

言い換え生成技術の課題と応用

藤田 篤* (名古屋大学)

Issues and Applications of Automatic Paraphrase Generation

Atsushi Fujita (Nagoya Univ.)

1 はじめに

次の例が示すように、言語には同じ情報を伝える表現がいくつもある。このような表現を言い換えと言う。

- (1) a. 某社の株価暴落に歯止めがかからない
⇔ ある会社の株価が急激に下がっている
b. 彼と代理人の会談が明日に予定されている
⇔ 彼は、明日代理人と話し合うことになっている

我々人間は、コミュニケーションを円滑に進めるために、伝える相手や状況に応じてある表現を別の表現に言い換える。ある表現から様々な言い換えを自動生成できれば、たとえば、文章を読者の読解能力にあわせて簡単化したり、文章を作成する際にうまい言い回しを提案したりできる。また、我々は、ある表現と別の表現が言い換えになっていることを認識しながら、情報を検索し、取捨選択し、整理している。与えられた2つの表現が言い換えであるかどうかを自動認識できれば、情報検索や質問応答、テキストマイニングなどの自然言語処理タスクにおける重要な問題が解決できる。

本稿では、言い換え技術のうち生成技術に焦点を当て、言い換え生成とそれに必要な知識の獲得における課題について述べる。これらは図1に示すような関係になっている。また、言い換え生成技術の現状をふまえ、言い換への応用例を提案する。

2 言い換え生成手法と必要な知識

言い換えと呼ばれる現象は多岐に渡り、広くとらえると「暑いですね」⇔「クーラーをつけてください」のように対話の状況や社会的慣習に根ざした現象も含む。一方、例(2)に示すような言い換えは、言語学的な知識や同義語対に基づいて実現できる可能性が高い。

- (2) a. 入学式当日に桜が開花した
⇔ 入学式当日に桜が咲いた
b. 司会をすることになっているのは彼です
⇔ 彼が司会をすることになっています
c. 金利の変動が株価に影響を与えた
⇔ 金利の変動が株価に影響した

(2a-c)では言い換える対象の範囲や言い換への種類が異なる。このため、言い換への実現に必要な知識や適切な言い換への手法も異なると考えられる。実際、先行研究においては、例(2a)を単純な語句の置換によって、例(2b)を統語構造の変換によって、そして例(2c)を統語構造変換に語彙的な情報を制約として加えることで実現した例がある。

以下では、上の例を参照しながら各手法の特徴を概観し、各々に必要な知識の整備に関する課題について述べる。

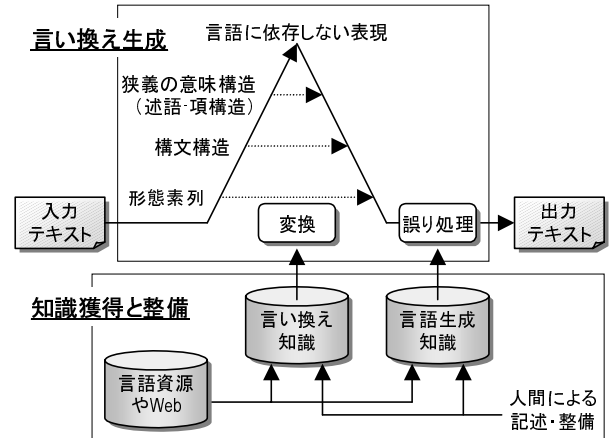


Fig. 1. Overview of paraphrase generation

2.1 語句の置換

内容語(例(2a))や「からには」、「に違いない」などの機能表現をそれと同義の語句に言い換える場合、与えられた表現を形態素解析し、対象となる形態素列を目的の形態素列に置換すれば良い。ここでは、様々な語句に対して頑健に言い換へを生成するための大規模な同義表現対をいかにして獲得するかという課題がある。英語に関しては、WordNet [1]などの十分細かい粒度で語の同義関係を整理した電子化辞書が存在するが、句動詞や慣用句については紙媒体の辞書に断片的に収録されているに過ぎない。一方、日本語に関しては、WordNetほど細かい粒度の無償の辞書はない。また、既存の言語資源やWebから同義表現対を安価に獲得する試みは色々報告されているが、どれだけ集めれば良いのか、どれだけ一般化できるのか、という疑問に対する解は示されていないのが現状である。

2.2 統語レベルのトランスファ方式

例(2b)のように統語構造が大きく変化するような言い換へは、統語レベルのトランスファ方式で実現できる。この方式では、依存構造と句構造の選択を含め、内部構造の設計に関する議論がある。我々は一つの実装例として、形態素をノードとする依存構造に基づく言い換へシステム KURA を開発した [5]。KURA は表現力の高い規則記述言語を提供している。これにより、副助詞や副詞などの任意の要素に関する制約や語順の指定などを可能にし、幅広い種類の言い換へ生成を支援する。

このレベルで言い換へを実現する際は、言い換へのカバレッジを最大化するように変換規則を汎化するという課題がある。たとえば、次の例を見てみよう。

- (3) This car was purchased by him ⇔ He bought this car

この例からどの程度一般的な変形規則が得られるだろうか？た

例えば“X be VERB-PP by Y ⇔ Y VERB X”という態交替の規則と“purchase ⇔ buy”という語彙的な言い換え対に分解できるが、そのような分解の基準は計算機上に実装できるだろうか？

2.3 語彙意味情報に基づく統語トランスファ

受動文と能動文の間の言い換えや例 (2c) のような言い換えは、統語トランスファだけでも実現できる。ただし、受身にできない動詞もあるし、例 (2c) と下の例が示すように、語と語の組み合わせによって言い換え先の表現が定まることもある。これらを考慮すると、語の意味的な共通性や違いをとらえる必要があることが分かる。

