

文章の表現形式に基づいた要約文章の生成について

田村 俊哉 田村 直良

横浜国立大学 工学部 電子情報工学科

E-mail: {tamtam, tam}@tamlab.dnj.ynu.ac.jp

概要

本稿では、論説文を対象として、文章の表現形式に基づいて要約文章を生成する手法について述べる。まず、本研究の文章要約モデルを示し、表現形式による規制を受けた論説文の構成を明らかにする。その上で、文章の構造、話題の連鎖において文章の構成を捉えることができるという観点から文章理解についての方針を示す。そして、文章の展開を考慮して、論説文の構成に基づいた論旨の収集規則を定める。さらに、収集した論旨を用いた要約文章の談話構造の組み立て方を示す。

A Summary Generation based on The Form of Text

Toshiya TAMURA and Naoyoshi TAMURA

Division of Electrical and Computer Engineering,

Faculty of Engineering, Yokohama National University

Tokiwadai, Hodogaya-ku, Yokohama, 240 Japan

E-mail: {tamtam, tam}@tamlab.dnj.ynu.ac.jp

Abstract

In this paper, we propose a method to generate a summary of an editorial, based on the *form of text*. At first, we show our model of the text summarization, and discuss the constitution of editorials which is restricted by the expression form of text. we explain an approach for the text understanding, from the viewpoint of the constitution of text according to the *text structure*, e.g. discourse structure, rhetorical structure and the *sequence of topics* in the text. Taking an expansion of the whole text into consideration, we define *extracting rules of theme*, based on the way of writing style of editorials. We also present a method to construct a discourse structure of a summary with the extracted theme.

1 はじめに

論説文などの論理的な文章は、筆者の主張を読者に効果的に伝えるという目的の文章である。そのため、一般的な事実から筆者の主張を正当化する論証を基調とした論旨の展開が図られており、比較的論旨が読み取りやすい文章である。文章の要約において、文章の論旨を捉えることが重要であるということは一般的な見解と思われる。文章構造から文章の論理的展開の把握を試みたものとして、内海ら [2] や岡村ら [5] の研究がある。また Schank らはスクリプトを用いた要約 [4] を示しているが、物語を対象としたものであり、部分的な情報の集約に基づいており論旨の把握は考慮されていない。

論旨を捉える上で、「理解の程度」についての問題を考える必要がある。言語表現とは、個々の書き手の主観的な脈絡を一般化して、言語形式に定着させたものであり、文章とは客観的な存在である。理解は、読み手がその客観的な文脈をたどって自らの主観的な脈絡を作り出すことである [6]。しかし、書き手の主観的な脈絡を作り出せる保証はない。理解というはたらしの特性上、読み手の主観により変容がなされる可能性があるからである。つまり、書き手の主観的な脈絡には客観的な文脈だけでは読み手が入り込めない領域が存在しているのである。

本研究では、扱う個々の情報の理解の深さには限度があることを認識した上で、その限度を客観的な文脈の理解に置いて文章の要約処理をモデル化する。文章表現には、文章の目的に応じ、説明的な表現や説得的な表現などの形式があり、これらは客観的な文脈を理解し、論旨を捉えるための手掛かりとなる。

以上のことを踏まえて、本稿では、論説文を対象に取り上げ、文章の表現形式に基づいて要約文章を生成する手法について検討する。また、原文章を 10% 程度に要約し、かつ結束性のある要約文章の生成を念頭におく。論旨を把握せずには適切な文章要約はできず、原文章を 10% 程度に要約した文章においてそのことが示されると考えられるからである。

2 論説文の文章要約

2.1 文章要約モデル

長文に対する大意や要旨を生成するためには、段落単位の構成を理解し操作しなければならず、文章の表現形式により大意や要旨の捉え方が異なると言われる [7]。

文章の表現形式には文章様式の特徴が現われる。論説文の場合には、筆者の主張を読者に伝えることを目的とするため、筆者の主張の正当性や妥当性を論理的、解説的に述べ、読者を説得させるように効果的な表現形式がとられる。人間は文章様式に応じた文章の表現形式に関する知識を持っていて、文章の要約においてこの知識を

