

用言等換言辞書を人手で作りました

山本 和英 吉倉 孝太郎

長岡技術科学大学 電気系

{yamamoto, yoshikura}@jnlp.org

1 研究の動機

筆者(山本)は2001年に、「換言知識は一つの重要な言語資源であるという認識を持ち、換言可能な表現を収集、整理して日本語言語処理研究者共有の財産とするのが望ましい」と書いた[1]。換言処理は言語初学者向け、文書校正などの需要の高い応用技術であると同時に、基盤技術として表現の類似度計算などの性能向上に大きく寄与するであろうとの認識は今でも何も変わらない。しかしながら、原稿執筆時において利用可能な汎用の日本語換言知識は存在しない。10年以上にわたって(本研究室を含めて)換言知識の獲得に関する様々な研究が行われてきたにも関わらず、その成果が言語資源として学界に蓄積されていない。これでは換言処理に関する議論がいつまでも深まらず、自然言語処理は進展しない。

以上の強い危機意識を動機として、我々は研究基盤整備のために網羅的で汎用の単語換言辞書を完全な手作業によって作成している。これまでに、動詞、サ変名詞、形容詞、副詞について概ね作業を終えたので、本稿ではこの作業について報告する。

2 関連研究

冒頭で述べた通り、日本語の一般的な語彙に対する研究利用可能な換言辞書は存在しない。

換言辞書に類似した言語資源としてシソーラス(is-a 関係オントロジー[2])がある。また、国語辞典の語釈文も同様に換言辞書に近く、実際に両言語資源を用いて換言処理を行う研究は少なくない(本研究室でも行っている[3])。確かに、我々がある語を別の語に言い換える際に「辞書的」な説明をしたり上位語を用いたりすることは一般的である。しかし、我々の直感として、人が行う換言はそれだけではないという印象を持っている。しかしこの仮説が正しいのかどうか、あるいは既存の言語資源で換言現象をどの程度網羅できているのかという検証報告は、我々の知る限り存在しない。このため本研究ではあえて国語辞典もシソーラスも用いずに換言辞書を作成する。これによって様々な疑問を(部分的に)検証できるのではないかと期待している。

3 辞書の記述内容

3.1 作業手順

まず、換言対象語を見てその語を使った例文を考える。次に、その例文から換言対象の語の意味を落とさないまま、どのような語にそのまま置き換えることが出来るかを考えてその換言候補を追加する。例えば「和える」という語に対して例文を考える(例えば「ごまと和える」)。この文で「和える」は「混ぜる」に置換可能なので辞書には「和える」の換言語として「混ぜる」と記載する。

和える → 混ぜる

換言候補は、日本語初学者からその言葉の意味を問われたときにどのように違う日本語で答えるかを念頭に置いて追加する。すなわち、作業者の感覚で明らかに難しい語には換言しない。

作業者は分からない換言対象の語があっても、国語辞書、シソーラス等は参照せずに作業を行う。

3.2 無記入を許す

作業においては無記入を許した。これは以下のような場合である:

- 用語の意味が明確に分からない
- 同一文脈で換言できる語を思いつかない
- 換言が可能であっても長い説明を要する(概ね内容語3語以上になる)

これは、作業の効率化のためでもあり、また無理に換言することで不自然な換言結果を作らないためでもある。ただし意味が明確に分からない語については、今後何らかの方法(例えば、別の作業者に依頼する)で入力を試みることを検討している。換言できる語が思いつかない場合も別作業者などの方法で検証が必要であるが、このような換言不可能な語の多くは基礎語彙に該当するのではないかと我々は考えており、そもそも全単語を換言する必要は必ずしもないと考えている。

3.3 多義語は文脈を付与する

換言対象の語に多義性があり、意味が大きく異なる

る場合は、換言対象の語を複製してそれぞれの語に対して候補を追加する。この際に、多義語に関しては考えた文を周辺文脈として記載する。例えば、「生かす」という語に対して「動物を生かす」という文と「長所を生かす」という文を考えた。前者では動物を「維持する」という意味となるが、後者は長所を「役立てる」という意味となる。従って本作業では、「生かす」に対して2項目を作成した。

