

動詞の振る舞いに基づく形容詞から動詞への言い換え*

藤田篤[†] 増野成章[‡] 佐藤理史[†] 宇津呂武仁[‡]

[†]名古屋大学工学研究科 [‡]京都大学情報学研究科

{fujita,ssato}@nuee.nagoya-u.ac.jp, {masuno,utsuro}@pine.kuee.kyoto-u.ac.jp

1 はじめに

例(1)に示すような種類の言い換えは、単純な統語構造変換で実現できる、生産性が高い言い換えである。

- (1) a. 暴風雨に見舞われた⇔暴風雨が見舞った (態交替)
b. 木々が風に揺れる⇔風が木々を揺らす (自他交替)
c. 友人の話に刺激を受ける⇔友人の話に刺激される
(機能動詞構文(名詞+動詞)⇔動詞)

表現を構成する語に依存して言い換える可否や言い換え方に違いがある(たとえば「感動を受ける」⇔「感動する」と(1c)の態の違い)ものの、ある部分は、語の振る舞いや意味特性に基づいて構成的に説明できる。したがって、言い換える種類ごとに語の特性の影響や文脈依存性を調査して言い換え生成モデルを構築すれば、認識・生成できる範囲を、個々の表現対を列挙するよりも効率的に広げられる可能性がある。

本稿では、このような言い換えるの一つである、例(2)~(4)のような形容詞から動詞への言い換えるを扱う。

- (2) s. 古いお寺の年老いた住職を訪ねた。
t. 古びたお寺の年老いた住職を訪ねた。
(3) s. 近年評価の高い蔵元の酒を取り寄せた。
t. 近年評価の高まっている蔵元の酒を取り寄せた。
(4) s. 得点が決定的だったに惜しいプレイだった。
t. 得点が決定的だったに惜しまれるプレイだった。

このような言い換えにおいては、品詞を変える際に生じる意味の差異をなくすために、語の特性に応じて適切な動詞接尾表現を付加する必要がある。たとえば上の例では、(2t)には「た」、(3t)には「ている」、(4t)には「られる」を後続することで適格な言い換えとなる。また、これらを組み合わせた「られた」、「ていた」などの表現を生成すべき場合もあると予測される。これをふまえ、我々は次の2点について調査・検討した。

1. 言い換え時に付加すべき動詞接尾表現の多様性
2. 言い換え生成時の動詞接尾表現の曖昧性解消方法

2 現象分析

所与の文脈における形容詞を動詞に言い換える際、形容詞の文脈中の意味を保存するためにテンス、アスペクト、ヴォイス表現などの動詞接尾表現が用いられる。そこでまず、形容詞とそれと言い換えるの関係にある動詞の対、およびそれらを適用した言い換え事例を収集し、付加すべき動詞接尾表現の多様性を調査した。

*Paraphrasing Adjectives into Verbs Based on Syntactic and Semantic Properties of Verbs.

Atsushi Fujita[†], Naruaki Masuno[‡], Satoshi Sato[†], and Takehito Utsuro[‡]

[†]Graduate School of Engineering, Nagoya University

[‡]Graduate School of Informatics, Kyoto University

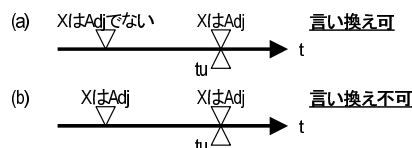


図1: いわゆる属性形容詞をいわゆる状態変化動詞に言い換える場合の適否の判定基準(X, Y: 対象物, Adj: 形容詞が表す属性, t: 時間軸, tu: 発話時間)

2.1 対象語彙

今回は、音(語根)および文字(漢字)を共有する形容詞と動詞の対を分析対象とした。たとえば「古い」⇔「古びる」、「速い」⇔「速まる」「速める」を含め、「冷たい」⇔「冷やす」、「忙しい」⇔「急ぐ」を含めない。より一般的な語を扱うために、日本語中核語彙JCore [8]の中核3レベル(A1, A2, B), IPAL形容詞辞書[4], IPAL動詞辞書[3]を対の収集源とし、次の手順で語対を収集した。

