

TR-0806

文の連接関係解析に基づく
日本語文章の構造解析

福本 淳一、安原 宏（沖）

September, 1992

© 1992, 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

文の接続関係解析に基づく日本語文章の構造解析

Japanese Text Structure Analysis based on Sentence Cohesion

福本 淳一 安原 宏

Jun-ichi FUKUMOTO, Hiroshi YASUHARA

沖電気工業(株) 総合システム研究所

Systems Laboratory, Oki Electric Industry Co., Ltd.

11-22, Shibaura 4-chome, Minato-ku, Tokyo 108, Japan

概要

本稿では、広い範囲の文章の各文間の形態的表現の解析から得られた知識を規則として記述し、それを用いた文章構造の解析手法及びその評価結果について述べる。論説文等の文章は、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る。このような文章においては、書き手の主張が読み手である読者に伝わるような論旨が展開され、この論旨の展開構造が文章の構造であると考えられる。我々は文章構造の解析のため、まず、新聞社説記事中に現われる文脈現象の解析から得られた知識を基に、文章の構造化のためのモデル及びそのモデルに基づく言語仕様を定義した。そして、この言語で記述した構造解析規則を適用することにより文章の構造解析を行うシステムを開発した。本システムでは、まず、文章の形態素解析結果から各文の文タイプ、主題提示語、接続詞語句、指示語、同語反復の各情報を抽出し、これらを用いて文章中の連接2文間の関係の解析を行う。次に、その関係に基づき文のグループ化を行い、各生成されたグループ間の関係の解析から文章構造要素を認識することにより文章全体の構造を得る。これまでに約400の規則を記述し、これを用いて新聞社説記事等の文章の構造解析結果の評価を行った。新聞社説記事30編についての評価では、平均で70.7の評価値を得た。また、科学技術文献等の文章15編については、平均で59.8の評価値を得た。

1 まえがき

従来まで行われてきた1文レベルの自然言語処理では、文章全体の意味するものや文の流れを判断することは困難である。このような1文レベルの処理を越える文脈処理技術は、文章要約などの自然言語応用システムや高度な機械翻訳システムの実現に不可欠な要素技術であると考えられる。

これまで文脈処理に関して幾つかの研究¹³⁾¹⁴⁾¹⁵⁾が行われてきたが、これらは限られた範囲の問題について予め準備された知識を用いたものであり広い範囲の文章に対応できないという問題があった。また、文章の構造解析についてこれまで幾つかの研究が行われている。辻井は文を6種類にタイプ分類を行い、このタイプ間の接続関係を規則として記述することで文章の修辞構造の解析を行っている¹¹⁾が、解析結果についての十分な評価がなされていない。住田他は各文間の接続語句等による修辞関係とそれらに対する制約を用いて文章の構造解析を行っている⁹⁾¹²⁾。しかし、この方法では修辞表現の少ない文章に対応できないといった問題がある。Mann他は、nucleusとsatelliteと呼ばれる文章構造unit間に20余りの関係を設定して文章の構造化を行っている¹³⁾が、具体的な解析方法は示されていない。

本稿では、広い範囲の文章の各文間の形態的表現の解析から得られた知識を規則として記述し、それを用いた文章構造の解析手法及びその評価結果について述べる。まず、文章構造の解析のため、我々は新聞社説記事中に現われる文脈現象の解析から得られた知識を基に、文章の構造化のためのモデル及びそのモデルに基づく言語仕様を定義した。そして、この言語で記述した構造解析規則を適用することにより文章の構造解析を行うシステムを開発した。

論説文等の文章は、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る。このような文章においては、書き手の主張が読み手である読者に伝わるような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ると考えられる。一つの文の中では、語及び文節間に明白な形態的指標が存在するのと同様に、複数の文からなる文章においても形態的指標が存在すると考えられる⁷⁾。また、論説文などの文章を構成する各文には、書き手の主張を表わす文や主張のために必要な客観的な事実を述べた文が存在する。日本語においてはこのような書き手の態度は主に文末の表現として現われ、特に文末の助動詞と動詞の機能によって決まる⁷⁾¹⁰⁾。我々は、各文間の関係を決定するための形態的な情報として、書き手の主張に基づいて分類した文タイプ情報を用いた²⁾。また、助詞「は」

等により提示される主題提示語、文章の流れを明示的に決定する機能のある接続語句、文章中の他の部分を参照する機能のある指示語、文章中で繰り返して用いられる同一名詞も文間の関係を決定するための形態的な情報として用いた³⁾。

本システムにおける文章の構造解析は、まず、文章の形態素解析結果から各文の主題提示語、文タイプ、接続語句、指示語、同語反復の各情報を抽出し、これらを用いて文章中の接続2文間の関係の解析を行う。そして、その関係に基づき文のグループ化を行い、各生成されたグループ間の関係の解析から文章構造要素を認識することにより文章全体の構造を得る。

以下、2章で文脈現象の調査結果について述べた後、3章で文章構造のモデル及び記述言語について述べる。次に、4章で本システムの文解析及び文章構造解析手法について述べ、5章で新聞社説記事を用いた文章構造解析の評価について述べる。

2 文脈現象の調査分析

文章中の各文間の関係を決定するため、新聞社説記事を用いた文脈現象の調査として文間関係と照応現象の調査を行った。

2.1 文間関係の調査

論説文などの文章を構成する各文は、書き手の考えや意見などの書き手の主張が現われている文（主張文）とそのような意見を表わすために必要な文（叙述文）とに分類できる。文タイプの連鎖からどのように文間関係が決定できるかを調査するため、朝日新聞社説記事70編より文間関係が存在すると考えられる2文を抽出し、2145件の文間関係データを収集した²⁾。そして、各文間関係データの2文に文タイプ情報を付与することにより、文タイプの連鎖と文間関係との関連を分析した。

