

用言等換言辞書を用いた換言の考察

吉倉 孝太郎[†] 山本 和英[†]

[†]長岡技術科学大学 電気系 〒940-2188 新潟県長岡市上富岡町 1603-1

E-mail: † {yoshikura, yamamoto}@jnlp.org

あらまし 汎用利用可能な換言のための言語資源の構築を目的としている。先行研究として我々は人の感覚でどのような語へ置き換えることで換言可能となる語であるかを念頭においた用言等換言辞書を構築した。本研究では、実際に用言等換言辞書を利用して換言を行った場合にどのような換言が起こるか、という点に着目して換言を行った。本発表では、毎日新聞中に含まれる1文に対して1語の動詞のみを換言した場合に起こる現象について報告する。

キーワード 換言、用言、動詞、言語資源

1. はじめに

ある言語表現をほぼ等価の別表現に変換することを換言という。換言は、自然言語処理において、様々な応用分野を解決する技術として求められている。以下に示した文は互いに等価表現で表された文である。

(a) 犬と散歩する

(a') 犬と歩く

文(a)と文(a')の差異は述部が「散歩する」と「歩く」という点であり、これらの表現は対応している。

本研究における換言とは、このような置き換えることで意味を大きく損なわずに換言可能となるようなものを想定とする。このような換言対を集めたとき、その集合は実際の換言処理において、汎用利用可能な言語資源であると考えられる。また、このような観点で構築された資源で換言を行った場合の結果から換言処理に必要な情報とは何なのかを調査することが可能となると考える。同時に、このような汎用利用可能な換言資源を人手で整備して資源として利用可能にすることが換言処理を進める上では大事である。従って、本研究では、実際に換言可能であろう語を表層的に置き換えて換言することで、どのような換言が起こるかを考察する。

2. 関連研究

語を換言する場合には、その解決しようとする課題に関わらず、国語辞書との対応やシソーラスを使って換言する方法が考えられ、実際に換言問題を解く上で良く用いられている。美野ら[1]は日本語に慣れていない外国人のために、国語辞書を用いてニュース内容の難易語とされる名詞を、日本語能力試験の難易度を基準として平易化している。同様に梶原ら[2]は小学生のための読解支援として、国語辞書における語釈文やシソーラスにおける上位語および同義語から平易な語を

獲得し、学習指導要領に含まれるか否かを基準に換言を行っている。確かに、ある語を換言する場合には、辞書で用いられている語を使っての換言やシソーラス中の同義語、上位語などを用いて正しい換言を行うことができる場合がある。しかし、そのような語以外への換言を人は感覚的に行うことがあると我々は考えている。語釈文やシソーラスの関係以外に現れる感覚的な換言が可能となる資源を構築し、傾向を調査することで、これからの換言処理に必要な資源・技術というものを考察することが可能となると考える。我々はその前段階として、用言等換言辞書を構築した。[3]

用言等換言辞書については、詳しくは割愛するが、動詞・サ変名詞・形容詞・副詞について、ある語をどのような語に置き換えて説明するか、という観点で辞書やシソーラス等を参照せず、人間の感覚・経験のみから構築した辞書である。

3. 手法

3.1 換言手法

本研究では、換言の対象として、毎日新聞コーパス1999年版(1)を使用し、用言等換言辞書から動詞のみを対象として換言を行うことにした。これは、用言等換言辞書に含まれるサ変名詞、動詞、形容詞、副詞の中で、動詞に最も多様な表現が存在していると考えたためである。用言等換言辞書に含まれる動詞から毎日新聞コーパスからランダムで文を抽出し、形態素解析器JUMAN(2)で形態素解析を行ったあと、形態素解析結果中に含まれる用言等換言辞書で換言可能な語(以下、換言対象語)を1語選定し、換言対象語と用言等換言辞書に含まれている語(以下、換言候補語)と入れ替えることにより換言を行うことにする。例として以下に示した文(b)を換言することを考える。