有効に活用すると思われる。論説文の場合には、筆者の主張を基調とした要約文章を作ることが一般的見解であろう。

そこで、本研究では論説文の文章要約に際し、特定の専門的知識に基づいた推論により筆者の脈絡を類推するのではなく、論説文の表現形式から客観的文脈レベルで論旨を捉え、その上で文章を要約することを考える。実際に、人間は文章中の事柄に関する専門的知識を持たなくても、個々の事柄の重要性を文章から読み取ること、ある程度望ましい文章の要約が可能であろう。

文章論の研究では、文章の構成が文章様式に応じた表現形式により規制されることが示されており [6] [8]、文章の表現形式は文章構成において扱うことができる。

以上のことを踏まえて、文章の要約過程を図 1 のようにモデル化する。まず、文章理解では、文章の表現形式による規制を受けた文章構成を把握する。そして、論旨の収集では、文章構成に基づいた抽出規則により理解結果から論旨を収集し、談話構造の構成において、収集した論旨を用いて要約文章の談話構造を組み立てる。

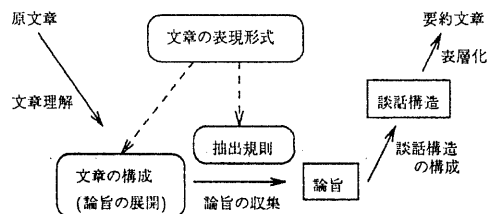


図 1 : 文章要約モデル

2.2 論説文の文章構成

文章の構成法は、文章を幾つかの部分に分け、各部分の機能の観点から文章の構成を考えるものと、文章に用いられる題材の性質を考え、その配列をもって文章の構成を考えるものに大別される [8]。一般に、両者は互いに排他的なものではなく、大局的には前者、局所的には後者というように重なることが多い。ここでは、表現形式により規制された論説文の構成を双方の見方において明らかにする。

2.2.1 大局的構成

論説文のトップレベルでは、導入部、論証部、結論部からなる定型的な構成が認められる。導入部では筆者の主張を想起した事物・事象に関する話題を提示し、論証部では提示された話題を受けて論証を行い、結論部では全体をまとめて結論付ける。

トップレベルの各部分の構成要素として、意味段落¹が

¹段落には一般に形式段落と意味段落があるが、形式段落は文字どおり形式的なものであり、特に意味を持つものではないので扱わない。

認められる。意味段落については諸説があるが、相原[8]にしたがい内容的にまとまりのある文の集まりを意味段落と呼ぶことにする²。意味段落は内容の面では思考の流れを文節的にまとめ[7]、文章を構造的に表現して読み手の便宜をはかるものである。特に、筆者の主張を基調とする論説文の場合には、意味段落は論証³の形態としてのまとまりであると思われる。また一般に、意味段落には中心的な話題が存在し[8]、話題の転換が文章の展開を形成する。話題が転換するといっても文章の展開上話題間には関連があるわけで、この意味で、意味段落間には話題によるつながりがあると考えられる。

(性質1) トップレベルの構成は、導入部、論証部、結論部からなる定型的な構成をとり、導入部から結論部に至る話題の展開がある。

(性質2) 意味段落は、論証の形態としてのまとまりである。

2.2.2 局所的構成

主張は筆者の主観によるものなので、読者を共感させるためには主張の正当性を示す必要がある。そのため、常識的な事実から主張を導き出す論証が行われる。

論説文の構成要素は、筆者の考えや意見を述べる主張文と一般的事実を述べる叙述文に分類できる[1]。さらに、主張文が中心で叙述文は主張文を補助する役割があり、主従関係という意味で叙述文は全ていずれかの主張文に従属すると考えられる⁴。局所的な構成では、接続関係により文がまとまって論証の形態をとる。

