

構文意味解析における文脈の利用方法

The Method of using Dialog Context in Integrating Japanese Dependency Analysis

池ヶ谷 有希^{*1}
Yuki Ikegaya野口 靖浩^{*1}
Noguchi Yasuhiro高木 朗^{*2*3}
Akira Takagi麻生 英樹^{*3}
Hideki Asoh小林 一郎^{*4}
Ichiro Kobayashi小暮 悟^{*5*6}
Satoru Kogure近藤 真^{*5}
Makoto Kondo小西 達裕^{*5*6}
Tatsuhiko Konishi伊東 幸宏^{*5*6}
Yukihiro Itoh^{*1}静岡大学 理工学研究科

Graduate School of Science and Engineering, Shizuoka University

^{*2}(株)CSK システムズ

CSK Systems Corp

^{*3}産業総合技術研究所 情報技術研究部門

Information Technology Research Institute, AIST

^{*4}お茶の水女子大学理学部

Faculty of Science, Ochanomizu University

^{*5}静岡大学情報学部

Faculty of informatics, Shizuoka University

^{*6}静岡大学創造科学技術大学院

Graduate School of Science and Technology, Shizuoka University

This paper describes how to use contextual information in parsing and semantic analysis. Utterances in a dialog involve various ambiguities, which should be disambiguated by contextual information. Our previous works proposed a framework for semantic representations in which the meaning of any predicate is converted into a semantic representation permitting only a single type of predicate (an identifying predicate "aru") and the contextual information is stored as a set of "attribute-value" pairs. Accordingly, we can compare the meaning of utterances by referring to the attribute-value pairs. Our system generates as many candidate trees as possible and chooses from the candidates by referring to the context. The process is applied each time a new word is introduced into a tree. We take dialogs of hotel search and reservation for example, and explain how our proposals disambiguate potentially ambiguous utterances.

1. はじめに

言語処理技術に応用したシステムでは、文脈の情報を考慮した解析が必要となることが多い。例えば、文章要約システム、機械翻訳システム、対話システムなどにおける文の解釈は、以前に言及された内容を踏まえた解釈を行うことが望ましい。だが、これまで文脈^{*1}の情報を効果的に利用する方法は十分に議論されてきたとはいえない。また、文脈をどのような形式で構成すべきかということも、明確な方法は確立されていない。

文を解析するための文脈情報は、単語、構文パターン、意味内容の3つに大別できる。これまで単語や構文パターンは文を解析するための文脈情報として用いられ、ある程度効果があることが確認されている[那須川 95]。一方、文脈の意味内容や知識を参照して文を解析する研究はほとんど行われていない。その理由としては、意味表現が表層文の多様な依存構造に直結して、極めて多様に変動してしまうことが挙げられる。同じ意味を表す同義文であっても、表層構造が異なると意味表現も異なる構造をもつため、着目している文と同一の意味を持つ文や、関連する意味を持つ文が文脈中に存在しても、単純な意味表現のマッチングでは、そうした文(の意味表現)を探ることが出来ない。そのため、文脈情報として意味表現を利用しても構文パターンを利用した場合と比べ、それほど大きな効果は期待で

きない。

このような問題を回避するためには、文の意味内容をできるだけ表層の表現によらない形で抽出できることと、複数の文や節に分けて表現されている意味内容を統合して意味理解できる意味表現形式が必要になる。このような条件を満たす意味表現形式の一つとして、高木の提案する意味表現[高木 04, Takagi06,]がある。

本稿では、文を構文解析・意味解析する際に、この意味表現形式で文脈や知識を構成しておき、意味表現同士の比較を行うこと文脈の意味内容と解析している文の意味内容を比較し解析に役立てる枠組みを提案する。^{*2}

2. 文脈の構成方法

2.1 意味表現

一般に、対象とする分野を限定することで、単語に関しては、どの語がどういう概念に属し、どのような属性を持ちうるのかということが定義できる。また、語と語の関係についても、全体-部分関係、同義関係、上位-下位関係、属性-値関係など整備することが可能である。

しかし、ある意味の語とある意味の語が依存関係で接続した場合には、その句全体の意味を形式的に評価することはできない。例えば、一般的な述語形式の意味表現を用いた場合であっても、意味内容を比較するためには、入力文構造の数と同等の数の解釈ルールが必要となる。これは、入力文を格フレームや意味ネットワークで表現した場合であっても同等で、語と語の

