

言い換えに適した意味表現について

A Semantic Representation Suited for Paraphrase Generation

麻生 英樹^{*1}
Hideki ASOH

伊東 幸宏^{*2}
Yukihiko ITOU

高木 朗^{*1,3}
Akira TAKAGI

^{*1} 産業技術総合研究所
AIST

^{*2} 静岡大学情報学部
Faculty of Informatics, Shizuoka University

^{*3} 言語情報処理研究所
NLP Research Laboratory

A semantic representation which is suited for paraphrase generation is proposed. The semantic representation is based on the idea that both modifying word and modified word share common semantic structures. Each lexical item including noun and adjective has complex semantic structure. A method for paraphrasing noun phrase including adjectives is described and applied to some examples in order to demonstrate the effectiveness of the semantic representation.

1. はじめに

自然言語では、近似的に等価な意味内容を多様な形で表現することができる。たとえば、「赤い花が咲いている」、「赤い色をした花が咲いている」、「咲いている花は赤い」、などの文は、ほぼ同じ意味内容を持つと考えられる。さらに、同じことを英語では "The red flower blooms", "The flower with red color blooms.", "The flower which blooms is red." などと表現できる。こうした同義文の表層構造の多様性、すなわち、「同義性」の問題は、これと相補的な「多義性」の問題、すなわち、同じ表層表現が文脈や知識によって多様な意味に解釈される問題とあわせて、自然言語に関する理論の中心的な課題であるとともに、自然言語処理における主要な困難の源泉でもある。

自然言語処理では、同義性に関連する課題として、「言い換え技術」、すなわち、入力文からそれとほぼ同義な出力文を生成する技術、が研究され、これまでに、さまざまな言い換えのための規則と手続きが提案されている[乾 06]。一般に、表層の表現をそのまま言い換える規則を書こうとすると、表層表現の多様性のために規則の数が爆発し、適用順序の問題などが生じやすいため、表層構造を、依存構造木や語彙概念構造などの別の中間構造に変換した後に、変換規則を適用し、その結果から表層構造を生成して言い換え結果を得る、ということが行われている。この場合、変換規則や、その適用基準がうまく書けるか否かは、中間構造に大きく依存する。

本発表では、言い換えのために適した中間表現の構造について検討した結果を報告する。特に、従来の研究であまり取り上げられてこなかった形容詞に関する言い換えを取り上げ、[高木 82, 84, 87]において提案した意味表現の構造が形容詞表現の言い換えに適していることを示す。

2. 言い換え技術

「言い換え」にはさまざまなものがある。たとえば佐藤[佐藤 99]では、それを、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の 3 つに分けている。乾ら[乾 06]は、その分類を踏襲しつつ、語彙・構文的言い換え、参照的言い換え、語用論的言い換え、と分類している。本発表で対象としているのは、このうちの語彙・構文的言い換え、すなわち、語の統語的特性と意味的特性の両方に基づいて構成的に説明可能な言い換えである。その中でも、特に、従来あまり取り上げられていない、形容詞を用いた表現の言い換えを取り上げる。

言い換え処理には何らかの構造変換が伴うが、それを実現する方法はいろいろ考えられる。たとえば、近藤ら[近藤 01]では、表層表現を格変換規則で変換している。また、佐藤[佐藤 99]は、表層文を依存構造木に変換し、そこで構造変換を行っている。高橋らによる言い換えシステム Kura [高橋 01]でも、同様に、表層文から依存構造木を生成し、それに構造変換規則を適用して変換を行っている。最近、藤田らは、動詞の意味構造を記述する意味表現である語彙概念構造を利用した言い換えを提案している[藤田 06]。

どのような方法を取るにせよ、構造変換規則を記述する場合には、表層文の依存構造および、各語彙の意味構造が重要な情報となる。そこで、我々は、それらの情報を融合して持つような意味表現に対して構造変換規則を適用するのが適切であると考えた。我々の枠組みにおける言い換え処理の流れは、1) 表層文を語彙分割し、依存構造を抽出するとともに、それを反映した意味表現を生成する。2) 意味表現に対して構造変換規則を適用する。3) 変換後の意味表現を語彙分割し、依存構造木を得るとともに、表層文を生成する、となる。

