

高階表現の意味表現と述語間依存関係の解釈について

Semantic Representation of Higher-order Statements and
Interpretation of Predicate-Predicate Dependence Structures麻生 英樹^{*1} 高木 朗^{*2,1} 小林 一郎^{*3} 橋田 浩一^{*1}
Hideki ASOH Akira TAKAGI Ichiro KOBAYASHI Koiti HASIDA^{*1}産業技術総合研究所 AIST ^{*2}言語情報処理研究所 NLP Research Laboratory ^{*3}お茶の水女子大学理学部 Ochanomizu University

We have proposed a semantic representation of natural language which enables semantic processing based on semantic structure mapping. Using the representation, object-predicate dependency structures can be interpreted by a simple matching algorithm. In this presentation the framework is extended to higher-order statements. Japanese higher-order statements are examined and classified into several categories. The proposed framework can be extended to some of them. An extension of the semantic representation is proposed, and effectiveness of the mapping algorithm for interpreting predicate-predicate dependence structures are demonstrated.

1. はじめに

人間の自然言語理解過程と、それをを用いた問題解決の過程をモデル化し、計算機上で実現する、という問題に対して、これまでに多くの研究が行われてきたが、処理対象である自然言語文の表層表現の大きな多様性に十分に対応できているとはいえない。特に、幅広い範囲の同義な文に対して、表層の依存構造の違いによらず、等価な意味解釈結果を与えるような、汎用的意味解釈処理アルゴリズムは未だに実現されていない。そのため、たとえば多くの対話システムでは、ユーザ発話から表層レベルのテンプレートを用いて情報を抽出して応答を返すことが行われており、適切に応答可能な入力文の多様性に大きな制限がある。

著者らは、「入力文の意味は、そこに含まれる情報が、文脈・状況・知識上へ位置づけられることによって理解される」という考え方[高木 87]に基づき、幅広い範囲の入力文に対して意味の位置づけによる理解を可能にする新しい意味表現と、それをを用いた意味解釈手法を提案してきた[高木 02, Takagi 06, 麻生 06, 野口 06]。また、ホテル検索, LAN 構築ヘルプ, 情報家電制御などのドメインで対話システムを構築し、意味表現および意味解釈処理の有効性を示してきた[池ヶ谷 02, 常盤 03, 内田 05]。

しかしながら、これまで提案してきた枠組みはいわゆる一階の表層表現のみを扱うものであり、述語が述語に係るような高階の表現を扱っていなかった。本発表では、著者らの枠組みを高階の表層表現に対して適用できるように拡張することを試みる。

以下では、まず、これまでに提案している意味表現および意味解釈処理について概説し、次に、本発表で対象とする高階表現について検討する。そして、ある種の表現に対しては、既存の枠組みを自然に拡張することによって、述語間依存関係の解釈が可能になることを示す。

2. 意味表現と意味解釈処理

自然言語処理を困難にしている要因の一つは、同じ意味内容が多様な表層表現で表され得ることである。たとえば、「Aさんは赤い車を持っている」、「Aさんの自動車の色は赤である」、「Aさんが持っている車は赤い」という3つの表層表現は、ほぼ同じ意味内容を持っている。こうした多様な表現が可能なのは、

単語レベルでの類義語の存在(「車」と「自動車」)、頻出表現の圧縮単語化(「赤である色を内包している」→「赤い」、「が持っている」→「の」)、そして、何を主体として表現するかに関連した依存関係(係り受け関係)や述語の選択(「持っている」、「である」、「赤い」)、などの組み合わせの結果である。

従来の意味表現では、単語レベルの類義語については、単語の意味を表す概念辞書を構築することによってある程度吸収することができているが、それを越えた句や節レベルの表現の違い、すなわち表層表現内の依存関係や述語の違いは、そのまま意味表現の中に残ってしまうことが多かった。たとえば、「東京を出発する」と「出発地は東京である」は、同じ意味内容を持つといえるが、この二つの文を素直に論理式による意味表現で表すと、それぞれ、出発(X, 東京)、である(出発地, 東京)のようになり、述語の選択の違いがそのまま意味表現に反映されてしまう。このように、表層表現の違いが、ほぼそのまま意味表現に反映され、意味表現中の同じ情報内容を持つ部分が異なる形態で表現されていると、それらに対応づけて、同じ情報を表している部分を同定し、対応づけることが難しい。