連絡先: 池ヶ谷有希, 静岡県浜松市城北 3-5-1, TEL:053-478-1592, E-mail:cs7005@s.inf.shizuoka.ac.jp

^{*1} 文脈とは、第一義的には文の脈絡を意味し、単語や文の間の論理関係の続き具合を指す。本稿では、ある文節や語に着目している場合に、同一文内の(文頭側の)語・句・節の解釈結果と前文以前に入出力された語・句・節・文の解釈結果をあわせたものを文脈と呼ぶ。我々の提案手法は、解釈結果は句や節が直接主張する内容とそれに関連するシステムの知識を統合した意味表現で表される。従って、文脈を参照する際に、システムの知識も同時に参照することになる。

^{*2} 我々は、文脈の利用方法の検討する状況として、ホテル検索予約対話を想定している。対話では一つ一つの文が比較的短く、一つの節で表現可能な情報でも前提節と主節の複文の形で表現されたり、複数の文や節に分けて表現されることが多い。従って、対話文の解析においては、文脈の意味内容や知識を参照することが重要となる。

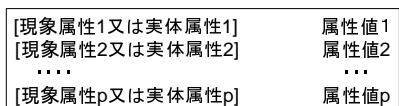


図 1: 意味表現の基本形

間の依存構造の解釈ルールを出現した数だけ記述することになる。そこで、見かけ上、依存関係の表現を出来る限り排除し、排除出来ない依存関係部分に関しては、表現しようとする意味によって変動しない共通の概念とそれが構成する依存関係のみを残す。また、意味表現同士で意味の比較を行なう際には、形式的に依存関係の解釈が可能な形式であることが望ましい。

このような条件を満たすために、現象を記述する文、実体を記述する文が、断定の述語「ある」を用いて言い換えが可能であるという性質を用いる。動詞句(連用修飾成分+述語)を「現象属性=値」、形容(動)詞句+名詞、連体助詞句+名詞を「実体属性=値」の形式で表現し、現象属性、実体属性を1つの属性名詞で表現する。これによって、修飾成分がP個あった場合述語の意味表現、名詞の意味表現は、図1のようになる。この意味表現形式では、文脈の参照は既に生成された意味表現の「属性=値」と着目している「属性=値」を比較することである。そこで、属性および値の比較を行うために、属性概念、実体概念、現象概念について、上位下位関係、全体部分関係に基づいて概念階層を準備を行う。

2.2 文脈の参照方法

この意味表現では、「属性=値」同士を比較することで意味内容の判別が可能である。「属性=値」同士の関係は次のように分類することができる。

- () 属性同士が一致し、値については両者が一致するか、上位下位関係にあるか、どちらか一方がNILである場合
- () 属性同士が上位下位関係にあり、値については両者が一致するか、上位下位関係にあるか、どちらか一方がNILである場合
- () 属性同士が主属性/副属性の関係*3にある場合
- () それ以外

() は、属性同士が同じで値についても矛盾がないため、全く同等の内容を述べていると解釈できる場合である。図2は一方が「駅前にあるホテルはAホテルです」と発話した後に、予約の話題に移り「期間はどのぐらいですか?」と宿泊期間を問い合わせたのに対して、もう一方が「駅前のホテルに3日間泊まりたい」と答えた様子を意味表現で示している。ここでは、「駅前のホテルに3日間泊まりたい」の「駅前のホテル」に対し、「ホテル」の意味表現中で「駅前」を値として取り得る属性を探して「存在場所が駅前である」と解釈して作られる意味表現「存在場所=駅前」と「駅前にあるホテルはAホテルです」の意味表現中の「存在」に含まれる「存在場所=駅前」が一致しており()に該当する。

() は、属性概念の意味体系において属性が上位下位関係にあることとそれらの値に矛盾が無いことに基づいて、全く同等の内容を述べてはいないが互いに整合性のあることを述べていると判定できる場合である。図2では「期間はどのぐらい

