

多重修飾された名詞句からの換言事例の自動収集

大竹 清敬* 児玉 充 増山 繁
豊橋技術科学大学 知識情報工学系

山本和英
ATR 音声言語通信研究所

kohtake@slt.atr.co.jp {kodama, masuyama}@smlab.tut.ac.jp yamamoto@slt.atr.co.jp

1 はじめに

多重修飾された名詞句に着目し、換言事例をコーパスから自動収集する試みについて述べる。一般に、換言とは「ある表現を意味内容を保ったまま別の表現に変換すること」である [1]。また換言を単言語内翻訳と捉えることもできる。しかし、本研究で扱う換言は、必ずしも同一の意味内容であることを要求せず、類似、あるいは関連する内容の換言事例を収集することを目的とする。また、本研究で扱う換言事例は、自動要約を目的として収集を試みるものであるが、収集された換言事例は自動要約のみならず自然言語処理のさまざまな応用分野において利用可能なものであると考える。

我々は、これまで自動要約の研究を行ってきた。要約を行うために換言によってより短い表現を得る処理は、文の大意を変更することなく、非常に頑健である。そのような換言処理による要約研究として、換言テーブルあるいは規則を用いた要約手法 [2, 3] がある。これらの要約手法は、規則やテーブルを手で整備しなければならない点が問題であるため、原文と要約結果を比較することにより換言規則を自動獲得する試みもある [4]。しかしながら、換言規則やテーブルを整備する、原文とその要約のコーパスを用意することは、いずれにしても、コストがかかる。そこで、このような換言テーブル・規則を自動的に収集することが望まれる。

我々は既に、直接引用表現に関してテキストコーパスから換言事例を収集できる可能性を示した [5] が、文脈処理の必要性や、換言規則の抽出の自動化が困難である点から、計算機上で直接的に利用可能な換言知識を得るに至っていない。本研究において、計算機上で直接的に利用可能である知識とは、複雑なルールを適用することなく換言を可能にする知識を指す。たとえば、若尾らや山崎ら [2, 3] が用いた換言規則や換言テーブルは、文脈などを考慮せずに適用できる。

言語間機械翻訳で採られてきた手法がそのまま、単言語内翻訳である換言へ適用できる可能性がある。規則を

*現在、ATR 音声言語通信研究所

主体とした機械翻訳から大量の翻訳事例を収集し使用する用例翻訳まで、さまざまな翻訳手法が存在する。換言処理も同様に規則を主体とした手法と、大量の換言事例を用いる手法が考えられる。しかしながら、長年精力的に研究されてきた言語間翻訳とは異なり、換言の研究はまだ萌芽的段階にある。

そのような背景から我々は、換言事例の収集は今後大量の換言事例を用いる換言手法を考案する際に有用になると同時に、計算機上で直接利用できる知識として応用できる点から非常に重要であると考え。一方で、このような知識の収集を試みる研究は、換言研究が緒に就いたばかりであることもあり、現在ほとんど存在しない。

換言事例の収集は、まさに言語間機械翻訳での翻訳事例の収集と同様のアプローチをとることが可能である。大量の平行コーパスがあれば、そこから収集することが可能である。当然のことながら、人手で記述することも可能である。しかしながら、ある文章全体をまったく別の表現で言い換えた文章を収集することや、人手で換言事例を記述することは非常にコストがかかる。一方で、ある程度類似した関連文書を大量に収集し、そこから換言事例を収集することが考えられる。たとえば、発行している新聞社が異なる同一事件に関する新聞記事を集めることにより、ある程度換言事例を収集することは可能であると考え。

しかしながら、本研究では、換言が単言語内で起こる現象であることに着目し、大量のテキストコーパスから、コーパス内に存在する換言事例を収集することを試みる。そのために、名詞にかかる修飾成分が複数存在する場合（多重修飾）に着目する。

2 多重修飾

文にみられる修飾関係には、さまざまな関係があるが、本研究では名詞を被修飾要素とする場合を扱う。ある名詞を修飾する連体修飾成分には、表 1 に示す種類がある [6]。これらの連体修飾成分が、一つの名詞に対して複数修飾している状態を多重修飾とする。特に、修飾成分が 2 つの場合を山本ら [7] の定義に従い二重修飾と