この困難を低減し、幅広い表層表現に対して、一定の手続きによって、その意味表現間の相互の位置づけを可能にするために、著者らは、多様な依存関係が「属性は値である」という「断定」の述語を用いた形に同義変形できることを根拠として、文を依存関係ごとに分解し、述語と連用修飾成分、及び、名詞と連体修飾成分に含まれる単語間の依存関係に含まれる情報を圧縮して一語の属性名の中に繰り込むことによって、各依存関係を、「属性名 = 値(ある属性が値である)」という均一な「断定表現」の形で表現する意味表現を提案した[Takagi 06]。

その意味表現は「属性名 値」(以下、意味表現中では=記号は省略する)を束ねて作られる素性構造[Sag 99]を基本構造とする。値部分は「[知識値]修飾値」という構造を持つ。知識値は、その属性に関する知識として、語彙辞書中に予め記載されている値であり、修飾値は、入力文で指定された値である。いずれの値も、名詞概念意味表現へのポインタや、数値、記号が格納される。

図1に述語概念意味表現の一種である現象概念意味表現、図2に名詞概念意味表現の一種である実体概念意味表現の基本形をそれぞれ示す。現象概念意味表現は、その主要な属性として、現象名・主体、現象名・場所、などの属性を持つ。一方、実体概念意味表現は、主要な属性として、その実体概念と、それに係りえる現象概念の属性を組み合わせ、実体名\$現象

名・主体、などの属性を持つ。例として、図 3 に、「Tさんは小金井に住んでいる」という文の意味表現を示した。

この意味表現により、「各連用成分+述語」あるいは「各連体修飾成分+名詞」によって表現される情報を、「属性名 値」という一定の形式で参照出来るようになる。さらに、そこで用いられる属性名は、対象問題領域の用例文に基づいて、出現する全ての概念およびそれに関する連用・連体修飾成分を抽出することによって、あらかじめ概念体系(オントロジー)として体系化しておくことができる。従って、対象問題領域に関する幅広い表層表現に対して、属性名をキーとした一貫した意味表現の位置づけが可能になる。

さらに、著者らは、提案した意味表現と意味の位置づけに基づく意味解釈として、節内の意味解釈処理と節間の意味解釈処理の二種類を提案している[Takagi 06]。節内の意味解釈処理とは、語と語の間の直接的な依存関係(係り受け関係)を手がかりとして、意味表現中の「属性名 値」を対応づけ、それらの値の情報を相互に伝搬させる処理である。

たとえば、図 3 の例では、「Tさん」が「居住 1」に主格で係っていることから、居住 1 意味表現中の「居住 1 場所」属性と、Tさん意味表現中の、「Tさん\$居住 2 場所」属性が位置づけ可能であり、相互に値の情報を伝搬させることができる。この結果、この文で言明された情報が、Tさん意味表現中および小金井意味表現中にも書き込まれる。これらの情報は、後で Tさんや小金井に関する質問に答えるために利用できる。

これはあたかも、ある現象や実体についての言及が増えるに従って、それについての情報が蓄積され、イメージが膨らんでゆく過程のモデルになっている。そして、イメージ中の同一視可能な属性は、相互に位置づけられて、情報が伝播される。

一方、節間の意味解釈処理は、直接的な依存関係が無い複数の節の間で、意味表現の対応づけを行って、情報を伝搬させる処理である。どちらについても、表層の依存構造によらない、汎用的な手続きによって、幅広い表現に対する意味解釈が可能になり、同じ情報を含む異なった構造の表現から同じ情報を抽出できることが示されている(詳細は[Takagi 06]を参照)。

しかしながら、これまでに提案した意味表現および意味解釈処理においては、「属性 値」の値には、実体概念や属性概念等の名詞概念意味表現へのポイントや、数、記号が入ることを想定していた。このことは、いわゆる「高階」の表層表現、すなわち、述語が述語に係るような表現には対応していないということの意味する。以下では、これまでの枠組みを拡張し、ある種の高階表現に適用する。

