

語構成を利用した言い換え表現の自動生成

麻岡 正洋 佐藤 理史 宇津呂武仁

京都大学情報学研究科

asaoka@pine.kuee.kyoto-u.ac.jp, {sato,utsuro}@i.kyoto-u.ac.jp

1. はじめに

ある表現を、その意味内容を変えずに別の表現に置換することを言い換えと呼ぶ。言い換えは、自然言語処理の分野で広く利用できる技術であり、機械的に言い換えを実現することは工学的に大きな意味がある。

言い換えは、実現に必要な知識や参照する情報の種類にもとづいて、構文的言い換え、意味的言い換え、プラグマティックな言い換えるの3つに分類できる¹⁾。このうち、構文的言い換えは、さらに、内容語の置換に重点を置いた言い換え(語彙的言い換え)と、構文構造の変換に重点を置いた言い換え(狭義の構文的言い換え)の2つに分類できる²⁾。

語彙的言い換えを実現する標準的な方法は、言い換えたい表現に対して代替表現を定義した「言い換え辞書」を作成し、この辞書を用いて語の置換を行なう方法である。この方法では、言い換えることができる表現の種類は、言い換え辞書のエントリーの数と等しいため、幅広い言い換えを行うには、言い換え辞書に多くの表現を登録する必要がある。このため、語彙的言い換えるの研究では、元の表現と代替表現の組(言い換えペア)をどのように大量に獲得するかということに焦点が当てられている²⁾。

しかし、語彙的言い換えるをもう少しよく観察すると、次のような複雑な対応関係が観察される。

- (1) a. 迅速な 対応 → すばやい 対応
b. 対応の 迅速さ → 対応の すばやさ
c. 対応が 迅速化した → 対応が すばやくなった
- (2) a. 明確に 答える → はっきり 答える
b. 答えは 明確だ → 答えは はっきりしている
c. 明確な 答え → はっきりした 答え
d. 答えを 明確にする → 答えを はっきりさせる
e. 答えが 明確になる → 答えが はっきりする

これらの例は、それぞれ、「迅速 → すばやい」と「明確 → はっきり」を基本とする言い換えである。しかしながら、それぞれの用法とそれぞれの語の性質に依存して、表層的な対応関係は、すべて異なった形となる。

これまでの研究では、活用する語を言い換える場合、言い換え元の語の活用形を言い換え先の語の活用形に写像する操作(活用形のマッピング)が行なわれてきた³⁾。この活用形のマッピングは、言い換え元の語とその代替表現が同一の語彙的性質を持つ場合には比較的うまく働く。

しかし、上記のような例は、単純な活用形のマッピングではうまくいかないのは明白である。そのため、先に述べた辞書を用いた方法で、このような語彙的言い換えを実現するためには、上記のような表層形での対応をすべて辞書に登録しておく必要がある。

一方、我々人間は、「迅速 → すばやい」と「明確 → はっきり」という基本的な対応関係を知っていれば、用法や語彙の性質に関する知識と組み合わせ、上記のような言い換えるをすべて行なうことができると考えられる。たとえば、(1c)の言い換えでは、「迅速」は形容動詞語幹(漢語)であるのに対し、「すばやい」は形容詞(和語)であるため、「～化する」を付加することができない。そのため、それに相当する「～なる」という表現で代用する。また、(2c)では、「明確」が形容動詞語幹であるのに対して「はっきり」は副詞のため、そのままの形では連体修飾に用いることができない。そのため、「～した」を補う必要がある。このような調整を、人間は無意識下で行なっていると考えられる。

本論文では、このような人間が行なっていると思われる処理を模倣した語彙的言い換えるの枠組を提案する。すなわち、この枠組は、核となる語(たとえば「明確」とその代替表現(「はっきり」)の組から、その語を含む表現群(「明確な」、「明確だ」、「明確になる」など)に対する代替表現を生成することを実現するものである。この枠組を実現するには、次の2つの問題を解く必要がある。

- (1) どのようにして、言い換えたい表現を、核となる語(コア)とそれに対する付加的な要素に分解するか。
- (2) ある表現にある付加的な要素(意味)を追加する場合には、どのような表現(表層形)を生成すればよいか。

