

視覚情報から多様な言語表現を生成するための意味表現形式

A Semantic Representation for Generating Various Linguistic Expressions from Visual Information

麻生 英樹^{*1} 野口 靖浩^{*2} 高木 朗^{*1,*3} 小林 一郎^{*4}
 Hideki ASOH Yasuhiro NOGUCHI Akira TAKAGI Ichiro KOBAYASHI

近藤 真^{*2} 三宅 芳雄^{*5} 岩橋 直人^{*6} 伊東 幸宏^{*7}
 Makoto KONDO Yoshio MIYAKE Naoto IWAHASHI Yukihiko ITOH

^{*1}(独)産業技術総合研究所知能システム研究部門
 Intelligent Systems Research Institute, 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

^{*4}お茶の水女子大学大学院人間文化創成科学研究科
 Graduate School of Humanities and Sciences, Ochanomizu University

^{*5}中京大学情報理学部
 School of Information Science and Technology, Chukyo University

^{*6}(独)情報通信研究機構
 National Institute of Information and Communications Technology

^{*7}静岡大学
 Shizuoka University

A semantic representation of natural language is proposed and the effectiveness of the representation for translating visual information into natural language sentences is shown. The proposed representation can naturally express relationships between concepts which exist in the recognition result of the target scene, and can generate various forms of linguistic expressions by simple algorithms for transforming semantic representation and dividing into lexical items.

1. はじめに

人間と自然言語を用いて円滑にコミュニケーションできる知的システムの実現は、人工知能研究におけるグランド・チャレンジのひとつであり、多くの研究が行われている。そこでの大きな問題のひとつは、自然言語表現のほぼ無限の多様性にどのように対処するかである。我々は、この問題を解決するために、柔軟な意味表現形式が鍵となるという立場から、概念間の依存関係をベースとした意味表現形式を提案し、「対話理解」、「問題解決」、「言い換え」、「言語獲得」等における有効性を示そうとしてきた[高木 87, 麻生 08a, 08b]。本発表では、提案する意味表現形式が、視覚等の感覚情報を多様な文体で言語表現するためにも有効であることを示したい。

以下、本稿では、本研究で扱う視覚情報と言語表現の関係づけの問題について述べ、我々が提案している意味表現を概観し、それを用いることで、簡単に汎用的な手続きによって、視覚情報から多様な構文の言語表現が得られることを示す。

2. 本研究の目的と関連研究

視覚情報は第一義的には網膜上で変化する光の輝度の束、すなわち、ある物理的な位置にある点のある時刻の輝度という情報の集合である。我々はその情報から「赤い花が咲いている」、「太郎が走っている」というような言語表現を生成することができる。また、逆に、そうした言語表現から具体的な視覚情報を

生成し、たとえば絵に描くこともできる。その情報処理プロセスを構成的に明らかにし、工学的に応用することが本研究の第一の目的である。

さらに、人間は、同じ視覚情報を多様な言語表現で表すことができる。たとえば「赤い花が咲いている」という単純な情報であっても、主語や主動詞の選択によって、「咲いているのは赤い花だ」、「咲いている花は赤い」、「咲いている花は赤い色をしている」等々の表現が可能である。こうした多様な言語表現の生成がなぜ可能なのかを明らかにするとともに、可能な表現をできるだけシステムティックに生成できるようにすることが、本研究の第二の目的である。