3. 高階の表層表現の検討

日本語に現れる高階的な表層表現、すなわち、述語が述語に係る表現にはさまざまなものがある。その中で、たとえば、「胃が痛いのでキャベジンを飲んだ」、あるいは、「LAN ケーブルを接続してから電源を入れる」のように、主節と従属節とが接続助詞で結合されている場合は、これまでの意味表現の範囲で、前述した複数節間の意味解釈手続きを用いて、複数の述語の意味を統合した意味解釈が可能であることが示されている。

これに対して、本稿では、二つの述語がより密接に関係している場合の典型として、「開発することを企画する」のように、「XX ということを XX する」という形を基本形とする表現について検討する。この中には、「朝食を食べたい=朝食を食べる(という)ことを希望する」のように助動詞によって表現されるものや、「食べるのが好き=食べる(という)ことを好む」のように「の」や「こと」によって表現されるものなども含まれる。

[Sag 99]に習って、「Tさんは朝食を食べることを開始した」と、「Tさんは朝食を食べることを希望した」という二つの例を取り上げる。これら二つの文は同型の依存構造を有している。すなわち、最初の例では、「食べる」が「開始する」に係り、後の例では「食べる」が「希望する」に係っている。これらの文に係り受け関係を分解し、そこに表現されている「属性 値」情報を取り出すと、下記のようになる。

「Tさんは朝食を食べることを開始した」 (1)

開始 1 主体は Tさんである。

開始 1 対象は摂取 1 である。

摂取 1 対象は朝食 1 である。

「Tさんは朝食を食べることを希望した」 (2)

希望 1 主体は Tさんである。

希望 1 対象は摂取 1 である。

摂取 1 対象は朝食 1 である。

これらは一見同じように見えるが、係り受け関係にある二つの述語の意味を限定する情報についてみると、大きな違いがあることがわかる。すなわち、(1)においては、摂取 1 主体も Tさんであると言える。一方、(2)においても、摂取 1 主体は Tさんであるのが自然だが、そうならない場合も起こり得る。たとえば、「Tさんは Sさんが朝食を食べることを希望した」という表現や「Tさんは Sさんに朝食を食べるように希望した」といった表現が可能である。さらに、「Tさんは 8時に食堂で朝食を食べ始めた」という文について考えるとわかるとおり、開始現象と摂取現象は、主体のみならず、場所や時点についても同じ情報を共有することがわかる。別の言い方をすれば、(1)では、開始現象は摂取現象と不可分であるが、(2)ではそうではない。

この例から、述語 A が述語 B に係るときに、述語 A と B に関する情報が幅広く共有される場合があり、どの属性の情報が共有可能か否かは述語 B のタイプに依存して決まる、ということがわかる。多数の属性の情報を共有できる述語 B のグループには、「開始する」、「継続する」、「終了する」など、述語 A が言及している現象のアスペクトを取り立てて表現するものが含まれる。

次に、「朝食を食べたい」という表現について検討する。これは、意味的には「朝食を食べることを希望する」と同じであるが、このように助動詞形式で表現された場合には、摂取 1 の主語と希望 1 の主語は、必ず一致していることがわかる。また、希望 1 の時点や場所といった属性についても、摂取 1 のそれらの属性と別に語るができない。

4. 意味表現の拡張と意味解釈処理

前節で検討したように、述語が述語に係ることにより、一方の述語の意味が他方によって制約されるということが起こる。これは、名詞が格助詞を介して述語に係る場合や、連体修飾節が名詞に係る場合と同様の現象である。そこで、この現象を意味表現上で実現するために、これまで提案している現象概念意味表現を拡張する。

これまでの現象概念意味表現は、図 1 に示したように、現象名・主体、述語名・時点、などの属性を持っていた。ここに、実体概念意味表現と同様に、この現象概念に係りうる現象概念の属性を組み合わせた、現象名 1%現象名 2 主体 のような属性を導入する。図 4 に、拡張された現象概念意味表現の基本形を示す。また、例として、「Aさんは朝食を食べることを開始した」の意味表現を図 5 に示す。この場合、摂取 1 現象の意味表現に、摂取 1%開始 2 主体、摂取 1%開始 2 時点、のような属性が追加されていることがわかる。