3. 意味表現

上記の枠組みにおいて、構造変換規則やその適用規準がうまく書けるか否かは、どのような意味構造を用いるかに大きく依存する。我々は、[高木 82, 84, 87]において提案した意味表現を取り上げて検討する。この意味表現は、依存構造をベースとし、ほぼ同一の意味を持つ表層文の意味が、ほぼ同一の意味表現となることを意図して提案されている。[高木 84]では、その意味表現を用いて、視覚認識結果から多様な表層構造の言語表現を生成している。従って、言い換えのための中間言語としても適していると考えられる。以下では、[高木 82]をベースとして形容詞および名詞の意味を中心に、意味表現の概要を説明する。記号などを少し変更しているが、本質的な違いはない。

3.1 基本的な考え方

「赤いボール」という名詞句は、たとえば「赤い色を持つボール」と言い換えることができる。ここで、二番目の表現における「持つ」はどこから生じたと考えればよいだろうか？

「赤い」が、「赤い色を持つ」と言い換えられることから、「持つ」は「赤い」の意味に含まれている、と考えるのは自然である。しかしその一方で、「ものは色を持つ」というのはボールを含む「もの」一般に関する基本的な知識であるため、「持つ」は「ボール」の意味の中にも含まれているとも考えられる。

連絡先: 麻生英樹, 産業技術総合研究所情報技術研究部門,
〒305-8568 茨城県つくば市梅園1-1-1中央第2,
h.asoh@aist.go.jp

日本語、英語、中国語等に対するこのような言い換えに関する考察に基づくと、「赤いボール」のように「赤い」が「ボール」を修飾できるのは、両者の意味表現に共通する情報が含まれており、それを糊しろとして、両者の意味表現を重ね合わせて接続することが可能であるためである、と考えることができる。具体的には、「赤い」と「ボール」の意味表現をそれぞれ図 1 のようにすることが考えられる。

図 1 では、「赤い」の意味が「赤に等しい色相を内包する色を内包するところの」であり、「ボール」の意味が「X に等しい色相を内包する色を内包するところのものであり、球に等しい形を内包するところのものであり、Y がそれを投げるところのものであり、…」である、ということが、それぞれ表現されている。この例からわかるとおり、形容詞も名詞も非常に豊かな内部構造を持ち、両者は、多くの部分を共有している。そして、「赤い」が「ボール」を修飾している「赤いボール」という表層表現に対応する意味表現は、二つの語彙の意味表現を、共有部分を糊しろとして接続した図 2 のような構造を取ると考える。

3.2 意味表現に使われる記号と構成規則

ここでは、意味表現に使われている記号とその間の接続規則について述べる。

- 実体や属性を表す名詞的な概念構成素を \equiv で表す
- 現象を表す動詞的な概念構成素を \rightarrow で表す
- 英語の関係代名詞相当の構成素を \leftarrow で表す
- ある \rightarrow が \leftarrow を指示することを $=$ で示す
- 格概念を (矢印) で表す
- 不定な概念構成素名を示すのに変数 X, Y などを用いる。

赤い(連体形)



