

# 直接引用表現を利用した要約知識の自動抽出の試み

児玉 充† 片岡 明† 増山 繁† 山本 和英‡

kodama@smlab.tutkie.tut.ac.jp, kataoka@smlab.tutkie.tut.ac.jp,

masuyama@tutkie.tut.ac.jp, yamamoto@itl.atr.co.jp

†豊橋技術科学大学 知識情報工学系

‡ATR 音声翻訳通信研究所

## 1 はじめに

近年、インターネット上で膨大な量の情報発信が行われている。また、各種出版物がCD-ROM等、機械可読媒体の形で流通するようになってきた。このような背景の下で、大量の情報の中から必要な情報のみを得るための技術として、テキスト要約は重要である。しかしながら、人手で多数の要約を作成することは困難であり、コストもかかるため、自動要約が望まれる。

自動要約には言語の表層情報を利用した様々な手法があり[1]、そのひとつとして原文とその要約結果の対応表を用いた要約手法がある。山崎ら[2]、および、若尾ら[3]は、ニュース文を対象として、文末の丁寧な表現からより簡潔な表現への言い換えや、文字数の多い語句から省略形など同意の語句への言い換え等、言い換え前の表現と言い換え後の表現の対応を示した換言テーブルを人手で作成している。しかし、多数の言い換えの表現を人手で記述することは困難である。

また加藤ら[4]は、コーパス(原文-要約文)のマッチングを行うことで、言い換えが行われている部分を特定し、それを要約知識として自動的に獲得する手法を提案している。この手法のように、原文とその要約結果が対応付けられたパラレルコーパスがあれば自動要約を行う上で有効である。しかし、そのようなコーパスを用意することは、一般には困難であるため、単一のコーパスから自動的に対応関係を発見し、それに基づき、要約知識を取得することが望まれる。

本研究では、単一コーパスからの要約知識抽出の試みとして、新聞記事から対応関係にある部分を特

定し、テキスト自動要約に利用可能な対応表を自動的に生成する方法を検討する。

新聞記事では、鉤括弧で括られた直接引用表現の直後の節が、その引用表現の要約である場合がしばしばある。

### 例1:

「小選挙区比例代表並立制であれば、野党や自民党の相当部分が賛同している。互いが話し合いながら妥協案を作ることができると述べ、並立制導入での各党合意は可能との見解を示した。

例1では、直接引用表現(鉤括弧で括られた部分)で述べられている内容が、記者によって要約され、直後の節(下線部)で簡潔に言い換えられている。このような人間(記者)による要約例と、その原文との対応関係を多く収集することで、人間が要約する際に使用した知識を収集することができると考えられる。

そこで、本研究では、直接引用表現に注目し、引用表現とその直後の節とが対応関係にあるか否かを判定する手法を提案する。さらに、取得した対応関係を自動要約に利用するために、対応関係から要約知識を抽出する方法を検討する。

引用表現の直後の節は、記者が内容を理解した上で言い換えている表現であることが多い。よって、本手法によって得られた要約知識を用いることにより、意味解析を実現するのと同等の品質を持った要約が実現できる可能性がある。さらに、上記の手法[2, 3, 4]の要約単位が語句であるのに対し、本手法で取得した対応関係は、節を単位とした要約にも利用可能である。

## 2 対応表の自動生成

ある文と、その要約文との対応関係を取得するために、記事中の直接引用表現に注目する。前述のように、新聞記事では、引用表現を用いて述べられた内容が、記者によって要約され、直後の節で再び述べられることが多い。ところが、以下の例2に示すように、すべての引用表現を含む文において、この現象が見られるわけではなく、要約が述べられているか否かの判定を行う必要がある。

### 例2:

「市内料金の引き上げでNTT株は百万円になる」と予言したとか、二月中から政治家が買い集めていたとかいうものだ。

直後の節が引用表現の要約である場合、どちらの節も、主部で示された人物が何らかの意見や考えを発言した、という意味を持つ。その結果、各節の末尾は、「～と述べる」、「～という見解を示す」といった陳述を意味する表現となることが多い。