視覚情報と言語情報の関係を扱う研究の代表的事例としては Winograd らによる SHRDLU が挙げられる [Winograd 72]。SHRDLU は、仮想的な積み木の世界について人間とコミュニケーションしつつ積み木の操作等を行うことができるシステムで、自然言語理解の研究に大きなインパクトを与えたが、その後の研究は、コンピュータビジョンのような視覚中心の研究と、自然言語中心の自然言語処理とに分かれて発展し、視覚等の感覚情報と豊かな言語表現との両方を扱う研究はあまり行われてこなかった [田中 00]。しかし、近年、ハードウェアとしてのロボットの発展等に伴って、再び研究が盛んになりつつある。たとえば、新山らは SHRDLU と同様に仮想的な空間でのイベントについてコミュニケーションするシステム「傀儡」を構築し、特に話者の視点の違いによる表現の漠然性や照応関係の解消について検討を行った [新山 01]。久野は、視覚と言語を対応づけて処理できるサービスロボットを提案している [久野 06]。また、岩橋や Roy は、視覚情報と言語表現との対応づけを学習するシステムの研究を行っている [Roy 02a, 02b, Iwahashi 04, 岩橋 09]。しかし、いずれの

連絡先: 麻生英樹, (独)産業技術総合研究所知能システム研究部門, 茨城県つくば市梅園1-1-1中央第2,
 h.asoh@aist.go.jp

研究においても、言語表現の文体の多様性に十分に対応できているとはいえない。

一方、言語学は Chomsky による変形生成文法の提案以来多様な言語現象の説明原理である普遍文法を探索してきているが、そこでの研究は言語現象、その中でも特に、統語的な現象に関するものが多かった。言語の意味に関しては形式意味論が高度に発展しているが [Portner 08, 戸次 10]、言語の意味の間の論理的関係が主題であり、言語表現と感覚情報との関係については扱われていない。しかし、1980年代から、たとえば Jackendoff らは感覚情報と言語表現とをつなぐ概念構造 (Conceptual Structure) の重要性を主張し、音韻構造、統語構造、概念構造がそれぞれ並列に生成的な能力を持ち相互作用する、というフレームワークを提案している [Jackendoff 83, 02]。また、認知言語学においても、人間の感覚情報と結びついた意味論の重要性が述べられる [Langacker 08] など、感覚情報と言語表現との関係に関する研究が少しずつ増え始めているが、工学的に有用なシステムの実現を目指しているものはまだ少ない。

3. 視覚情報の認識

視覚情報を言語化するためには、視覚情報中のオブジェクトを認識する必要がある。一定の光の束のまとまりをオブジェクトとして認識し、その属性を推定する問題はコンピュータビジョンの研究課題であり、セグメンテーションなどと呼ばれている。以下では、簡単のために、ほぼ単色の比較的単純な形状をした、あまり変形しないオブジェクトを対象とする。その場合、同じ色の部分が、同じ速度・角速度で動いていることを用いて、視覚情報中のオブジェクトを切り出すことが可能である。その結果、視野内の個々のオブジェクトについて、色、形状、大きさ、そして、重心点の一定時間ごとの位置座標を得ることができる。以下では、これらの情報を基本となる視覚認識結果として用いる。

ここで注意すべき点は、上記の認識結果の中に、既に、様々な概念とその間の関係が現れているということである。すなわち、物体1の色が赤(と呼ばれる周波数)であるという認識は、物体1という概念が、色という概念を内包しており、その色がさらに明度や彩度という概念を内包し、その彩度が赤に等しい、ということの意味している。すなわち、視覚情報の認識処理とは、ある点のある時刻での輝度が X に等しい、という情報の束を、上のような概念間の関係の束に変換する処理である。

4. 概念間依存構造表現

われわれは、多様な自然言語表現と対応づけることが可能な意味表現形式として、概念間依存構造表現を提案してきている [高木 87]。この意味表現形式の特徴は、

- 表現の多様性に対処するため、プリミティブ概念を用いて、概念とその間の依存関係をできるだけ詳細に表現する。
- 概念間の依存関係を素直にプリミティブ概念の隣接接続関係として表現する。
- 表現中の節構造のまとまりを明示する。

という点である。

以下では、プリミティブ概念として、以下のものを用意する。

- (1) 実体概念 ○
「もの」とその部分
- (2) 属性概念 ○
「色」、「形」、「大きさ」、「速度」、「方向」など
- (3) 値概念、具体値 [値]
「赤」、「円形」、「40 pixel/sec」など、「色相」、「形」、「速度」等の属性概念が取る値