*3 主属性/副属性の関係とは、ある属性の値が他の属性の値により決定される場合である [池ヶ谷 02]。

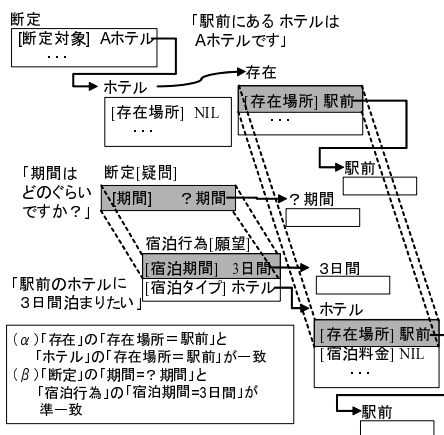


図 2: () , () を含む意味表現の例

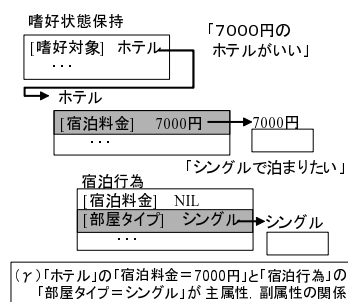


図 3: () を含む意味表現の例

ですか」という文の意味表現における「期間=?期間」の「期間」が「駅前のホテルに3日間泊まりたい」という文の意味表現中の「宿泊行為」の「宿泊期間」と上位下位関係にあるため、()に該当する。

() は、主属性と副属性の関係にある場合で、値が互いに影響し合うため文脈と関連のある事柄を言及しているとみなすものである。図3は「7000円のホテルがいい。シングルで泊まりたい」という入力を解釈した様子を示している。ここでは、第1文の意味表現中の「宿泊料金=7000円」と第2文の意味表現中の「部屋タイプ=シングル」があらかじめ用意した主属性/副属性のリスト中に存在することから()に該当する。

() は、()~()のいずれにも該当せず「属性=値」の関係だけからは直接意味的な関連性が見出せない場合である。

これらのうち「属性=値」同士が()~()の関係にある場合には文脈との整合性が高いとみなせる。

3. 解析の枠組み

3.1 解析方法

構文解析と意味解析は、それぞれ別個の処理と捉えるのではなく、1語ずつ構文構造を決定する度に意味解析を行い、意味表現を生成していくことが望ましい。これは、構文解析を単独で行い、あらゆる可能な構造をつくり出し、それを意味解析の段階で絞り込むという方法では、構文解析結果の解の数が膨大になり効率が悪くなるためである。ここでは、構文解析と意味解析を同時に行い、文脈を参照しながら意味表現を生成して

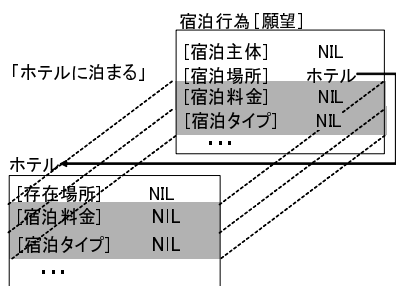


図 4: 「ホテルに泊まる」の意味表現

いく方法について説明する。

解析はスタックを用いて、可能な構文構造を全て生成し、その意味表現を生成する。着目している語がそれ以前の生成された部分木意味表現のヘッドの語と係ると判定された場合^{*4}に、意味解析を行う。意味解析は、係る語が受ける語のどの属性に格納されるのかを、意味素性に基づく選択制限によって決める。格納される属性が複数存在する場合には、その全ての候補を生成する。

着目語を処理して「属性=値」が生成できた場合、それと文脈の意味表現内の「属性=値」が 2.2 節の () ~ () の関係にあるかを調べる。文脈は基本的に次のような順序で参照する。

1. 同一文内の意味表現

- (a) 着目語を修飾する語
- (b) 着目語を修飾するほかの語
- (c) 主節の提題表現
- (d) 前提節

2. 前文以前の意味表現

このような順序で参照を行う理由について述べる。この意味表現において定義されている属性には係りが圧縮されている。つまり属性が定義されている場合はその元となる文が存在することを表している。例えば「ホテル」には「宿泊料金」や「宿泊タイプ」のような属性が定義されている。これは「ホテル」が「宿泊」現象に対して特定の連用修飾成分となりうることを表す。そこで「ホテルに泊まる」という文で「ホテル」と「泊まる」の「属性=値」を比較した場合には、() ~ () の関係にある「属性=値」が存在することを理由に、係る可能性が高いと判定できる(図 4)。よって、最初に (a) として修飾語内の「属性=値」を参照する。次に、(b) のように同じ語を係り先とする語に含まれる「属性=値」を調べる。これは同一の語と係りをもつ語同士には、関連する現象属性が定義されている可能性が高いためである。修飾語については、提題助詞により提題化されたり、前提節のなかで表現されることがある。そこで、(c), (d) の順に参照を行っている。