以上の観察に基づき、本研究では、陳述表現を形式的に定義し、引用表現とその直後の節の末尾の表現がそれを含むか否かの判定を行う。各節の末尾がともに陳述表現であれば、二つの節が同一の文に存在していることから、それらに意味的な対応関係があると予想できる。以下の例3では陳述表現である「と述べ」、「要望した」から、引用表現とその直後の節が対応関係にあると判定する。

### 例3:

「この地域のほとんどの国が日本が他の国や国連と一緒に活動するほうが安心できると感じる」と述べ、日本が近隣諸国と積極的に意見交換することを要望した。

以下に示す定義(1),(2)は、日本経済新聞1993年1月-6月の記事(約10万)を参考にして作成し、また、{名詞}、{動詞}のリストも、それらの記事から人手で抽出した。

文が以下に示す形式であれば、対応関係を取得する。

#### (1)

「…」(など) { とし  
として  
と(の){陳述表現} } …{陳述表現}

陳述表現の形式を以下に示す。

#### (2)

{ ( { 名詞 } · ) { 動詞 } } { 名詞 } { している。  
する。  
した。  
た。  
す。  
る。  
。  
して  
し  
、  
だ。  
。 }

陳述表現の記述の詳細について以下に示す。

- ·: 任意の助詞
- …: 任意の文字列
- (): 省略可能

ただし、{名詞}は以下に示す語のみが該当する。

意向、認識、見解、見方、意欲、見通し、方針、姿勢、考え、構え、必要性、可能性、重要性、問題点、結果、狙い、判断、批判、弊害、判決、決意、懸念、期待、主張、訴え、政策、措置、期待、評価、決意

また、{動詞}は以下に示す語のみが該当する。

設立、求め、決定、要求、要請、語り、述べ、指摘、説明、主張、強調、評価、示、下、促、囑、明らかに、明確に、示唆、訴え、答え、明言、言明、力説、提言、証言、発言、認め、歓迎、指示、批判、判断、表明、否定、設立、懸念

## 3 実験

前節の定義を作成するのに使用した記事(日本経済新聞1993年1月-6月)を対象として予備実験を行った。鈎括弧を含む約8000記事に対して以下の条件をともに満たす文を抽出する。

- 鉤括弧内の文字数が5文字以上
- 鉤括弧を含む文が(1)を満たす

抽出の結果、765個の対応関係を取得した。しかし、取得した対応関係には鉤括弧内の節とその直後の節が意味的に同値でない場合も含まれる。取得した対応関係の評価を行うために、765個から無作為に選んだ95個の対応関係について、人手により以下の3つに分類した。

鉤括弧内の節とその直後の節の関係が

- 意味的に同値 (A)

例 4 :

「OPECのすべての加盟国がすでに反対の立場を明確にしている」と述べ、OPECが今後一致して新税導入に反対していく考えを表明した。

- 理由あるいは原因 (B)

例 5 :

「民主的な選挙の洗礼を受けた新政府ができれば、ポト派がそれにチャレンジするのは難しい」と述べ、選挙をできるだけかく乱するというポト派の姿勢は変わっていないという判断を示した。

- 意味的な関係がない (C)

例 6 :

「プルトニウムの在庫調整炉」として、高速炉の開発を続ける考えを表明している。

取得した対応表を分類した結果を表1に示す。

表1：対応関係の分類結果

	個数	%
意味的に同値 (A)	67	70.5
理由あるいは原因 (B)	15	15.8
意味的な関係がない (C)	13	13.7
計	95	100

## 4 考察

### 4.1 取得した対応関係の分類結果

表1の分類結果の中でAとBに属する対応関係は、他の文章の要約に何らかの形で利用できると考えられる。AとBに属する対応関係は、8割を占めた。

### 4.2 取得されなかった対応関係

取得されなかった対応関係の中にも妥当であると思われるものもあった。

- {動詞}, {名詞}の定義中に語が存在しなかったために取得されない場合  
以下の例では、「非難」が{動詞}に存在しないため取得されない。

例 7 :