JCore: 当該レベルの9,767語中、形容詞は248語、和語動詞は1,801語含まれている。各形容詞に動詞を、各動詞に形容詞を、内省に基づいて付与する。

IPAL辞書: IPAL形容詞辞書(135語)の『派生動詞』の項、およびIPAL動詞辞書(和語動詞は802語)の『転成形容詞』の項を参照して形容詞と動詞の対を抽出する。ただし、形容詞と動詞の少なくとも一方がJCoreの中核3レベルに含まれている対のみ、上のすべての和集合を取り、形容詞と動詞の対127対(形容詞85語)を得た。

2.2 言い換えとしての文の適否

所与の文中の形容詞を動詞に置き換えるだけでは必ずしも適格な言い換えにはならないし、どのような動詞接尾表現を付加しても適格に言い換えられない場合や評価者間で適否が揺れる場合もある。そこで今回は、

「元文と言い換え文の各々の解釈に重複があればこれらの文は言い換えるの関係にある。」

という基準で言い換えるの適否を判定することにした。

より詳細な基準の例を1つ示す。2.1項の語対の大半を占めるいわゆる属性形容詞といわゆる状態変化動詞の対(たとえば「緩い」⇔「緩む」「緩める」)については、図1のように時間軸を設けて解釈を整理した。図の(a), (b)の各々に相当する言い換えるの例と適否の判定理由を(5), (6)に示す。

- (5) s. 緩いねじを締め直した。
t. 緩んだねじを締め直した。

表 1: 形容詞の用法の分布 (毎日新聞 1995 年版)

用法	形容詞全体	対象形容詞
基本形の連体修飾用法	63,739	21,302
基本形以外の連体修飾用法	11,522	3,150
「する」、「なる」を後続しての動詞用法	7,113	2,295
「さ」を後続しての名詞用法	17,203	7,537
その他 (連用修飾用法, 述語用法を含む)	106,192	27,665
合計	205,769	61,949

説明 s., t.ともに「(ある特定の)ねじ」が「緩くない状態」から「規制が緩い状態」に変化した結果状態を指しうるため言い換え可.

(6) s. 地球が丸いことを知らなかった.

t.*地球が丸まっ{ている, た}ことを知らなかった.

説明 s. が「地球が丸い」というある定常的な性質のみを表すのに対し, t. は「地球」が「丸くない状態」から「丸い状態」に変化した結果状態のみを表す. これらに共通の解釈がないので言い換え不可.

2.3 事例の収集

文中での用法ごとに形容詞を動詞に言い換える際の言い換えの可否, 言い換え方が異なると考え, それらを調査した. 毎日新聞 1995 年版における形容詞の用法ごとの頻度は表 1 に示す通りである. この中で最も大きな割合を占める基本形の連体修飾用法を第一の調査対象とし, 比較のために述語用法を取り上げる.

まず, 基本形の連体修飾用法の言い換え事例を収集した. 表 1 の 21,302 件から無作為に 960 文取り出し, 形容詞を対応する動詞に言い換えた文の適否を作業者が判定した¹⁾. 結果, 219 文に対する 253 件の言い換え事例を得た. この中には, 被修飾名詞が実際は複合辞や形式名詞である 77 文に対する 91 事例も含んでいた.

次に, 述語用法の言い換え事例を収集した. 機械的には連用修飾用法と区別できないこと, および連体修飾用法での知見をふまえ, IPAL 形容詞辞書の例文を事例の収集源とした. 2.1 項で収集した形容詞のべ 364 語義の例文について, 同じく 2.1 項の範囲で対応する動詞を『派生動詞』に持ち, かつ当該形容詞が述語として使われている文を取り出したところ 157 文 (語義) を得た. これらを連体修飾用法の場合と同様に自動的に動詞に置換し, 作業者が適否を判定した結果, 70 文に対する 84 件の言い換え事例を得た.