ボール

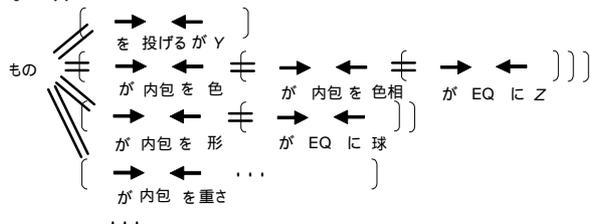


図 1 「赤い(連体形)」と「ボール」の意味表現

赤いボール

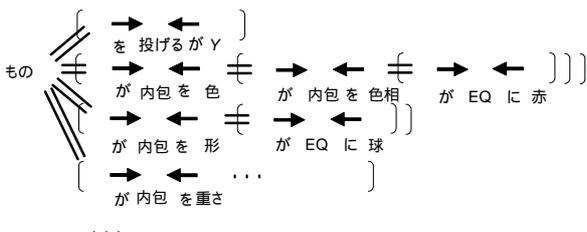


図 2 「赤いボール」の意味表現

これらの記号は、下記のような制約を満たすように接続される。

- \rightarrow , \leftarrow , 変数は単一の \equiv の始点に接続することができる
- \rightarrow は複数の \equiv の終点に接続することができる。(ただし、の種類によっては一つしか接続できない。)
- \rightarrow は複数の \leftarrow と $=$ で接続することができる

図 1, 2 に示すように、意味構造は一般に、 \rightarrow や \leftarrow が \rightarrow や $=$ で接続された木構造となる。節に対応する意味構造のまとまりを () によって表す。上述したように、表層表現において、語(あるいは文節)A が B を修飾しているときには、A の意味表現と B の意味表現は共有する構造を持ち、その部分を糊しろとして A の意味表現を B の意味表現のヘッドに接続することで全体の意味表現が形成される。

概念構成素として具体的にどのようなものを用意するかは、概念体系の設計による。以下、本稿の議論の範囲では次のような概念構成素を使用する。

- \equiv : もの, 色, 形, 色相, など
- \rightarrow : 内包, EQ(値保持), など
- \leftarrow : 主格(が格), 目的格(を格), 対象格(に格), など

ただし、上記の意味表現の構造は、具体的な構成素として何を用いるかは独立に、意味表現の構造および接続関係を規定していることに注意しておきたい。

3.3 終止形容詞の意味表現

「赤いボール」と同じ情報内容を、終止形容詞を用いて「ボールが赤い」と叙述形で言うこともできる。英語では、形容詞は活用しないため、日本語の終止形の形容詞にあたるものは「be + 形容詞」という形で表現される。また、日本語の終止形容詞も、「赤くある」のように、連辞の「ある」を用いて言い換えられる。これはまた、「赤い色を持つ」とも同義である。こうした現象について考察すると、終止形容詞の意味表現を、関係節内部の動詞句を取り出して動詞化する働きを持つ特殊な概念構成素("be"や"ある"に対応する)を使って表現するべきであることがわかる[高木 82]。「赤い(終止形)」の意味表現を図 3 に示す。

3.4 「の」の意味表現

「赤いボール」は「赤い色のボール」とも言い換えられる。日本語の助詞「の」は英語の"of"同様、二つの名詞をつないで非常に多様な意味を持つことで知られている。たとえば「太郎の本」という表現は、文脈によって、「太郎が持っている本」、「太郎が書いた本」、「太郎のことを書いてある本」、など多くの意味を持つことができる。こうした現象についての考察に基づくと、「の」の意味表現は図 4 のようになることがわかる[高木 82]。図で、 \rightarrow は任意の名詞句に対応づけられ、その名詞句を節構造の外に出す働きを持つ概念構成素である。

赤い(終止形)

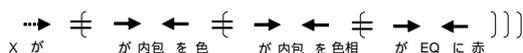


図 3 「赤い(終止形)」の意味表現

の

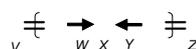


図 4 「の」の意味表現

4. 言い換え処理

4.1 語彙辞書および意味構造変換規則

上記の意味表現を用いて、[高木 84]における日本語文生成手続きを参考にしつつ、言い換えを行うための具体的な手続きについて述べる。語彙辞書として、図1, 3, 4 に示したものに加えて、図5 にその一部を示すような、語彙と意味表現のペアが用意されているとする。図5の中の点線の矢印は、その上にさらに格概念構成素が重ねられることを示している。一部の語彙項目は、日本語での表層表現を持たないことに注意。

意味構造変換規則として、下記の3つを用意する。

- 1) 関係節構造と形容詞終止形の変換
 $\dots = (\dots) \dots = (\dots)$
- 2) 関係節構造と叙述形の変換
 $\dots = (\dots) \dots$
- 3) 関係節構造と「の」を使った表現の変換
 $\dots = (\dots) \dots = (\dots) = \dots$