(動物を)生かす → 維持する
(長所を)生かす → 役立てる

また、実際に用いる際の利便性を考え、複数の周辺文脈を思いついた場合は思いつくだけ記載する。

3.4 格変化は記述する

換言可能な場合であっても、格関係を変化させる必要のある語も存在する。例えば、「背く」という語の換言は「裏切る」が妥当と思われるが、そのままの形で両者の語が置換可能ではない。すなわち「上司に背く」に対しての換言は「*上司に裏切る」ではなく、「上司を裏切る」である。自動的な換言の際にこのような格変化も実現するために、換言前後において格変化が必要となる語に対しては下記のような情報を付与した。

(～に)背く → (～を)裏切る

3.5 慣用的表現はまとめて換言する

慣用句や慣用的な表現は構成単語単独で換言することができない。例えば「入れる」という語に対して「ギターを手に入れる」という文を考える。この場合、「入れる」のみを別の語に換言して同じ意味の文を作成することはできないが「手に入れる」は「得る」と換言できる。以上の情報を辞書に持たせるため、「手に入れる」という辞書項目を新しく作り、当該表現の換言候補として「得る」という単語を記入する。これは3.3節の周辺文脈を付加することと似ているが、周辺文脈も含めて換言する点が異なる。

手に入れる → 得る

3.6 意味の重複可能性は付記する

「片付ける」という語の換言としては「綺麗にする」とするのが妥当に感じる。例えば、「部屋を片付ける」に対して「部屋を綺麗にする」と問題なく換言できる。しかし、実際には「部屋を綺麗に片付ける」と言うこともできるので、これを同様に換言すると「*部屋を綺麗に綺麗に片付ける」という文になってしまう。すなわち、

「片付ける」の意味の中に「綺麗に」の意味を含む場合と含まない場合があり得る(と判断した)場合に、このような語の換言は、意味が重複して表出する可能性がある旨の付記をした。

片付ける → (綺麗に)する

3.7 その他

サ変名詞は名詞であるが、意味的にはサ変名詞の動作等をするという意味で用いられる場合が多いと考えた。このためサ変名詞の換言はすべて「する」を付加したサ変動詞として換言を行った。例えば、「息抜き」という語は「息抜きする」という語を換言することを考えて「休憩する」などとした。多くの場合、サ変動詞の換言結果の末尾に「こと」を付加すればサ変名詞としての意味になる(「息抜き」は「休憩すること」と換言できる)。

副詞の換言については、副詞を副詞に換言することにこだわらずに換言可能な表現を入力した。用言を修飾する副詞の換言については、副詞以外にも用言の連用形で表現できる場合がありうる。

4 作業内容と現況

今回作成した作業について述べる。実際に作った辞書の一部は付録に示す。なお、辞書作成に要した期間は概ね5カ月である。

4.1 作業対象と作業者

換言対象語は、形態素解析器 JUMAN(1)の形態素辞書から見出し語を抽出したもののうち、動詞、サ変名詞、形容詞、副詞の合計 12,813 語とした。ただし、3.2 節で述べた通り一部の語は換言しなかった。作業者は日本語を母語とする成人男性1人(著者である吉倉)である。

4.2 辞書の記述形式

今回作成した辞書の形式について述べる。

辞書は動詞、サ変名詞、形容詞、副詞の4つに分割して作成した。辞書はスプレッドシート上に、JUMAN の形態素辞書における見出し語(換言対象の語)、JUMAN の形態素辞書の記述情報、換言後の語、多義性のある語の場合に追加した用例、備考(タグ)から構成されている。備考欄は、作業者が作業しやすくなるためにつけたものや、実際に利用する際に注意する必要がある格変化について付与した。

3.2 節で述べた無記入については、実際に無記入だと未作業項目と区別がつかなくなって作業管理が難しくなるため、無記入である旨の情報を付与した。

また、3.5 節で示した慣用的表現については、使用

時に形態素解析の手間を省くため、あらかじめ当該表現を単語ごとに分割した状態で辞書に記述した。それらの語以外は全て単語分割せずに作成した。

4.3 作業結果

作業対象の語数と実際に換言を行った項目数、及び無記入とした語数を品詞別に表1に示す。3.3節で述べた通り作業対象1語に対して換言結果が1語とは限らないので、表中の「換言作成」欄と「無記入」欄を合計しても「作業対象」欄の数とは一致しない。表に示す通り、作業は約1万3千語を対象に行い、その結果無記入約3千語を除いた約1万語に対して換言辞書を作成した。