さらに、前節で述べたように、助動詞形の表層表現と、一般動詞形の表層表現では、意味解釈処理が異なる場合があることを実現するために、助動詞形を取り得る現象概念については、その意味表現中に、助動詞形か一般動詞形かのフラグを持つこととする。意味解釈処理アルゴリズムは、このフラグも参照して、どの属性が位置づけ可能であるかを定める。

この意味表現を用いると、これまでと同様の、属性名をキーとした位置づけのアルゴリズムによって、述語と述語の依存関係の解釈が可能になる。たとえば、図 5 の例の場合、摂取 1 現象が開始 1 現象に係ることによって、開始 1 現象意味表現中の、開始 1 主体や開始 1 時点といった属性が、摂取 1 意味表現中の開始 2 主体、開始 2 時点といった属性に位置づけ可能になり、それらの属性の間で値の伝播を行えることになる。図 6 に意味解釈処理後の意味表現を示した。

こうした意味解釈によって、たとえば、「Aさんは8時に朝食を食べ始めた」という文の後に、「Aさんは何を食べたか?」という質問が来た場合に、ヘッダの現象概念が異なる(最初の文では「開始」、二番目の文では「摂取」)にもかかわらず、回答することができるようになる。また、「Aさんは風呂場で朝食を食べ始めた」という文が入力されたときに、ここに含まれる「風呂場で食べ始めた」(開始 1 場所 [場所 1] 風呂場)という情報が、朝食に関する辞書知識中に書かれている、「朝食は通常食堂で食べる」(朝食\$摂取%場所 [食堂])という情報と矛盾することを確認し、何かおかしいことが言われている、と気づくことができる。

このように、述語が述語に係る場合でも、名詞が述語に係る場合などと同様に、依存関係を単位とする意味表現と意味解釈処理によって、現象概念あるいは実体概念意味表現中に、入力文に含まれる情報を蓄積してゆくことができることがわかる。

5. おわりに

高階の表層表現の意味表現とそれを用いた意味解釈処理について述べた。我々が提案してきている意味表現形式および汎用的な意味解釈処理が、高階の表現の意味処理に対しても有効であることを確認することができた。

依存関係を単位とする意味表現と、意味の位置付けによる意味理解という考え方は、[Schank 75] の Script などと同じものである。しかし、Schank の用いた意味表現は、表層表現に含まれる依存構造をほぼそのままの形で残しているものであり、位置づけ方法はテンプレートマッチングをベースとするものであった。このため、Script 上に位置付け可能な言語表現及び意味解釈可能な言語表現の幅は狭いものに限定された。これに対して、ここで用いている意味表現は、依存関係を単一の述語に帰着させることで、より広範囲の言語表現を受容出来る可能性を示したものである。

「属性 値」表現の束による表現形式は、Generalized Phrase Structure Grammar(GPSG)や Head Driven Phrase Structure Grammar (HPSG)[Pollard 87, Sag 99]などで、素性構造として用いられている。それらの研究と我々の研究は、使用している表現の構造およびその上の基本的な操作(Unification)の点では類似しているが、本稿で用いた意味表現の重要性は、意味表現の素性を具体的にどのように決めれば、その後の意味解釈や問題解決を、入力文の範囲を限定することなく、汎用的な手続きで柔軟かつ安定に行えるようになるか、という問題に対して、二つの表現が「同義である」という言語現象に立脚し、用例に基づく属性名の体系化に基づいて、ad hoc ではなく、答えているところにある。

今回の検討では、様々な高階表現のうち、最も基本的なものと考えられる形態を取り上げた。今後はより広い範囲の高階表

現に対しての検討を進めてゆく必要がある。また、今回は意味解釈処理のみを検討したが、係り受け解析処理との融合[池ヶ谷 07]も重要な課題である。さらに、学習的な対話システムや、学習的なオントロジーマッピングによる Web サービス連携システムなどへの適用を行ってゆきたい。