本研究では、1の問題を語構成を利用することによって、2の問題を「見立て」という概念を用いることによって解決する。

2. 語構成を利用した言い換え

本論文で提案する枠組の概要を図1に示す。本枠組は、大きく、還元、言い換え、生成、の3つの処理から構成される。それらの処理では、次の3つのコンポーネントを用いる。

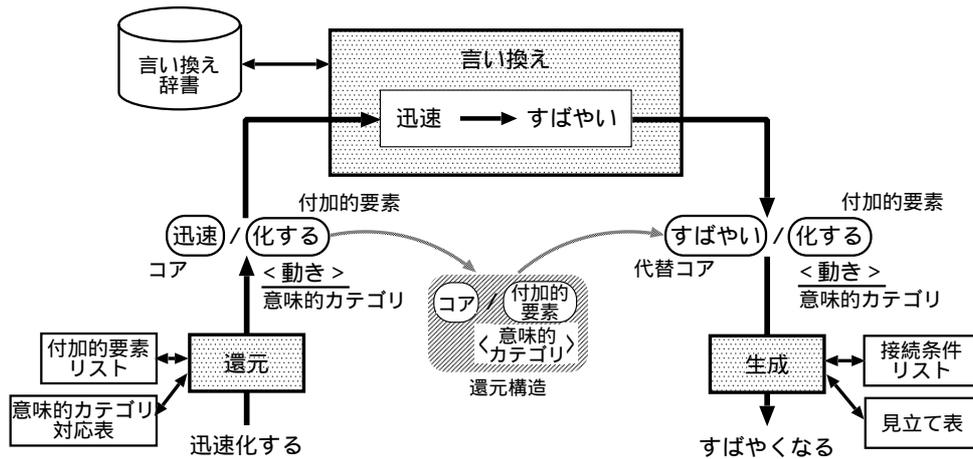


図 1 語構成を利用した言い換えの枠組

表 1 言い換え辞書の概要

	語数	ペア数
動詞	1837	3303
形容詞	339	421
形容動詞	1253	1454
サ変名詞	1900	2043
副詞	989	1954
	6318	9175

- (1) 言い換え辞書：核となる表現の組を定義した辞書。言い換え辞書の概要を表 1 に、例を図 2 に示す。この図に示すように、それぞれの表現（語）には、意味的カテゴリと呼ぶラベルを付与してある。意味的カテゴリは、現在、16 種類を設定している。
- (2) 付加的要素規則表：述語相当語から複合表現を構成する際に付加される付属語等に関する情報を記述した表。具体的には、(a) そのような付属語自身、(b) その付属語がどのような意味を付加するかということを表した意味的カテゴリ（見立て）、および、(c) その付属語がどのような語に接続可能かということを表した接続条件、からなる。付加的要素規則表の例を図 3 に示す。現在、この規則表には 93 語の付属語相当表現が登録されており、そのうち 33 語に意味的カテゴリが付与されている。なお、図 1 では、付加的要素規則表を、付加的要素リスト、意味的カテゴリ対応表、接続条件リストの 3 つに分けて記述している。
- (3) 見立て表：ある表現にある付加的要素を接続することができない場合に、その付加的要素が表す意味を、元の表現に追加するにはどうしたらよいかということ記述した表。この処理を、ある表現を、ある意味カテゴリに 見立てる という操作と捉えるため、見立て表と呼ぶ。見立て表の一部を表 2 に示す。

なお、本枠組が現在対象としているのは、次の 5 種類の述語型表現である。

- (1) 述語相当語

- | | |
|------------|--------------|
| 〈状態〉迅速 | 〈状態〉すばやい |
| 〈状態〉巨大 | 〈状態〉とても大きい |
| 〈状態〉みずみずしい | 〈状態〉新鮮だ |
| 〈状態〉未恐ろしい | 〈状態〉将来がおそろしい |
| 〈動き〉縮まる | 〈動き〉短くなる |
| 〈動き〉言い洪る | 〈否定〉なかなか言わない |
| 〈動き〉失職 | 〈動き〉職業を失う |
| 〈動き〉婚約 | 〈動き〉結婚の約束をする |
| 〈様態〉じめじめ | 〈状態〉暗い |
| 〈様態〉重々 | 〈様態〉かさねがさね |

図 2 言い換え辞書の例

- (接尾辞
 (動詞性接尾辞
 ((見出し語 過ぎる すぎる) (読み すぎる)
 (活用型 母音動詞)
 (見立て *)
 (接続条件 ID=78)))
- (接尾辞
 (形容詞性述語接尾辞
 ((見出し語 ない) (読み ない) (活用型 イ形容詞アウオ段)
 (見立て 否定)
 (接続条件 ID=150)))

図 3 付加的要素規則表の例

- (2) 述語型複合表現：述語相当語に接尾辞や助動詞が接続したもの
 例)「すばやい」+「さ」 「すばやさ」
 「読む」+「せる」 「読ませる」
 「助ける」+「だろう」 「助けるだろう」
 「不自然だ」+「のだ」 「不自然なのだ」
- (3) サ変名詞の動詞的用法：サ変名詞に形式的な用言が接続したもの
 例)「研究」+「する」 「研究する」
 「復元」+「不可能だ」 「復元不可能だ」
- (4) 「副詞+する」型の動詞
 例)「はっきりする」=「はっきり」+「する」