表1 換言対象語数と作業結果

| 品詞 | 作業対象 | 換言作成 | 無記入 |
|------|----------|----------|---------|
| 動詞 | 3,608 語 | 3,206 語 | 481 語 |
| サ変名詞 | 5,627 語 | 4,494 語 | 1,144 語 |
| 形容詞 | 2,335 語 | 1,851 語 | 496 語 |
| 副詞 | 1,243 語 | 785 語 | 463 語 |
| 合計 | 12,813 語 | 10,336 語 | 2,585 語 |

5 換言結果の観察

換言結果については、様々な面からの考察及び検証(必要に応じて辞書の修正)が必要であるため、ここではこれまでに気がついた点のみ述べる。

5.1 無記入について

無記入については、意味が分からない語が3分の2ほどあり、3分の1程度は簡単な語にできないものであった。これらの中にはビジネスやスポーツの特定分野でのみ用いられる語が多く含まれているために無記入としたものが含まれている。このような特定の分野の文書でのみ用いられる語については換言しにくい傾向にあるようだ。

5.2 和語と漢語の傾向

動詞は基本的に和語であり、和語を換言する際には修飾語+和語にする場合とサ変名詞を伴う漢語にする場合がある。基本的には簡単であろう語を選択しているので、和語にする場合が多かった。従って、和語でなく漢語へと換言した語については、換言された漢語はかなり基本的な漢語として用いることができるのではないかと予想する。

サ変名詞についても当然、同様なことが起こると考えられたが、即座に思いついた語を感覚的に追加しているため、動詞に比べてサ変名詞への変換が多く

見られる。従って、サ変名詞を実際に換言する場合には、複数回適用することでより簡単な和語へと変換可能だと考える。このことから基本的には和語のほうが簡潔な文章を構成するという点で使い勝手が良い語であり、それでも使われる漢語というものは基礎語彙として使われると考える。

5.3 副詞の換言について

副詞の中には「副詞+する」で用言として換言可能になる語も見受けられた。例としては「あつぶあつぶ」などが含まれ、「あつぶあつぶする」で「苦しい」という意味になる。同様に一般に呼応の副詞と呼ばれる副詞については、今回は例文を考えた際に呼応関係がすぐに出てくるものはそれをタグ付けしている。この呼応については、動詞にも存在していて、「必ずしも」という副詞や「否める」という動詞が当てはまる。

6 まとめ

用言等換言辞書を人手で作りました。JUMANの形態素辞書を作業対象として、国語辞典等を使用せずに動詞、サ変名詞、形容詞、副詞の合計約1万3千語の約8割にあたる約1万語の換言対を得た。

今回の作業は作業効率を優先したためにいくつかの問題点があるのは承知しているが、これだけの規模で汎用的な一次資源を得たことの意義は大きい。今後はこの辞書の品質向上及び拡充に努めていくと同時に、当該言語資源を利用した成果を実際に示すことで、筆者(山本)の10年来の主張である換言処理の重要性を訴えていきたい。

参考文献

- [1] 山本 和英. 換言処理の現状と課題. 言語処理学会第7回年次大会併設ワークショップ, pp.93-96 (2001.3)
- [2] 柴木 優美, 永田 昌明, 山本 和英. カテゴリ名と記事名の意味属性分類に基づくWikipediaからの上位下位関係オントロジーの構築. 自然言語処理, Vol.19, No.4, pp.229-279, 言語処理学会(2012.12)
- [3] 梶原 智之, 山本 和英. 小学生の読解支援に向けた語釈文による換言. NLP若手の会 第7回シンポジウム, (発表1) (2012.9)

使用した言語資源及びツール

- (1) 形態素解析器 JUMAN Ver.7.0. 京都大学大学院情報学研究科 知能情報学専攻
<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

付録

今回作成した辞書の一部を以下に示す。「慣用」タグは複数形態素の換言を示し(3.5 節)、換言語中の(...)は換言によって(...)内の意味が重複して表出する可能性があることを示す(3.6 節)。