文タイプ情報としては、朝日新聞社説記事の各文末表現を調査することにより、主張文について[問掛文, 断定文, 推量文, 要望文, 判断文, 意見文, 理由文, 義務文]の8つに、叙述文について[現在, 過去, 可能, 伝聞, 様態, 叙述, 存在, 継続, 状態, 使役]の10に再分類した²⁾。以下にそれぞれの文末表現の例を表1に示す。

表1 文末表現の例

文間関係としては、永野⁷⁾や市川⁸⁾による分類がある。我々は、これをも

とに図1に示す20種類の文間関係を設定した。

図1 文間関係名

収集した文間関係データを主張文、叙述文の文タイプの連鎖により分類したものを表2、表3、表4に示す。

表2 主張文-主張文関係

表3 主張文-叙述文関係

表4 叙述文-主張文関係

文間関係データの分析から、主張文-主張文、主張文-叙述文の連鎖に関しては、文タイプの連鎖によりいくつかの文間関係を決定することができた。例えば、表1より意見文-理由文間の関係は根拠であり、意見文-義務文間の関係は累加となる。しかし、叙述文-叙述文、叙述文-主張文の連鎖に関しては、文タイプ情報のみでは明確な関係を決定することができなかった。そこで、これらについては、さらに主題提示情報等を用いることによりいくつかの関係を決定した。

2.2 照応現象の調査

文章中の指示語及び同一名詞の反復からどのような文間の関係が決定できるかを調査するため、文間関係の調査に用いた文章を用いて文章中の照応現象の調査を行った。この調査では、以下の項目について3003件のデータを収集した²⁾。

- (1) 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
- (2) その文間の距離

照応の距離として、照応関係の存在する文間の距離をまとめたものを表5に示す。

表5 照応関係の存在する文間の距離

照応現象データの分析から以下のことが明らかになった。まず、指示語の照応に関しては約8割がその文内又は前文を指しているため、指示語が用いられている文は、その前文の内容の一部を指し示すことにより話題が展開されていると考えられる。文間関係データとの関連として同一文間に照応現象

データと文間関係データが共に存在するものについてまとめたものを表6に示す。

表6 照応現象と文間関係の関連表

表6から、指示語については累加、背景、補充関係が比較的多く現われ、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は、前文から引き続いた内容が述べられる傾向があることがわかる。

名詞語句については、並列、転換関係が多く現われる傾向があった。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現われているため、同一名詞句の反復情報のみでは単純に話題の転換と判定することはできず、その他の情報が必要であると考えられる。

3 文章構造のモデル化

ここでは、まず文章構造を定義する。次に、文章構造のモデルについて述べた後、そのモデルに基づいて文章の文脈を記述するための記述言語の仕様について述べる。

3.1 文章構造

書き手の主張に基づき文章を構造化するためには、書き手の主張を表わす文とその前後の文がどのような関係で結ばれているかを捉える必要がある。また、そのような関係によって結ばれた文のまとまり間の関係も捉える必要がある。この関係を一方が主で他方が従である2項関係と捉えることで文章全体を木構造として構造化できる。しかし、実際の文章中には、主従を決定しにくいものがいくつか存在する。図2に示す例のように、文章中である事柄について述べられた文が順に列挙される場合、これらの文間に主従の関係はないと考えられる。

図2 文脈現象の例 (1)

また、図3に示す例のように、「第1に」「第2に」「第3に」のような表現を用いて、いくつかの文を序列的に並べるものもある。これらの文間にも主従

の関係はないと考えられる。

図3 文脈現象の例 (2)

なお、図2、図3において数字はパラグラフ番号－文番号を示す。

このような文間の関係を主従の関係のないn項関係と捉えることにより、これらの文を1つのまとまりとして表現することにした。また、主従の関係のあるものについても、同じ話題について述べられているものを1つのまとまりとして表現する。このようないくつかの文をまとめる操作をグループ化と呼ぶことにする。文章の構造は、文章中の各文間の関係のうち主従の関係があるものについては同じ話題について述べられているものをグループ化し、主従の関係がないものについてはそれらをグループ化されたものとする。また、グループ化によってまとめられたものの間にも同様にある関係が成り立つものとする。

3.2 文章構造モデル

文章構造をモデル化するため「。」で区切られた1文を1つのノードとして表わし、これらの文間に主従の関係が認められる時、その関係をノード間のアークで表わす。このとき、アークには従となるノードから主となるノードの方向へ矢印をつける。また、いくつかのノードをまとめてグループ化したものも1つのノードとして扱う。本文章構造モデルで扱うノードには、文を表わすSノード (sentence node) とグループを表わすGノード (group node) の2種類がある。各ノードの性質はノードの属性として表わす。Gノードには、以下の3つのタイプがある。

○連続タイプ

ある事柄が連続して述べられているものをグループとして表現したノード

○並列タイプ

並列的な内容や序列的に述べられているものをグループとして表現したノード

○スコープタイプ

文間関係が存在するノードのうち同じ話題に関するものをまとめてグループとして表現したノード

Gノードの各タイプを表現したものを図4に示す。図中、ノードは「○」で、アークは「←」で表わされている。また、連続タイプでは、ノードは事柄が

連続している順に「→」で結ばれて表わされている。

図4 Gノードの種類

文章構造は、ノード(Sノード, Gノード)間にアークを張ったり、ノードをグループ化することにより表現される。アークを張るための条件は、各ノードに付与された属性やそのノードから張られているアークの情報により、また、Gノードの場合にはその要素ノードの属性によっても決まる。

3.3 記述言語

文脈現象の分析結果を文章構造解析のための知識とみなし、これを規則として記述するため、前節で述べた文章構造モデルに基づく記述言語を定義した。

本言語で記述される規則は、以下のようにルール名、ルールグループ名、パターン部、条件部、実行部から構成されている。

<ルール名> : <ルールグループ名> {
 <パターン部> <条件部> <実行部> }

規則は、ルール名の後にそのルールグループ名を付与することにより、いくつかをまとめてグループとしてモジュール化できる。パターン部では、文章構造のノード及びアークに関する条件を記述する。条件部では、ファクト形式の情報とのパターンマッチ及び任意の述語を記述することができる。パターン部及び条件部に記述された全ての条件が満たされた場合その規則は適用され、実行部の記述にしたがって、ノード、アーク及びファクト形式の情報の追加・修正・削除を行う。これによって文章の構造解析を行う。