- (4) a. ベテランのプレーが若手に刺激を与えた
 ⇔ ベテランのプレーが若手を刺激した
 b. この映画が彼に感動を与えた
 ⇔ この映画が彼を感動させた

我々は、語の意味的な特性を統語的振る舞いの観察に基づいて分析する語彙概念構造[4]に基づいて、例 (2c) と (4a) では言い換え先の動詞が能動態「する」になり、例 (4b) では使役態「させる」になるメカニズムを説明した [2]。語彙概念構造が提供する語彙意味情報は動詞の特性の一部であり、言い換えの種類によっては他の意味的な特性をとらえる枠組と実際に語彙資源の構築する際の効率的手法の確立が課題となる。

なお、例 (4) では「刺激」⇔「刺激する」などの知識も用いている。このような語間の関係に関する知識もより広範な種類の言い換えを実現するには欠かせない。名詞と動詞の語対は 10,000 対以上存在するが、活用形の変更や「する」を接続するだけで容易に獲得できる。一方、形容詞と動詞 (e.g. 「暖かい」⇔「暖まる」)、名詞と形容詞 («悲しみ」⇔「悲しい») などの語対は、少数ながらも形態的な特徴の多様性が高いため、網羅的に収集するには何らかの工夫が必要である。

2.4 言い換えの適格性を判定する

与えられたテキストの一部を局所的に同義な表現に置き換えると、言い換えた箇所と周辺の表現のつながり方が不自然になるなどの問題が生じる。このため、誤りの検出と修正の処理が必要である (図 1 を参照)。

誤り検出・修正手法のアプローチは、文法規則に基づく手法とコーパスに基づく手法の 2 つに大別できる。活用誤りや接続助詞の選択誤りなどは、文法規則に基づいて検出・修正できる。たとえば、「泳いでいる」と「歩いている」の接続助詞「で」、「て」の違いは動詞の活用型を参照すれば説明できる。一方、例 (5) の 2 例の適否の違いは、動詞とヲ格名詞の共起の適否によって説明できるが、このようなオープンクラスの語と語の共起に関する知識を人手では記述できない。

- (5) a. 収入が 500 万円を超える ⇔ 収入が 500 万円を上回る
 b. 情報が時空を超える ⇔ * 情報が時空を上回る

このような問題には、生成した表現の文法適格性を評価する言語モデルが有効である。過剰生成した候補をランキングするという手法は、確率的構文解析、統計的機械翻訳などの諸タスクにおいても有効性が示されている。

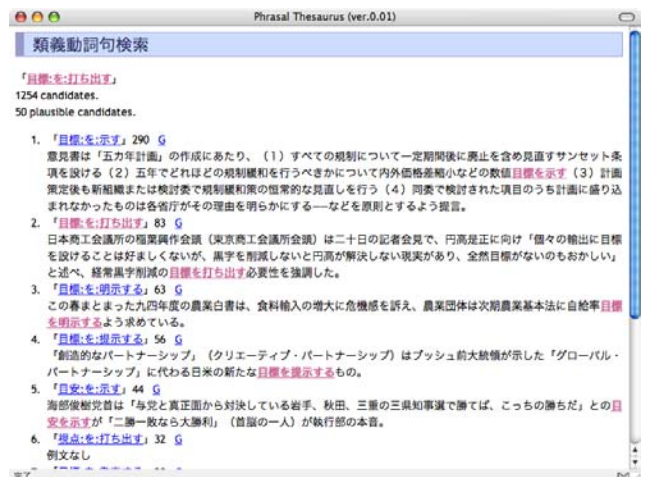


Fig. 2. Snapshot of an interface for Phrasal Thesaurus

3 言い換え生成技術の応用

言い換え生成技術は、現時点では、あらゆる入力に対して多様な言い換え表現を生成できるほどは成熟していない。言い換え生成に必要な知識が十分に整備されていないためである。ただし、現在の技術でも様々な応用が考えられる。我々は現在、より広範な種類の言い換しの生成を考慮しながら、言い換しの応用技術の開発に取り組んでいる。たとえば、我々自身が持っている文章作成の際のニーズをふまえ、与えられた表現の類義表現を提案する文章読解支援システム、フレーザルシソーラス (語よりも大きな単位の類義表現辞書、図 2) の実現方法について研究している。このシステムは、現状では、例 (6) に示すような動詞句の言い換えを生成する機能を備えている。

- (6) a. 実験結果の再現可能性を検証する
 ⇔ 実験結果を再現できるかどうかを調べる
 b. 表舞台に再登場する可能性が高い
 ⇔ 再び表舞台に登場するに違いない

4 おわりに

本稿では、言い換しの生成手法および各手法に必要な知識について概観し、言い換え生成の応用例を一つ示した。

我々は言い換えに関する文献、および様々な種類の言い換しを下記 URL にて継続的に収集・整理している。文献 [3] とあわせて読者の参考になれば幸いである。

<http://sslab.nuee.nagoya-u.ac.jp/~fujita/paraphrase/>

文献

[1] C. Fellbaum. *WordNet: an electronic lexical database*. The MIT Press, 1998.
 [2] 藤田篤, 降幡建太郎, 乾健太郎, 松本裕治. 語彙概念構造に基づく言い換え生成—機能動詞構文の言い換しを例題に. *情報処理学会論文誌*, Vol. 47, No. 6, pp. 1963–1975, 2006.
 [3] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. *自然言語処理*, Vol. 11, No. 5, pp. 151–198, 2004.
 [4] R. Jackendoff. *Semantic structures*. The MIT Press, 1990.
 [5] T. Takahashi, T. Iwakura, R. Iida, A. Fujita, and K. Inui. KURA: a transfer-based lexico-structural paraphrasing engine. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPWS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 37–46, 2001.