表 1: 連体修飾の種類

文法的性質	例
こそあど詞連体形	この話
連体詞	ある話
形容詞の現在形・過去形	おもしろい話
形容動詞の連体形・過去形	変な話
名詞+接続助詞「の」	昔の話
名詞+格助詞+「の」	昔からの話
名詞+取立て助詞+「の」	ここだけの話
副詞+「の」	突然の話
関係節	私が聞いた話
内容節	子狸が少年と仲良くなる話

関係節：被修飾名詞が修飾節の格要素（内の関係）

内容節：被修飾名詞が修飾節の格要素ではない（外の関係）

呼ぶ。たとえば、「私が聞いたおもしろい話」という名詞句では、「私が聞いた」という連体修飾節と「おもしろい」という形容詞がともに名詞「話」に係っている。

3 換言事例収集手法

この節では、換言事例収集手法について説明する。本研究において検討する換言事例収集手法は、テキストコーパス中から、換言可能な表現を収集するものである。そのような換言可能な表現として、二重修飾された名詞句に着目する。

3.1 着目する二重修飾

Webの検索エンジンで「二重修飾」を検索すると興味深い結果が得られる。本研究で定義する二重修飾という用語とは別に、一般的な二重修飾は「大きなビッグニュース」や「年老いた老人」のような誤用を指す語として用いられる傾向がある。このことが示唆しているのは、まさに換言の可能性である。つまり、このように誤用として用いられる二重修飾という現象は、名詞に対する形容表現の換言事例を内包している。

さらに、別の言い方をすると、誤用に対して用いられる二重修飾は、狭義の二重修飾であり、本研究で我々が定義した二重修飾は広義の二重修飾と呼ぶことができる。そこで、狭義の二重修飾の拡張である広義の二重修飾も換言事例を内包している可能性があると考えた。実際、同一の名詞に2つの修飾成分がかかる場合に、そ

れらの修飾成分が非常に類似している場合や、一方がもう一方の要約ととれる場合がある。たとえば、「変装したあやしい集団」という名詞句は、「変装した」という連体修飾節と、「あやしい」という形容詞がともに「集団」にかかる二重修飾である。この場合「変装した」と「あやしい」の間になんらかの関連性を見いだすことができる。つまり、「変装した」と「あやしい」を相互に換言できると考える。

そこで、本研究では、連体修飾節と形容詞、連体修飾節と形容動詞¹の2種類の二重修飾をコーパスから収集する。二重修飾を収集するにあたり、その語順を考慮しなければならない。日本語においては、一般的に長くて複雑な構造をもった成分を、文の前方に置こうとする傾向がある[8]。そのため、本研究では連体修飾節のあとに形容詞・形容動詞が続く語順の二重修飾を扱う。

3.2 換言の可能性の検討

収集する二重修飾された名詞句内の連体修飾成分が全て換言可能ではないと予想する。そこで、あらかじめ考慮しなければならない点を検討する。

- 被修飾名詞に対する連体修飾節の役割（内容的か、補足的か）：

2節で、例として「私が聞いたおもしろい話」という名詞句を取り挙げたが、この場合「私が聞いた」という連体修飾節は名詞「話」の内容を示している。そして、このような連体修飾節と形容詞の間には、換言できるほどの関連性を感じられない。また、このような連体修飾節には内容節が多いと予想されるが、「私が聞いた話」の場合は関係節である。一方で「～という」によって導かれる連体修飾節は、その被修飾名詞の内容を示している場合が多いにもかかわらず「～という...な名詞」という名詞句では、「～という」と「...な」の間になんらかの関連性を期待できる。したがって、このように連体修飾節とその被修飾名詞の関係を無視して換言可能性を考えることはできない。

- 形容詞・形容動詞の性質：

名詞を修飾するために頻繁に用いられる形容詞は、被修飾名詞を仲立ちとした連体修飾節に対する関係において換言可能な関係を構成するとは考えにくい。たとえば、「新しい」という形容詞はさまざま