(b) “二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」と意気込んでいる。

文(b)中には履く、起こす、意気込むという3語の動詞が含まれている。これらの動詞のうち、用言等換言辞書の中には履く、意気込むが換言対象語として換言対を持っている。起こす、は別の語に換言することが困難であるとして換言対を作っていない。従って、今回用言等換言辞書を用いて換言可能な語はこの2語である。実際の換言をする場合には評価の都合上、どちらか一方のみを換言、評価する。従って、文(b)を換言する場合には、以下の文(b₁)と文(b₂)を別のもととして評価する。

(b₁) “二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」とやる気になっている。

(b₂) “二足のわらじ”を身につけて「関西にネットベンチャーの波を起こす」と意気込んでいる。

文(b₁)においては「意気込む」という換言対象語が「やる気になる」という換言候補語へと換言されている。文(b₂)においては「履く」が「身につける」という語へと換言されている。これらの各文について、それぞれの換言対象語を中心に換言前後の結果がどのようになっているかを評価する。

また、用言等換言辞書内においては、語義が複数あると考えたものについては換言対が複数存在している。例えば、「仰ぐ」という語の場合には、「見上げる」、「尊敬する」、「求める」という3つの換言対が存在している。本研究において、これらの語義をどのように選択すれば良いか、という問題と換言という事象は別問題だと考える。従って、今回の評価では、あらかじめ上手く語義が選択できたものとして評価する。換言対を持つ換言対象語は必ず換言するものとする。

3.2 評価手法

3.1 節で構築した換言手法に対する評価は全て人手で行った。この節以降、出てくる例はすべて実際の作業対象となった例である。作業者は論文筆者の吉倉を含む20代の日本語母語話者である男性2名である。評価においては、作業者間による評価の差異を失くすため、作業開始時に両者の感覚を合わせるために同じ換言データを対象として換言を行う練習を行った。また、文章全体を見て評価することは評価効率を著しく悪化させるため、換言対象語を中心とした複数形態素のみを見て評価する。もし、評価中にその形態素数で評価できない場合には評価しない。たとえば文(c)を対象に評価を行う。

(c) ないか」と案じる(→考える、心配する)人もいる

作業者は、換言対象として下線を施した「案じる」の直後の()内の換言候補(文(c)中では考える、心配する)の中から換言候補としてふさわしいものを選び、合うように活用させて、入れ替えた場合に文がどのようになるかを評価し、その評価を記録する。評価は以下の5つに分ける。

・換言可能

直接置き換えても意味が大きく変わらず、換言可能であると評価できるもの。例えば以下の換言対象の文(d)は換言可能である。

(d) 釈迦が示した(→見せる)人間としての極限最高の

また、用言等換言辞書内では、可能動詞は原形に換言して、可能形であるとタグをつけているため、文中で可能動詞として用いられているものは全て換言可能であるとする。具体的には「換言候補語+ことができる」と補って換言する。例えば、文(e)も換言可能である。

(e) から産地直送の新鮮なタラバガニを思う存分楽しむ(→楽しむ)のが、このお店

文(e)の場合、「タラバガニを思う存分楽しむことができるのが」と換言されると考えて評価する。

・条件付換言可能

換言した場合、換言対象語以外に何らかの補正を施すことで、換言可能であると評価できるものを条件付換言可能であると評価する。例えば以下の文(f)は条件付換言可能である。

(f) 退任するのに伴い(→一緒に)、参院自民党は30日、

文(f)では活用を揃えて換言した場合、「退任するのと一緒に、参院自民党は30日、」となる。この例において、下線を施した助詞「に」が連続していることに対して、評価者は違和感を覚えた。しかしながら、この文章は、例えば「退任するのと一緒に、参院自民党は30日、」と補正することで自然な文章となる。このように助詞を変換する、または助詞や形式名詞などを追加する、などの補正を文に施すことによって文が換言可能となる場合は条件付換言可能だと評価する。

・不自然な換言

文を換言した場合に、文が表現していることは評価

者に理解できるが、日本語として、そのような表現は使わない、というように評価者が考える場合はこの不自然な換言となる。条件付換言と似通った評価であるが、文の機能表現の補正によって換言可能とならないものが不自然な換言となる。例えば、文(g),(h)が不自然な換言となる。