動詞辞書

| 換言対象 | 換言候補 | 用例 | タグ |
|----------|---------|------------|----|
| 合う | 同じだ | (形,鍵,条件)が- | |
| 会う | (無記入) | | |
| 仰ぐ | 見る | (空,天)を- | |
| 仰ぐ | 尊敬する | (師匠,師)を- | |
| 仰ぐ | 求める | (指示,助力)を- | |
| 明かす | 過ごす | (夜)を- | |
| 明かす | 明らかにする | (種,真実)を- | |
| 案じる | 考える | (一計,策)を- | |
| 案じる | 心配する | (両親,身)を- | |
| 案ずる | 案じる | | |
| 弄る | 触る | (耳,吹き出物)を- | |
| 弄る | 変える | (庭,装丁)を- | |
| 歌う | (無記入) | | |
| 打ち明ける | 話す | | |
| を_侵す | に_勝手に入る | | |
| 収まる | 弱くなる | (騒ぎ,嵐,雨)が- | |
| 収まる | 入る | (胃,タンクス)に- | |
| 掛かり付ける | よく行く | | |
| 一線_を_画する | 違う | | 慣用 |
| 画する | 考える | (テロ,買収)を- | |
| 片付く | (綺麗に)なる | | |
| 片付ける | (綺麗に)する | | |
| 嘯む | (無記入) | | |
| 聞く | (無記入) | | |
| 聴く | 聞く | | |
| を_けなす | の_悪口を言う | | |
| 志す | 目指す | | |
| 心する | 覚悟する | | |
| 否める | 否定する | | |
| 下りる | (下)に行く | | |
| 下げる | (下)にする | | |
| 介する | 通す | | |
| 会する | 会う | | |
| 一堂_に_会す | 集まる | | 慣用 |

サ変名詞辞書

| 換言対象 | 換言候補 | 用例 | タグ |
|------|--------|----|----|
| アシスト | 助ける | | |
| 雨漏り | 雨水が垂れる | | |
| 依拠 | 基づく | | |
| 育成 | 育てる | | |
| 一喜一憂 | (無記入) | | |
| 賛同 | 同意する | | |

| | | | |
|----------|--------|--------|--|
| 榮進 | 出世する | | |
| 世話 | (無記入) | | |
| 出庫 | 出る | (車)が- | |
| 出庫 | 出す | (商品)を- | |
| シミュレーション | 試す | | |
| に_君臨 | を_支配する | | |
| オーダー | 頼む | | |
| 解雇 | 辞めさせる | | |
| 御碎儀 | 挨拶する | | |
| 御供 | ついていく | | |
| コンパイル | (無記入) | | |
| エラー | 失敗する | | |
| 改善 | よくする | | |
| 概算 | 計算する | | |
| ガイド | 案内する | | |
| 賛成 | 同意する | | |
| 増大 | 増える | | |
| ディベート | 議論する | | |
| 着手 | とりかかる | | |

形容詞辞書

| 換言対象 | 換言候補 | 用例 | タグ |
|--------|--------|----|----|
| 暖かだ | 暖かい | | |
| 温かだ | 温かい | | |
| 暖かい | 気温が高い | | |
| 温かい | 温度が高い | | |
| 詭え向きだ | 似合っている | | |
| 危ない | (無記入) | | |
| 危なかしい | 危ない | | |
| 危なげない | 危なくない | | |
| 危なっかしい | 危ない | | |
| 相容れない | 合わない | | |
| 青い | (無記入) | | |
| 青白い | (無記入) | | |
| 赤い | (無記入) | | |
| 悪質だ | たちの悪い | | |
| アクティブだ | 活動的だ | | |

副詞辞書

| 換言対象 | 換言候補 | 用例 | タグ |
|-----------|--------|----|----|
| 相次いで | 次々と | | |
| あつぷあつぷ_する | 苦しい | | 慣用 |
| 生き生き_する | 元気だ | | 慣用 |
| いずれ | いつか | | |
| あわよくば | もしかしたら | | |
| いきなり | 突然 | | |
| 敢えて | わざと | | |
| 改めて | (無記入) | | |
| 予め | 前もって | | |
| 有らん限り | 全部 | | |
| ありあり | (無記入) | | |
| 如何 | (無記入) | | |