同一文内の文脈の参照手順については、着目語と修飾語のタイプにより、参照方法が多少異なるため、その順序について表 1 に示す。

文脈を参照した結果、() ~ () の関係にある「属性=値」が発見できた場合には、意味表現候補にスコアを与える。このような処理を文末まで行い、正解可能性の高い候補を残しながら

*4 意味表現は品詞情報を有しており品詞規則に基づいて係りを決めている。

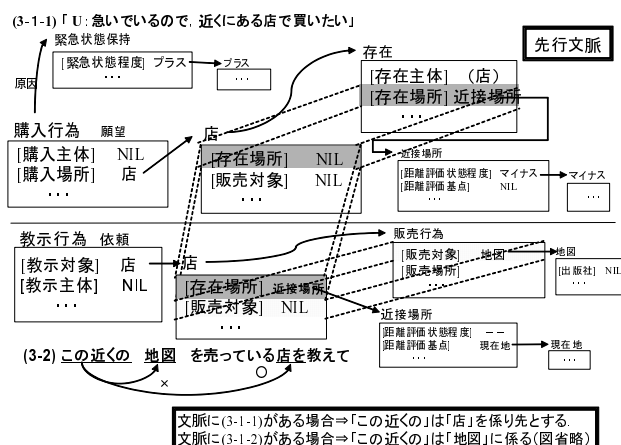


図 5: 連体助詞句の曖昧性を解消する例

ら部分意味表現木を生成していく。生成した意味表現は全て文脈として蓄積を行う。

3.2 解析事例

ここでは連体助詞句と名詞の係りにおいて、文脈に重なる「属性=値」を発見することで、正しい係り先を選択する事例を挙げる。

(3-1-1) 「急いでいるので、近くにある店で買いたい。」

(3-1-2) 「この近くの地図が欲しい。」

(3-2) 「この近くの地図を売っている店を教えてください。」

(3-2) は「この近くの」の係り先候補として「地図」と「店」が考えられ曖昧性を含む文である。しかし、文脈に (3-1-1) がある場合には「この近くの店」、(3-1-2) がある場合には「この近くの地図」という係りを考えるのが妥当である。図 5 では、文脈に (3-1-1) の文が存在する場合に、(3-2) の文で連体助詞句「この近くの」の係り先として「店」を選択する様子を示している。(3-1-1) では、「近くにある店」の「店」の中の「存在場所」属性の値が連体修飾節中の「存在場所=近接場所」の「近接場所」であると判断でき、(3-2) の「この近くの」が生成する「存在場所=近接場所」と () の関係にある。これによって、先行文の文構造が解析対象文と同じ連体助詞句ではないにもかかわらず「この近くの」が「店」を修飾するという可能性を高く見積もり、「この近くの店」という遠方への係りを優先的に選択することを可能としている。

4. 実験

4.1 実験方法

実験の目的は構文・意味解析において文脈の情報を効果的に利用できていることを検証することである。この枠組みでは、最終的な出力は意味表現であることから、文脈を利用する場合としない場合で、意味表現の生成精度がどの程度向上するかを調べた。また、構文構造(この意味表現からは、文の構文構造をどう解析したか抽出可能)についても解析結果についても調査した。

実験データには、情報系学部、大学院の学生 23 人から収集したホテル検索予約をタスクとする対話データを用いる。対話は、ゲスト役とオペレータ役の 2 人で行い、ジェスチャーな

着目語	修飾語句	参照手順
述語	格助詞 + 名詞句	1. 述語内の「属性 = 値」と名詞句内の「属性 = 値」 2. 述語と名詞句から生成される「属性 = 値」と述語に係る他の名詞内の「属性 = 値」 3. 述語内の「属性 = 値」と提題助詞で述語に係っている名詞句 N 内の「属性 = 値」 (但し、この名詞句は述語内のいずれかの属性の値になりうる名詞概念であること) 4. 述語内の「属性 = 値」と述語に係る前提節中に存在する名詞 N 内の「属性 = 値」 (但し、この名詞 N は述語内のいずれかの属性の値になりうる名詞概念であること)
	副詞句, 連用形 形容 (動) 詞句	述語内の「属性 = 値」と副詞句や連用形形容 (動) 詞句がもつ「属性 = 値」
名詞	連体修飾節	(連体修飾節内の述語内の値が空の属性 E1 へ着目語の名詞が格納できることを条件とする) 1. 名詞内の「属性 = 値」と述語の「属性 = 値」 2. 述語内の E1 以外の属性値の値である名詞内の「属性 = 値」と E1 3. 述語内の値が空の「属性 = 値」と主節の述語で提題化されている名詞内の「属性 = 値」 4. 名詞内の「属性 = 値」と名詞の修飾する述語へ係る前提節中の名詞内の「属性 = 値」
	連体助詞 + 名詞句	1. 一方が属性名詞で他方が (部分) 実体名詞の場合, (部分) 実体名詞内の「属性 = 値」と属性名詞 2. 一方が値名詞 (N1 とする) で他方が属性名詞 (N2 とする) の場合, 値名詞からそれを値として取り得る属性を調べ, 属性名詞 N2 による「属性 (N2)=NIL」と N1 を値とする「属性 = 値 (N1)」
	連体形 形容 (動) 詞句	連体形形容 (動) 詞句内の「属性 = 値」と名詞内の「属性 = 値」