参考文献

- [麻生 06] 麻生 他: 複数節間の意味の位置づけによる意味解釈について, 人工知能学会第 20 回全国大会予稿集 1F3-02, 2006.
- [池ヶ谷 02] 池ヶ谷 他: 文脈への意味の位置付けに基づくホテル予約対話システムの構築, 人工知能学会研究会資料, SIG-SLUD-A202-11, pp.63-70, 2002.
- [池ヶ谷 07] 池ヶ谷 他: 対話文脈を利用した構文意味解析, 人工知能学会論文誌, (to be published), 2007.
- [内田 05] 内田 他: 意味の位置づけを可能にする意味表現を用いた情報家電操作のための対話的インタフェース, 第 19 回人工知能学会全国大会, 予稿集 2005.
- [野口 06] 野口 他: 意味の位置づけを考慮した意味表現方式と依存関係の解釈, 人工知能学会第 20 回全国大会予稿集 1F3-01, 2006.
- [Pollard 87] Pollard, C., Sag, I. A.: Information-based Syntax and Semantics, Vol.1: Fundamentals, CSLI, 1987. (郡司訳: HPSG 入門 制約にもとづく統語論と意味論, 産業図書 1994.)
- [Sag 99] Sag, I. A. and Wasow, T.: Syntactic Theory: a Formal Introduction, CSLI Publications, 1999. (郡司, 原田訳: 統語論入門(上, 下)岩波書店, 2001.)
- [Shank 75] Schank, R. C. : Conceptual Information Processing, North-Holland, 1975.
- [高木 87] 高木, 伊東: 自然言語の処理, 丸善, 1987.
- [高木 02] 高木, 中島, 伊東, 近藤, 今仁, 三宅: 文脈への意味の位置づけを重視した対話意味表現, 人工知能学会 言語・音声理解と対話処理研究会 SIG-SLUD-A202-10, pp.55-62, 2002.
- [Takagi 06] Takagi, A. et al.: Semantic representation for understanding meaning based on correspondence between meanings, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.10, pp.876-912, 2006.
- [常盤 03] 常盤 他: 学生ネットワーク構築実験支援対話システム, 人工知能学会全国大会第 17 回大会, 2G2-01, 2003.

現象概念 G(マーカー類)

G 現象属性 A1 [知識値 A1] 修飾値 A1
G 現象属性 A2 [知識値 A2] 修飾値 A2
...
G 現象属性 B1\$内包対象 [程度 B1] NIL
...
G 現象属性 C1(原因等) [NIL] 修飾値 C1
...
G 部分現象 [知識値 D1] NIL
...
属性限定句 [NIL] 修飾値 E1
...