(性質3) 論説文は、主張文と叙述文で構成される。

(性質4) 局所的構成は、文間の接続関係を用いて1つの構造で表現することができる。主従関係という意味で、基本的には叙述文は主張文に従属する。

3 文章理解

文章理解は、表現形式による規制を受けた論説文の構成の把握であり、文章構造と語句の連鎖に着目した話題構造の観点から捉える。ここでは、文章理解の方針について述べる。

3.1 文章構造

論説文の文章構造は、2.2節に基づいて図2に示した構造とする。大局的なレベルでは、論証の部分構成する意味段落が不確定なことを除けば、定型的な構造であ

²あるいは、幾つかの形式段落の集まり。

³ここでは、論証とは論理に基づいて結論を導き出すことである。

⁴直接従属しない叙述文でも、主従関係をたどり間接的に主張文に従属するものも含む。

る。この文章構造は、性質の違いはあるが、ある程度の長文であれば論説文以外の文章様式にも当てはまる。

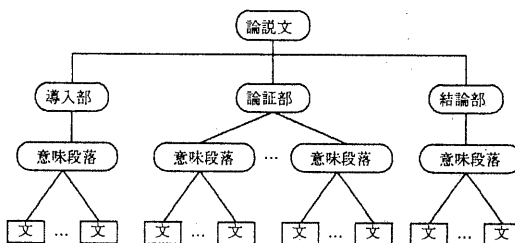


図2：論説文の文章構造

修辭構造理論[3]は、文章の構造を記述するための枠組であり、文間の修辭関係に基づいて文章構造を木構造で表現する。性質4から、意味段落内の構造を表現するために修辭構造理論の基本的な考え方が利用できる。

しかし、修辭構造理論では、定義上接続関係を適切に表現できないという問題があり、文章理解には適していないと思われる。原因は、木構造の定義上子ノードが親ノードを複数持たないことにある。論説文の場合、主張文と叙述文間の修辭関係は主張文が nucleus、叙述文が satellite と考えるのが妥当である⁵。一般的事実を示した叙述文が2つの主張文に参照される場合、叙述文は各々の主張文と接続されるべきである。しかし、修辭構造理論では一方の関係を切らなければならない、適切な接続関係が切れてしまう可能性がある。

文章を構造化する場合、文章の展開を考慮して考えられる接続関係の中から適切なものを選び、見方を限定すべきと思われ、修辭構造の定義が見方を制限してしまうことは適切ではない。

また、修辭構造理論では、スパンに含まれるユニット全体がそのスパンの関わる修辭関係に関与すると解釈される。つまり、スパン全体が均一化される。そのため、修辭関係に関わるものがスパン内の一部のユニットなのか、スパン全体なのか表されない。このことも、適切に接続関係を表現できない原因の1つである。

このような問題を考慮して、本研究では修辭構造理論を改良して用いる。

定義1(修辭構造の基本単位) 基本単位は、修辭関係で接続した2つのユニット(単文)に着目した2項関係とし、修辭関係を rl、nucleus、satellite に相当するラベルを n、s とし次のように記述する。

$$(rl, n/segment, s/segment)$$

また、並列関係のように主従関係が認められないものは次のように記述する。

$$(rl, n1/segment, n2/segment)$$

⁵叙述内容を補足説明する主張文のような例外もある。

segment は、ユニット、もしくはグループ化したユニットの集まりである。グループ化は、修辭構造理論のスパンとは異なり、まとまりが認められるものに対して行なう。

[グループ化の条件]

1. 主従関係の認められない文の集まり。
2. 重文で表される文の集まり。
3. 指示語により参照される文の集まり。

定義 2(修辭構造ネットワーク) 修辭構造は、木構造のように階層化せずに、基本単位の集合で表されるネットワーク構造とする。図 3 は修辭構造ネットワークを图示したものである。ノードにはユニットノードとグループノードがある。アークは修辭関係を表し、アークの先が nucleus、アークの根が satellite である。また、主従関係が認められないものは、双方の先を持つアーク(⇔)で表す。さらに、性質 3 を考慮して、主張文と叙述文を区別する。

