　がある　　はない　　にある

・継続　　つつある

・状態　　ている　　てくる　てしまう

・使役　　させる　　せる

(2) 関係名

文間関係の分類としては、永野[4]や市川[5]によるものがある。これらをもとに、次のような文間関係を設定した。

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

以上の記入要領にしたがって朝日新聞朝刊社説記事70編から2145件の連接関係データの収集を行った。データを関係名ごとに分類したものを表２に示す。

表２　文間関係分類表

関係名	件数
背景	１１８
累加	３６２
対立	４２
根拠	７８
前提	５７
序列	１９
転換	４５０
制約	１２
結果	１２
追加	５６
反復	７８
補充	２１８
目的	５
並列	７７
例示	８０
連係	２８３
逆接	１６０
比較	７
限定	１６
呼応	１５
合計	２１４５

収集したデータのうち接続詞等の語句によって決まった関係を除いたものを、主張文－叙述文、叙述文－主張文、主張文－主張文の文間関係についてまとめたものをそれぞれ表５、表６、表７に示す。叙述文－叙述文に関して

は、タイプ間の関係から得られる関係に傾向がなかったため、ここでは除外した。

これらの文間関係データの解析およびこれらの関係の見直しを行うことで、文タイプ情報によって文間の関係を決定することのできるものがいくつか存在した（主張文－主張文について15件、主張文－叙述文について 6件）。その他のものについては文タイプ情報だけでは文間の関係を決定することがでなかった。

２．３　文脈現象データの解析

文の連接関係と照応現象の関連を調査するため、同一文間に照応応現象データと文間関係データが共に存在するものについて調査を行った。調査結果をまとめたものを表３に示す。また、各連接関係での各照応関係の割合を示したものを表４に示す。

表３　文間関係と照応現象との関連表

関係名	全体	指示語	ゼロ代	同一名詞	全体-部分	部分-全体	下位-上位	上位-下位
結果	8	6	1	1	0	0	0	0
目的	2	2	0	0	0	0	0	0
連接	68	23	7	32	0	1	4	1
序列	4	0	1	3	0	0	0	0
追加	15	6	3	5	0	0	0	1
並列	11	1	2	8	0	0	0	0
累加	144	67	15	62	4	1	4	4
対立	13	5	0	7	0	0	0	1
比較	3	1	0	2	0	0	0	0
転換	97	12	1	72	5	2	2	3
限定	3	2	0	1	0	0	0	0
反復	23	9	5	9	0	0	0	0
換言	33	9	4	16	0	0	2	2
転約	2	0	0	1	0	0	1	0
補充	120	33	12	66	3	0	3	3
連体	65	23	8	33	1	1	1	1
呼応	3	1	0	2	0	0	0	0
前提	27	8	3	14	0	0	0	2
背景	52	25	3	20	3	1	0	0
例示	22	4	3	13	0	0	2	0
合計	729	237	66	367	16	6	19	18

（数字は、件数を示す）

部分－全体、全体－部分、上位－下位、下位－上位、ゼロ代名詞の各関係については、データ数が少ないため連接関係を決定することができなかった。指示語については、累加、背景関係が比較的多く表れ、転換関係が少ないという傾向があった。このことから指示語が用いられた場合、その指示語の参照している文に関して話題が変わらず、前文を引き継いだ内容が述べられる傾向がある。同一名詞

表４　各照応現象の割合

関係名	指示語	ゼロ代	同一名詞	全体-部分	部分-全体	下位-上位	上位-下位
結果	75	13	13	0	0	0	0
目的	100	0	0	0	0	0	0
連接	34	10	47	0	1	6	1
序列	0	25	75	0	0	0	0
追加	40	20	33	0	0	0	7
並列	9	18	73	0	0	0	0
累加	47	10	43	3	1	3	3
対立	38	0	54	0	0	0	8
比較	33	0	67	0	0	0	0
転換	12	1	74	5	2	2	3
限定	67	0	33	0	0	0	0
反復	39	22	39	0	0	0	0
換言	27	12	48	0	0	6	6
転約	0	0	50	0	0	50	0
補充	28	10	55	3	0	3	3
連体	35	9	49	2	2	2	2
呼応	33	0	67	0	0	0	0
前提	30	11	52	0	0	0	7
背景	48	4	38	4	2	0	0
例示	18	14	59	0	0	9	0
合計	33	9	50	2	1	3	2

