

日本語支援動詞構文の述部に対するサ変動詞への換言

山本 和英 高橋 寛治 楳澤 優希

長岡技術科学大学

{yamamoto, takahashi, gumizawa}@jnlp.org

1 研究の動機

ほぼ同一の事物を表すための表現が自然言語には複数存在し、これが自然言語の計算機処理の難しさを生み出す一つの原因であると考えられる。そこで、我々は計算機処理のための言語表現の集約を目指して、表記揺れの吸収 [1] や同義語集合によるまとめあげ [2] を行っている。それでは、表記揺れや単語の集約の次に有効だと考えられる言語表現の集約とは何だろうか。

佐藤ら [3] が議論している言い換えによる表層的な多様性を計算機処理のために考慮することで、言語表現の集約を行うことができる。計算機処理のための言い換えによる表層的な多様性の吸収は、日本語の正規化と捉えることができる。日本語の正規化を目的とした研究は、泉ら [4] による機能動詞表現の述部を言い換えることにより述部の表層的な多様性の正規化を行ったものがある。述部のみで評価した場合の換言精度は94%と高いが、文で判断した場合は精度が53.0%との報告である。そこで本稿では文を考慮した「サ変名詞+を+支援動詞」を対象として、サ変動詞へと言い換えることを提案する。

サ変動詞へと言い換えることにより、各種応用処理の性能向上が期待出来る。例えば「運動する」、「運動をする」、「運動を行う」という表現は同義であると認識できる。情報検索では、これらの同義性の認識により再現率の向上が期待できる。また依存構造を用いた応用処理でも性能向上が期待できる。例えば「私が運動をする」を係り受け解析すると、「私」と「運動」が「する」に係る。「私が運動する」を解析すると、「私」が「運動する」という1つの動詞に係る。「運動する」という一つの単語に正規化することにより、「する」に係るのではなく1単語の動詞に係ることにより、係り受け関係が明確になりテキストマイニングを始めとする応用処理での性能向上が期待出来る。正規化による上記のような効果を期待し、本稿では「運動をする」や「運動を行う」と言った「サ変名詞+を+支援動詞」の一部に着目し、サ変動詞「運動する」へと言い換

ることにより言語表現の正規化を行う方法を提案し、換言を行う。

2 換言として扱う範囲

2.1 単語の単位と品詞について

換言は日本語解析システム「雪だるま」(1)の出力を対象に行う。雪だるまは、独自の品詞体系¹で分割された単語に対して表記揺れの吸収や複合語処理、形態素の結合を行った結果を出力する。本稿の換言に特に関わる形態素の結合、および雪だるまにおける品詞(助詞、形容詞、副詞)について説明する。

雪だるまにおける助詞とは、一般的な助詞に加え「に関する」や「とは」のような助詞として振る舞う複数の形態素から構成される機能表現を1単語の助詞とする。

次に形容詞だが、一般的には「無理な」は名詞「無理」と助動詞「な」の2形態素として出力される。雪だるまでは「無理な」という形容詞1単語で出力される。

「すごく」は形容詞の活用として取り扱われるのが一般的だが、雪だるまでは「副詞」として扱う。本稿での品詞体系は一般のものとは異なるため注意されたい。

2.2 換言対象の選定

「サ変名詞+を+支援動詞の一部」を換言対象として取り扱う。換言後はサ変動詞1単語とする。ここで、支援動詞の一部とは「する」「やる」「行う」である。これらの支援動詞はほぼ同義であるため同一視して取り扱う。

また、取り扱うサ変動詞は日本語解析システム雪だるまの単語辞書に含まれているサ変動詞15,598語とする。言い換えは対象とするサ変名詞の前の語に基づき、規則を作成することにより行う。サ変名詞前の単

¹詳細は <http://snowman.jnlp.org/part-of-speech> に記載。

語を調査し、さらに換言可能性が高い部分の換言を行うことを念頭に置く。まず、BCCWJ内で「サ変名詞+を+対象とする支援動詞(する)」の前に出現する品詞を調査した。そのうちの上位5件を換言対象として取り扱う。その割合を表1に示す。

表 1: 換言対象直前の語の品詞

品詞	頻度	総頻度に対する割合
助詞	43,786	55%
形容詞	10,685	13%
記号	6,918	9%
助動詞	4,075	5%
副詞	3,525	4%
その他	11,207	14%
合計	80,196	100%

助詞に対してはさらに細かく分類する。高頻度語から「を」の削除による換言の可能性を持った語8語を対象とした。しかしながら、最も高頻度語である助詞「の」(助詞のうち約37%)は事前の調査より、換言の可能性が低いいため今回対象外とした。