連体修飾用法と述語用法では, 事例の収集源における形容詞の出現頻度などが異なるが, どちらも概ね 4~5 トークンに 1 つしか言い換えられなかった. 形容詞の多義性と実際の文脈における品詞交替の難しさが, 他の種類の言い換えに比べて言い換えられる割合を少なくしていると考えられる. また, 形容詞ごとに言い換えやすさに違いが見られた. 文脈に依存せず必ず言い換えられる語はなかったが, 「可愛い」⇒「可愛がる」, 「頼もしい」⇒「頼む」など, 事例収集の範囲では言い換え可能な文脈が見つからない語があった. このよう

¹⁾ 人目の作業者も言い換え可能と判定した事例を正例とした. なお, 言い換え事例の適否判定にあたり, 動詞接尾表現を付加する以外にも格助詞を変更する, 副詞を追加するなどの修正を認めた.

表 2: 形容詞と動詞の対ごとの動詞接尾表現の分布

動詞接尾表現	C_{ad-c}	C_{ad-o}	C_{pr-1}	C_{pr-2}
る	9	2	16	0
ている	5	10	42	0
られる	14	13	8	0
られている	2	1	5	0
た	57	23	0	7
ていた	2	2	0	2
られた	6	2	0	1
られていた	0	0	0	1
た, ている	4	1	0	0
た, る	1	1	0	0
てある	0	0	2	0
合計	100	55	73	11

な文脈における可換性についての知識も言い換え生成・認識のために蓄積すべきである.

2.4 動詞接尾表現の分布

次節以降で述べる言い換え生成実験のため, 便宜的に, 基本形の連体修飾用法の言い換え事例の中から被修飾名詞が複合辞である文を除く. そして残りの 142 文 (162 件) を分析用の 91 文 (C_{ad-c} , 事例 105 件) と評価用の 51 件 (C_{ad-o} , 事例 57 件) に分けた. また, 述語用法の言い換え事例を原文の形容詞が「た」を後続する 10 文 (C_{pr-2} , 事例 11 件) とそれ以外の 60 文 (C_{pr-1} , 事例 73 件) に分けた. 各々における言い換え先の動詞ごとの動詞接尾表現の分布 (表 2) から以下に示す特徴が観察された.

- 連体修飾用法では, テンスについては「る」と「た」, アスペクトについては「ている」の有無, ヴォイスについては「られる」の有無が組み合わさった表現のみ用いられた. 述語用法ではこれに加えて「てある」も用いられていた.
 - 連体修飾用法では「た」と「られる」が大半を占める. これは影山 [5] の『動詞が「た」形を取る場合に結果状態の意味を表すため, 完了形容詞的な性質を帯びる』という分析とも類似する.
 - 「た」を後続しない述語用法 (C_{pr-1}) では「ている」と「る」が事例の大半を占め, 「た」を付加すべき場合はなかった. 一方, C_{pr-2} の 11 例を見る限り, 「た」を後続する述語用法の形容詞は原文の形容詞を基本形にして言い換える際の動詞接尾表現に「た」を加えることで構成的に表現できる.
 - 文脈によって同じ動詞が異なる接尾表現を要求することがある一方で, 複数の動詞接尾表現を取りうる例 (7) のような事例は少なかった. これより, 動詞そのものの性質に加えて, 連体修飾用法の被修飾名詞および述語用法の主格からも語彙的に強い制約がはたらいていると予想される.
- (7) s. 長いモデル論争に一つの結論をつけた.
t. 長び{く, いた}モデル論争に一つの結論をつけた.

3 言い換え生成手法

前節の調査結果をふまえ, 所与の文脈における形容詞をそれに対応するある動詞に言い換える際の動詞接尾表現を推定する, という問題に取り組む.