（数字は、割合 "%" を示す）

については、並列、転換関係が多く表れる傾向があった。並列関係は同じ内容が並列的に述べられているという関係であり、このような場合には同じ語句が用いられる傾向があることが分かる。しかし、同じ名詞が使われた場合でも転換関係が多くあらわれたことから、同一名詞語句の反復情報のみでは単純に話題の転換とすることはできなく、その他の情報が必要である。

また、話題の変わる部分である転換関係のみに注目して文の連接関係データの解析を行った。１４社説分の連接関係データについて、各文タイプの連接関係のうち転換関係は以下の割合であった。

主張文－主張文	１７％ (9/53)
主張文－叙述文	５４％ (40/74)
叙述文－主張文	９％ (8/91)
叙述文－叙述文	２２％ (24/111)

このことから、主張文に続く叙述文において話題の転換が行われる傾向が強いことがわかる。また、文においては、「は」等により主題の提示が行われるが、この情報を加えた上で主張文－叙述文について更に分類すると、

○叙述文で主題提示あり 転換関係 27例　　その他 16例 ○叙述文で主題提示なし 転換関係 13例　　その他 17例

となる。したがって、主張文に続く叙述文において主題の提示が行われた場合、話題が転換すると扱うことができる。

3　文章構造のモデル化

これまでに我々は、文の連接関係を一方が主で他方が従であるような主従関係として捉えることで文章の構造化を行ってきた。しかし、実際の文章中には以下のように主従を決定しにくいものとして、文章中で並列関係となる文や連続してある事柄が述べられる文がある。文章中で並列関係になるものとしては、ある文の連接関係が並列関係とされたものや、また「第１に」「第２に」「第３に」のような接続的表現を用いることで、いくつかの文がまとまった形で序列的に並んでいるものもある。（数字はパラグラフ番号 - 文番号を示す）

5-1	第１に、水の節約が、水道料金の値上げにつながらぬようにしなければならない。
8-1	第２に、給水配管の漏水対策を積極的に進めてほしい。
9-1	第３に、最も重要なのは、今回の水不足が、東京集中によってひき起こされた構造的な問題だ、という認識を持つことである。

（昭和62年 8月30日付け朝日新聞社説記事より）

また、文間関係に主従関係がなく連続しているものとしては、以下のものがある。

1-1	ジョギング中の突然死や社長の急死などの不幸な事件が、このところ目立っている。
1-2	産業構造が変わり技術革新が進んで、働く人のストレスもつのってきた。
1-3	高齢化への歩みが速まるなかで、働き盛りの中高年の健康管理が特に重要な問題になっている。

（昭和62年10月 2日付け朝日新聞社説記事より）

この例では、いくつかの事柄が順に列挙されており、文間には直接は主従の関係に当るものは考えられない。

文の連接関係の中で主従関係は決定できないが、なんらかの関係があるものについてはそれらをグループとしてまとめて表現することで扱うことのできる。

また、従来まで部分構造として表現されていたものもこのグループ化によって表現することができる。これは、文章中の文の連接関係を主従関係として表現された木構造において、同じ内容について述べられている部分をグループ化することで表すことができる。このグループ化を部分構造内部にも取り入れることで、ある内容について述べられている部分がまとめられるため、木構造中の係り受けのスコープが表現される。

以上のように文章構造を表現するためには、各ノード間の連接関係を主従関係と認識することで文章を木構造の形に構造化することができる。その際、同じ話題について述べられているものはグループとしてまとめることでノード間の係り受け関係のスコープを表現することが可能である。また、文間に連接関係は認められないが、話題の変化がないものや並列関係にあるものも同様にグループとしてまとめることができる。このようにして、文章の構造をネットワーク構造として表現することができる。

4　おわりに

本稿では、論説文などの文章を書き手の主張という観点から構造化するため、新聞社説記事を用いて文章中に表れる照応現象及び文の連接関係をデータとして収集し、その解析を行った。この解析から文の連接関係を決定する為のいくつかの規則性を見つけることができた。また、これらの調査を基にしてこれまでの木構造形式の文章構造をグループ化を用いることでネットワーク構造への拡張について考察した。