- (4) 現象概念 ◎
「存在する」、「内包する」、「等しい」、「変化する」などの基本的な現象
- (5) 代名詞、代動詞概念 ● など
実体・属性概念と等値される代名詞概念、現象概念と等値される代動詞概念
- (6) 格概念 → + 格の名前
「主格」、「対象格」、「時間格」、「場所格」など、現象とそれに参画する実体との関係を表す概念
- (7) 接続助詞概念 ←→
現象間の関係を表す概念

これらの概念を下記のように相互接続することで、複雑な意味を表す概念間依存構造を構成することができる:

- 実体概念、属性概念は、現象概念と格概念を介して接続することができる。
- 現象概念は、現象概念と接続助詞概念を介して接続することができる。
- 代名詞概念は照応先の実体概念、属性概念と=で接続することができる。
- 複数の現象概念を含む場合、従属節のまとまりを()で囲んで表示する。

この概念間依存構造表現を用いれば、前節で述べた視覚情報の認識結果を素直に記述することができる [高木 84, 野口 10a]。図 1 に概念間依存構造表現の例を示す。この概念間依存構造は、視野内の一つの物体(物体1)に関して、

- 色を内包し、色が色相を内包し、その色相の値が [赤] に等しい
- 形を内包し、その形の値が [四角] に等しい
- 存在する位置が [Y 軸] に沿って、[+] に等しい方向に向かって、[5 pixel/sec] に等しい速度で変化する

という概念間の関係を表している。

5. 言語表現の生成

概念間依存構造表現は、(1) 依存構造自体を構造変換する、(2) 単語に分割して単語間の依存構造木を得る、(3) 得られた単語を、依存構造を考慮しつつ、統語的に正しい文になるように一次元の列に並べる、という手続きを経て多様な構文構造の表層言語表現に変換される [高木 84, 野口 10a]。このうち (3) は通常 of 自然言語処理において行われている処理と同じであるため、以下では、(1) と (2) について簡単に説明する。

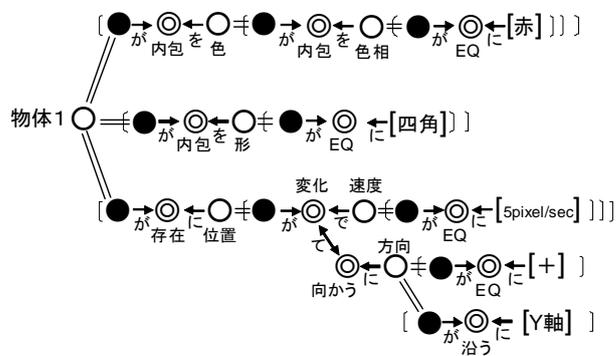


図1 概念間依存構造表現の例

5.1 依存構造変換

構文構造が異なる(たとえば主動詞が異なる)文を生成するためには、概念間依存構造を変形しておく必要がある。例えば、図1の構造を素直に日本語に変換すると、「赤に等しい色相を内包する色を内包し、四角に等しい形を内包し、Y軸に沿って+に等しい方向に向かって5pixel/secに等しい速度で変化する位置に存在する物体1」という名詞句になるが、これをたとえば、色を主語とし、色が主格で依存関係を持っている内包現象を主動詞として「物体1が内包する色が赤に等しい色相を内包する」のように表現しようと思った場合、該当する部分を従属節の()の外に出す操作が必要になる。さらに、連体助詞の「の」と終止形容詞を用いて「物体1の色が赤い」のように表現することも可能である。また、「物体1の色が赤い」は、「物体1は色が赤い」、「物体1が色に関して赤い」のように、属性からそれが所属する物体へと主語を変換して表現することもできる。

このような表現を生成するための概念間依存構造の変換規則として、

- 主節と従属節の入れ替え規則
- 連体修飾節と連体助詞句の変換規則
- 値を示す表現と程度表現の形容詞との変換規則
- 主語の変換規則
- 連体形容詞と終止形容詞との変換規則