3.1 素性

文献 [2] では、例 (1c) のような言い換えにおけるヴォイス、格助詞の非決定性を語彙概念構造の変換によって解消している。一方、本稿での言い換え対象である形容詞については、語彙概念構造のような意味構造に関する議論や資源がなく、同じモデルを構築することは困難である。ただし、言い換えの生成対象は動詞であるため、語彙概念構造の情報を少なからず活かせる可能性はある。和語動詞を広く収録した利用可能な LCS 辞書はいまだ存在しないが、文献 [6, 9] などの成果に期待し、我々は、語彙意味論で指摘されてきた動詞の統語的振る舞いの違いを動詞接尾表現選択の素性として用いる。

文献 [5] などを参考に、動詞に関して下記の振る舞いを観察し、辞書を作成した。なお、振る舞いの観察のために、各動詞について IPAL 動詞辞書で『転成形容詞』を持つ語義の例文を 1 文取り上げ、掲載されていない語については 1 文作例した。

- V_a : 自動詞か他動詞か
- V_b : 時間副詞「T 時間」との共起の可否
- V_c : 時間副詞「T 時間で」との共起の可否
- V_d : 「である」との共起の可否
- V_e : 「ている」の解釈が結果状態か動作の継続か
- V_f : 「た」に完了の意味があるか否か
- V_g : 「わざと」「いやいや」などの意志性をはかる副詞との共起の可否

また、次に示す素性として用いる。

- D : 形容詞と動詞の語根以外の部分のパターン
- N : 分類語彙表における名詞の意味クラスの選言
- C : 対象の形容詞が節の主辞か単なる修飾句か

3.2 動詞接尾表現選択アルゴリズム

複数の動詞接尾表現を取りうることは稀であり、多様性も多くはなかった。これをふまえ、ある入力に対して、ある動詞接尾表現 (の集合) を割り当てるアルゴリズムを実装した。このアルゴリズムでは、表 3 に示す規則の上方から順に適用を試みる。表 3 の規則は、連体修飾用法の事例 C_{ad-c} を訓練事例として、適用条件を調整しながら人手で作成したものである。個々の規則およびそれらの適用順序は C_{ad-c} における表現の分布や事例の性質に特化している可能性はあるが、できる限り直感から逸脱しないように努めた。

規則の解釈の例として ID2-1 を取り上げる。この規則は文献 [1] における分析を参考に作成した。

- 被修飾名詞が本来は動詞の対格であれば「られる」を付加すべきと仮定した。まず、 C_{ad-c} において他動詞のみが「られる」となっていたため、「 V_a =他動詞」を「られる」の適用条件に加えた。一方、被修飾名詞のギャップの解析は直接行わず、「 N =主体以外」によって近似した。

表 3: 動詞接尾表現選択規則

適用順序	条件 「素性ラベル=素性値」の連言	推定結果
1	「 V_a =他動詞」かつ「 V_b =可」かつ「 V_f =持たない」かつ「 N =人間 (1.10) 以外」かつ「 $D=A_{ui}-V_{bumu}$ または $A_{i}-V_{mu}$ または $A_{asii}-V_{uu}$ 」	られる
2	「 V_a =他動詞」かつ「 V_b =可」かつ「 V_f =持たない」かつ「 V_d =不可」かつ「 N =心-心情・態度 (1.303)」	た
3	「 V_a =自動詞」かつ「 V_g =可」	た
4	「 V_a =自動詞」かつ「 V_f =持つ」かつ「 $D=A_{i}-V_{migakaru}$ 」	た、ている
5	「 C =節」かつ「 $D=A_{i}-V_{maru}$ 」	る
6	「 V_a =自動詞」かつ「 V_f =持つ」	た
7	「 V_a =自動詞」かつ「 V_b =可」かつ「 V_f =持たない」	た
8	「 V_b =可」かつ「 V_f =持たない」かつ「 V_c =可」かつ「 V_d =可」かつ「 V_e =最中」かつ「 N =主体 (1.2)」	ている
9	*	言い換え不可

- C_{ad-c} で「られる」を付加すべき動詞がいわゆる感情動詞であることから、継続動詞 (「 V_b =可」かつ「 V_f =持たない」) および D の条件を列挙することで感情動詞を近似的に定義した。

4 評価実験と考察

2 節で収集した事例を用いて、動詞の振る舞い辞書、動詞接尾表現選択規則の有効性を検証する。まず、連体修飾用法について、 C_{ad-c} を用いたクローズドテスト、 C_{ad-o} を用いたオープンテストの結果を表 4 に示す。