本言語仕様に基づいて記述した文章構造解析規則の記述例を図5に示す。この規則のパターン部では、連続する2つのノードがそれぞれ主張文、叙述文であり、叙述文に主題提示があり、それらの間にアークがないことを表わす。条件部では、叙述文の主題提示語が主張文に現われないことを示す。実行部では、それらの間に転換関係のアークを張ることを表わす。

図5 文章構造解析規則記述例

4 文章構造解析システム

文章構造解析システムのシステム構成を図6に示す。本システムでは、まず、文解析部において入力された文章の形態素解析を行い、その結果から構

造解析に必要な情報を抽出し、ワーキングメモリにセットする。次に、トランスレータにより文章構造解析規則を変換し、ルールメモリにセットする。文章構造解析部では、ルールメモリにセットされた文章構造解析規則とワーキングメモリにセットされた文解析情報とのマッチングを繰り返し、その内容を書き換えていくことにより文章の構造解析を行う。

図6 システム構成

4.1 文解析部

文解析部では、文章中の各文についての形態素解析結果をもとに、文章の構造解析のための入力となる情報として、主題提示語、文タイプ、接続語句、指示語、及び同語反復の各情報を抽出する。

1) 主題提示語情報抽出

主題提示語情報の抽出処理においては、主題提示語として、助詞「は」「も」が付属する語句の抽出を行う。一般に、主題提示語は、自立語に助詞「は」等が付属する形を持つが、この形の中には、「特には」「～とはいえない」のように主題提示の機能を持たないものも存在する。そこで、新聞社説記事を用いた調査から、以下のものを主題提示語として抽出した（「も」も同様）。

名詞+は	名詞+においては	名詞+では
名詞+とは	名詞+にとっては	名詞+としては
名詞+には	名詞+については	
名詞相当語+は	名詞相当語+には	

2) 文タイプ情報抽出

文タイプ情報の抽出処理においては、2.1で述べた文タイプとして、主張文については[問掛文, 断定文, 推量文, 要望文, 判断文, 意見文, 理由文, 義務文]の8種類を、叙述文については[現在, 過去, 可能, 伝聞, 様態, 叙述, 存在, 継続, 状態, 使役]の10種類の文タイプの細分類を抽出する。

3) 接続語句情報抽出

接続語句情報の抽出処理においては、文中で接続詞等の文間の関係を決める語句、及び「第1に」「1つは」のように文章の流れを決める機能を持つ語句を抽出する。接続詞情報としては、品詞が接続詞であるものを取り、接続的な機能を持つ語句は、新聞社説記事中の接続語句の分析で得られた結果をもとに決定した。

4) 指示語情報抽出

指示語情報の抽出処理においては、文中の指示代名詞及び指示連体詞を含

む文節を抽出する。但し、人称代名詞の1人称、2人称のものや、「その半面」、「このところ」等の直接指示参照するものがないものは抽出しない。後者の抽出しないものは、新聞社説記事の指示語情報の分析で得られた結果をもとに決定した。

5) 同語反復情報抽出

同語反復情報の抽出処理においては、文章中で繰り返し用いられている名詞語句及びその名詞語句に対して修飾語が係っているかどうかの情報も抽出する。

4.2 文章構造解析部

文章構造解析部では、文解析部で抽出された情報と文章構造解析規則との適用を繰り返しながら文章構造の解析を行う。この解析では、まず、文章中の隣接文間の関係の解析を行い、次に、解析された文間関係を基にノードのグループ化を行い、文章全体をいくつかのGノードとして構造化する。最後に、文章を構成するために必要な要素を生成されたGノードより認識し、それぞれの要素として認識されたGノードをさらにグループ化し、文章全体の構造を得る。

1) 隣接文間解析処理

隣接文間の関係は、主として新聞社説記事の解析結果から得られた文のタイプ間関係によって決定する。但し、文間に接続詞情報が存在した場合、また、指示語情報や主題提示情報などを用いて記述された規則に適用した場合、そちらを優先して文間関係を決定する。文間関係としては、文脈現象の調査で用いた20の関係を分類・整理し、次の17の関係を設定した。これらの関係は、ノードの主従の順序と結び付きの強さにより以下の3つのクラスに分類されている。

○クラス1（前文が主で結び付きの強い関係）

例示・補足・呼応

○クラス2（後文が主で結び付きの強い関係）

背景・前提・根拠・結果

○クラス3（結び付きの弱い関係）

逆接・転換・序列・累加・反復

並列・対比・継続・展開・連係

2) ノードグループ化処理

ノードグループ化処理では、隣接文間解析処理で得た文間関係を基に、まず、継続、連係関係で結ばれたノードを連続タイプのGノードとしてグループ化する。次に、結び付きの強い文間関係で結ばれたノードをスコープタイ

ブのGノードとしてグループ化する。そして、以上の処理でグループ化されなかったノードについては、文間関係に強度を設定し、強度の大きい関係で結ばれているノードを先にグループ化する。連続する3つのノードをa, b, cとし、ab, bc間の関係をそれぞれR1, R2とした時の、R1, R2の強度関係を以下に示す（ $R1 > R2$ のとき、R1が強い強度を持つとする）。

(a) 連続する3つのノードが主張文の場合

クラス1 > クラス2 > クラス3

(b) 連続する3つのノードが順に主張文, 叙述文, 叙述文の場合

クラス1 > クラス2 > クラス3

(c) それ以外の場合

展開 > クラス1 > クラス2 > クラス3

また、グループ化により生成されたGノード間の関係についても、その主ノードの情報等を用いることでグループ間関係の解析処理を行い、さらに上位のGノードとしてまとめる。