¹本研究の実験では構文解析としてKNPを用いており、KNPは形態素解析器としてJUMANを要求する。これらのツールにおいて形容詞はイ形容詞に、形容動詞はナ形容詞にそれぞれ対応する。

まな連体修飾節とともに二重修飾を構成するが、それらの連体修飾節と形容詞「新しい」がすべて換言可能とは考えにくい。この現象はむしろ、その形容詞が連体修飾節を補足する形で被修飾名詞に係っていることにより説明されるかもしれない。

しかしながら、上述した点を考慮して、計算機上で直接利用可能な換言事例のみを自動的に取得するためには、多くの問題を克服しなければならない。そこで、我々は、換言事例候補から計算機で直接利用可能な換言事例のみを自動的に収集するために収集した換言事例候補に対して制約を設ける。具体的には、二重修飾の修飾成分それぞれが単体の修飾成分として同一の名詞にかかっている名詞句がコーパス中に存在した場合には換言可能とする。連体修飾節の動詞がさまざまな格要素を取りうることから、同一の連体修飾節の判定が問題となるが、比較的容易に実現可能である。たとえば、「耐寒力のある丈夫な一年草」という名詞句における「耐寒力のある」と「丈夫な」が換言可能かどうか判定する場合を考える。この場合、コーパス中に二重修飾ではない単体の修飾成分として「耐寒力のある一年草」と「丈夫な一年草」という名詞句がコーパス中に存在するかどうかで判定する。ここで、「一年草」という名詞がコーパス中に数多く出現するとは考えにくいので、シソーラス等を用いて汎化する必要がある。

その他、言語学的知識を総動員して、規則による判定を行うことも考えられるが、今後の検討課題である。

4 換言事例候補の収集

本研究で提案する換言事例収集手法を用いて実験を行った。対象としたコーパスは日本経済新聞の1990年版である。形態素解析器にはJUMAN、そして構文解析器にはKNPを用いた。コーパスから任意に抽出した10万文に対して構文解析を行った。この際、構文解析器が異常停止してしまう極端に長い文を除くため121文字以上の文を除いて解析した。その結果99146文の解析結果が得られた。この構文解析結果から、連体修飾節と形容詞、連体修飾節と形容動詞による二重修飾を持つ名詞句を抽出したところ1129の名詞句が得られた。一部を表2に示す。

5 考察

まず、収集を試みた換言事例が量的に十分得られるのかどうかについて考察する。この報告では約10万文から換言事例の候補を抽出した。日経新聞の1990年版に

は約158万文存在するため単純に見積もっても今回得た換言事例候補の15倍を新聞記事1年分から得られることになる。つまり概算で日本経済新聞の記事1年分から約1万5千事例を獲得できることになり、量的には十分であると考えられる。ただし、今回収集した換言事例候補の中には、構文解析誤りによって連体修飾節と認定されている箇所があるのでこの部分をどのように排除するかは、今後の課題である。

つぎに、表2から予想されることは、収集した換言事例の候補を3つに分類できる点である。ひとつめは、換言事例として不適切なもの。たとえば、「今まで知らなかった いろいろな」がこれに該当する。2つめは、連体修飾節と形容詞、あるいは形容動詞の間で一対一で換言可能なもの。たとえば、「耐寒力のある 丈夫な」がこれに該当する。そして、最後に連体修飾節とその被修飾名詞を一緒に考えなければ換言可能と考えにくいものである。たとえば、「子どもを幼稚園などに送迎する(主婦) 若い」が該当する。この最後の分類は、「子どもを幼稚園などに送迎する大きなバス」という名詞句と比較することによってよりはっきりすると考える。

本研究では、名詞を中心としたその形容表現である二重修飾に着目し、その修飾成分間になんらかの関連性があると仮定し、実際に収集し検討した。これは、ある言語単位を中心としてそれに係る表現の間にはなんらかの関連があるとする思想に基づいている。この思想に基づく、名詞を中心とした係り受け構造から動詞を中心とした係り受け構造における格要素間の換言にまで拡張することが可能である。ただし、この場合は、格要素は格という動詞に対するはっきりとした役割をもつことと、参照する範囲がさらに広がるため名詞を中心とした場合の手法をそのまま適用することはできない。そこで、複数の関連する動詞を収集し、格要素の換言を考える。具体的には「東京で言語処理学会を開く」「言語処理学会で発表する」という2つの文がコーパス中に存在するとき、「言語処理学会で発表する」という文を「東京で発表する」と換言するようなことを考えている。