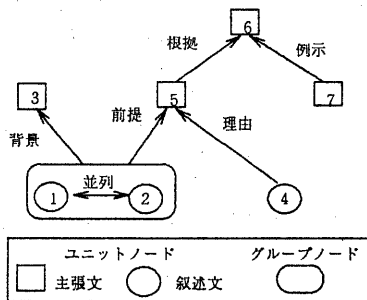


図 3 : 修辭構造ネットワーク

3.2 話題構造

本研究では、主張の対象として取り立てられる事物を表す名詞句や事象を表す動詞句を話題と呼ぶ。

性質 1 のように、論説文では導入部から結論部に至るまで、中心的な話題に関連する話題や対照的な話題を取り込んだ展開が見られる。永野 [6] は、語句が文の連鎖の中に位置付けられ文章の構成の支えとなっているという見解から、語句の連鎖に着目し、文章構造の解明の 1 つの手段としている。このように、語句も文脈を形成する構成要素と考えられる。話題の展開に着目する目的は主張で取り立てられる事柄を捉えることであり、このことは文章の論旨を把握する手掛かりになる。

定義 3(話題構造) 話題構造は、文章中の話題の展開を表すもので、ネットワーク構造で表現する。ノードは話題、アークは話題間の関係を表し、アークの根が展開元の話題である。話題間の関係については、同義関係、派生関係、提起関係の 3 種類の関係がある。

● **同義関係 (Equivalent Relation)** 同義関係は、同一語句、もしくは同義語の話題間の関係。

例) 「方法」⇔「手段」、「償う気持ち」⇔「謝罪の意」

● **派生関係 (Derivation Relation)** 派生関係は、派生した話題と派生元の話題の関係。

1. 一方の話題が他方の構成要素の一部となっている。

例) 「中国」⇔「中国の影響力」、

2. 現象を名詞化したもの。

例) 「株主優遇を求める声が出ている」→「株主優遇を求める声が出ていること」

3. 一般的知識によるもの⁶

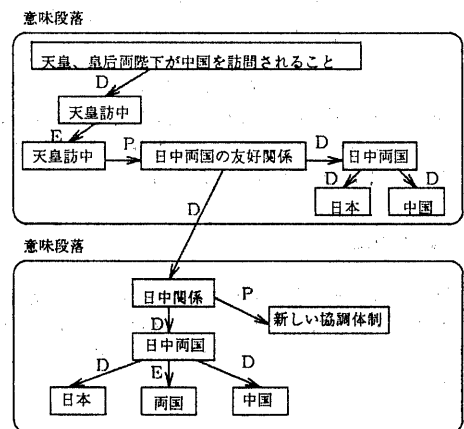
例) 「証券」⇔「株式」

● **提起関係 (Presentation Relation)** 提起関係は、一文内での既出の話題とそれにより提起される新出の話題の間の関係⁷。

例) 以前に「天皇訪中」が提示されていて、

天皇訪中を契機に、日中両国の友好関係がさらに発展することを期待する。

という文で「日中両国の友好関係」という話題が提起される場合、「天皇訪中」→「日中両国の友好関係」



注) '文'に関しては簡単化のため省略している

図 4 : 話題構造

話題構造は、主張文で取り立てられる話題を起点とし、次の規則を関係付けられる話題に対して繰り返し適

⁶これについては、一般的知識に関するシソーラスを想定している。

⁷これは、いわゆる問題提起文と呼ばれる文に相当する。

用して構成する。ただし、同義関係を派生関係より優先し、同じ関係については距離の近いものを優先する。

- 意味段落内に注目している話題と上記の関係を満たす話題が存在すればそれらを関係付ける。
- 意味段落内に上記の関係を満たす話題が存在しない場合には、以前の意味段落に上記の関係を満たす話題が存在すれば関係付ける。複数の意味段落について可能ならば、全てを関係付ける。

4 論旨の収集

2.2節で明らかにしたように、文章表現は文の接続による論理的な展開により段落を形成し、さらに段落を単位とした大きな展開が文章全体を形成することで完成している[7]。したがって、文章の論旨を捉えるためには、文章の展開上不可欠である段落を意識するべきであろう。そこで、論旨の収集は文章の展開に基づいて、大局的な構成を捉えたいうえで局所的な構成に視点を移して行なう。ところで、ここで扱う情報は文章ではないので命題と呼んで区別する。