3) 文章構造認識処理

文章構造認識処理においては、文章を構成するために必要な要素となるGノードを調べ、構成要素として認識されたノードを1つのGノードとしてまとめたうえで文章構造の要素名を認識する。現在、文章全体の構造の要素として認識しているものには、[序論, 本論, 結論]がある。また、序論についてはさらに[事例紹介, 問題提起]に細分類される場合もある。構造解析を行った文章を図7に、その構造解析結果を図8に示す。

図7 文章構造解析例文

図8 文章構造解析結果

5 文章構造解析結果の評価

本システムによる文章構造解析結果の評価のため、本記述言語を用いて記述した約400の文章構造解析規則を用いて、朝日新聞社説記事30編及び科学技術文献等の文章15編について構造解析結果の評価を行った。評価は、以下の3項目について○△×(各2, 1, 0点)の3段階で行った。

(a) ノード間の関係の妥当性

文と文、文とグループ、グループとグループの関係が良いかどうかを評価する。

(b) 生成されたGノードの要素のノードの妥当性

いくつかの文を1つのまとまりとしたもの（グループ）について、そのまとめ方が良いかどうかを評価する。

(c) 文章構造要素名の妥当性

文章構造要素として認識された序論・本論・結論といった文章のおおまかな構成について、そのグループの範囲、名称が良いかどうかを評価する。

文章構造解析の評価値は、各項目の合計値をすべてが○であった場合に100となるように規格化したものとした。1名の評価者により3段階の評価を行った結果、新聞社説記事30編については、最高86.6、最低53.8、平均70.7であった。科学技術文献等の文章15編については、最高79.4、最低23.9、平均で59.8という評価値を得た。それぞれの結果を表にまとめたものを表7、表8に示す。

表7 文章構造解析の評価結果（新聞社説記事）

表8 文章構造解析の評価結果（科学技術文献等）

得られた評価結果を基にして、いくつかの観点から文章構造解析の誤りの原因を分析した。

(1) 文間関係名の誤り

文間関係のうち叙述文に関するものの誤りが多く存在した。これは、これまでの文間関係の解析からは叙述文に関するものについて明確な規則性が得られなかったためである。叙述文間の解析規則は、一部の文タイプ情報と主題提示情報等を用いて規則化しているが、名詞の反復情報を利用することで正しい文間関係を得ることができると考えられる。また、文章中で短い文を用いることにより、前文の内容に対して補足を行ったり前文で述べられた内容についての主張を行う表現があり、そのような現象も規則として取り入れる必要がある。

(2) グループ間関係名の誤り

グループ間の関係を解析するために主としてグループ内で中心となる文のタイプ情報とグループの先頭の文の主題提示情報を用いているが、これだけでは不適切なものがいくつかあり、他の情報を用いるように検討を行う必要がある。また、タイプにとってグループ間の関係を決定する場合、文間関係の解析規則をそのまま適用しているため、グループについて不適切な場合がある。今後、これについて検討する必要がある。また、グループの先頭にある文ノードの情報が、その前にあるグループとの関係を解析するための重要

な鍵となる。このようなノードの情報の利用について検討する必要がある。

(3) グループ化の誤り

グループ化の場合、グループ間の関係情報だけを用いるのではなく、まとめられるノードで述べられている内容の関連性についても考慮する必要があると考えられる。しかし、述べられている内容の関連をどのようにして判断するかは今後の課題である。

(4) 文章構造要素名の誤り

要素ノード名の付与については、事実の列挙の認識がうまくいっていないため、文章構造要素名の決定規則の見直しを行う必要がある。

(5) その他の誤り

その他の誤りとしては、まず、一般の科学技術文献の解析から、これまでに抽出するものとして設定していなかったものが接統語句として表れ、このため関係が正しく解析されないものがあった。次に、文タイプの抽出方法について、現在は新聞社説記事の分析に基づき分類を行っているが、科学技術文献では主張を行うために別の表現を用いていることから、文章の種類によって主張表現の抽出方法を検討する必要がある。「第1に」のような序列表現については、抽出できていない表現が存在するため、抽出語の修正を行う必要がある。また、序列表現のうち、最後の要素の範囲の認識が正しく行われないものがあった。これは、グループ化の問題点とも関連している。序列表表現の列挙の前に「以下の3つの例がある」のような序列表表現の導入表現があり、これも扱う必要がある。

6 あとがき

本稿では、新聞社説記事中の文脈現象の分析から得られた知識を基に、文章の構造化のためのモデル及びそのモデルに基づく言語仕様について述べた。そして、これを用いて記述した規則を適用することにより文章の構造解析を行うシステムおよびそのシステムを用いた文章構造解析の評価結果について述べた。

本システムでは、文の表層情報として現われた書き手の主張を述べた文を中心として文章の構造解析を行っている。そのため、新聞社説記事等の論説などのタイプの文章の構造化については約7割という評価結果がえられた。しかし、一般の新聞記事のように単に事実が羅列してあるタイプの文章の構造化は困難である。このような問題を解決するためには、文章中で用いられている言葉の連鎖情報や表層情報だけでなく意味的な情報も用いる必要があると考えられる。

本文章構造解析技術の応用としては、文章の構造情報を用いた文章の要約処理の実現や文章における省略部分の補完による機械翻訳システムの訳質の向上が考えられる。要約処理の実現のためには、得られた文章構造からどこを重要部分として抽出すべきかを検討する必要がある。また、省略語の補完のためには、どのような文間関係の場合に省略現象が起こるかの関連を調査する必要がある。