4.1 クローズドテスト

頻度が高い「た」、「られる」を中心に推定の条件を調整したが、現状の素性では必ずしも現象の全てをカバーできず、再現率、精度ともに 60% 強に留まった。その原因について考察する。

同じ振る舞いの動詞に言い換える場合でも、連体修飾節のギャップと節内の格要素の影響で例 (8) と (9) のように言い換え方が異なる。

- (8) s. 近所の知人が温かいおかゆを持って来てくれた。
t.*近所の知人が温まったおかゆを持って来てくれた。
(正解:「温まる」には言い換えられない)
- (9) s. 低気圧の影響で暖かい空気が流れ込んだ関東地方。
t. 低気圧の影響で暖まった空気が流れ込んだ関東地方。
- しかしながら、それらの情報を動詞接尾表現推定の際に参照していないため、(8s) に対しても誤って「た」と推定した。状態変化を表す自動詞への言い換える可否は対象物の内在的コントロールの有無 [5] にも左右されるため、今後は動詞格構造のみでなく名詞と動詞の関係についての知識の導入を検討すべきである。
- いわゆる感情形容詞を対応する感情動詞に言い換える際、形容詞の感情主がほのめかされているならば「られる」とは言い換えられない。しかし、現状の 1 番目の条件ではこれをとらえられておらず、(10t) のように誤って「られる」と推定してしまった。
- (10) s. (感情主にとって) 望ましい結論はまだ出ていない。
t.* (感情主に) 望まれる結論はまだ出ていない。
(正解:「(感情主が) 望んでいる結論」)

この言い換えを実現するにあたっては格解析技術が必要である。また、感情主を含む主格名詞はこの例のよ

表 4: 言い換え生成実験結果

動詞 接尾表現	C_{ad-c}		C_{ad-o}	
	再現率	精度	再現率	精度
た	45/57	45/66	19/23	19/30
られる	12/14	12/19	7/13	7/11
る	3/9	3/6	0/2	0/5
ている	1/5	1/7	2/10	2/5
た, ている	2/4	2/2	0/1	0/1
規期末実装	11 事例分		6 事例分	
合計	63/100 (63%)	63/100 (63%)	28/55 (51%)	28/52 (54%)

うに省略される場合が多いため、ゼロ代名詞を補完する技術も採り入れる必要がある。

「た」、「られる」以外の動詞接尾表現については現状の素性だけでは識別不可能な例が多かったため、クローズドテストにも関わらず精度良く推定できていない。ゆえに、今後事例を増やしながらか表現推定の素性およびそれらの蓋然的な組み合わせ方について検討する必要がある。なお、文献 [7] では、これらの形容詞を区別するための検査項目として (i) 「がる」の接続の可否, (ii) 主語に関する選択制限, (iii) 「～は～が」構文による対象語の表出の有無を挙げており、参考になる。

4.2 オープンテスト

C_{ad-o} に対する動詞接尾表現推定の再現率と精度はともに C_{ad-c} に対するそれより低い 50% 強に留まった。

C_{ad-c} にも出現した形容詞については、動詞接尾表現を正しく推定できた現象と、誤って推定した理由に大きな傾向の違いはなかった。ただし、素性 N の素性値および「ている」を推定するための 8 番目の条件のように制約が C_{ad-c} に出現する事例に特化していたことが原因で、再現率を大きく下げってしまった。

C_{ad-o} で新出した形容詞 (12 語, 事例 12 件) については 4 件しか正しい動詞接尾表現を推定できなかった。主な誤りの理由は素性 D の素性値が網羅されていなかったことにある。「煙たい」 \leftrightarrow 「煙たがる」、「危ない」 \leftrightarrow 「危ながる」のような「A_i-V_{garu}」のパターンは、 C_{ad-o} においてはすべて「られる」とすべきだが、 C_{ad-c} で多くの誤りを生じたため 1 番目の条件から除いていた。