また、依存構造木と表層表現との間の関係に関する統語規則（主に語順の決定に関するもの）も与えられているとする。

4.2 意味表現生成

入力文に対して、以下のように意味表現を生成する。

- 1) 入力文を語彙に分割する
- 2) 語彙辞書を引き、それぞれの語彙の意味表現を得る
- 3) 統語規則と意味表現構成規則を使って、入力文から依存構造木を得る。それと並行して、語彙の意味表現を組み合わせ、文の意味表現を得る。

4.3 意味構造変換と分解

適用可能な意味構造変換規則を用いて意味構造の変換を行う。それに伴い、概念構成素の接続規則を満たさない場所が現れた場合には、さらに上記の変形規則を適用して、を受けるとを新たに生成する、あるいは、=で結ばれたとを融合する、()をつけかえる、などの必要な制約違反解消処理を行う。

それと並行して、意味表現のヘッドから語彙を割り当ててゆき、意味表現を語彙項目に分解する。このとき、と接続している関係節構造で、完全にカバーできない枝が生じた場合には、その枝の意味を表す修飾表現が必要であるということなので、その枝全体をヘッドから切り出してさらに分解する。可能な変換や分解が複数通りある場合もある。語彙項目間の接続関係から語の間の依存関係が決まるので、依存構造木が得られる。

4.4 表層表現生成

統語規則を用いて、意味表現および依存構造木から表層表現を生成する。

5. 言い換え事例

上記のプロセスに従って、「赤いボール」という入力文から「赤い色のボール」および「ボールの色が赤い」を生成する様子（意味構造変換）を図6に示した。図には「ボール」や「色」意味表現のうちの関連する部分のみを記している。例にあげている「赤いボール」の場合、記号は3つ含まれており、そのそれぞれについて、意味構造変換規則を適用する可能性がある。また、同じ意味構造に対して、形容詞連体形を用いて表現するか、関係代名詞（日本語では無表層）を用いて表現するかを選択でき