表 2 見立て表の一部

	イ形容詞〈状態〉	サ変名詞〈動き〉
動き	K2=連用形 なる c /K2=連用形 する c	する c
状態 (述語)	c	している c
状態 (連体)	K2=基本形	している/した
様態	K2=連用形	していて/して
こと	こと/K2=語幹 さ	
さま	さま/K2=語幹 さ	
ひと	ひと	者
いきもの	いきもの	生物
もの	もの	物
とき	とき	時
ところ	ところ	場所
ほう	ほう	方向
否定	K2=連用形 ない c	しない c
可能	K2=連用形 なれる c /K2=連用形 できる c	できる c
不可能	K2=連用形 なれない c /K2=連用形 できない c	できない c
使役	K2=連用形 ならせる c /K2=連用形 させる c	させる c
受け身	K2=連用形 なれる c /K2=連用形 される c	される c

注) :そのまま, c: 直前の語は活用する, K2=:直後の活用形に変換する, /:複数方法の分割記号, 空白:形態素区切り

(5) 可能動詞

例)「許せる」=「許す」+〈可能〉

本枠組では、文節単位に処理を行なう。すなわち、入力、言い換えるべき述語型表現を含んだ文節であり、出力は、それに対する代替表現（文節）の候補リストである。

3. 還元

還元処理では、入力された文節を分解して、その構造を表す表現（還元構造）を作成する。この過程で、付加的要素規則表を使用する。

入力される文節は、すでに形態素解析によって、形態素に分割されていると仮定する。まず、文節の末尾から、ひとつづつ付属語を取り除く。取り除いた付属語が、付加的要素規則表に含まれていない場合は、そのままの形で残し、付加的要素規則表に含まれている場合は、その付属語に対して定義されている意味のカテゴリの情報を付与する。このような処理により、たとえば、「迅速化しなかった。」という文節に対して、次のような一連の構造が得られる。

- (1) 迅速化しなかった /。
- (2) 迅速化しない / 〈文末〉[夕形] /。
- (3) 迅速化する / 〈否定〉ない / 〈文末〉[夕形] /。
- (4) 迅速化 / 〈動き〉する / 〈否定〉ない / 〈文末〉[夕形] /。
- (5) 迅速 / 〈動き〉化 / 〈動き〉する / 〈否定〉ない / 〈文末〉[夕形] /。

なお、「はっきりする」などの「副詞+する」は、形態素解析により「副詞」+「する」に分割されないの、例外的に処理する。また、可能動詞（可能形）も、原形の

動詞と〈可能〉というダミーの付属語に還元する。

4. 言い換え

言い換えでは、還元によって得られた構造（還元構造）に対して、その先頭部にあるコアを、言い換え辞書によって置換する。この処理は、還元によって得られた複数の構造に対して適用する。このため、言い換え辞書には、たとえば、「迅速化しない」という複合表現に対して代替表現を定義することもできる。

いま、「〈状態〉迅速」に対して「〈状態〉すばやい」という代替表現が定義されていると仮定しよう。このとき、先に示した還元構造のうち、(5)に対して、置換を行なうことが可能であり、次の構造が得られる。

(5b) 〈状態〉すばやい / 〈動き〉化 / 〈動き〉する / 〈否定〉ない / 〈文末〉[夕形] /。

5. 生成

こうして得られた構造に基づいて、代替表現を生成する。基本的には、言い換えられたコアに、付属語を一つづつ結合していくことにより、代替表現を生成する。付加的要素規則表に記載されている接続条件をチェックして、付属語が接続できる場合は、活用形などの調整を行なって結合する。接続できない場合は、付属語に付加されている意味のカテゴリを利用して、代替表現を生成する。

たとえば、「すばやい」には「~化」を接続することができない。ここで、「~化」には、「〈動き〉」という意味のカテゴリを表すという情報が付与されている。そこで、「すばやい」という〈状態〉を〈動き〉に見立てるという操作を行なう。一般に、〈状態〉を表す表現は、「~になる」を付加することにより、その状態へ遷移する〈動き〉を表すことができる。このような情報は見立て表にまとめられているので、この表を見て、接続する付属語等を決定する。

上記の例の場合、

- (5b) 〈状態〉すばやい / 〈動き〉化 / 〈動き〉する / 〈否定〉ない / 〈文末〉[夕形] /。
- (5c) 〈動き〉すばやくなる / 〈動き〉する / 〈否定〉ない / 〈文末〉[夕形] /。
- (5d) 〈動き〉すばやくなる / 〈否定〉ない / 〈文末〉[夕形] /。
- (5d) 〈否定〉すばやくならない / 〈文末〉[夕形] /。
- (5e) 〈文末〉すばやくならなかった /。

