

時制と態を考慮したサ変名詞の動詞化

土田 雅之, 大橋 一輝, 山本 和英

長岡技術科学大学 電気系

E-mail:{tsuchida,ohashi,ykaz}@nlp.nagaokaut.ac.jp

1 はじめに

昨年, 我々は「サ変動詞+名詞」を複合名詞にする換言を提案した [1]. この換言は換言前に付属していた「する」等の助動詞が省略されていても人間ならば補完出来るという性質を利用している (例「ガンを治療する施設」「ガンの治療施設」). しかし, 何か処理しなければ機械には本来付属していた助動詞が何であるか補完することはできない. そこで我々は「名詞1+助詞+サ変名詞+名詞2⁽¹⁾」の形を「名詞1+助詞+サ変名詞+助動詞+名詞2」へ換言し, 時制と態を補完する手法を提案する. 時制と態がわかることでさらなる換言や翻訳, 意味理解に役立つと期待できる. 換言例を以下に示す.

例 1) がんの予防効果 も期待できるという。
がんを予防する効果 も期待できるという。

例 2) 容疑者の侵入経路などについて調べを進める。
容疑者が侵入した経路 などについて調べを進める。

2 関連研究

関連研究として, 佐藤は換言対象を論文表題に限定して換言規則を作成した [2]. 本稿は換言対象を論文表題に制限しないかわりに「名詞1+助詞+サ変名詞+名詞2」と制限して時制・態を考慮した. また, 近藤らは「サ変名詞+する」の型を動詞に換言するという本稿の手法の後に適用できる手法を提案している [3].

3 提案手法の内容

3.1 換言対象

本稿では換言対象に「名詞1+助詞+サ変名詞+名詞2」の型 (以下, この型になっている部分を換言対象と呼ぶ) を選択し, さらに助詞等を制限して換言を行なった. ここで名詞1及び名詞2は複合名詞でも構わない. 換言対象の助詞は「の, が, に, も, で, へ, からの, での, までの, との, への」の11種類である. 名詞2にサ変名詞が含まれている, つまりサ変名詞が連続している場合, 全てを換言すると文が判りにくくなり, いずれか

一つを換言する場合にもどこを換言するべきかが違ってくるので対象外とした (例3). また, 名詞2に接尾辞が含まれている場合には単純な換言を行なうことができず, さらなる換言が必要となるのでこれも対象外とした. (例4)

例3) 気象庁の対策実施本部 *気象庁の対策する実施する本部

例4) 科学技術が飛躍的 *科学技術が飛躍する

3.2 換言処理手順

本稿で提案する手法は以下に示す流れで処理する.

1. 換言対象を含む文を検索する (例5).

例5) 金融当局が、認可の判断基準 として～

2. 換言対象のサ変名詞をサ変動詞化する (例6).

例6) 金融当局が、認可の判断する基準 として～

3. 換言対象の助詞を換言する (例7).

例7) 金融当局が、認可を判断する基準 として～

3.3 サ変名詞のサ変動詞化

サ変名詞をサ変動詞化するため, 「する, した, される, された, している」のいずれかを追加する「している」は追加される場合がほとんど無いと考えたので使用していない. 5種類の助動詞のうちどれを追加するかを判定する必要があるが, 決まった助動詞が追加される場合や, 前後の文によって追加される助動詞が変化する場合があり, 規則を作成して照合するのでは処理することができない. そこで Web 上のサーチエンジン Google (以下 Google) を用いて処理を行なう.

最初に, 「サ変名詞+(する, した, される, された, している)+名詞2」という連続した文字列として Google で検索し, その照合件数を記録する. 次に, 得られたそれぞれの照合件数に応じて以下の処理を行なう.

3.3.1 照合件数が全体的に少ない場合

照合件数が全て30件以下の場合には換言を行なわない. この場合はサ変名詞と名詞が分離できない (例えば「捜査本部」を「捜査する本部」とは換言できない), 「する」等の助動詞を追加するだけでは不足である (例

(1) 本稿では名詞1と記した語はサ変名詞を含むが, 名詞2と記した語はサ変名詞を含まない.

