

# 動詞名詞対応辞書の構築と敬語の常体への換言

山本 和英      高橋 寛治      西山 浩気

長岡技術科学大学

{yamamoto, takahashi, nishiyama}@jnlp.org

## 1. はじめに

日本語における敬語は待遇表現の一つであり、語が表す事柄は同じだが敬意や丁寧さを示している(菊池 1997)。言語処理の観点からは同一の事象を表す表現の一つにまとめること、すなわち正規化を行うことにより各種応用処理の性能向上が期待できる(泉ら 2009)。例えば「お話になる」が「話す」と同義と認識できればテキストマイニングや情報検索で有用である。また日本語学習者にとって敬語の扱いは一般的に困難で、常体へと換言された文を提示することにより理解の促進が期待される。しかしながら、敬語の換言を言語処理の対象として取り扱った研究は(大野ら 2003)や(太田, 笠井 2006)など少ない。他に(山本ら 2005)が要約において部分的に敬語表現の換言を行っている。以上の研究では規則を元に削除や換言を行っているがいずれもツール化はされていない。

敬体を常体へと変換するのは一見簡単に見えるが実は語義の曖昧性や資源不足といった問題がつきまとう。例えば、「いらっしゃる」は「来る」「居る」「行く」の意味を持つ。語義を特定しなければ正しい換言は行えない。また「お知り合いになる」のような表現の単語解析結果を言い換える際には、名詞「知り合い」が動詞「知り合う」と対応しているという辞書が必要である。そこで、敬体の常体への換言のために必要な動詞-名詞対応辞書の構築及び換言処理を行う。

本稿では敬語の常体への換言処理、及び構築した資源について報告する。また本稿の成果を我々が独自に研究・開発している日本語解析システム「雪だるま」(Yamamoto et al. 2015)に組み込み、システムとして公開する。

## 2. 対象とする敬語表現

本稿で対象とする敬語表現を列挙する。(大野

ら 2003)にあるように、これらは規則変化と不規則変化に分かれる<sup>1</sup>。なお、雪だるまでは内部処理をすべて ID で行っているため、システム内部で換言後の活用調整をする必要はない。

### 2.1. 規則変化

何らかの敬語表現が付加されることにより規則的に敬語表現となるものである。

- ・接頭辞が付属し動詞化するもの

例1) お話になる→話す  
例2) お聞きする→聞く

- ・接頭辞が付加されるもの

例3) お風呂→風呂  
例4) お弁当→弁当

なお、「～される」や「～られる」は規則的に敬語表現を生成できるが、語義曖昧性解消が必要なため本稿の対象外とする。

### 2.2. 不規則変化

前項とは異なり、換言によって全く異なる表現になる場合である。下記に例を示す。

例5) おっしゃる→言う  
例6) いらっしゃる→来る、居る  
例7) ご覧になる→見る

ただし、不規則変化の一部表現は語義曖昧性解消を行わなければ言い換えることができないため、本稿では対象外とする。

<sup>1</sup> (大野ら 2003) らはそれぞれ付加形式と置き換え形式と呼んでいる。

- 例8) いらっしゃる→来る、居る、行く  
 例9) いただく→貰う、食べる、飲む

大野らと同様に、常体で言い換えることができない表現は言い換えない。下記に示す例のように、日本語には少数ではあるが敬体しか存在しない表現が存在する。

- 例10) ごめんください  
 例11) おしゃべり

### 3. 敬語表現の換言

#### 3.1. 雪だるまによる単語解析

単語解析<sup>2</sup>に日本語解析システム「雪だるま」を用いる。雪だるまの特徴として、表記ゆれの吸収および複合語処理を行う。本稿で取り扱う単語は雪だるまが出力した単語単位で処理を行う。表記ゆれの吸収によって例えば以下の表現を同一視できる。