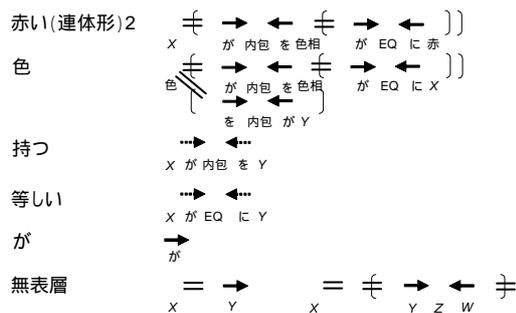


図5 語彙項目の例

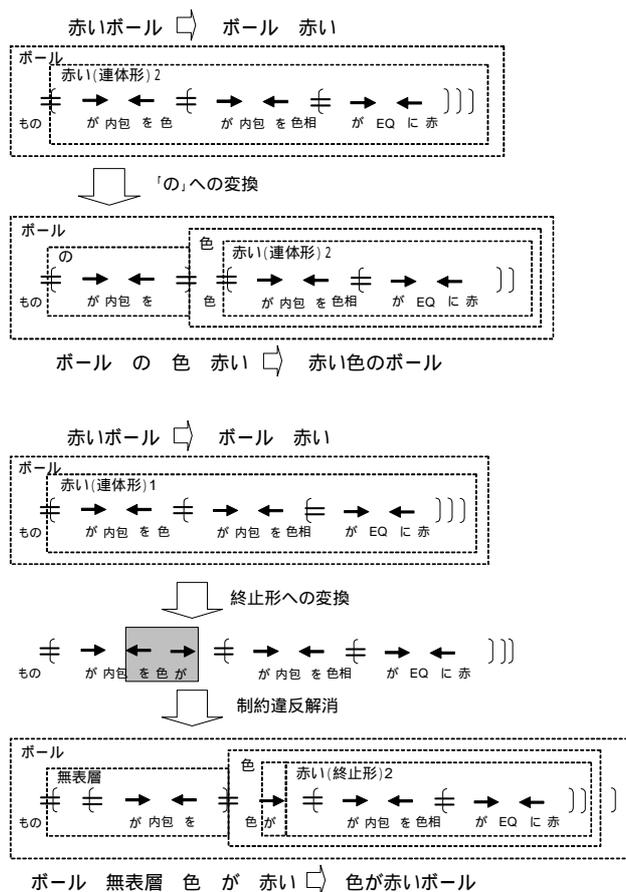


図6 言い換の例

る。したがって、生成可能な表層表現は、名詞句を叙述形(文)に変換するものも含めて、多岐に亘る。制約ベースなので、双方向の変換が可能である。また、必要な語彙項目を用意すれば、英語や中国語への言い換えも可能である。

6. 関連研究

6.1 意味表現形式について

Schank の概念依存構造[Schank 75]をはじめとして、語彙の内部構造を、概念構成素を用いて分解、記述しようとする意味表現は多い。しかし、その多くは、動詞の意味を記述するものであり、形容詞や名詞の意味について、その内部構造を詳細に記述しているものはほとんど無い。たとえば、語彙概念構造では動詞の構造を対象としているし、Jackendoff による概念意味論の

表現でも, "a red ball" の意味は以下のように単純に記述されている[Jackendoff 83, 05].

[Object BALL [Property RED], INDEF]

これに対して, 本稿で述べた意味表現の最大の特徴は, ある語彙の意味表現が, その語彙と依存関係を持ち得る語彙の意味表現と相互作用するための共通な内部構造を必然的に持つ, という考え方であり, 従って, 動詞のみならず, 形容詞, 名詞も, 修飾を行う, あるいは修飾を受けるための豊富な内部構造を持つ点にある. このことはまた, 意味構造における概念間の依存関係(接続関係)が, 表層における依存関係(係り受け関係)の源になっているということを明示的に示している.

Pustejovsky は, 名詞の意味表現中に, その名詞に関する目的クオリア(TELIC QUALIA)を含めることを提案し, good typist などの, 修飾先によって good の意味が変わる現象や, begin a book のように, 意味的項に取る名詞によって動詞の意味が変わる現象を説明しているが[Pustejovsky 95], これは, 名詞意味表現中にそれが係り得る述語の意味表現を含めるという意味で, 我々の意味表現のアイデアと共通性がある.

提案した意味表現は, 知識表現としてしばしば用いられるフレーム表現[Minsky 75]と類似した構造を持っている. しかし, フレーム表現と自然言語表現との対応は十分に考慮されていない. ここで示した意味表現形式は, 言語表現の観察に基づいて, フレーム表現をより精緻にしたものと考えられる.

また, 多くの意味表現では, 述語を関数によって記述し, ある名詞が述語に係る場合に, その名詞が関数の引数に代入されるという形で意味表現を構成してゆく. これに対して, 我々の意味表現では, 意味表現同士の接続は, 重複する部分を糊しとして重ね合わせることで行われ, 通常の間接の代入とは異なる.

6.2 言い換え技術について

意味表現レベルの構造変換に基づく言い換えプロセスの研究としては, 既に述べたように, たとえば藤田らによる, 語彙概念構造に基づく言い換え計算モデルの提案がある[藤田 06]. この研究に代表されるように, これまでの研究では, 動詞まわりの言い換えを取り上げたものが多い. 野上ら[野上 00] は連体修飾節の言い換え規則を提案しているが, そこでも形容詞は扱われていない. また, 言い換え手続きは依存構造木の構造変換に基づくもので, 各語彙が内部構造を持つような意味表現は用いられていない.

7. おわりに

言い換えに適した意味表現を提案し, それを用いた言い換え処理の手続きを示した. 言い換の具体例として, 形容詞表現の言い換えが行えることを示した. 本稿で述べた言い換え処理の最大の特徴は, 言い換えに必要な規則が, 汎用的な意味構造変換規則と, 意味表現合成, 意味表現分割規則だけであり, 表層の表現や依存構造に固有な言い換え規則を用意する必要が無いという点である.