【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, pp.113-120 (1990).
- 2) 福本, 安原：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報告 85-11, pp.81-88 (1991).
- 3) 福本, 安原：文の接続関係解析に基づく文章構造解析, 情報処理学会自然言語処理研究会報告 88-2, pp.9-16 (1992).
- 4) 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回全国大会論文集 (1990).
- 5) 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回全国大会論文集 (1991).
- 6) 斎藤, 柴田, 福本：文章構造化のための文の接続関係の解析, 情報処理学会第43回全国大会論文集 (1991).
- 7) 永野 賢：文章論総説－文法論的考察－, 朝倉書店 (1986).
- 8) 市川 孝：国語教育のための文章論概説, 教育出版 (1978).
- 9) 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).
- 10) 山梨 正明：発話行為, 大修館書店 (1986).
- 11) 辻井潤一：文脈処理, 「自然言語処理技術」シンポジウム, pp.75-87 (1988).
- 12) Sumita, K., Ono, K., Chino, T., Ukita, T., and Amano, S. : A Discourse Structure Analyzer for Japanese Text, Proceedings of International Conference on Fifth Generation Computer Systems, Vol.2, pp.1133-1140 (1992).
- 13) Mann, W.C., Thompson, S.A., : Rhetorical Structure Theory: A Framework for the Analysis of Texts, USC/Information Science Institute Research Report RR-87-190 (1987).
- 14) Schank, R.C. and Abelson, R.P. : Scripts, Plan, Goals and Understanding, John Wiley and Sons (1977).
- 15) Grosz, B., Sidner, C.L. : The structures of Discourse Structure, Technical Report, CSLI, CSLI-85-39 (1985).
- 16) Sidner, C.L. : Focusing in the Comprehension of Definite Anaphora, in Brady, M. and Berwick, R.C. (eds.), Computational Models of Discourse, MIT Press, pp.267-330 (1983).

【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る。研究の機会を与えて頂いた ICOT 淵所長，内田部長，有益な助言を頂いた ICOT 第6研究室田中室長，及び ICOT 自然言語理解ワーキンググループの皆様には感謝します。また，研究協力頂いた(株)沖テクノシステムズラボラトリの皆様にも感謝します。

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

図1 文間関係名

Figure 1 Relationships between Sentences

- 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)

図2 文脈現象の例 (1)

Figure 2 An Example of Context Phenomena (1)

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）

図3 文脈現象の例 (2)
Figure 3 An Example of Context Phenomena (2)

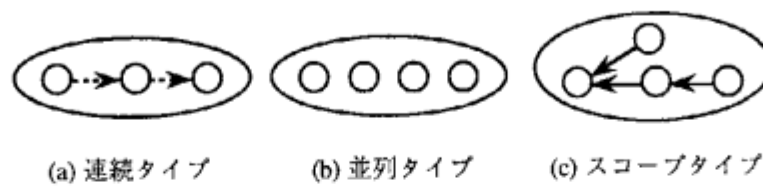


図4 Gノードのタイプ
Figure 4 Three Types of G-nodes

rule1:sent_sent {	ルール名：ルールグループ名	
pattern: top:[* x y *];		パターン部
x#stype0 == '主張文';		
y#stype0 == '叙述文';		
x#topic != 'yes';		
y#topic == 'yes';		
!arc(x,y);		
condition:		条件部
topic_word(y#name,WORD,_,_);		
!rep_word_rel(x#name,y#name,WORD);		
action:		実行部
make_arc(x,y);		
set(arc(x,y)#name,'転換');		
}		