- (g) 新たな仕事に一步踏み出した(→はじめる)。
- (h) なかで、カザルスホールにエールを送る(→ついていく、渡す)など積極策を展開している

評価者は文(g)を換言した「新たな仕事に一步はじめた」という文章が意図するところは理解可能であるが不自然であると評価した。このとき、条件付換言可能となる例であげた、機能表現の補正のみによってこの文章を自然なものとするのは難しい。また、文(h)中においては「エールを渡す」という表現は評価者が実際に換言する場合「渡す」という表現に換言しないが、文章の意図は理解可能と評価できるため、このような換言を不自然な換言とした。

・ 換言不可能

換言することによって換言対象の文の意味が読み取れなくなってしまうと評価者が判断した場合、換言不可能であるとした。例えば文(i),(j)が換言不可能とした例である。

- (i) フセイン大統領が絡む(→巻きつく)「政治事件」に発展する
- (j) やや小太り(→大きくなる)のミハイル・ネポロジネフさん(29)は

・ 例外

評価対象中に何らかの原因により評価することができないものについては例外として評価しないことにした。例えば、文(k)のようなものが例外に含まれる。

- (k) 4兆円のノシ袋つきで(→はみ出る)国有・長銀が外資系投資集団に譲渡される。

文(k)中の「つきで」という語に対する換言候補「はみでる」は「突き出る」という語の換言候補となっている。しかし、文(k)中での「つきで」は明らかに「突き出る」という意図で使われていない。このようなものが評価対象に含まれている場合、評価者がこれを取り除いて評価データに含めないことにする。

3.3 評価データの構築

本節では、本研究で実際に構築して、評価に利用したデータについて述べる。まず、毎日新聞コーパス99年版から無作為に6000文を抽出し、JUMANによる形態素解析を行う。その6000文中から1文につき1語の換言候補を持つ文を構築した。構築した文は8646文となり、これらの文から1000文を抽出したものを、評価者2名の感覚を合わせるための練習に用いた。構築した8646文に用いられている換言対象語の特徴を調べたところ換言対象語は1246種類出現していた。これらの語の出現頻度の分布を調べたとき、198種類の語が10回以上出現していた。この出現回数が多い198種類の語は全体の50%ほどを占めている。出現回数が多い語は特定の表現としてのみ新聞中で出現することが多い傾向にある。例えば「述べる(147回出現)」という語は、新聞中ではほぼ「言う」という表現と換言可能な文脈で用いられる。このようなある特定の語を一定の表現としてのみ多く含んだものを評価することは、今回の換言という現象の考察、という目的を鑑みれば必要ないと考え、全ての語を一律10回までを上限に評価対象とすることにした。この評価対象には評価の練習に用いたデータは含まない。練習データを除いた7646文から評価するデータとして3500文を抜きだし、それぞれの評価者が個別に評価した。3500文のうち、212文は評価者2名で評価するデータとなっている。この212文の一致度が二人の作業の一致の信頼性を表すことになる。

4. 結果

4.1 評価作業の一致度

作業者2名が個別に行った同じ換言対象語に対する作業の一致度を表1に示す。また、表2に評価者による評価の分布を示す。

表1. 評価作業一致度

	換言対象語数	作業一致数	一致度[%]
全換言対象	212	178	84.0
例外を除く	194	166	85.6

表2. 評価者による換言の分布

換言種類	評価者 A	評価者 B
換言可能	94	98
条件付換言可能	8	8
不自然な換言	25	21
換言不可能	68	72
例外	17	13

表 1.中の例外を除く、とはどちらか一方でも例外として飛ばしたものを除いた評価の結果の一致度を表している。

4.2 評価者間の差異

それぞれの評価者が同じ換言対象語を個別に評価した 212 文のうち、一致しなかった 28 文の評価がどのような分布となっているかを表 3.に示す。二人の評価がそれぞれどのように違っているかを示している。

表 3.評価者で一致しなかった組み合わせ

評価の組み合わせ	一致しなかった文数
換言可能-不自然な換言	11
換言可能-条件付換言可能	4
換言可能-換言不可能	3
条件付換言可能-換言不可能	1
条件付換言可能-不自然な換言	3
不自然な換言-換言不可	6