4.1 意味段落の選択

意味段落間の話題のつながりは、話題構造では意味段落間にまたがるアークが表している。性質1に基づいて、話題のつながりに着目することで、意味段落を単位とした大まかな文章の展開を断片的に捉えることができる。話題構造における導入部から結論部に至る話題のつながりを話題の連鎖と呼ぶことにする。そこで、話題の連鎖に基づいて意味段落を選択する。

話題には、新出の話題や既出の話題から派生した話題がある。話題構造において展開元の話題を原始話題と呼ぶ。各原始話題の優先度を判定するために、展開度を定義する。

定義4(話題の展開度) 話題の展開度は、話題構造においてその話題が同義、派生関係をたどることで得られる子孫ノード数の総和とする。

また、特に展開度が高くなくてもタイトルを構成する話題⁸は重要と思われ、この話題を中心話題と呼ぶことにする。

定義5(意味段落の選択) 意味段落の選択は、要約化率⁹を考慮して結論部から導入部まで逆戻りして行なう。選択した意味段落に入る連鎖に着目し、原則としてその意味段落と連鎖が最も密な意味段落を選択する。ただし、連鎖に偏りがない場合には各話題の優先度や中心話題を考慮する。

⁸これは、文章の主題(中心思想)が表出したものと考えられる。

⁹凡そ原文章と要約文章における文数の比。

4.2 意味段落内の命題の選択

4.2.1 修辞関係に関する命題の選択条件

ここでは、修辞構造ネットワークにおけるユニットノードを単一命題、グループノードを複合命題と呼ぶ。

意味段落ごとに中心的な主張が存在し、意味段落はそれをゴール¹⁰とする論証の形態をとる。このような観点から、修辞関係は論証に寄与する修辞関係(typeA)、複合命題を形成する修辞関係(typeB)、論証に寄与しない修辞関係(typeC)の3種類に分類できる。ただし、特殊な場合としてグループ化の条件2.、3.を満たす場合には、typeAの修辞関係も複合命題を形成しうる。

タイプ	修辞関係
typeA	根拠、理由、展開、前提、背景
typeB	並列、換言、選択、対比、比較、typeAの特殊な場合
typeC	説明、補足、例示、累加

表1: 論証性に基づいた修辞関係の分類

typeAの修辞関係を持つ2項関係を論証規則とみなす¹¹と、論証規則をゴールまでたどった軌跡に含まれる命題が重要な命題と考えられる。

定義6(修辞関係に関する命題の選択条件)

R1 選択した命題をnucleusとするtypeAの修辞関係のsatelliteに相当する命題を選択する。

ただし、satelliteに相当するものが複合命題の場合の選択は、複合命題のタイプにより2つに分類できる。

R2 typeBの修辞関係による複合命題の場合、複合命題中の単一命題を単位として選択する。

R3 typeA、及びグループ化の条件3.を満たすtypeBの修辞関係による複合命題の場合、複合命題を単位として選択する。つまり、複合命題中の各単一命題は連動して取捨されなければならない。

4.2.2 話題に関する命題の選択条件

修辞関係に基づいた命題の選択条件は、意味段落内での局所的な文脈のみに関するものであり、文章全体の展開も考慮する必要がある。

まず、意味段落内のみで結束性のある文章が構成できても、意味段落間に結束性がなければ文章全体の展開が不自然になるので、意味段落間の結束性の要因となる話

¹⁰文章全体からみると、意味段落内のゴールはサブゴールに相当する。

¹¹実際の論説文の場合、完全に論理的に説明しきれものではないので、具体的な論証の戦略に当てはめるわけではなく、単に2項関係($r1, n/p1, n/p2$)を論理式 $p2 \stackrel{r1}{\rightarrow} p1$ のように抽象化して考えている。

題を選択する必要がある。また、文章のキーワードに相当する中心的話題も選択すべきである。

また、これらとは異なるものであるが、談話のダイクシスを考慮しなければならない。つまり、選択した命題の中に他の未選択の命題を参照しているものがある場合、参照先の命題を選択する必要がある。

定義 7(話題に関する命題の選択条件)