図 1 現象概念意味表現の基本形

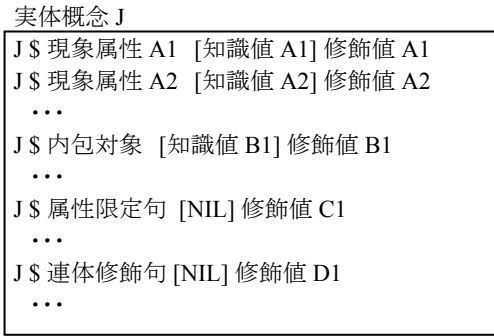


図2 実体概念意味表現の基本形

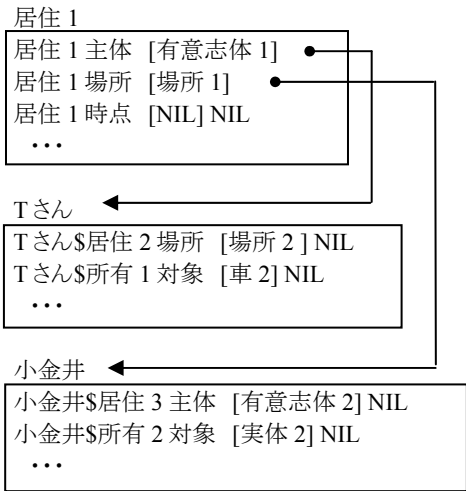


図3 「Tさんは小金井に住んでいる」の意味表現

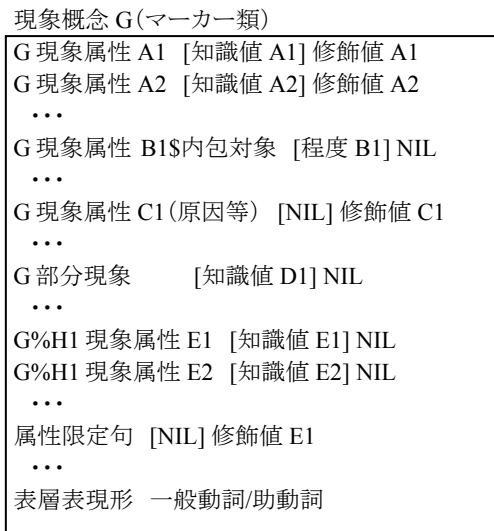


図4 拡張された現象概念意味表現の基本形

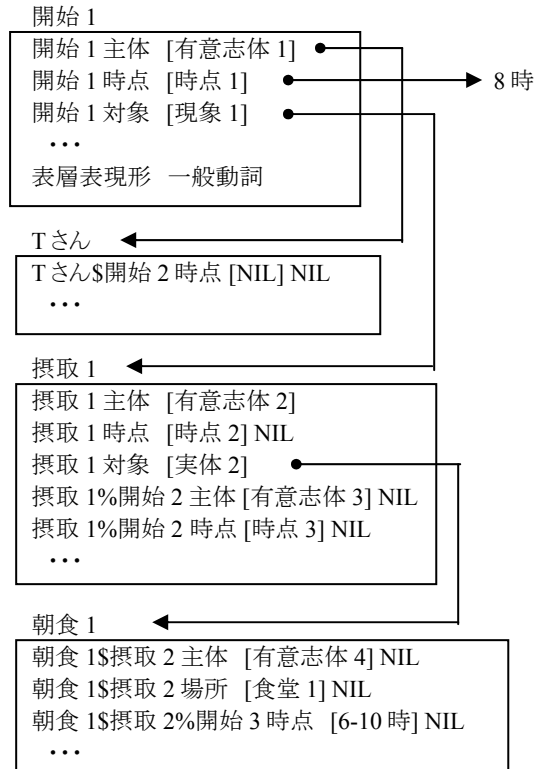


図5 「Tさんは8時に朝食を食べることを開始した」意味表現

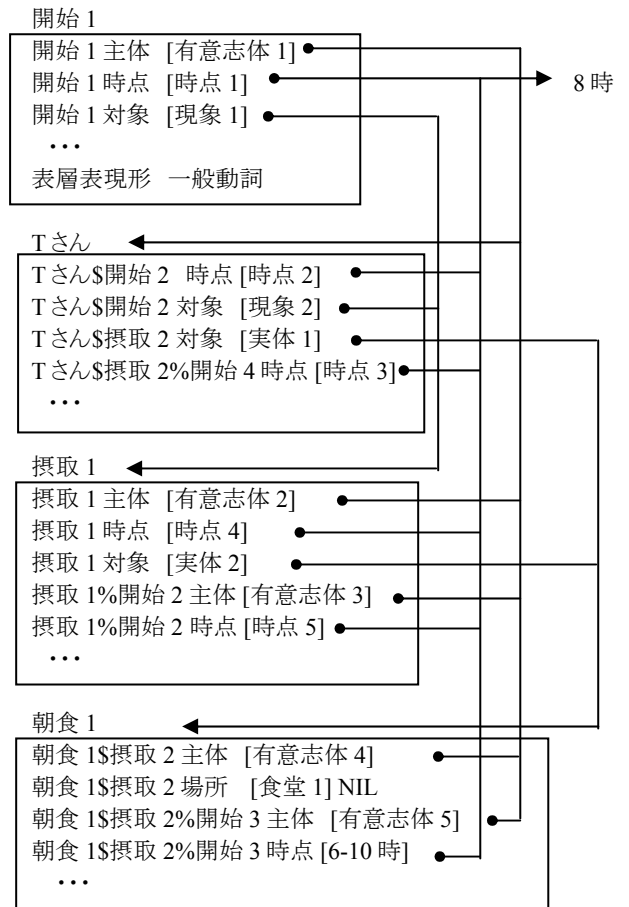


図6 意味解釈処理後の意味表現
位置づけ可能な属性がリンクで接続されている。