4.3 評価の結果

2 名による独立した評価の結果を表 3.に示す。独立した評価の結果とは、4.1 節、4.2 節に示した作業一致度の評価に用いた 212 対を除いた語の評価結果の分布である。() 内の数字は練習に用いた 1000 語を含んだ結果である。

表 3. 評価結果の分布

評価の種類	評価語数	割合[%]
換言可能	1671(2093)	53.9(51.8)
条件付換言可能	187(266)	6.0(6.6)
不自然な換言	402(537)	13.0(13.3)
換言不可能	840(1143)	27.1(28.3)
全数	3100(4039)	
例外	188(249)	

5. 換言の考察

本章では、実際の換言結果を随時例示しながら考察する。

5.1 評価者間の差異の考察

4.1 節に示した結果により、評価は 80%を超えて一致しているため、評価者による感覚の差異は際立って存在していない、と言える。

一致しなかった 28 文について、どのような差異があるのかについて一部考察する。

以下に表す文(i),(ii)が評価者間で差異の現れた例である。

(i) 電話の向こうから喜びがひしひしと伝わって(→渡る)きた(換言不可-不自然な換言)

(ii) 1機の価格が200億円を超える(→越える)ため、代わりに9機分1133億(換言可能-換言不可能)

文(i)はどちらの評価者もそのままでは日本語としておかしい、と評価したが、「伝わる」と「渡る」で文章の意味が変わるのか、という点で差異が現れた。この差異はこれらの動詞に対する表現の感覚の違いであり、どちらの評価者も「あるものが伝わる」と「あるものが渡る」ではあるものが移動している、という点は一致している。しかし、どのように移動しているか、という点を許容できるか否かが揺れてしまう。このような揺れを吸収して、より自然な換言候補を選ぶためには、動詞「伝わる」と「渡る」の移動の仕方をそれぞれ定義する、などの手段が考えられる。

文(ii)の例では、漢字の異表記による意味感覚の違いが表れている。「<金額>を超える」の場合、「<金額>を上回る」という意図であると評価者はどちらも評価している。しかし、「<金額>を越える」の場合、「<金額>を(物理的に)通り過ぎる」のように捉えて換言不可とする評価者と「<金額>(の基準)を過ぎる」というように捉えて換言可能としている評価者に分かれた。<数量、金額>を超える、という点を無視して換言して良いのか、という点について評価者の感覚が異なる例である。このような問題はたとえば、「変える」の異表記群「替える、代える、換える」などにも起こると考える。より意味を限定的にするために用いる後者群と意味を曖昧にした前者を区別するかどうかは、換言の目的により判断する。そのために、資源としては、(i)の例で示したことと同様に、どのような変化があるのか、を記述しておくことが考えられる。

5.2 評価による特徴

4.3 節より、実際に換言可能だと考えて構築された換言辞書で、動詞の 50%ほどの換言対象語は適切な換言と評価されている。本節では、その他の適切な換言ではないと評価された語の特徴を考察するとともに、適切な換言がなされるためにはどのような情報が必要かについて考察する。

5.2.1 換言不可能語

換言不可能とした語については概ね以下のような傾向が見られた。

- ・用言等換言辞書に適切な換言がない

例えば「解く」という語は「答える」という語のみ換言候補として持っている。したがって、「拘束を解く」

などという文脈で用いられる場合にも全て「答える」と換言されてしまい、全て換言不可能と評価される。これらのような理由で換言不可能だと評価されたものについては、換言辞書に改めて換言を登録することで対応が可能である。

・分割すべき対象でない換言対象語

複数形態素のまとまりとして一つの意味を為すような表現の一部を換言した場合、意味が通じない換言が為される場合がある。例えば、話し合い (→同じだ)、取り上げる (→渡す、上にやる)、過ぎ (→経つ、通る) 去る、連れ去る (→出る) などの複合動詞の一部のみを換言してしまう例である。しかし、複合動詞の中でも、換言可能な例も存在する。花が咲き (→開き) 始めるという場合、「花が開き始める」と換言可能であると評価できる。このような差異があることから、複合動詞を扱う際には、動詞間の構成要素がどうなっているかを考慮して換言を行う必要がある。また、それぞれの複合動詞が一意の意味を持つ傾向にあるのであれば、例えば、「3時間が過ぎ去る」を「3時間が経つ」と換言 (複合動詞「過ぎ去る」を併せて「経つ」と換言) することで、複合動詞の一部を換言する問題は解決可能となる。