今後の課題としては、このような構造化のための文章の構造化のためのモデルおよびそのモデルに基づく言語仕様を定義し、この言語仕様で記述した規則を用いて文章構造の解析を行うことがある。そのための言語処理系の実現も必要である。また、文章構造を解析する規則を記述するため、文脈現象データの更に詳細な解析も必要である。

【謝辞】

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ICOTからの委託で行われたものであり、研究を進めるにあたり有益な助言を下さいましたICOT第６研究室の田中室長に感謝いたします。また、研究協力頂いた沖電気工業(株)の甲斐氏、(株)沖テクノシステムズラボラトリの柴田、田中、斎藤各氏に感謝いたします。

【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 78-15, 1990.

[2] 柴田, 田中, 福本："新聞社説記事における照応現象",情報処理学会第40回全国大会, 1990.

[3] 上野, 正保, 田中, 菱沼, 日向："日本語と外国語との照応現象に関する対照研究", 国立国語研究所報告79, 1984.

[4] 永野賢："文章論総説－文法論的考察－",朝倉書店, 1986.

[5] 市川孝："国語教育のための文章論概説",教育出版, 1978.

[6] 木下, 小野, 浮田, 天野："日本語テキスト理解における文脈構造抽出法",「談話理解モデルとその応用」シンポジウム, pp.125-136, 1989.

[7] B.Grosz, C.L.Sidner："The Structures of Discourse Structure", Technical Report, CSLI, CSLI-85-39, 1985.

後＼前	可能	保留	推量	叙述	比況	存在	連続	状態	使役	現在	過去
関係文		呼応1	転換1 真加1	転換2 真加1 逆接1 補足1 呼応1		転換1 真加1 補足1		転換4 逆接1 呼応1 規範1		転換1 真加1 逆接4 補足1 呼応1 規範1	逆接1 転換1 真加1
断定文	真加1 逆接1	転換2 真加1 補足1	転換1 真加4 規範1 例示1	転換4 背景2 真加1 逆接1 並列1 補足5 追加1		転換2 真加1 逆接1 補足1 追加1		転換4 背景2 真加4 逆接1 並列1 補足4 例示1 追加1		転換9 真加1 補足3 規範4 例示1	転換9 真加1 補足4 例示2 追加3
推量文	並列1	転換1 並列2	真加1 規範1	転換3 真加1 逆接1 補足1 例示1		転換1 背景1 追加2		転換3 真加1 補足1 例示1	転換1	転換7 真加5 逆接1 補足2 規範1 例示4	転換4 補足1 規範1
要望文			転換3 規範1	転換5 逆接1 反復3		転換2 背景1 補足1		転換9 補足1		転換4 逆接1 並列1	転換9
判断文	転換1 逆接1 規範1	補足1 対立1	転換2 逆接1	転換2 真加1 補足1		転換3		転換7 真加3 並列1 補足1 規範1 例示1		転換7 背景1 真加1 規範1 追加1 例示1	転換4 真加1 補足1 例示1 規範1
意見文		転換1	転換2 規範1	転換2 並列1		転換2 規範1	転換3	転換3 補足1		転換2 真加1	転換4
理由文			転換1 規範1	真加2 逆接1		真加1		転換2 真加1		転換1 背景1	転換4 補足1
感動文		転換2	転換1 逆接1			転換1 真加1		転換1 逆接1 補足1 例示1		転換1 真加1 補足1 並列1 反復1	補足1