表 1: 同一文脈の参照手順

ど音声情報以外の伝達は出来ないように非体面の環境で行ったものである。この対話データ (対話数 80, ゲスト発話数 1886) に対し, 話し言葉を書き言葉に直し, 言い淀みや言い直しの削除, 方言の訂正等を行い。さらに, 文節数が 2 以下の文を除いた 680 文 (平均形態素数 11.7) を実験データとした。

意味表現生成精度及び構文解析結果については人が目で判定を行った。与える文脈は, 解析対象とする文から遡り, 提題格で言い立てられた語が存在する範囲としている。

4.2 実験結果

意味表現生成精度は, 文脈の参照無しで 59.4%(404/680), 文脈を参照することで 64.3%(437/680) に向上し, 少なくとも文脈が正しく解釈できれば 71.2%(484/680) まで向上し。文脈を利用する効果があることが分かった。これは, 文脈内の「属性=値」の情報を参照することで正しい意味表現が生成できたためである。

一方, 正しい意味表現が生成できなかった主な理由は, 解釈に必要な意味表現が文脈上に存在しなかったことであった。これは, 対話が進んでいく場合には, 全ての内容が文として表現されないために起こるものである。

構文構造については, 文脈の意味表現を参照することで 80.1%(545/680) から 82.9%(564/680) へ約 3%向上した。これについても, 文脈の表層の情報ではなく, 文脈の関連のある「属性=値」の情報を優先することで文脈にある構文構造を選択できたためである。

解析に用いている辞書や概念階層はホテル検索用に作りこまれたものだが, 文脈を参照することで解析の精度が向上することで, 提案する文脈の利用方法が有効であることを示している。

5. まとめ

本稿では, 一定の意味表現形式で文脈を構成することを想定している。しかしながら, 構文・意味解析における文脈利用

の枠組みや, 文脈の参照順序の考え方は, 意味表現を拡張した場合でも, 共通に利用できるものであると考えている。

また, ここで提案した手法は, 文頭から順に逐次意味表現を生成していく方式をとっている。従って, 音声での入力に対し, 聞き取れた順に文脈に適合する意味解釈結果を生成することができ, 音声入力向けの構文・意味解析システムとして適した枠組みといえる。今後, さらに頑健性を向上させるためには, 文脈で直接言及されない情報についての知識をいかに整備して解析に役立てていくかという問題を検討したい。

参考文献

- [那須川 95] 那須川哲哉: 文脈を用いた曖昧性解消のアプローチ, 情報処理学会論文誌, Vol36, No.10, pp.2362-2370 (1995)
- [高木 04] 高木 朗, 麻生 英樹, 中島 秀之, 伊東 幸宏, 小林 一郎, 意味の対応付けと依存関係の解釈を考慮した対話意味表現, 情報処理学会研究報告, 2004-SLP-54-49, pp.289-294 (2004)
- [Takagi06.] A. Takagi, H. Asoh, Y. Itoh, M. Kondo, I. Kobayashi: "Semantic Representation for Understanding Meaning Based on Correspondence Between Meanings", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol.10, No.6, PP.879-912 (2006)
- [池ヶ谷 02] 池ヶ谷有希, 野口靖浩, 鈴木夕紀子, 伊藤敏彦, 小西達裕, 近藤真, 高木朗, 中島秀之, 伊東幸宏: 文脈への意味の位置付けにもとづくホテル予約対話システムの構築, 言語・音声理解と対話処理研究会資料, SIG-SLUD-A202, pp.63-70 (2002)