複合動詞以外の例としては、小太り (→大きくなる)、早指し (→示す) などの接頭辞+動詞の場合に接頭辞を考慮して換言されない結果、換言不可能と評価される例が見受けられた。接頭辞による修飾が考慮されていない例では接頭辞を、全て動詞の程度・数量を修飾する形へと併せて換言することで自然な換言となると考える。例えば、「小太りの男」の場合接頭辞「小」を「少し」へと換言することで「少し大きい男」(この場合条件付換言可能となるが) と換言することができると考えられる。

また、慣用的に用いられる表現を分割して換言した結果換言不可能であると評価された例も見受けられた。例えば、祖父が息を引き取る (→もらう)、特定の状況に的を絞る (→出す) などが見受けられた。これらの語は動詞1語のみを換言していること自体が誤りとなっている例であると考え。つまり、「祖父が息を引き取る」に含まれる「引き取る」という表現は「祖父が子どもを引き取る」という表現と違い、「息を引き取る」という句で「死ぬ、亡くなる」と換言すべきである。このような表現については、単語単位で表層的に見て換言すると誤りとなる。従ってこれらの慣用的表現については、表現をまとめた状態で資源が持っている必要がある。今回の場合は「息を引き取る」が「死ぬ」となる、という例を用言等換言辞書に追加することで対応できる。

・換言すべきでない例

動詞の中には機能語のように用いられる動詞が存在する。「に対する (→向かい合う)」や「にとって (→持つ、行う、捕まえる、取る、食べる、盗む)」などが挙げられる。これらの語が機能語のように用いられた場合、代替表現が用言内に見つからないものとなる。

「選挙に対する不満」という表現の場合、「選挙についての不満」というように換言することが可能であるが、「ついでに」は用言としての換言ではない。また、「自分にとって最良の方法」という表現の場合も同様に、「とって」は動詞であるが、換言候補の動詞で換言して同様の意味を作るとは難しい。これらの語は前項に示した適切な換言のものと換言可能となる換言候補を定義できない点で異なり、また、分割せずにまとめて換言することで換言可能になる問題ではないという点で分割すべき対象でない換言対象語と異なる。

5.2.2 条件付換言可能語

どのような補正を行うことで換言可能となるか、という点について、以下のようなものが見受けられた。

・動詞の名詞形をサ変名詞へ換言

「見通す」という換言対象語は「予想する」という換言候補を持っている。例えばこのとき、「総務省の見通しによると」という表現がある。このとき、評価基準に依れば、「総務省の予想しによると」となり、不自然な文章となる。このような場合「見通し」を用言としてではなく、名詞として換言することで、換言が可能となる例が見受けられた。この現象は動詞の換言対象語がサ変名詞の換言候補語を持つ場合のみ起こると考えられるため、補正は難しくないと考える。

・助詞を補正する

「疑心暗鬼に陥る」という表現は「疑心暗鬼から抜け出せなくなる」という表現に換言可能である。「300台を上回る」という表現が「300台より上になる」という表現などが存在している。これは格助詞の変換が補正として必要であるものである。この補正を施すためには、ある用言がどのような格を持つか、という情報が必要である。これを解決するためには、動詞の格フレーム情報を用いて換言することが考えられる。

・意味を付与する補正

自然な表現へと換言する場合に、可能の助動詞、否定の助動詞、形式名詞等を補正する必要な換言が見受けられた。例えば、以下のような表現である。「経済が分かる(→理解する)」という表現に対しては「経済が理解できる」と可能の意味をくわえて補正をする必要がある。「注意を怠ら(→しない)ない」という表現で用いられる換言対象語「怠る (→しない)」は換言候補語がすでに否定の形になっている。したがって、「注意を