T1 話題構造において選択した意味段落間にまたがるアークによって関連付けられた話題を含む命題を選択する(意味段落間の結束性)。

T2 最も優先度の高い話題、及び中心話題を含む命題を選択する(中心的話題)。

T3 選択した命題が未選択の命題を参照している場合、参照先の命題を選択する(談話のダイクシス)。

4.2.3 命題の選択規則

一般に、意味段落内で局所的な話題を用いてさらに論旨の展開が図られるため、ゴールに至る軌跡は1つとは限らない。軌跡に含まれる話題を考慮することで各軌跡を評価して、論旨に相当する軌跡に絞ることができる¹²。

定義 8(命題の選択規則)

S1 意味段落内のゴールに相当する命題から **R1** を満たす命題を選択して可能な軌跡を全て選択する。ただし、軌跡中に typeB の修辭関係による複合命題がある場合には、図5のように軌跡が分裂する場合がある。分裂しない場合には、複合命題を1つの命題として選択する。

S2 **T1**、**T2** を満たす命題を含む割合で各軌跡を評価する。

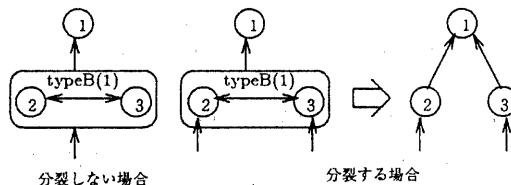


図5：軌跡の分裂

その上で、要約化率を考慮して、以下の規則にしたがって高い評価の軌跡から順次命題を選択する。

S3 ゴールから **T1** を満たす命題までと、残りのうちの **T3** の条件を満たす命題を選択する。

¹²先述したように、語句は文の連鎖に位置付けられ文章の構成の支えになる[6]ということから、命題の選択規則では修辭関係に関する選択条件を話題に関する選択条件より優先する。

S4 複合命題を含む場合には、**T1**、**T2** の満たす命題を重要性の判定基準として **R2**、**R3** を適用する。

S5 選択された命題について **T3** を満たす命題があれば選択する。ここで選択する命題は、typeA の修辭関係である必要はない。

5 談話構造の構成

トップレベルの構成は、要約文章の構成も原文章と同様とする。理解では、このレベルの接続は、修辭関係ではなく話題によるつながりであると考えた。しかし、生成ではトップレベルの構成要素間の接続関係は、生成のためのみに用いる“転換”という修辭関係で表して、要約文章の談話構造を1つの木構造で表現する。また、論証部を構成する意味段落は1つとは限らず、要約化率によっては複数でありうる。

局所的な構成では、論説文の論証性が現れる。これは筆者の主観に基づいたものであり、論証の手順については原文章を維持すべきと考えられる。そこで、命題間の接続関係は原文章の修辭関係をを用い、強い接続関係を優先して見方を限定することで、意味段落ごとに部分木を構成する。

定義 9(部分木の構成規則) 選択した命題が全てなくなるまで、以下の適用規則にしたがい部分木を構成する。

1. ゴールを含む2項関係を root として、2. の操作を繰り返す。
2. 部分木に含まれる命題と同じ命題を nucleus にもつ2項関係が存在するならば、部分木の該当する命題にその2項関係を挿入する。

要約文章の談話構造は、トップレベルの定型的な構成に意味段落ごとの部分木を組み入れて構成する。

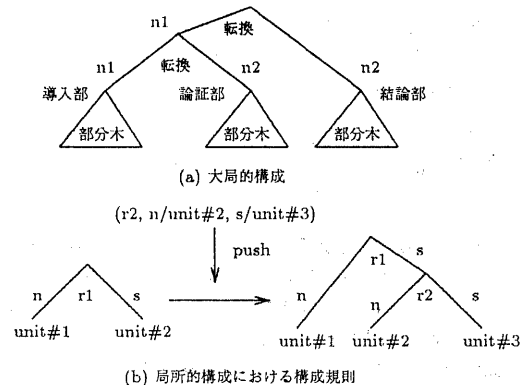


図6：話題構造の構成

6 思考実験による検証

6.1 思考実験

本稿では思考実験により本手法を検証する。対象とした文章は、平成4年8月26日付け毎日新聞社説記事の「新しい日中関係を築くために」であり、約1400文字からなる文章である。この文章は、4つの意味段落(論証部が2つの意味段落からなり、ここでは論証1、論証2とする)から構成されており、定義5により選択される意味段落は「導入、論証1、結論」である。文章例1は「論証1」の部分であり、各文頭の数字は修辞構造ネットワークにおけるユニット番号を表す。また、図7は「論証1」の修辞構造ネットワークである。