- 例1) 運動会の準備をする。 換言不可
- 例2) 午後の運動をする。 換言不可
- 例3) ピアノの練習をする。 ピアノを練習する。

影山 [5] は「NPのVNをする」という構文において、NPは修飾語として解釈可能であるが、「VNする」へと言い換えると修飾語としての意味が消滅すると分析している。換言処理においては、助詞「の」が前後の単語をどのような関係で結んでいるかを解析する必要がある。最も頻度が高い語であるため換言できれば正規化の効果は高いと考えられるが、助詞「の」の解析は現状では困難と判断し、今回は対象外とした。

助動詞も助詞と同様に高頻度語のうちの換言の可能性が高い「に」と「ず」を取り扱う。助動詞「に」とは、「部分的に検査をする」の「に」である。「部分的に」という形容詞的用法であるが、換言後は副詞的用法となる。助動詞「ず」は「思わず拍手をする」の「ず」である。

まとめると本稿では「サ変名詞+を+対象の支援動詞」のうちの約43%の表現を換言対象として取り扱う。

3 換言規則

「サ変名詞+を+支援動詞」をサ変動詞へと言い換える。この際に、助詞「を」を削除するというのが基

本的な方針だが、サ変名詞前の単語を削除もしくは言い換える場合がある。この節では、直前の単語の特徴に応じた換言規則を列挙する。

3.1 サ変名詞前の単語に基づいた換言規則

サ変名詞の前の単語に着目して換言処理を適用する。規則は以下に示した15個である。

サ変名詞の前の単語が

- 助詞「に」、助詞「で」、助詞「て」、助詞「が」、
- 助詞「は」、助詞「も」、助詞「を」、副詞
- 助動詞「に」、助動詞「ず」、記号

例4) ほとんどが歯磨きをしてこないのです。
ほとんどが歯磨きしてこないのです。

例5) 悪いというよりもフォローをしない。
悪いというよりもフォローしない。

例6) こういうふうに行動をして~
こういうふうに行動して~

サ変名詞前の単語が助詞「と」かつ

助詞「と」の前の単語がサ変名詞以外のもの

例7) 姉と勉強をする
姉と勉強する

例8) いったん売主と顔合わせをする
いったん売主と顔合わせする

ここで助詞「と」の前の単語がサ変名詞以外の場合とするのは、動詞を並列に並べた文の場合に換言規則を適用できないからである。「運動と勉強をする」という文の場合、「勉強と運動する」となり不適切な換言となる。

サ変名詞前の単語が形容詞かつ

- 形容詞が「~に」形のもの
- 形容詞が「~で」形のもの

例9) 十分に反省をした結果として、~
十分に反省した結果として、~

例10) 日銀が単独で介入をするということ
日銀が単独で介入 するということ

サ変名詞が文頭

例11) 問い合わせをしようと思うのですが、~
問い合わせしようと思うのですが、~

3.2 する以外の支援動詞

2.2 節でも述べたように、「する」以外には「行う」と「やる」を対象とした。これらの動詞は「する」と同様の換言規則を適用したあと、「行う」「やる」を「する」へと言い換えることにより対応する。

例 12) 運動を行う 運動をする 運動する

例 13) 普通にテストを行い、採点してください。
普通にテストし、採点してください。

例 14) 勉強をやる 勉強をする 勉強する

例 15) ドリフトをやってみようと思うのですが
ドリフトしてみようと思うのですが

4 換言実験

換言対象は、現代日本語書き言葉均衡コーパス(BC-CWJ) (2) 内に含まれる「サ変名詞+を+対象とする支援動詞(する)」とする。換言対象を日本語解析システム雪だるま(1)により単語解析し、これに換言規則を適用する。それぞれの規則が適応される文を100文ずつ無作為に選択し、換言規則の妥当性の判断を行う。換言結果が文法的に正しいか、また意味が保たれているかを判断する。また、「サ変名詞+を+やる、行う」といった文章に関してもそれぞれ50文ずつ無作為に選択し、換言規則が正しく適用できるかどうかを判断する。

5 結果と考察

実験結果を表2に示す。「サ変名詞+を+(する、やる、行う)」を含む文、80,196件のうちの34,832文、つまり換言対象のうちの43%が換言された。また、妥当性を判断した1,293文のうちの6文が誤りであり、高い精度で換言を行うことができた。

5.1 正解例

実際に換言規則を適用して得られた正解例をいくつか示す。

サ変名詞直前が助詞「に」

例 16) 私の呟きに兄は苦笑いをする。
私の呟きに兄は苦笑いする。

表 2: 規則ごとの換言対象の数と100文中の誤り

規則(サ変名詞前)	換言対象数(文数)	誤り(文数)
助詞「に」	4,804(4,761)	0(0)
助詞「で」	4,442(4,414)	2(2)
助詞「て」	2,779(2,766)	0(0)
助詞「が」	2,661(2,618)	2(2)
助詞「は」	1,753(1,749)	0(0)
助詞「も」	927(923)	0(0)
助詞「を」	565(546)	0(0)
助詞「と」かつ前がサ変名詞でない	2,322(2,313)	2(2)
副詞	3,526(3,511)	0(0)
助動詞「に」	398(397)	0(0)
助動詞「ず」	59(59)	0(0)
形容詞「～に形」	1,834(1,824)	0(0)
形容詞「～で形」	48(48)	0(0)
文頭または記号	9,007(8,903)	0(0)
合計	35,125(33,987)	6(6)