という過程を経て、最終的に、「迅速化しなかった」に対する代替表現「すばやくならなかった。」が出力される。

6. 実験と検討

提案する枠組を実装したシステムを用いて、次の2つの実験を行なった。

6.1 実験 1

ここでは、本システムが、付加的要素を伴った述語型

くっきりとした見取り図を示す
はっきりとした見取り図を示す
 にわかに緊迫化している にわかに差し迫っている
 株を購入できる 株を買える
 粉飾を見抜けず、被害が広がった
 粉飾を見破ることができず/見通せなく、被害が広がった

図 4 言い換え成功例

表 3 実験 1: 付加的要素を含む述語型表現の言い換え結果

	文節のみに注目	文全体を考慮
成功	113 (86%)	96 (73%)
失敗	19 (21%)	36 (27%)
合計	132 (100%)	132 (100%)

表 4 言い換えが失敗した原因

原因	
A. システムの不備	19 (53%)
B. 格助詞が不適当	6 (17%)
C. 代替表現が不適当	5 (14%)
D. その他	6 (17%)
合計	36 (100%)

表現を、どのくらい適切に言い換えられるかを確認する実験を行なった。

まず、京大コーパスから、核となる表現が言い換え辞書に含まれており、かつ、以下のいずれかの条件を満たす述語型表現を含む文を収集した。

- 意味のカテゴリを持つ付加的要素 (33 語) が接続する
- 可能動詞となる
- 「副詞 + する」となる

こうして収集した文から、各付加的要素については最大 6 文まで、可能動詞、「副詞 + する」、「副詞 + と + する」については、最大 10 文まで取り出して得られた合計 132 文を実験に使用した。実験では、本システムを用いて、これらの文中の、上記の条件を満たした述語型表現を含む文節を言い換えた。

実験結果を表 3 に示す。この表で、「文節のみに注目」とは、言い換え対象の文節だけを見て、元の文節の言い換えとして適切な代替文節を出力した場合に成功とする評価である。一方、「文全体を考慮」とは、周りの文節との関係も含めて適切な代替文節を出力した場合のみ成功とする評価である。なお、本システムは、一つの文節に対して、複数の代替文節を出力するので、その中に一つでも適切なものが含まれていればよいとして評価した。図 4 に、言い換えの成功例 (文全体を考慮した場合) を示す。「文節のみに注目」した場合の精度は 86%、「文全体を考慮」した場合の精度は 73% であった。

「文全体を考慮」した場合の失敗の原因を分析した結果を表 4 に示す。A は、本システムの生成処理の不備によるものである。その一部は、見立て表や活用形調整に修正を加えることで解決できると考えられる。なお、文

表 5 実験 2: 単純な置換システムとの比較

	本システム	単純な置換システム
成功	90 (82%)	83 (75%)
失敗	20 (18%)	27 (25%)
合計	110 (100%)	110 (100%)

節のみに注目した場合の失敗は、すべてここに分類される。B は、本システムが文節を単位にして、言い換えを行っていることに起因するものである。C は、言い換え辞書に適切な代替表現が存在しないことが原因である。以下に失敗例を示す。

- システムの出力ミス
 例) 広々とした敷地 たいへん広くなった/たいへん広くした敷地 (「たいへん広い」のように、〈動き〉に見立てない方がよい)
- 格助詞が不適当
 例) 故植村直己の本の 愛読者 故植村直己の本の 好きでよく読むひと (直前の「の」は「が」が適切)
- 代替表現が不適当
どっしりした門構えの邸宅 目方の多くなった/落ち着いた門構えの邸宅 (「重々しい」が適切)
- その他
 例) マーケットの 透明性 を向上させる マーケットの透き通っているさま/透き通っていること を向上させる (文節の言い換えとしては適切だが、この文脈ではおかしい)

6.2 実験 2

京大コーパスに含まれる 100 文 (実験 1 とは異なる) を実験対象文として、単純な置換による言い換えシステム (ベースライン) と本システムの性能の比較を行なった。その結果を表 5 に示す。この表での成否の評価は、前節の実験で言うところの、「文全体を考慮」した文節の言い換えの成否である。

ベースラインの単純な置換システムの精度が 75% であるのに対して、本システムの精度は 82% であった。

なお、この精度が、実験 1 における精度 (73%) より高い値を示しているのは、実験 1 が、付加的要素を伴った述語型表現のみを実験対象としているのに対して、実験 2 では、述語相当語 (単体) も言い換え対象となっているからである。

参考文献

- 1) 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, 1999.
- 2) 乾健太郎. コミュニケーション支援のための言い換え. 言語処理学会第 8 回年次大会ワークショップ「言い換え / パラフレーズの自動化」, pp.71-76, 2002.
- 3) 近藤恵子, 佐藤理史, 奥村学. 「サ変名詞 + する」から動詞相当句への言い換え. 情報処理学会論文誌, Vol.40, No.11, pp.4064-4074, 1999.