図5 文章構造解析規則の記述例
Figure 5 : An Example of Text Structure Analysis Rules

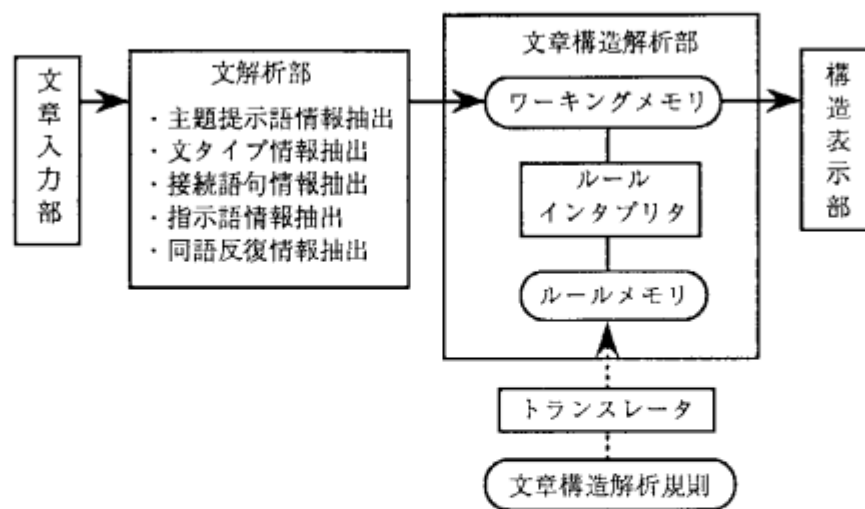


図6 システム構成
Figure 6 An Overview of the System

- 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ものってきた。
- 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。
- 2-1 業務上の病気に限れば、近年、減ってきている。
- 2-2 昨年は1万5000人足らずで、この十数年の間に半減した。
- 2-3 が、定期健康診断で見つかる一般の病気の率は毎年上昇している。
- 2-4 昨年は9.8%で、定期健康診断を受けた1090万人のうち、107万人弱が病気とわかった。
- 2-5 高度成長期を含めて最高の率だ。
- 3-1 その半分近くを高血圧が占める。
- 3-2 突然死に関連が深い心臓の病気も、高血圧の10分の1ほどあった。
- 3-3 高齢化が進めば、病気の発見率はさらに高まるだろう。
- 3-4 気がかりである。
- 4-1 しかし、病気が早く見つければ、治る率も高い。
- 4-2 定期的な健康診断をもっと広め、定着させることが望ましい。
- 4-3 雇われて働いている人の、4人に1人しか受診していない現状でよいはずはない。
- 4-4 中小企業や客細な商店の人も受診できるよう、保健所などを中心に地域ぐるみの取り組みを強めてほしい。
- 5-1 突然死やストレス関連の病気については、その実態がどうなっているのか、国が本格的に調査すべき時期にきている。
- 5-2 これまで公表されたいくつかの部分的な調査でも、事態の深刻さをうかがい知ることができる。
- 6-1 東京都監察医務院が、昨年6月までの5年半ほどの間に扱ったスポーツ中の突然死は、107件にのぼる。
- 6-2 中高年では心臓や血管系の疾患が大半だが、40歳未満でははっきり病が1番多い。
- 6-3 全国的な実態調査にもとづいた原因の速やかな究明を期待したい。
- 7-1 マイクロ・エレクトロニクス機器の普及につれて増えているストレスについては、東京都労働研究所がソフトウェア技術者650人近くから集めたアンケート調査がある。
- 8-1 「二者択一的な思考」「機械への過剰な親しみ」「強迫神経症の傾向」「仕事をほうり出して逃げたいと考える」といった状態に陥った人が、まわりには「かなりいる」「少しいる」という答えが7～8割にも達した。
- 9-1 しかし、自分自身もそうだと答えた人は3～5割にすぎない。
- 9-2 一般にはあるが、自分は大丈夫、というわけだ。
- 9-3 この種の症状は自覚されにくい傾向がある。
- 10-1 対策としては、職場の人間関係を改善したり残業を減らしたりすることに加え、心とからだの日常的な健康管理が大切だと思う。
- 11-1 ふえる海外勤務のストレス予防策も、もっと組織的に取り組まれてよい。
- 11-2 海外赴任から帰国して1年以内のビジネスマン300人を対象にした国際電信電話会社の調査では、4人に3人が海外でストレスを感じている。
- 11-3 現地従業員とのコミュニケーション、ビジネス習慣の違い、情報不足などが原因だ。
- 12-1 そのため、酒やたばこがふえ、眠れなくなったり病気になったりする人も多い。
- 12-2 解消策は、スポーツや酒、日本人同僚との交流など個人的な努力が目立つ。
- 12-3 国際化が進むにつれて海外進出は今後もふえる一方なのだから、事前の健康診断から現地でのストレス解消策まで、企業は本格的に取り組むべきだ。
- 13-1 石綿公害をはじめ化学物質による障害の予防策にも力を入れる必要がある。
- 13-2 中小企業経営者へのセミナーや中堅管理者への「ヘルス・リスナー（上手な聴き役）」研修会を、中央労働災害防止協会が労働省と連携して今年度からはじめた。
- 13-3 結構なことだ。
- 14-1 働く人の健康は地道な努力で守るほかはない。
- 14-2 1日からの全国労働衛生週問を機に、身の回りの健康を点検し直してみたい。
- (昭和62年10月2日付け朝日新聞社説記事 朝日新聞社提供)