表５　主張文－叙述文の文間関係

表6　叙述文－主張文の文間関係

前＼後	問題文	断定文	推量文	要望文	判断文	意見文	理由文	概略文
可能	逆接1 転換1 其加1	背景1 逆接1	逆接1	背景1 逆接1	根拠1 反復1 逆接1			転換1
伝聞	逆接2	反復1	比較1 補足1	逆接1				
推聖	逆接1 転換1 例示1 逆接2	逆接2 根拠2 転換2 背景1 其加1 逆接2	其加2 追加1 転換2 背景3 逆接2	背景1 補足1 其加1 逆接2	転換1 補足1 逆接3	前提2 其加1 逆接2	根拠1 背景1	其加1
叙述	転換1 背景1 反復1 其加2 逆接2	追加1呼応1 前提1対立1 転換1反復3 其加4逆接3	逆接2結果1 追加1転換3 比較1補足1 其加1例示1 逆接3	逆接1 転換1 其加1 逆接10	其加1逆接1 逆接1根拠1 前提1追加1 背景1補足2	逆接1 前提1 転換1 背景1 其加1 逆接2	転換1 逆接1	前提1 補足1 其加3 逆接3
主況								
存在	其加1 逆接1	引用1 逆接1 転換1 反復2 例示1 逆接5	逆接2 転換1 背景1 其加1 例示1 逆接4	前提1 逆接1 背景1	転換1 補足1 其加2	逆接1 前提1 其加1	補足1	反復1
推測			逆接1					
状態	前提2 転換1 背景2 其加2 逆接3	引用1逆接3 反復1根拠2 追加1例示1 対換4背景1 並列1補足6 其加10 逆接5	引用1逆接3 転換6背景3 追加3補足1 其加4 逆接7	逆接1追加1 転換2 並列1 其加2 背景1 逆接3	逆接1 背景4 補足1 其加3 逆接6	反復2 其加4 逆接1	逆接1 根拠2 転換2 反復1 逆接1	限定1 転換1 背景1 逆接5
使役		反復4						
現在	逆接1 前提2 転換1 反復1 補足2 其加2 逆接3	根拠3前提2 対立1転換1 背景1反復3 其加1例示1 逆接4	逆接1追加1 転換6背景3 並列1補足1 其加7 逆接3	引用1 前提4 転換1 背景1 逆接6	逆接3 限定1 転換1 背景1 其加4 逆接4	転換1 背景1 其加3 逆接2	根拠7 対立1	逆接1 前提1 転換1 反復1 其加1 例示1 逆接2
過去	逆接3 転換1 補足3 逆接2	逆接1前提1 追加2転換2 背景3反復3 補足4 其加4例示1 逆接9	逆接2 対立1 転換1 其加2 逆接5	転換1 背景1 逆接1	逆接1 転換1 補足1 其加1 逆接4	追加1 反復1 其加1	転換1 補足1	引用1 背景1 其加1 逆接1

表7　主張文－主張文の文間関係

前＼後	問題文	断定文	推量文	要望文	判断文	意見文	理由文	概略文
問題文	呼応1 転換1 反復1 並列3 其加4	結果1 呼応1 転換1 並列1 其加2	呼応3 転換3 其加1 逆接3	前提1 逆接1	呼応3 前提1 逆接2	転換1 其加1 逆接1	対立1 根拠1	呼応1 転換2
断定文	転換1 背景1 補足2 其加1 逆接1	根拠1 前提2 追加1 転換3 反復1 補足2 其加3 逆接1	転換1 反復2 補足2 逆接1	前提1 転換1 並列1 其加3 逆接4	前提1 転換1 其加2 逆接2	前提3 並列1 逆接3	反復1	
推量文	追加1 転換3 反復1 其加1	前提1 転換3 並列2 其加4 逆接3	逆接1 前提2 転換1 補足1 其加1	補足2 逆接4	根拠1 其加1 逆接5	前提1 追加2 転換1 背景1 並列1 目的1 逆接5	根拠3 転換1	前提1 背景1
要望文	反復1 其加1	転換6 逆接1	根拠1 追加1 逆接1	追加1 転換1 並列1 目的1 其加1 逆接1	補足2	目的1	根拠2	転換1 其加5
判断文	追加1 逆接1	転換1	転換1 補足1 其加2 逆接2	背景1 補足1 逆接5	転換1 背景1 逆接2	転換1 逆接1		前提2 対立1
意見文	逆接1	転換1 目的1 其加1	転換2 並列1 補足1	前提2 転換1	補足1		根拠4	其加2
理由文	呼応1	転換1	逆接1	転換2 逆接2	逆接1			
概略文	転換2	転換2 例示1	転換1 反復1	並列1 補足1 逆接2	根拠2 転換2	根拠1 転換2 背景1 並列1 其加3 例示1 逆接1	根拠1	並列1 其加1