「問題は全世界的なもの」と述べ、強く北朝鮮の姿勢を非難した。

- 引用表現の直後に陳述表現が存在しない  
以下の例では、引用表現を受ける陳述表現が存在しないため取得されない。

例 8 :

「直接、販売に結び付かなくても、消費者に新商品を提案する機会になる」と、客とのつながりを重視している。

### 4.3 対応関係の利用方法

取得した対応表から、他の文章の要約に利用することができる言い換え規則を抽出する。今回は試行として、人手によって言い換え規則を抽出した。以下に抽出された規則を示す。以下の例文において、下線部に着目して規則を抽出した。

例 9 :

「欧州の農業問題が経済・社会の死活にかかわる問題 なのに対し 日本のコメは日本固有の心理的問題という側面がある」と述べ、欧州の農業問題 と事情が違う との考えも示した。

規則 :

“Xが……なのに対しYは……”  
→ “(Yは)Xと事情が違う”

例 10 :

「金丸氏への盆、暮れの裏献金は あいさつ代わり だった」などと証言し、ヤミ献金が 日常的 に行われていたことを認めた。

規則 :

“あいさつ代わり” → “日常的”

## 5 むすび

本研究では、直接引用表現に注目し、単一コーパス(新聞記事)から、自動要約に利用可能な対応表を生成することを試みた。また、対応表の自動要約への利用を検討するために、取得した対応関係から人手で要約知識を抽出した。

以上の実験から明らかになった本手法に関する今後の課題を以下に示す。

- 陳述表現の充実  
{名詞}, {動詞}に該当する語の数を増やすことで、語が登録されていないために取得されなかった対応関係に対処する。
- 対応関係における情報の不足  
取得した対応関係の中には以下に示すように情報が不足する場合もあるため、何らかの後処理を行い情報の不足を補う必要がある。
  - 代名詞の出現により情報が不足する場合  
例 11:  
「憲法にそれを禁止する明文規定はない」として、発動は可能との考えを強調する。
  - 鈎括弧の直後にある情報が鈎括弧内にない場合  
例 12:  
「OPECのすべての加盟国がすでに反対の立場を明確にしている」と述べ、OPECが今後一致して新税導入に反対していく考えを表明した。
- 要約知識の自動抽出  
取得した対応関係からの要約知識の抽出を現在は人手で行っている。従って、今後は要約知識の自動抽出を検討する必要がある。自動抽出の方法としては、類似した対応関係を収集することで自動化の際に要約知識として抜き出す部分を特定する方法を検討している。
- 引用表現以外の対応関係の取得  
鈎括弧内の節とその直後の節以外の部分から対応関係を取得できないかを検討する。対象として想定しているのは以下の部分である。
  - 記事のリードとそれ以降の対応部分
  - 関連記事の重複部分

## 参考文献

- [1] 奥村学, 難波英嗣: テキスト自動要約に関する研究動向, 自然言語処理, Vol. 6, No. 6 (1999).
- [2] 山崎邦子, 三上真, 増山繁, 中川聖一: 聴覚障害者用字幕生成のための言い替えによるニュース文要約, 言語処理学会 第4回 年次大会論文集, pp. 646-649 (1998).
- [3] 若尾孝博, 江原暉将, 白井克彦: テレビニュース番組の字幕に見られる要約の手法, 情報処理学会研究報告 NL-122-13, pp. 83-89 (1997).
- [4] 加藤直人, 浦谷則好: 局所的な要約知識の自動獲得手法, 自然言語処理, Vol. 6, No. 7, pp. 73-92 (1999).