を用意した。もちろん、これらがすべてを尽くしているわけではないが、これだけでもかなり広範囲な文体をカバーすることが可能である。図2に、図1の概念依存構造を、「変化する」が主動詞となるように構造変換した例を示す。

5.2 単語分割

表層言語表現に現れる単語は、プリミティブ現象概念を複数組み合わせるとともに、その値部分に特定の値を代入することによって意味を限定したものであると考えられる。

例えば「大きさ」などの属性の変化に関する概念間依存構造表現を考えると、「変化する」という抽象度の高い動詞概念を「+に等しい方向に向かって」という連用修飾成分によって限定することによって、「増大する(=+に等しい方向に向かって変化する)」という単語が生成される。さらに、「大きさに関して増大する」というまとまりは、「膨らむ」という動詞に対応することになる。

このように、各単語の意味もまた、概念間依存構造によって表現しておくことができる。これを辞書として用いて、変換後の概念間依存構造に対して、単語辞書に存在する単語の割当てを試みて行き、過不足無く割り当てることができる単語の組み合わせを求めることで、概念間依存構造から単語間の依存構造を

表す木を得ることができる。

単語への分割処理は、ヘッドのプリミティブ現象概念から順に単語を割当て、更にその割当てから余った枝に対して再帰的に単語割当てを実施すればよい。ただし、概念間依存構造へ割り当てる単語の差異によって、複数の単語間依存構造が生成可能な場合があるため、その全てのパターンを検証する必要がある。

図3は図2に単語分割処理を行った例である。具体的には「変化する」をヘッドとする概念間依存構造に対して、その主節の主動詞として「移動する」を割り当てた場合を表している。以下、その手順のごく簡単な概要を示す。

- (1) 概念間依存構造表現のヘッドのプリミティブ概念に着目する。
- (2) 着目したプリミティブ概念を意味表現のヘッドとする単語を単語辞書から検索し、単語候補リストにする。
- (3) 単語候補リストから、着目したプリミティブ概念をヘッドとする部分木(A)に割当て可能な単語((A)が単語意味表現(B)を包含する関係にある単語)を調査する。包含関係にある場合、その単語を割当て、割当てから外れた概念間依存構造の枝のヘッドのプリミティブ概念に着目して(2)へ。包含関係にない場合、この単語は棄却し次の単語を調査する。
- (4) 概念間依存構造全てに単語を割当てることができれば、その単語の組み合わせで単語間依存構造を生成する。割当てられない概念間依存構造が残る場合にはその単語の組み合わせは棄却する。

図2の例の場合には、手順(1)においてヘッドのプリミティブ概念「変化する」に着目する。次に、手順(2)において、「変化する」をヘッドとする意味表現を持つ単語、例えば「変化する」、「移動する」、「上がる」、「下がる」・・・などを単語候補リストに挙げる事ができる。手順(3)において、ここでは仮に「移動する」を単語候補とする。この場合、図2の概念間依存構造表現の「変化する」をヘッドとする部分木と、図3の「移動する」の単語意味表現の関係を調査する。両者は包含関係にあるため、「移動する」を割り当てることができる。

更に処理を進め、この「移動する」の割当てにおいて、余った以下の枝に対して、更に単語割当てを実施していく。

- 「赤い色をした、四角い形をした物体が」
- 「Y軸に沿う+に等しい方向に向かって」
- 「5pixel/secに等しい速度で」

例えば、余った枝「Y軸に沿う+に等しい方向に向かって」に関しては、図3のように「上」「に」「向かう」「て」のように単語を割り当てることができ、その結果、この枝の概念間依存構造全てに