例えば「建築廃材」は「建築した後に残った廃材」のように換言するのが適切)等, 妥当な換言が出来ない可能性が高いからである。

例 8) 「減産+期限」の場合

表 1 に Google での検索結果を示す。

表 1: 「減産+(助動詞)+期限」の照合件数

する	した	される	された	している
0	0	0	0	0

全ての場合で照合件数が 0 なので「換言しない」と判断する。

3.3.2 照合件数が偏っている場合

ある言葉の照合件数が他の言葉の照合件数の 5 倍以上であった場合, その言葉は使用方法が決まっている, または偏っていると考えられるのでその言葉を換言先の語とする。これを「使用方法は限定」とする。

例 9) 「判定+基準」の場合

表 2 に Google での検索結果を示す。

表 2: 「判定+(助動詞)+基準」の照合件数

する	した	される	された	している
863	17	31	4	6

「判定する基準」で検索した場合に, 他の場合の 5 倍以上照合するので追加する助動詞を「する」に決定する。

3.3.3 上記に当てはまらない場合

上記 2 つのどちらにも当てはまらない場合, 使用方法が限定されないが換言自体は存在するとして, ここでは能動か受動かだけを照合件数で決定しておき, 具体的に使用する助動詞の決定は保留しておく。これを「使用方法は様々」とする。

例 10) 「発生+状況」の場合

表 3 に Google での検索結果を示す。

表 3: 「発生+(助動詞)+状況」の照合件数

する	した	される	された	している
2280	827	4	0	2150

「発生する状況」が一番多いので能動に決定する。

3.4 助詞の決定

3.4.1 助詞が「の」以外の場合

換言対象の助詞が「の」以外の場合には, サ変名詞をサ変動詞に換言したとしても基本的には助詞を変更する必要は無く意味が保たれる。これは「の」はその他の様々な助詞の意味を内包しているが, 換言対象の助

詞は意味が一意に決定されているためである。例外として, 助詞「からの, までの, での, への, との」のような, 本稿では助詞と一括りにしているが実際には「助詞+の」で構成されている助詞がある。このような助詞の場合には後の「の」を取り除くことで換言できる。

1. 換言対象の助詞が「が, に, も, で, へ, と」の場合
助詞の換言は行なわない。

例 11) パートナーを迎えた方が繁殖確率は高くなる。
パートナーを迎えた方が繁殖する確率は高くなる。

2. 換言対象の助詞が「からの, までの, での, への, との」の場合

後の「の」を取り除き「から, まで, で, へ, と」という助詞にする。

例 12) 図解や語句説明など市民の視点からの補足資料も付け、～

図解や語句説明など市民の視点から補足する資料も付け、～

3.4.2 助詞が「の」の場合

換言対象の助詞が「の」の場合には, サ変名詞をサ変動詞に換言する前後で「の」も換言しないと意味が保持できないことが多い。これは助詞「の」が含む意味が多様な為である。以下に例を示す。

例 13) 公式戦の出場選手 公式戦に出場する選手

例 14) イスラムの武装集団 イスラムの武装した集団

例 15) ゲリラの掃討作戦 ゲリラを掃討する作戦

例 16) 動植物の生息空間 動植物が生息する空間

「の」を意味で分類しようとしてもその数が多すぎて分類が難しい。そこで, 我々は統計的に換言する助詞を決定することとし, Google での照合件数を用いる。以下に処理の手順を示す。

1. 3.3 節で「使用方法は限定」とした場合

検索条件 1: 「名詞 1 (名詞 1 が複合名詞の場合は最後の名詞)+助詞 (が, の, に, を)+サ変名詞+決定された助動詞+名詞 2」という言葉で Google で検索し, ヒット件数が最も多い言葉を換言先の言葉とする。

検索条件 2: もし, 全ての場合で照合件数が 0 件の場合, 名詞 2 を削除し, 名詞 1 が複合名詞の場合それら全てを用いて Google で検索, 照合件数を比べる。

検索条件 3: それでも照合しない場合は名詞 1 を最後の名詞のみにして Google で検索, 照合件数を比べる.

検索該当無し: 上記の検索全てで照合しない場合には助詞は「の」のままにしておく.