図7 文章構造解析例文

Figure 1 A sample text for text structure analysis

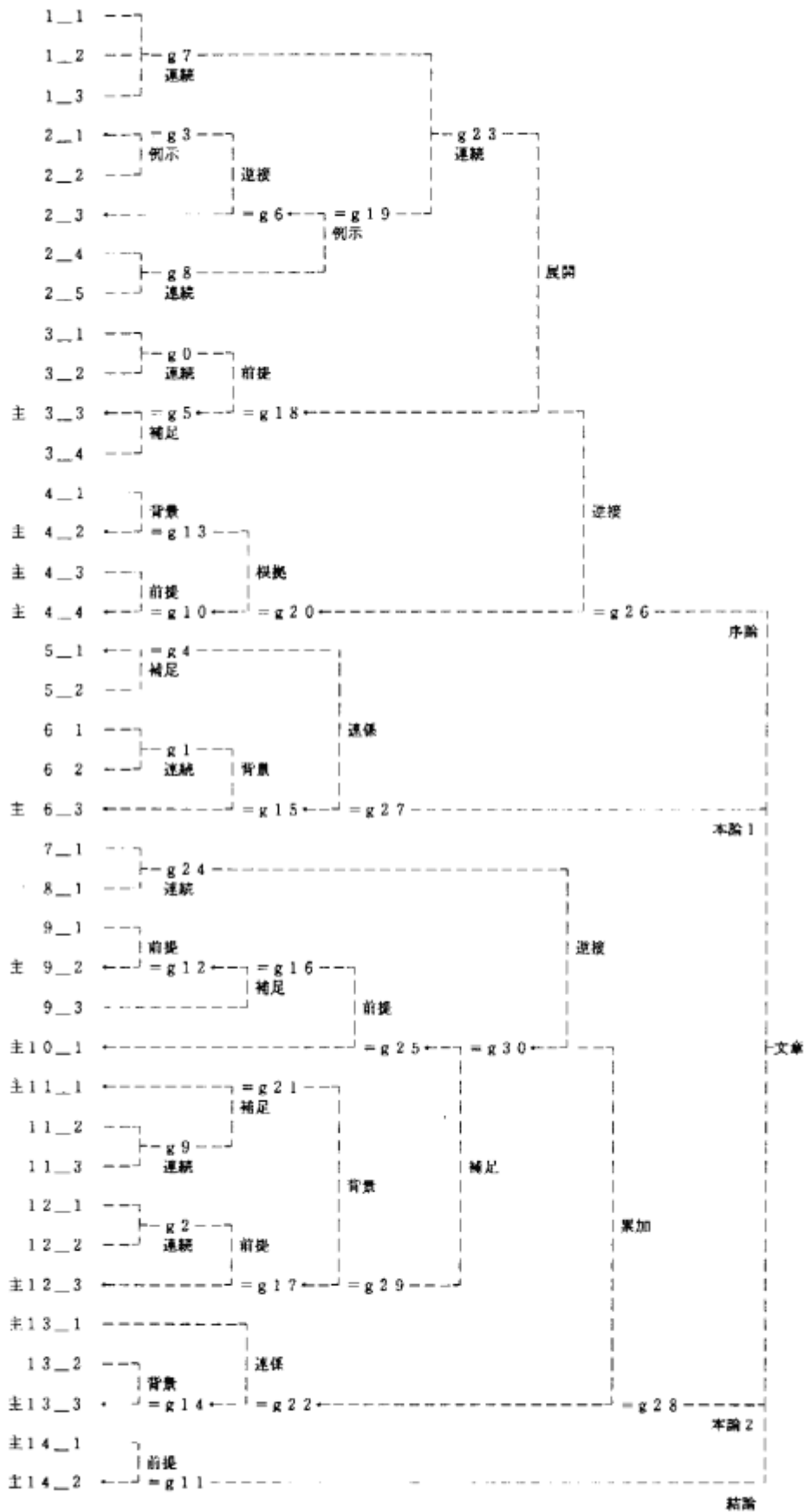


圖8 文章構造解析結果
Figure 8 A result of text structure analysis
- 23 -

表1 文末表現の例
Table 1 Examples of tail expression of a sentence

文タイプ名	文タイプ名 (細分類名)	文末表現の例		
主張文	問掛文	のではあるまいか どう～だろうか	どう～のか ありはしないか	恐れはないか あるのだろうか
	断定文	である 明らかである	ことである のではない	のである のだ
	推量文	だろう べきだろう	ではあるまい かもしれない	とはいえまい やむをえまい
	要望文	～たい てもらおう	期待する てもらいたい	～したい てほしい
	判断文	はずだ といえる	はずはない とはいいがたい	と考える ているようだ
	意見文	てはならない 必要である	必要がある 当然だ	望ましい 大切だ
	理由文	ているからである わけではない	からだ ためだ	わけだ
	義務文	なければならない べきである	べきだ	ねばならない
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	

表2 主張文-主張文關係
Table 2 Relationships between Assertive type sentences

前 \ 後	問掛文	断定文	推量文	要望文	判斷文	意見文	理由文	義務文
問掛文	呼応1 転換1 反復1 並列3 累加4	結果1 序列1 転換2 並列1 累加2	呼応3 転換2 累加1 連係3	前提1 連係1	呼応2 前提1 連係2	転換1 累加1 連係1	対立1 根拠1	呼応1 転換2
断定文	転換1 背景1 補充2 累加1 連係3	根拠1 前提2 追加1 転換3 反復1 補充2 累加3 連係2	転換1 反復2 補充2 連係3	前提1 転換1 並列1 累加3 連係4	前提1 転換1 累加2 連係2	前提2 並列1 連係3	反復1	
推量文	追加1 転換3 反復1 累加1	前提1 転換3 並列2 累加4 連係3	逆接1 前提2 転換1 補充1 累加1	補充2 連係4	根拠1 累加1 連係3	前提1 追加2 転換3 背景1 並列1 目的1 連係3	根拠3 転換1	前提1 背景1
要望文	反復1 累加1	転換6 連係1	根拠1 追加1 連係1	追加1 転換1 並列1 目的1 累加1 連係1	補充2	目的1	根拠2	転換1 累加3
判斷文	追加1 連係1	転換1	転換1 補充1 累加2 連係2	背景1 補充1 連係3	転換1 背景1 連係2	転換1 連係1		前提2 対立1
意見文	連係1	転換1 目的1 累加1	転換2 並列1 補充1	前提2 転換1	補充1		根拠4	累加2
理由文	呼応1	転換1	連係1	転換2 連係2	連係1			
義務文	転換2	転換2 例示1	転換1 反復1	並列1 補充1 連係2	根拠2 転換2	根拠1 転換2 背景1 並列1 累加2 例示1 連係1	根拠1	並列1 累加1

表3 主張文-敘述文關係
Table 3 Relationships between Assertive type and Declarative type sentence

後 前	可能	傳聞	樣態	敘述	比況	存在	継続	狀態	使役	現在	過去
問掛文		呼応1	転換1 累加1	転換2 累加1 連係1 補充1 呼応1		転換1 累加1 補充1		転換4 連係1 呼応1 根拠1		転換5 累加1 連係4 補充1 呼応1 根拠1	連係1 転換1 累加1
断定文	累加1 連係1	転換2 累加1 補充1	転換1 累加4 根拠1 例示1	転換6 背景2 累加1 連係1 並列1 補充3 追加1		転換2 累加1 連係1 補充1 追加1		転換4 背景2 累加4 連係1 並列1 補充4 例示1 追加1		転換9 累加1 補充3 根拠4 前提3	転換9 累加3 補充4 例示2 追加3
推量文	並列1	転換1 並列2	累加1 根拠1	転換3 累加1 連係3 補充1 例示1		転換1 背景1 追加2		転換3 累加1 補充1 例示1	転換1	転換7 累加5 連係1 補充2 根拠1 例示4	転換8 補充1 根拠1
要望文			転換3 根拠1	転換5 逆接1 反復3		転換2 背景1 補充1		転換9 補充1		転換4 連係1 並列1	転換9
判断文	転換1 連係1 前提1	補充1 対立1	転換2 連係1	転換2 累加1 補充1		転換3		転換7 累加3 並列1 補充1 根拠1 例示1		転換1 背景1 累加1 根拠1 追加1 制約1	転換6 累加1 補充1 例示1 根拠1
意見文		転換1	転換2 根拠1	転換2 並列1		転換2 根拠1	転換2	転換3 補充1		転換2 累加1	転換4
理由文			転換1 根拠1	累加2 連係1		累加1		転換2 累加1		転換2 背景1	転換4 補充2
義務文		転換2	転換1 連係1			転換1 累加1		転換1 連係1 補充1 例示1		転換2 累加3 補充1 前提1 反復1	補充1

表4 敘述文-主張文關係
Table 4 Relationships between Declarative type and Assertive type sentence