ここで、命題の収集規則(定義8)にしたがい「論証1」における命題の選択を行なってみよう。

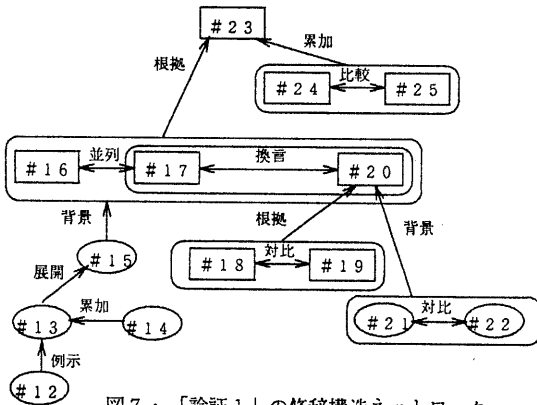


図7: 「論証1」の修辞構造ネットワーク

1. S1により選択される軌跡、及びS2による各軌跡の優先度は表3のようになる¹³。ただし、定義7の条件となる話題は、T1について「日中関係」(導入部との結束性)、「新しい日中関係」と「アジアの平和」(結論部との結束性)、T2について「新しい日中関係を築くために」(中心話題)、「日中関係」(最も優先度の高い話題)である(表2参照)。

話題	話題を含む命題
日中関係	#15, #16, #23
アジアの平和	#17, #20
新しい日中関係を築くために	#23

表2: T1、T2の条件となる話題とそれを含む命題

¹³ 評価は、軌跡に含まれる各命題が満足する個々の話題に関する条件数の和の比較で行なう

優先度	選択される軌跡
1	[#23, (#16, #17), #15, #13]
2	[#23, #20, (#18, #19)]
2	[#23, #20, (#21, #22)]

表3: 選択される軌跡と優先度

2. 次に、最も評価の高い軌跡について、S3からS5を行なう(表4参照)。この場合には、「アジアの平和」と「新しい協調体制」はともに重要であるので、S4では命題15、16とも選択される。

規則	選択される命題
S3	[#23, (#16, #17), #15, #13]
S4	[#23, (#16, #17), #15, #13]
S5	—

表4: 各規則での選択される命題

3. この結果、論証1では[#23, #17, #16, #15, #13]という命題が選択される。

本稿では述べなかったが、本手法では選択された命題に対しさらに概念単位の選択を行なう。これは、T1、T2に関する話題を含むか否かで各命題の任意格を選択するものである。

全体としては、「導入[#3, #2]、論証[#23, #17, #16, #15, #13]、結論[#40, #37]」のように命題が選択される。最後に、構成規則(定義9)に基づいて、図8に示す談話構造を組み立てる。この談話構造から、文章例2に示した要約文章が期待できる。

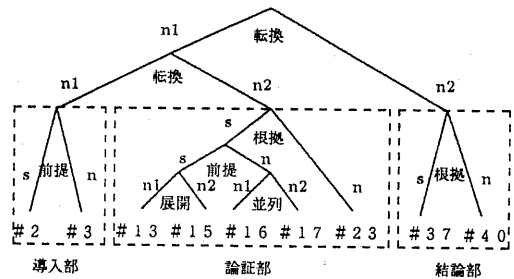


図8: 要約文章の談話構造

6.2 要約文章の評価

文章例2の要約文章について、我々は重要な情報を含み、かつ結束性のある文章であると評価しており、本手法の妥当性が示されていると考えている。この文章では原文を約17%に要約しているにとどまっている。その理由としては、新聞社説は紙面の都合上無駄なく簡潔にまとめ挙げられていること、本手法では簡潔な言葉での言い換えなどの情報の集約処理を取り入れていないこと、が挙げられる。