6 今後の展望

今回収集の可能性を示した換言事例は実にさまざまな分野へ応用が可能だと考える。一対一で換言可能な事例を大量に集めることができたならば、要約の分野において、連体修飾節をより短い形容詞・形容動詞に換言することが可能となる。その他、文脈に依存した要約知識も得られるが、適用は比較的容易ではないかと考えてい

表 2: 二重修飾を持つ名詞句の例

連体修飾節	形容詞・形容動詞	被修飾名詞を含む文節
子どもを幼稚園などに送迎する	若い	主婦,
今まで知らなかった	いろいろな	世界が
耐寒力のある	丈夫な	一年草で
いずれもクワイの花に似た	白い	三弁の
リトアニアとモスクワとの三カ月半に及ぶ	厳しい	対決は,
音楽ファンの期待に答える	本格的な	音楽番組も
平安朝から続く	雅な	三箇夜餅の
三〇度を超す	暑い	日が
女性オペレータが監視する	クリーンな	鋳物工場も
八五年秋のプラザ合意をきっかけとした	急激な	円高の

る。たとえば、「子どもを幼稚園などに送迎する」を文脈を無視して「若い」とすることはできないが、「子どもを幼稚園などに送迎する主婦」を「若い主婦」と要約することは可能である。

要約以外にも、連体修飾節と形容詞・形容動詞の間の換言事例が収集されるのでさまざまな応用が期待される。これまでは、換言事例の応用として要約を中心に考えてきたが、ある形容詞・形容動詞を文脈上の手がかりからさらに詳細な記述である連体修飾節へ換言することも可能である。たとえば、「冬を越すことができる丈夫な一年生植物」を「冬を越すことができる(ほど)耐寒力のある一年生植物」などである。このような文脈を考慮した換言処理のためには野上ら [9] のように、文脈を考慮してデータベース化する手法は非常に有用である。

今回報告した換言事例は、二重修飾のうち連体修飾節と形容詞・形容動詞に着目したものであるが、それ以外の修飾要素間の換言事例の収集も検討する必要がある。たとえば、「おもしろいこの話」のような二重修飾の場合、「この」というこそあど詞連体形が指す表現を「おもしろい」と換言できる可能性がある。

7 むすび

本研究では、大量のテキストコーパスから換言事例を収集するために、多重修飾された名詞句に着目し、連体修飾節と形容詞・形容動詞の間で換言可能な事例を収集できる可能性を示した。今回収集した以外の換言事例を単言語テキストコーパスから収集することは今後の課題である。

参考文献

- [1] 佐藤理史：論文表題を言い換える，情報処理学会論文誌，Vol. 40, No. 7, pp. 2937–2945 (1999).
- [2] 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 97-NL-122, pp. 83–89 (1997).
- [3] 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換えによるニュース文要約，言語処理学会 第 4 回 年次大会発表論文集，pp. 646–649 (1998).
- [4] 加藤直人：ニュース文要約のための局所的な要約知識獲得とその評価，電子情報通信学会信学技報 NLC98-7, pp. 7–14 (1998).
- [5] 児玉充，片岡明，増山繁，山本和英：直接引用表現を利用した要約知識の自動抽出の試み，言語処理学会 第 6 回年次大会 発表論文集，pp. 241–244 (2000).
- [6] 寺村秀夫：日本語の文法（下），大蔵省印刷局 (1981).
- [7] 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，自然言語処理，Vol. 2, No. 1, pp. 39–55 (1995).
- [8] 野田尚史：語順を決める要素，言語，Vol. 29, No. 9, pp. 22–27 (2000).
- [9] 野上優，藤田篤，乾健太郎：文分割による連体修飾節の言い換え，言語処理学会 第 6 回年次大会 発表論文集，pp. 215–218 (2000).