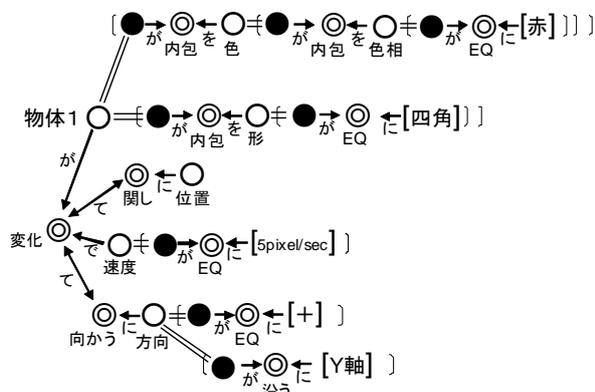


図2 依存構造変換後

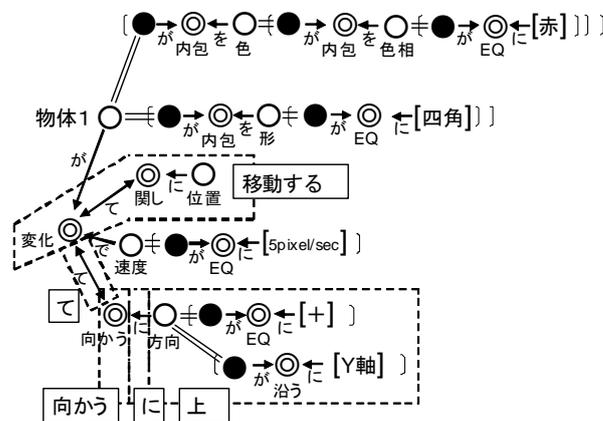


図3 単語分割の例

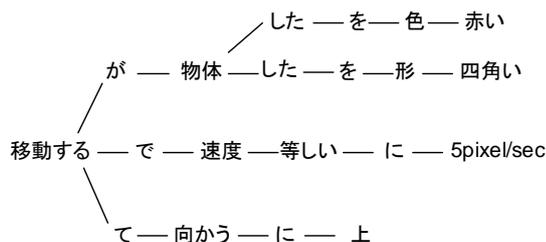


図4 単語間依存構造木の例

単語を割り当てることができる。更に、他の余った枝についても、同様の単語割当てを試み、枝ごとに余りの無いように単語間依存構造を生成していく。こうして得られる単語間依存構造の例を図4に示した。

概念間依存構造の「変化する」をヘッドとする部分木に対して、最初に「上がる(=Y軸に沿う+に等しい方向に向かって位置に関して変化する)」という動詞を割り当てることができる。この場合、「上」「に」「向かう」「て」「移動する」という単語分割の代わりに「上がる」という一語の分割が得られることになる。こうした異なる分割操作によって、異なる単語間依存構造が得られる。

6. まとめ

視覚現象を多様な言語表現に変換するために有効な意味表現形式=概念間依存構造表現について述べた。視覚認識結果が概念間依存構造表現によって素直に表現できること、および、そうして得られた概念間依存構造表現から、少数の変換規則と簡単な汎用的な単語分割手続きにより、多様な構文の自然言語文をシステムティックに生成できることを示した。

ここで扱っている問題と、[麻生 08a]で扱ったような自然言語処理における「言い換え」問題との違いは、視覚情報認識結果からの変換を含んでいる点にある。逆に言えば、視覚情報認識結果を概念間依存構造表現に変換した後については、「言い換え」と同等の処理である。したがって、本稿の最大の貢献は、視覚情報の認識結果の中に、既に概念とその間の依存関係が含まれているという指摘にある。

言語との対応づけを考慮していない通常の視覚情報処理やそれを用いたアプリケーションでは、それらの依存関係は明示的には意識されず、省略された形で表現される。しかしながら、そこに存在する依存関係を詳細に明示的に表現することが、視覚情報と自然言語表現との対応づけにおいて重要であるという点の本稿で示した洞察である。

今後の課題としては、まず、提案した手法の有効性の検証が挙げられる。我々は、本稿で示した視覚情報を言語表現に変換する手続きを実装したシステムを構築するとともに[野口 10a]、その出力を、人間が生成する文と比較することで、提案した意味表現形式と変換手続きの有効性を評価している[野口 10b]。そうした評価結果に基づいて、今回の方法で生成できない表現に対しても対応できるようにシステムを改良してゆきたい。