文章例

文章例1(意味段落「論証1」の原文)

(12) 中韓国交樹立やカンボジア和平の動きにみられるように、(13) いまや冷戦構造はアジアでも崩壊している。(14) 20年前の日中国交正常化当時に比べて、国際政治の枠組みは一変した。(15) そのことは当然、日中関係にも大きな影響を及ぼさずにはおかない。(16) 日中両国は、新しい協調体制を築くとともに、(17) アジアや世界の平和のために積極的な役割を果たしていかねばならないだろう。(18) 日本は世界経済の15%を占める経済大国であり、(19) 12億弱の人口を擁する中国は巨大な潜在的可能性をもつ大国である。(20) 両国は冷戦後のアジアや世界の平和秩序と繁栄のためにその能力を活用すべきである。(21) 朝鮮民主主義人民共和国(北朝鮮)の核査察受け入れなど朝鮮半島をはじめとするアジアの緊張緩和は、中国の影響力を抜きには考えられない。(22) 日本の経済力や外交力も問われている。(23) 未来を志向する新しい日中関係を築くには、歴史の過ちを反省し、誠意をもって償う気持ちがなければならないだろう。(24) 他国からやかましくいわれて渋々認めるのではなく、(25) 国のモラルの問題として自らそれを対外関係の基本に据える姿勢が必要である。

文章例2(要約文章 - 約17%)

日中国交正常化二十周年の節目に天皇訪中が実現する。これを契機に、日中両国の友好関係がさらに発展することを望みたい。

冷戦構造はアジアでも崩壊しており、そのことは日中関係にも影響を及ぼさずにはおかないので、日中両国は新しい協調体制を築くとともに、アジアや世界の平和のために積極的な役割を果たしていかなければならない。新しい日中関係を築くためには、歴史の過ちを反省し、誠意をもって償う気持ちがなければならない。

日中関係はアジアの平和と安定にとって重要なので、天皇訪中が互いにものを言い合える新しい日中関係を築く契機となることを期待したい。

7 おわりに

本稿では、論説文を対象として、文章の表現形式に基づいて要約文章を生成する手法について検討した。思考実験による検証とはいえ、本手法の妥当性を示し、文章様式に依存した表現形式を手掛かりとした客観的な文脈の理解によって、ある程度適切な要約文章を生成できることを示した。

文章の論旨を捉えるためには、様々な制約を考慮する必要がある。本手法は文章論的な観点からの制約を考慮したものであり、本手法によりある程度論旨を絞り込むことができる。我々は、本手法が文章の要約処理において必要十分であると言うわけではなく、意味論的な観点からの制約などとの併用も必要となると考えている。

我々は、システムの実現に向けて、論説文の調査分析を進め、要約化率を含めた論旨の収集規則、及び話題構造の定義付けについてさらなる検討を行なっている。また、同時に文章構造、話題構造の解析手法についても着手している。

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. 情報処理学会研究報告, Vol. 78, No. 15, Jul. 1990.
- [2] 内海功朗, 重永実. 英語文章の大意生成. 情報処理学会研究報告, Vol. 54, No. 8, 3 1986.
- [3] W. C. Mann. Rhetorical structure theory : Description and construction of text structure. In G. Kempen, editor, *Natural Language Generation*, pp. 279-300. Martinus Nijhoff Publishers, 1987.
- [4] R. Schank and R. Abelson. *Script Plans Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
- [5] 岡村秀一, 田村直良. 要約のモデル化及び、そのプランを用いた実現についての検討. 日本ソフトウェア科学会第8回大会論文集, pp. 565-568, 1991.
- [6] 永野賢. 文章論総説 - 文法論的考察 -. 浅倉書店, 1986.
- [7] 所一哉. 日本語思考のレトリック. 匠出版株式会社, 1986.
- [8] 相原林司. 文章表現の基礎的研究. 明治書院, 1984.