前 \ 後	問掛文	斷定文	推量文	要望文	判斷文	意見文	理由文	義務文
可能	連係1 轉換1 累加1	背景1 連係1	連係1	背景1 連係1	根據3 反復1 連係1			轉換1
傳聞	連係2	反復1	比較1 補充1	連係1				
樣態	逆接1 轉換1 例示1 連係2	逆接2 根據2 轉換2 背景1 累加1 連係2	累加2 追加1 轉換2 背景3 連係2	背景1 補充1 累加1 連係2	轉換1 補充1 連係2	前提2 累加1 連係2	根據1 背景1	累加1
敘述	轉換1 背景1 反復1 累加2 連係2	追加1/序列1 前提1/對立1 轉換3/反復3 累加4/連係3	逆接2/結果1 追加1/轉換3 比較1/補充1 累加1/例示1 連係3	逆接1 轉換1 累加1 連係10	累加1/連係1 逆接1/根據1 前提1/追加1 背景1/補充2	逆接1 前提1 轉換1 背景1 累加1 連係2	轉換1 連係1	前提1 補充1 累加3 連係3
比況								
存在	累加1 連係1	引用1 逆接1 轉換1 反復2 例示1 連係5	逆接2 轉換1 背景1 累加1 例示1 連係4	前提1 逆接1 背景1	轉換1 補充1 累加2	逆接1 前提1 累加1	補充1	反復1
綜統			連係1					
狀態	前提2 轉換1 背景2 累加2 連係3	引用1/逆接3 反復1/根據2 追加1/例示1 轉換4/背景1 並列1/補充6 累加10 連係5	引用1/逆接2 轉換6/背景3 追加1/補充1 累加4 連係7	逆接1/追加1 轉換2 並列1 累加2 背景1 連係3	逆接1 背景4 補充1 累加2 連係6	反復2 累加4 連係1	逆接1 根據2 轉換2 反復1 連係1	限定1 轉換1 背景3 連係3
使役		反復4						
現在	逆接1 前提2 轉換2 反復1 補充2 累加2 連係5	根據3/前提2 對立1/轉換1 背景1/反復3 累加1/例示1 連係4	逆接2/追加1 轉換6/背景3 並列1/補充1 累加7 連係3	引用1 前提4 轉換1 背景3 連係8	逆接2 限定1 轉換1 背景1 累加4 連係4	轉換1 背景1 累加2 連係2	根據7 對立1	逆接1 前提1 轉換1 反復1 累加1 例示1 連係2
過去	逆接3 轉換3 補充2 連係2	逆接1/前提1 追加2/轉換2 背景2/反復3 補充4 累加4/例示1 連係9	逆接2 對立1 轉換1 累加2 連係3	轉換1 背景3 連係1	逆接1 轉換3 補充1 累加1 連係4	追加1 反復1 累加1	轉換1 補充1	引用1 背景1 累加1 連係1

表5 照応関係の存在する文間の距離

Table 5 Distance of two sentences which have anaphora

文間の距離	名詞語句による 照応	指示語による 照応
-2以下	1	0
-1	8	11
0	241	117
1	542	239
2	364	47
3	224	20
4	184	10
5	134	3
6	134	2
7	112	1
8	63	1
9	71	2
10	62	1
11以上	405	9
合計	2545	463

表6 照応現象と文間関係の関連表

Table 6 Relation between anaphora data and cohesion data

関係名	名詞語句による 照応	指示語による 照応
結果	2	6
目的	0	2
逆接	45	23
序列	4	0
追加	9	6
並列	10	1
累加	77	67
対立	8	5
比較	2	1
転換	85	12
限定	1	2
反復	14	9
根拠	24	9
制約	2	0
補充	87	33
連係	42	23
呼応	2	1
前提	19	8
背景	27	25
例示	18	4
合計	492	237

表7 文章構造解析の評価結果（新聞社説記事）
Table 7 Evaluation of Text Structures Analysis (newspaper editorials)

文章名	グループの妥当性	関係名の妥当性	文章構造の妥当性	総合評価
1001_1	86.4	80.4	60.0	81.0
1002_1	85.0	72.7	50.0	76.1
1002_2	90.3	76.5	87.5	83.3
1003_2	80.6	61.1	33.3	86.6
1004_1	91.3	72.0	83.3	81.5
1005_1	90.9	73.1	75.0	80.8
1007_1	68.8	56.3	60.0	62.2
1007_2	77.3	61.1	66.7	68.3
1008_1	84.5	66.7	37.5	73.0
1010_2	80.8	65.5	50.0	71.2
1011_1	78.8	56.7	50.0	65.6
1011_2	84.6	70.6	100.0	78.8
1012_1	78.3	67.3	100.0	74.5
1013_1	81.0	54.3	100.0	70.4
1013_2	77.6	54.8	25.0	63.3
1014_1	80.0	78.8	33.3	76.9
1014_2	61.5	65.4	80.0	62.9
1015_2	76.0	63.3	37.5	66.9
1016_1	66.1	46.8	33.3	54.8
1016_2	76.2	52.1	100.0	65.6
1017_1	65.9	50.5	20.2	53.8
1017_2	94.4	72.7	100.0	83.7
1024_1	81.5	70.3	30.0	71.9
1030_1	92.0	71.2	66.7	79.8
1030_2	70.0	55.9	50.0	61.8
1001_2	75.0	56.7	100.0	67.2
1003_1	91.2	72.5	50.0	78.0
1004_1	72.7	68.0	0.0	66.0
1005_2	82.6	60.3	100.0	71.8
1024_2	81.8	46.2	60.0	62.3

表8 文章構造解析の評価結果（科学技術文献等）
Table 8 Evaluation of Text Structures Analysis (technical articles)

文章名	グループの妥当性	関係名の妥当性	文章構造の妥当性	総合評価
tabako	90.0	56.8	-	72.6
0151	71.0	53.8	0.0	56.2
0152	69.4	47.3	0.0	54.9
p90-9	92.9	71.4	66.7	79.4
p91-9	83.3	58.3	-	70.8
p92-3	77.8	50.0	-	63.9
zen91	87.5	50.0	-	68.8
seika91	73.3	62.5	50.0	65.7
ibunka	31.3	17.1	-	23.9
chigai	53.8	38.5	-	46.2
5g	41.7	42.5	-	42.1
mazu	44.7	31.6	-	38.2
icot90-9	81.8	75.5	-	78.3
icot91	66.7	68.8	-	67.7
rdg92	75.0	61.9	-	68.3