5 性能比較実験

毎日新聞 1991 年版～2003 年版から名詞とそれを修飾する動詞 (接尾表現込み) の組を網羅的に収集し、ある言い換え候補の尤度を推定する確率モデルを構築した。そして、所与の動詞と被修飾名詞の組に対して、表 2 に示すあらゆる動詞接尾表現の中で最も尤度が高い候補を選択するモデルの性能を検証した。相互情報量を尤度関数として生成実験を行った結果、 C_{ad-c} に対しては再現率 34%, 精度 28%, C_{ad-o} に対しては再現率 16%, 精度 28% と、性能は非常に低かった。提案手法の再現率, 精度はそれほど高くないように見えるが、単純にコーパスベースで最尤の表現を選択するのに比べると十分高い精度を得ていたことが分かる。

また、提案手法で推定した動詞接尾表現から、上の確率モデルで推定される尤度が低い候補を棄却する実験を行った。紙面の都合上図の掲載は割愛するが、閾値を変化させたときに得られた再現率-精度曲線において、再現率を犠牲にすれば精度を向上させられることを確認した。たとえば、相互情報量 0 以下の候補を棄却した時、 C_{ad-c} に対して再現率 29%, 精度 85%, C_{ad-o} に対して再現率 22%, 精度 81% と、十分高い精度で言い換えを生成できている。今後は事例数を増やして尤度関数などについても定量的に評価したい。

6 おわりに

本稿冒頭で見たように、形容詞を動詞に言い換える際には、形容詞, 動詞および周辺の語に応じて動詞接尾表現を付加する必要がある。本稿では、動詞の統語的振る舞いの違いに基づいて動詞接尾表現を定める規則について検討し、非常に単純なアプローチにも関わらず少数の規則である範囲の現象を説明できる可能性を示した。また、現状での誤り分析を通じて、連体修飾節のギャップやゼロ代名詞を含む動詞の格要素も連体修飾用法の形容詞の言い換えの可否および言い換え方に影響していることが明らかになった。

幅広い現象をカバーする語彙資源とモデルを構築するためには、今後、より多くの事例の分析に基づいて規則を高度化するとともに、連体修飾用法と他の用法について同じ原理で動詞接尾表現を決定しうるか否かを調査する必要がある。ここでは、2.4 項で示したように表現の分布に基づいて戦略的に差分を整理することが考えられる。また、「広い」 \leftrightarrow 「広がる」「広げる」「広まる」「広める」のような言い換え先の動詞の選択に関わる要因についても調査したい。

参考文献

- [1] 坂東美智子, 松村宏美. 心理動詞と心理形容詞. 影山太郎 (編), 日英対照 動詞の意味と構文, pp. 69-97. 大修館書店, 2001.
- [2] 藤田篤, 乾健太郎, 松本裕治. 語彙概念構造で表現できる言い換え. 言語処理学会第 11 回年次大会発表論文集, pp. 1129-1132, 2005.
- [3] 情報処理振興事業協会技術センター. 計算機用日本語動詞辞書 IPAL (Basic Verbs) - 解説編 -. 情報処理振興事業協会技術センター, 1987.
- [4] 情報処理振興事業協会技術センター. 計算機用日本語形容詞辞書 IPAL (Basic Adjectives) - 解説編 -. 情報処理振興事業協会技術センター, 1990.
- [5] 影山太郎. 動詞意味論 - 言語と認知の接点. くろしお出版, 1996.
- [6] 加藤恒昭, 島山真一, 坂本浩, 伊藤たかね. 日本語和語動詞に関する語彙概念構造辞書構築の試み. 言語処理学会第 11 回年次大会発表論文集, pp. 871-874, 2005.
- [7] 西尾寅弥. 国立国語研究所報告 44 形容詞の意味・用法の記述的研究. 秀英出版, 1972.
- [8] 佐藤理史, 土屋雅稔, 村山賢洋, 麻岡正洋, 王晴晴. 日本語文の規格化. 情報処理学会研究報告, NL-153-18, pp. 133-140, 2003.
- [9] 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央, 阿部修也. 分類の根拠を明示した動詞語彙概念構造辞書の構築. 情報処理学会研究報告, NL-169-18, pp. 123-130, 2005.