その代償として, 意味表現の構造が複雑になり, 語彙の内部意味構造の推定および意味表現辞書の構成が困難になっている. また, 意味表現の構成や分割の手続きが組み合わせ的になり, 処理時間がかかることも問題である. 後者の問題に対処するために, 意味表現同士の対応づけを容易にするような表現形式を提案している[Takagi 06].

今後の課題としては, まず, 今回提案した言い換え手続きを評価用のコーパスなどの実際のデータに適用して, その性能を評価することがあげられる. たとえば, 今回の言い換え規則では, 「ボールは色が赤い」という提題 + 主語の構文は生成できない

ように思われる. このような, 生成できない言い換えについて, さらに検討を行いたい.

構造変換規則の適用順序によっては, 接続規則違反の解消ができないケースや, 意味構造が語彙分解できない場合もある. 望ましい形の言い換え表現を効率的生成するためには, そうしたことができるだけ起こらないように, さまざまな言い換え表現の間の変換関係を整理し, 変換規則適用や意味表現分解の制御を適切に行うことが必要である. これは一種の組み合わせ計画問題になると考えられる. そうした場合に, 他の中間表現を用いる場合と比べて, 変形規則や適用順序の制御がどれくらい簡素化されるかも比較検討したい.

[高木 82]では, 色を表す形容詞だけでなく, さまざまな属性形容詞についても意味構造を検討している. また, [高木 84]では「ボールが大きく膨らむ」のような自動詞や副詞句を含む言語表現の意味構造についても検討している. 今回示した構造変形規則および意味表現の分解手続きが, 色を表す形容詞まわり以外の表現についても汎用的に適用可能かどうか検討してゆきたい.

参考文献

- [乾 06] 乾, 藤田: 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.
- [藤田 06] 藤田, 降幡, 乾, 松本: 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えを例題に, 情報処理学会論文誌, Vol.47, No.6, pp.1963-1975, 2006.
- [Jackendoff 83] Jackendoff, R.: Semantics and Cognition, MIT Press, 1983.
- [Jackendoff 05] Jackendoff, R.: Simpler Syntax, Oxford Univ. Press, 2005.
- [佐藤 99] 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.
- [近藤 01] 近藤, 佐藤, 奥村: 格変換による短文の言い換え, 情報処理学会論文誌, Vol.42, No.3, pp.465-477, 2001.
- [Minsky 75] Minsky, M.: A framework for representing knowledge, in The Psychology of Computer Vision, McGraw-Hill, 1975.
- [野上 00] 野上, 藤田, 乾: 文分割による連体修飾節の言い換え, 言語処理学会第6回年次大会予稿集, 2000.
- [Pustejovsky 95] Pustejovsky J.: The Generative Lexicon, MIT Press, 1995.
- [Schank 75] Schank, R. C.: Conceptual Information Processing, North-Holland, 1975.
- [高木 82] 高木, 小原: 属性形容詞の意味構造 - 意味表現方法の一つの試み -, 電子情報通信学会論文誌 D, Vol. J65-D, No.11, pp.1427-1434, 1982.
- [高木 84] 高木, 伊東, 六沢, 北岡, 清水, 小原: 二次元図形世界における視覚情報からの日本語文の生成, 電子情報通信学会論文誌 D, Vol. J67-D, No.2, pp.216-223, 1984.
- [高木 87] 高木, 伊東: 自然言語の処理, 丸善, 1987.
- [Takagi 06] Takagi, A., Asoh, H., Itoh, Y., Kondo, M., and Kobayashi, I.: Semantic representation for understanding meaning based on correspondence between meanings, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.10, No.6, pp.876-912, 2006.
- [高橋 01] 高橋, 岩倉, 飯田, 乾: Kura: 統一的かつ宣言的記述法に基づく言い換え知識の開発環境, 信学技法 TL01-12, 2001.