例 17) 扶助対象の活動期間 (追加する助動詞は「する」が選択されていたとする)

検索条件 1: 対象+(が, の, に, を)+活動する期間

検索条件 2: 扶助対象+(が, の, に, を)+活動する

検索条件 3: 対象+(が, の, に, を)+活動する

2. 3.3 節で「使用方法は様々」とした場合

基本的には「使用方法は限定」とした場合と同じだが, 助動詞が能動とされた場合には Google で検索する際に使用する助動詞を「する, した」とし, 助動詞が受動とされた場合には使用する助動詞を「される, された」とする.

3.5 特殊な処理

1. 地名+助詞「の」+サ変名詞+「事件」の場合

この場合は換言対象の助詞を「で」にする.

例 18) 浦和市の J R 浦和駅の爆発事件 について ~
浦和市の J R 浦和駅で爆発した事件 について ~

2. サ変名詞+「自体」の場合

この場合は換言後の「サ変名詞+助動詞+名詞 2」の助動詞と名詞 2 の間に「こと」を追加する.

例 19) この問題は 我々の存在自体 を問う.

この問題は 我々が存在すること自体 を問う.

4 評価実験

毎日新聞 2000 年版 (2) 約 23 万文のうち, 換言対象となりえる文 (名詞 2 にサ変名詞が含まれていたり接尾辞で名詞 2 が終了している文も含む) を含む文 10608 文を入力し, 実験を行なった. 換言された文は 3447 文, 換言されなかった文は 7161 文であった. 換言数は Google の状況によって ± 150 程度変化するので安定しない.

換言された文のうち 400 文を手で評価した. 評価の基準は換言後の助動詞の時制・態が適切か, 意味が保持されているか, 助詞が適切かである. ただし, 人間が換言した場合に換言先が複数あり, 一意に換言出来ない場合はそれら全てを正解としている. 助詞が適切かの判断は, 例えば「トラブルの発生状況」という言葉は「トラブルの発生した状況」と換言するより「ト

ラブルが発生した状況」と換言したほうが良いので適切でないとしている. 表 4 に結果を示す.

表 4: 400 文の評価結果

使用方法	限定	様々
正解	127	116
助動詞の誤り	6	26
助詞の誤り	79	
助詞と助動詞の誤り	2	5
人間なら判別可能	19	
適切な換言不可能	23	

400 文に含まれていた換言対象は 403 個, 正解率は正解数 (127+116)/換言対象の数 (403) = 60.3%であった. 換言対象の助詞が「の」の場合, サ変名詞のサ変動詞化よりも助詞の換言誤りが多かった. 当然のことではあるが検索条件が厳しい時に Google で照合した時の方が正解率は高かった. 換言しなかった文のうち約 1100 文は名詞 2 が接尾辞で終了したり, 名詞 2 にサ変名詞が含まれる文であった.

5 考察

5.1 換言誤りについて

換言誤りの例を以下に示す. なお, これらは全て助詞の換言の際に検索条件 2,3 で Google での検索に照合したものである.

例 20) 首相の諮問機関「経済審議会」が ~
*首相が諮問する機関「経済審議会」が ~
首相の諮問する機関「経済審議会」が ~

例 21) グループの成長部分 を担う高収益企業に育て、
*グループに成長した部分 を担う高収益企業に育て、
グループの成長する部分 を担う高収益企業に育て、
グループを成長させる部分 を担う高収益企業に育て、

統計的手法が有効なのは使用方法が偏っている場合, 言い換えれば限定されているような場合である. しかし, 同じ名詞 1, サ変名詞が使われていても, 最後の名詞 2 が変われば助詞は候補に様々なものが上がってくる. よって, 検索条件 2 で検索した場合には, 適切でない候補が数多く含まれることになり, 助詞の誤りが発生する. また, 同じサ変名詞, 名詞 2 が使われていてもサ変名詞は名詞 1 によって様々なサ変動詞へ変化する場合も多い.