- 例12) お知り合いになる = お知りあいになる

#### 3.2. 不規則変化辞書の作成

「拝見する」や「おっしゃる」のような不規則変化する語は置換により換言を行う。このため、該当する語に対して辞書を作成する。辞書は(菊池 1997)を参考にしながら作成した。現在は、お気に召す、お目にかかる、おっしゃる、申し上げる、思し召す、存ずる、拝見する、拝受する、拝借する、拝聴するの 10 語が登録されている。表 1 に辞書の一部を示す。

表 1 置き換え形式辞書

敬語	常体
お気に召す	気に入る
拝見する	見る
おっしゃる	言う

<sup>2</sup>日本語文の単語分割を行い、表記ゆれを集約した上で品詞などの様々な情報を付加することを単語解析と呼ぶ

### 3.3. 動詞名詞対応辞書の構築

「お考えになる」は規則的に変化するため、常体への換言は「考える」である。この処理を実現するには、名詞「考え」と動詞「考える」の対応付けが必要なので、動詞名詞対応辞書を構築する。作業は、簡単な規則で自動的に辞書を作成した後に、著者 1 名(西山)の主観により全確認した。作業対象は「雪だるま」の単語辞書に含まれるサ変動詞を除く動詞である。以下に単語辞書からの抽出規則を示す。

- i. 雪だるまの単語辞書から「～する」で終わる動詞を除く、いわゆる和語動詞を取得する。
- ii. 2文字の動詞の場合、2文字目のひらがなを「う」の音から「い」の音へ変換する。  
 例13) 「動く」→「動き」  
 例14) 「働く」→「働き」
- iii. 3文字以上の動詞の場合、手順 ii と同様に最後の1文字の音を変換する。もしくは、最後の1文字を削除する。  
 例15) 「決まる」→「決まり」  
 例16) 「生まれる」→「生まれ」
- iv. 手順 ii 及び iii で生成した単語が雪だるまの単語辞書内に名詞として存在するものを採用する。

このようにして得た動詞性名詞対に対して人手による整備を行った。現在、雪だるまの単語辞書には 26,945 語の動詞が含まれており、そのうち 3,235 語が上記規則により見つかった。これを人手によって整備した結果、2,700 対が獲得できた。構築した動詞名詞対の一部を表 2 に示す。

表 2 動詞名詞対応辞書の一例

動詞	名詞
動く	動き
考える	考え
送り出す	送り出し
足搔く	足搔き
痛む	痛み

### 3.4. 敬語表現の換言規則

入力には雪だるまを用いて単語解析された単語

列とする。敬語表現は下記の規則に基づいて検出を行う。以下に規則を示す。

- 1) 助動詞「ます」の削除
- 2) 置き換え形式の語の換言
- 3) 接頭辞「お／ご」+動詞+「になる」の動詞への換言
- 4) 接頭辞「お／ご」+名詞+「する／いたす」の動詞への換言
- 5) 接頭辞+名詞の接頭辞の削除

検出及び換言は規則 1～5 の順に行う。

### 3.4.1. 助動詞「ます」の削除

助動詞「ます」は丁寧表現である。本稿では「ます」を削除しても文の意味は保てるものとする。

例17) 彼が来ました。→彼が来た。

### 3.4.2. 置き換え形式

3.2 節で作成した不規則変化辞書に登録している語に対して換言を行う。

例18) ご覧になる→見る

### 3.4.3. 「お／ご～になる」形式

例19) お申し込みになる→申し込む

接頭辞+動詞+「になる」形式の場合、動詞を抽出することにより換言である「申し込む」を得ることができる。

### 3.4.4. 「お／ご～する／いたす」形式

例20) お知らせいたします→知らせる

例21) おしゃべりする→しゃべる

接頭辞+名詞+「する／いたす」形式の場合、接頭辞と「する／いたす」を削除した上で、名詞の単語を動詞へと言い換える必要がある。ここで、0 節で作成した動詞名詞対応辞書を用いる。上記