また、今回提案した意味表現形式と変換手法は、視覚以外の聴覚、触覚等の感覚情報の言語化に対しても有効であると考えられる。さらに、感覚情報から言語表現への変換と逆の方向の変換、すなわち、言語表現を入力として、それと矛盾しない感覚情報を生成することにも使えらる。そうした拡張を行い、最終的には人間と同じように多様な自然言語表現を生成できるシステムを構築してゆきたい。

参考文献

- [高木 87] 高木朗, 伊東幸宏: 自然言語の処理, 丸善, 1987.
- [麻生 08a] 麻生英樹, 伊東幸宏, 高木朗: 言い換えに適した意味表現について, 第22回人工知能学会全国大会, 1F2-03, 2008.
- [麻生 08b] 麻生英樹, 伊東幸宏, 高木朗: 言語獲得に適した意味表現について, 第22回人工知能学会全国大会, 3E3-03, 2008.
- [Winograd 72] Winograd, T.: Understanding Natural Language, Academic Press, New York, 1972.
- [田中 00] 田中穂積: 言語理解—SHRDLUの先にあるもの—, 人工知能学会誌, 15(5), 821-828, 2000.
- [新山 01] 新山祐介, 徳永健伸, 田中穂積: 自然言語を理解するソフトウェアロボット: 傀儡, 情報処理学会論文誌, 42(6), 1359-1367, 2001.
- [久野 06] 久野義徳: サービスロボットのための視覚と対話の相互利用, 情報処理学会論文誌, 47, SIG15 (CVIM16), 22-34, 2006.
- [Roy 02a] Roy, D. K., Pentland, A.: Learning words from sights and sounds: A computational model, Cognitive Science, 26(2), 113-146, 2002.
- [Roy 02b] Roy, D. K.: Learning visually-grounded words and syntax for a scene description task, Computer Speech and Language, 16(3), 353-385, 2002.
- [Iwahashi 04] Iwahashi N.: Active and unsupervised learning for spoken word acquisition through multimodal interface, Proceedings of 13th IEEE Workshop Robot and Human Interactive Communication, 437-442, 2004.
- [岩橋 09] 岩橋直人: LCore: 言語と動作によるコミュニケーションを学習するロボットの知能化技術, 第12回情報論的学習理論ワークショップ, 2009.
- [Portner 02] Portner, P. and Partee, B. H.: Formal Semantics, The Essential Readings, Blackwell Publishing, 2002.
- [戸次 10] 戸次大介: 日本語文法の形式理論, 活用体系・統語構造・意味合成, くろしお出版, 2010.
- [Jackendoff 83] Jackendoff, R.: Semantics and Cognition, MIT Press, 1983.
- [Jackendoff 02] Jackendoff, R.: Foundations of Language, Oxford Univ. Press, 2002 (郡司訳: 言語の基盤, 岩波書店, 2005).
- [Langacker 08] Langacker, L. W.: Cognitive Grammar, A Basic Introduction, Oxford University Press, 2008.
- [高木 84] 高木朗, 伊東幸宏, 六沢一昭, 北岡和憲, 清水正朗, 小原啓義: 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, J67-D(2), 216-223, 1984.
- [野口 10a] 野口靖浩, 麻生英樹, 高木朗, 小林一郎, 近藤真, 三宅芳雄, 岩橋直人, 伊東幸宏: 視覚情報から言語表現を生成するシステムの試作, 人工知能学会研究会資料, SIG-SLUD-A903-08, 43-48, 2010.
- [野口 10b] 野口靖浩, 麻生英樹, 高木朗, 小林一郎, 三宅芳雄, 岩橋直人, 伊東幸宏: 視覚情報から多様な言語表現を生成するための意味表現形式, 第24回人工知能学会全国大会 2G1-OS3-8, 2010