怠らない」という表現の場合、「注意をする」というような補正を施す必要がある。用言等換言辞書が換言候補語を否定となっている、という情報を持っていれば補正は容易くできると考える。形式名詞を補う例では、「問い合わせ」などが見受けられた。これらの語は5.2.2 節で触れた動詞の名詞形をサ変名詞と換言する例と類似のものであるが、今回の場合、換言対は「問い合わせ（→聞く）」となっている。したがって、「問い合わせは事務局まで」などの文があった場合、「問い合わせ」を同活用の連用形としてではなく、連体形＋形式名詞と補正して換言することで可能であると考えられる。例文は「聞くときは事務局まで」と換言される。しかし、全ての名詞として用いられる動詞の連用形を補正して換言することは適切でないと考える。例えば、以下のような例の場合である。「失踪騒ぎ（→うるさくする、言う）を起こした」この場合、「うるさくすること」と補正しても適切な換言とはいえない。全ての名詞として用いられる連用形を換言する場合には、動詞ではなく、名詞として同等の表現を用いて換言しなければならないと考える。

5.2.3 不自然な換言

不自然な換言だと評価された文については、以下に示したものが見られる。

・過度の抽象化・具体化

明確に意味が違う、と断じることができないが、ある大事な語の持つ言外の意図が確かに欠落してしまう場合である。

例えば「男が殴った」という表現を「男が叩いた」とした場合、目に浮かぶ大まかな光景は同じであるが、表現の差により行動の程度が欠落している。状況に応じては自然な換言の範囲に収まることが多いのがこの場合である。また、逆の状態も考えられる。例えば、「だます」という語に対しては、「うそをつく」という換言候補が存在する。「だます」という行為には確かに「うそをついている」状態を含んでいる場合があるが、常にうそをついている状態か、という点では語弊を招く可能性がある換言となる。

・主語、目的語などから連想する語への感覚の違い

これらの例は、「破る（→引き裂く、負かす）」という表現である。「破る」という語を使った表現としては換言可能と評価できる「ライバルを破る（→負かす）」や「紙を破る（→裂く）」と言った表現が存在する。しかし、「鼓膜を破って」という表現の場合、一転して不自然な換言となる。「鼓膜を破って」の場合、対象物である鼓膜に対して何らかの物がぶつかって、穴が開くような感覚を評価者は持つから、「裂く」とは言わない、と評価者は考える。「紙を破る」の場合、張力によ

って対象である紙を二つにする、という感覚を対象者は持つため、「裂く」が妥当な換言と評価する。このような感覚は用いられる用言、今回は「破る」の持つ語の感覚と名詞に対する感覚の関係によって妥当であるか否かが評価できると考えられる。

6. 今後の課題

考察を踏まえ、今後の課題として用言等換言辞書に追加すべき情報を挙げる。まず、換言可能と判断される可能性が低い語の中では分割して換言できない語を句として換言できるように換言対象語を拡充することが必要である。例えば、換言対象語として、複合動詞リスト・レキシコン[4]を追加することが挙げられる。補正が必要となる条件付換言の中では、統計情報である格フレーム情報と併せて換言することで補正を出来る場合がある。この格フレーム情報で補えない換言については人手で整備する必要がある。

7. まとめ

本研究では、汎用的に換言可能である資源の構築を目指して、用言等換言辞書を使って動詞を表層的に置き換えることによる換言実験を行った。現状として、動詞のみで50%の換言は自然であるとの結果を得た。換言できていないものを換言するためには、換言対があるものに対しては複数形態素にまたがった句として一つの意味を持つものをまとめて換言する必要があることを示した。

使用した言語資源及びツール

- (1) 毎日新聞社,CD-毎日新聞データ集 1999 年版,1999
- (2) 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>

文 献

- [1] 美野秀弥,田中英輝.国語辞典を使った放送ニュースの名詞の平易化言語処理学会第16回年次大会発表論文集,pp.760-763,2010
- [2] 梶原智之,山本和英.小学生の読解支援に向けた複数の換言知識を併用した語彙平易化と評価,言語処理学会第19回年次大会発表論文集,pp.272-275,2013
- [3] 山本和英,吉倉孝太郎.用言等換言辞書を人手で作りました,言語処理学会第19回年次大会発表論文集,pp.276-279,2013
- [4] 神崎享子,『複合動詞レキシコン』ver.1—形態的・統語的・意味的情報付与—,言語処理学会第19回年次大会発表論文集,pp.761-764,2013