例 20 の場合では、名詞「知らせ」に対して動詞「知らせる」を付与することが可能となる。

### 3.4.5. 接頭辞の削除による換言

接頭辞「お」や「ご」を用いた単語は、接頭辞を削除するだけで常体へと換言することが可能である。

例22) お買い物→買い物

例23) ご指摘→指摘

ただし例外として、お子さん、お産、お蔵入り、お辞儀、お吸い物、おしゃべりの6語については接頭辞「お」を取り除くと意味が変化してしまうか、意味をなさない。これに対して我々は、雪だるまの単語辞書にこれら表現を追加する、すなわち「単語化」することで対応した。この結果これら表現は自動的に本処理の対象外となる。

## 4. 評価実験と考察

3 節であげた換言規則がどの程度の精度で正しく換言できるのかを評価する。換言対象には現代日本語書き言葉均衡コーパス (BCCWJ) を用いる。換言規則が適応できる文を規則ごとに 100 文抽出し、換言が適切かどうかの評価を行う。結果は規則別に表 3 に示す。

表 3 換言処理の結果

規則	誤り
置き換え形式	6
お／ご～になる	3
お／ご～する／いたす、ます	22
接頭辞の削除	6

BCCWJ 全体で規則が適応可能な換言対象は BCCWJ 中の約 16% の文であった。

### 4.1. 失敗例と考察

例24) もしお望みでしたら、～  
→ もし望みでしたら、～

常体へと換言を行うとすると、「もし望んでい

るのであれば、～」と機能表現を言い換える必要がある。

例25) いいおさらいになる

→ いいさらいになる

これは、「おさらい」が単語辞書に登録されていないため、接頭辞「お」+名詞「さらい」と解析される。この接頭辞を削除すると「さらい」となる。この問題は単語辞書の増強により改善できる。

## 4.2. n-gram 統計量での比較

敬語表現の換言を行うことにより、言語表現がどれほど集約されたか n-gram を用いて確認する。

雪だるまでの単語解析では複合語処理を行っている。そのため複合語の取り扱い、例えば「合法的」や「動的」といった語は「～的」というように複合語を示すシンボル「～」と末尾の形態素で表す。このように複合語の縮約を行ったうえでの単語 1~3gram の異なり数を表 4 に示す。

表 4 n-gram 統計量による比較

n-gram	異なり数	換言後 異なり数	縮約率
1gram	296,986	295,103	0.994
2gram	7,353,271	7,324,711	0.996
3gram	30,683,400	30,597,200	0.997

このように敬語表現を言い換えることにより縮約率は 0.5%程度とわずかであった。地道な言語表現の正規化が言語処理タスクの性能向上へと寄与するのではないかと期待している。統計的機械翻訳などの応用タスクでの効果を今後検証する。

## 5. おわりに

敬語表現の常体への換言規則を列挙し換言する方法を提案した。また換言を行うために必要な動詞名詞対応辞書を作成した。

今後の課題としては換言可能な敬語表現の拡充と曖昧さのある敬語表現の語義曖昧性解消および換言である。また今後の展望として敬語への換言(逆変換)も取り組む予定である。なお、本稿の敬語の常体への換言システムは「日本語解析

システム雪だるま」に組み込んで公開する。

## 使用したツールと言語資源

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/>

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

菊池 康人, 敬語, 株式会社講談社 (1997.2)

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, pp.264-267 (2009.3)

大野 満, 横山 晶一, 西原 典孝, 日本語敬語表現の変換・解析システム, 言語処理学会第 9 回年次大会, pp.218-221 (2003.3)

太田 ふみ, 笠井 俊信, 機械処理を意識した敬語判定法とその応用システム, 言語処理学会第 12 回年次大会, pp.1051-1054 (2006.3)

山本 和英, 安達 康昭, 国会会議録を対象とする話し言葉要約, 自然言語処理, Vol. 12, No. 1, pp.51-78, 言語処理学会 (2005.1)

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 (2015.10)