

## 文章構造に着目した要約に関する研究\*

3K-1

大島 隆義

田村 直良†

横浜国立大学 工学部‡

## 1 はじめに

実際に文章を要約するためには、人は意味のレベルまで踏み込んだ種々の解析を行なっている。しかし、市川[3]は「文の接続・配列関係」や「繰り返し語句」などを手掛かりにした客観的な要約が、ある程度可能だとしている。そこで本研究では文章構造が比較的明確で、筆者の主張、主題を把握しやすい新聞社説を対象として、その文章構造や語句の連鎖に着目することにより、文章を要約する手法を示す。また、本研究では要約文章の評価の基準として要約率だけでなく、話題の連鎖、文章構造的な面からも文章評価を行なう。

## 2 要約モデル

要約文章を生成する過程は大きく3つの過程、文章理解過程、文章再築過程、評価過程に分けられる。本研究では、要約文章の評価を行なう評価過程を設けることにより、品質の高い要約文章の生成を目指している。生成された要約文章の評価によっては要約文章の練り直しが行なわれる。(図1参照)

## 3 文章理解過程

要約を目的とした文章理解に限らず、文章を理解する上で、次の3項目に着目することが最低限必要ではないかと考えられる。

## (1) 中心文 (2) 文章展開 (3) 話題

論説文や小説、詩など文章が属する分野によって、どういった文が中心文にあたるかなど、それぞれ異なるが、本研究で対象としている新聞社説では比較的明確に文の表層に表れており、現段階で解析することが十分可能だと思われる。

## 3.1 中心文の把握

筆者の主張を伝えることを目的とする新聞社説では中心文は主張文にあたる。したがって中心文は文章中に1

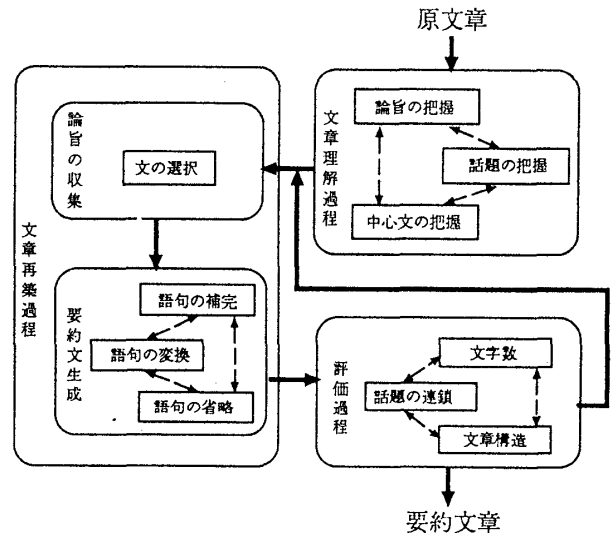


図1: 要約モデル

つとは限らず、複数個存在する場合もある。

筆者の主張を表す主張文と、事実や現象を表している文のタイプを区別することは、「～べきだ」などの文末表現と「大切なのは～」など、ある特定の単語を含む文のパターンマッチングを行なうことで可能である[2]。

## 3.2 論旨の把握

主張の正当性を読者に伝えるための論旨の展開は、文章の展開となって表れる。そして文章の展開は修辭構造として捉えることができる[1]。本研究において、修辭関係を同定するために使用される文の情報は(1) 接続表現 (2) 文のムード の2項目である。これらは文の表層に表れており、計算機上での実現を考えた場合に使用できる情報といえる。

## 3.3 話題の把握

文章中で接続表現が表れていない場合でも、我々は文間に何らかの継りを見出し、文章を理解している。それは文章中に表れる語彙が継りを持っている、語彙結合

\*Summary Generation based on The Text Structure

†Takayoshi OSHIMA, Naoyoshi TAMURA

‡Faculty of Engineering, Yokohama National University

のためであると考えられる。我々が特に前後の段落間で意味の継りを感じるのは、話題の連鎖によるところが大きい。また、以前に表れた語句、内容を先行詞とする照応語を用いることで文間、段落間の継りを保証することもある。

新聞社説を対象として形式段落内における照応語を含む文の位置と、照応語の先行詞の関係を調査した結果、以下のような傾向が得られた。

#### 【傾向 1】

形式段落の第 1 文目に照応語が存在する場合、先行詞は直前の形式段落中に存在することが多い。

#### 【傾向 2】

形式段落の第 2 文目以降に照応語が存在する場合、先行詞は同一形式段落中に存在することが多い。

## 4 文章再築過程

筆者の論旨の展開を表す修辞構造などから、要約文章として取り上げる文を決定する。

### 4.1 論旨の収集

どの文を要約文章に取り上げるかを決定する過程である。以下のような制約の下での文の選択が考えられる。

#### 【修辞関係からの制約】

筆者の主張の正当性を論証するためにはさして重要ではない文を、修辞関係の satellite を削除することにより、論旨の収集を図る。重要度の低い修辞関係としては、例示、換言、補足、強調が挙げられる。

#### 【省略・照応からの制約】

候補文中に省略・照応表現が存在している場合には、候補文の省略・照応の先行詞が特定されるか、その先行詞が存在している文もしくは形式段落も採用されなければならない。

#### 【文内の制約】

同一文で複数の事象が述べられている場合、文の後半ほど重要度が高い場合が多い。

## 5 要約文生成過程

文の結束性を高めるため、および要約率を向上させるための処理を行なう。

#### 【語句の補完】

文間の関係を明確にし、文章の読み易さを高めるために接続詞を補う。

#### 【語句の変換】

現象文の名詞化など、簡潔な語句へ変換する。

#### 【語句の省略】

省略表現とすることにより要約文章の結束性、読み易さを高める。また不要な修飾語句を削除する。連体修飾語句では「外の関係」にある連体修飾節は、「内の関係」にある連体修飾節より、要約文章中に残り易い傾向にある [4]。

## 6 評価過程

生成された要約文章を評価する過程である。評価の結果により再び文章再築過程に戻る場合もある。

#### 【文字数からの評価】

本研究では原文の  $\frac{1}{3} \sim \frac{1}{4}$  の要約率を目指している。

#### 【話題連鎖からの評価】

形式段落は筆者の主張を成り立たせるための論証単位なので、原文における形式段落内の話題は、連鎖していると考えられる。しかし要約文章では文の削除により、話題が途切れる場合が考えられる。話題連鎖を保つことで、文章の首尾一貫性を評価する。

#### 【文章構造からの評価】

原文の文章構造と比較することにより、要約文章の段落構成などを評価する。

## 7 おわりに

本研究では新聞社説を対象として、その要約モデルおよびその実現に際しての着眼点を示した。本モデルにおいては要約率のみではなく、生成された要約文章を話題の連鎖、文章構造から総合的に評価、再生成することにより品質の高い要約文章を目指したものである。しかし評価基準の設定、要約文章を再生成する場合の文選択基準の変化など検討課題は多く残されている。

## 参考文献

- [1] 福本淳一, 安原宏. 文の接続関係解析に基づく文章構造解析. 情報処理学会研究報告, Vol. 88, No. 2, 1992.
- [2] 山本和英, 増山繁, 内藤昭三. 文章内構造を複合的に利用した論説文要約システム green. 情報処理学会研究報告, Vol. 99, No. 3, 1994.
- [3] 市川孝. 国語教育のための 文章論概説. 教育出版社, 1978.
- [4] 佐久間まゆみ (編). 文章構造と要約文章の諸相. くろしお出版, 1989.