実験として助詞の換言の際に「検索該当無し」となった場合, さらなる処理として「名詞 1+助詞 (が)+サ変名詞」と「助詞 (が)+サ変名詞+決定された助動詞」

の照合件数を合計し、他の助詞(の、に、を)についても同様に照合件数を見て、それらの数を比較して換言先を決定した。しかし上述したように検索条件が緩くなることで適切でないものが大量に紛れ込み、結果としてむしろ誤りが多くなった。よって、検索条件を緩くした際にはさらに別の制約を加える必要がある。

5.2 サ変名詞が連続した場合の換言について

本稿ではサ変名詞が名詞2に含まれていたり、名詞2が接尾辞で終わっている場合には換言しなかったが、4節で記述したように半数以上が換言されない結果となっている。しかし、全てのサ変名詞を単純にサ変動詞にした場合、意味が全く別のものになってしまったり、日本語として成立しない可能性が考えられ、容易には判断することが出来ない。この部分はさらに検討が必要である。

5.3 接尾辞の換言について

接尾辞が換言対象の末尾に付いている場合にはその接尾辞に対して新たに別の換言を行わないと文としておかしくなると考えた。しかし、接尾辞が末尾に付いていても、その直前の名詞がサ変名詞で無いならば換言可能と思われる場合があった(例22)。

例22) 毎日新聞社内の選考委員会
毎日新聞社内の選考する委員会

5.4 その他の問題点について

換言対象外との助詞の整合が合わないことがあった。これは構文解析器を用いることで解決できると考えている。

助詞「の」の換言において、換言後の助詞に「の」が使用されていて意味は保持できているが、より適当な換言がある場合というものが目立った。これはほとんどの場合、助詞を決定する際に、最初の検索条件では照合できず、検索条件を緩くして助詞を決定した場合に起こっていた。「の」は意味が多様多様なので人間ならば同じ意味と捉えることができるが、機械が同じ意味と捉えられるかは定かではないので出来るだけ「の」の使用を避けたい。この部分は現在の手法では改善が出来ないので、他に新たな手法を講じる必要がある。

サ変名詞のサ変動詞化の部分で換言しないと判断される場合にも、通常使用はしないが換言することは可能ではないかと思われる文があった(例23, 例24)。

例23) 支配地区 支配する地区、支配された地区 等

例24) 伝送速度 伝送する速度

新聞の見出しに使われるような助詞や名詞で終わっ

ている表現に換言対象が含まれている場合、現在の手法ではうまくいかなかった。

例25) 国民党の協力必要に
*国民党が協力する必要に
国民党の協力が必要に

サ変名詞のサ変動詞化の処理の部分で使用する動詞を決定する際に、5倍以上ならば限定、30件以下ならば換言しない等とパラメータを設定したが、値を変更して換言を行ない、適当と思われる値として設定したものののでまだ改善の余地がある。

6 終わりに

「名詞1+助詞+サ変名詞+名詞2」という型の文を「名詞1+助詞+サ変名詞+助動詞+名詞2」という型へ換言し、サ変複合名詞に対して時制、態を追加する手法を提案した。手法を実装し、評価を行なったところ、正解率は60.3%程度であった。

謝辞

本研究の一部は、科学研究費補助金 若手(B)「高密度表現を利用したまとめ型要約に必要な言語変換技術」課題番号16700134、及び科学研究費補助金 基盤(A)「円滑な情報伝達を支援する言語規格と言語変換技術」課題番号16200009によって実施した。

使用した言語資源及びツール

- (1) 形態素解析器「茶筌」, Ver.2.3.3, 奈良先端科学技術大学院大学 松本研究室, <http://chasen.aist-nara.ac.jp/hiki/ChaSen>
- (2) 毎日新聞全文記事データベース 2000年版, 毎日新聞社.
- (3) サーチエンジン Google, <http://www.google.co.jp/>

参考文献

- [1] 大橋一輝, 山本和英:「サ変動詞+名詞」の複合名詞への換言, 言語処理学会第10回年次大会, pp.693-696, 2004.
- [2] 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.
- [3] 近藤 恵子, 佐藤 理史, 奥村 学:「サ変名詞+する」の動詞への言い換え, 情報処理学会自然言語処理研究会, NL127-24, 1998.
- [4] 寺村 秀夫: 日本語のシンタクスと意味 III, くろしお出版, 1991.