表 3: 規則ごとの換言対象の数と誤り

種類	換言対象数	誤り
行う	479	0
やる	121	4
合計	600	4

サ変名詞直前が助詞「も」

例 17) 女房も仕事をしていたが、～
女房も仕事していたが、～

サ変名詞直前が副詞

例 18) ちょっと調査をしたのでございますが、～
ちょっと調査したのでございますが、～

5.2 誤り例と考察

サ変名詞前の助詞が「が」の場合

例 19) そして夜遅く、暖炉の前で伯父さんが話をします。

「話する」(はなしする)は不自然で、この場合には「話す」もしくは「話しする」と言い換えるほうが自然である。これを例外的な表現であると捉え、この種類の語の言い換えは別の処理で行うことにより換言できると考える。

サ変名詞前の助詞が「で」の場合

例 20) ここで教授をする

「教授する」というサ変動詞が存在するため、誤って換言された。「教授をする」と「教授する」のように

言い換えることによって全く意味が異なる例である。今回の実験では「教授する」の一語のみ、このような現象が見られた。

サ変名詞前の助詞が「と」かつ、

助詞「と」の前がサ変名詞以外の場合

例 21) 夕方は畑に出かけて水遣りと収穫を
するとほとんど時間を使ってしまい、～

例 22) ディスクのクリーンアップとデフラグを
しておいた方が

これは、助詞「と」が動作の並列を表し、換言することができない。例 21 の場合は助詞「と」の前が「動き」のような動詞を名詞化した語の場合に、換言不可とする規則を追加することで解決できると考える。例 22 の場合は、「クリーンアップする」という単語を辞書に登録することで換言対象から除外されるため解決できる。

サ変名詞前が記号の場合

例 23) やれ有効だのアジア共同体だのとこれまた「政治色の強い」発言をするんですね。

助詞「を」を削除すると『「政治色の強い」発言するんですね』となり不自然な日本語となる。今回はサ変名詞前が記号の場合に換言するとしたが、カギ括弧の場合は書き換えられないほうがいい。

「やる」の誤り

例 24) 先生方のなかで、臍摘をしないで血液型不適合をやられた先生はいらっしゃいますか。

これは「やる」が支援動詞としてではなく自立語として扱われているからである。「肩をやる」のような場合は、「肩する」というサ変名詞がないため、換言は起こらない。「やる」が「する」と同じ意味で用いられているかどうかを解析しなければ換言できない。

6 おわりに

本稿では、サ変動詞へと換言可能な表現を列挙し、サ変動詞へ換言する方法を提案した。サ変動詞への換言処理のための規則を列挙した。本手法によって支援動詞構文の 43% が換言され、検証の結果「する」は 99% 以上、「やる」「行う」については 96% 程度で正確に換言できることを確認した。今回は、高頻度から確実に換言できる部分に絞って規則を構築し適用したため、換言可能な文数は少ない。今後の換言の課題とし

て、換言できる支援動詞の数を増やし、更なる日本語の正規化を目指す。また実際に応用タスクに適用した際に、どのような効果があるかを検証する。

なお、今回作成したサ変動詞への換言規則は日本語解析システム「雪だるま」に組み込み公開する。

使用したツールと言語資源

1. 日本語解析システム「雪だるま」, (Yamamoto et al. 2015), 長岡技術科学大学 自然言語処理研究室, <http://snowman.jnlp.org/>
2. 現代日本語書き言葉均衡コーパス (BCCWJ), Ver.1.1, 国立国語研究所.

謝辞

本研究は、平成 27～31 年科学研究費補助金基盤 (B) 課題番号 15H03216、課題名「日本語教育用テキスト解析ツールの開発と学習者向け誤用チェッカーへの展開」の助成を受けています。

参考文献

- [1] Kazuhide Yamamoto, Yuki Miyanishi, Kanji Takahashi, Yoshiki Inomata, Yuki Mikami, and Yuta Sudo. What we need is word, not morpheme; constructing word analyzer for japanese. *Proceedings of the International Conference on Asian Language Processing (IALP 2015)*, pp. 49–52, Oct 2015.
- [2] Kazuhide Yamamoto and Kanji Takahashi. Construction of japanese semantically compatible words resource. *Proceedings of the International Conference on Asian Language Processing (IALP 2015)*, pp. 61–64, Oct 2015.
- [3] 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, jul 1999.
- [4] 泉朋子, 今村賢治, 菊井玄一郎, 藤田篤, 佐藤理史. 正規化を指向した機能動詞表現の述部言い換え. 言語処理学会第 15 回年次大会発表論文集, pp. 264–267, Mar 2009.
- [5] 影山太郎. 文法と語形成. ひつじ書房, 1993.