

# 対話文脈を利用した構文・意味解析手法の検討

## Integrated Japanese Dependency Analysis Using a Dialogue Context

池ヶ谷有希\*1  
Yuki Ikegaya

野口靖浩\*1  
Yasuhiro Noguchi

鈴木夕紀子\*1  
Yukiko Suzuki

伊藤敏彦\*2  
Toshihiko Itoh

小西達裕\*1  
Tatsuhiko Konishi

近藤真\*1  
Makoto Kondo

高木朗 \*3\*4  
Akira Takagi

中島秀之 \*4\*5  
Hideyuki Nakashima

伊東幸宏\*1  
Yukihiro Itoh

\*1 静岡大学  
Shizuoka University

\*2 北海道大学  
Hokkaido University

\*3 (株)CSK  
CSK Corporation

\*4 産総研サイバーアシスト研究センター  
Cyber Assist Research Center, AIST

\*5 公立はこだて未来大学  
Future University Hakodate

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. In our previous study, we proposed a framework for semantic representations that circumvented problems of interpreting dependency structures. Within the proposed framework, the meaning of any predicate is converted into a semantic representation which only permits a single type of predicate: an identifying predicate "aru". Proposals to be made in this paper are based on this semantic representation. We aim at solving three kinds of ambiguities: structural ambiguity, lexical ambiguity and ambiguity of attribute selection. The last ambiguity arises from semantic analysis based on our frame work for semantic representations. We take dialogs of hotel search and reservation for example, and explain how our proposals disambiguate potentially ambiguous utterances.

### 1. はじめに

自然言語文には、構造的な曖昧性や意味的な曖昧性が含まれている。これらの曖昧性を解消することは、自然言語理解の重要な課題である。特に、対話のやりとりの中で入力される文には、入力された一文だけに着目しては、曖昧性が解消できない文が多い。また、構造的に単純な文であっても、文脈に応じてその意味が変化する場合も多く存在する。これらの文に対しては、文脈情報を利用して、曖昧性を解消する必要がある。

これまで、文脈情報を利用して自然言語文の曖昧性を解消する研究として、文脈内における語の振舞いに着目した手法がある [那須川 95]。この手法は文脈内の語に対して、1箇所でも係り受けが決定できた場合には、その係り受けを文脈中の同じ語の係り受けが決定できない箇所に適用し、構造的曖昧性を解消する。

しかし、文脈情報を利用して構造的な曖昧性を解消しても、意味的な曖昧性は依然残ったままである。これまで意味的な曖昧性を解消する手法はほとんど提案されてない。そこで、今回、文脈情報を利用して構文的な曖昧性と意味的な曖昧性を解消する枠組を提案する。

文の意味的な曖昧性を解消が困難な理由は、先行文脈及び知識との意味の比較をどのように行うか、またそのために、どのように表現形式で知識等を記述しておけばよいのか明らかになってないためである。我々は、語と語の依存関係の意味を属性概念の中に繰りこみ、任意の述語で構成される節の意味を断定の述語「ある」で統一して表現する意味表現形式を提案してきた。この表現形式では、句や節の意味は「属性値」の組で表現でき、文脈、知識を全てこの形式で構成すれば、属性の比較という単純な処理で先行文脈や知識との意味の比較を行うことができる [Noguchi 02]。我々は、これまでホテル検索予約という対話ドメインでこの意味表現形式の有効性を示してき

た [池ヶ谷 02]。今回、提案する対話文脈を利用した構文・意味解析手法も、このドメイン内で検証を行っている。

### 2. 対話文の曖昧性

対話においては、先行文脈の意味内容が前提となって、文の入力が行われる。そのため、1文だけでは正しい構文構造が決めづらい場合が存在する。例えば、

「沖縄の料理がおいしいホテルを探して」

のような文は「沖縄の」の係り先が「料理」なのか「ホテル」なのか判別することができない。どちらに係るかは先行する文脈に依存する。先行文脈において、「沖縄の料理」に関して言及をしていれば、「沖縄の」は「料理に」係るであろうし、「沖縄にあるホテル」に泊まるということが言及されているならば「沖縄の」は「ホテル」に係る。

また、意味的な曖昧性も文も多く存在する。意味的な曖昧性を生じさせる理由は少なくとも次の2つが存在する。

(2-a) 語義の曖昧性

(2-b) 依存関係にある2語の意味関係が決定不可能

(2-a) は、例えば、「つく」などに見られる。「つく」には「着く」や「付く」などの語義が存在する。「Aホテルにつきますか」という文を解析する場合、「Aホテルに(例えば、朝食が)付きますか」とも「Aホテルに着きますか」とも解釈できる。この場合、この文だけでは判断がつかない。語義の曖昧性を解消する研究には、コーパスの中の統計情報に基づくものがあるが [福本 02]、対話の中で出現する文においては、文脈に応じた語義の選択をする必要がある。

(2-b) は、意味解釈を困難にする原因の1つである。例えば、「[名詞A]の[名詞B]」の場合、名詞A、名詞Bのそれぞれの意味は、単語辞書から推定できても、その意味関係を一概にきめることはできない。「太郎のメガネ」では所有、「大学の先生」では限定修飾、「ホテルの部屋」では全体一部分の関係にあり、表層表現から依存関係の意味を推定することがで

連絡先: 池ヶ谷有希, 静岡大学情報学部伊東研究室, 静岡県浜松市城北 3-5-1, cs7005@cs.inf.shizuoka.ac.jp

きない。意味的な曖昧性を生じさせる原因には、この他にも、数量子・否定・副詞のスコアの曖昧性や、比喩の解釈の問題等がある [長尾 96] が、本稿で扱う曖昧性は、構文的曖昧性と、(2-a)(2-b) の意味的な曖昧性である。次章では、これらの曖昧性を解消する構文・意味解析の方法を述べる。

### 3. 構文・意味解析の枠組

我々は構文解析と意味解析を別個の処理と捉えるのではなく、構文構造を決定する度に意味解析を行い、意味表現を生成するという方針をとる。その理由は、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る場合がある。また、構文解析結果から n-best 解を得た場合は、構文的情報だけから解を求めるため、文脈に適合する解が落ちてしまう可能性があるからである。構文解析と意味解析を融合すれば、構文解析の各段階（2 語間の係りを決定する時など）で文脈を参照し、文脈の意味内容と適合する解を選択しながら解析が行える。

構文解析は、品詞の組に基づく依存関係判定規則を用いて、形態素列を前から順にボトムアップで処理し、一語処理する度に、意味解析を行う。意味解析は、語の意味素性に基づく選択制限と、構文解析結果から意味表現生成処理を行う（図 1）。生成した意味表現は全て文脈として蓄積を行う。

意味表現は、先行文脈に蓄積された意味内容と現在処理中の文の意味内容とを、文体の差異にかかわらずに比較できるようなものでなければならない。ここでは、[高木 02] で提案されている意味表現を用いるものとする。この意味表現は、現象の記述を「現象属性 = 値」、実体の記述を「実体属性 = 値」という形式で表現する。これは、表層文に現れる多様な係り受け構造を、すべて均一的に「= (である)」を用いた表現に変換して記述したものであり、前述の条件を満たしている。なお、属性概念については、予め対話ドメイン内の範囲で、上位 - 下位関係、全体 - 部分関係などが整理されているものとする。

構文・意味解析を進めながら、例えば格助詞を伴う名詞句が述語に係ると判定しようとする場合や、形容詞や連体助詞句などが名詞に係ると判定しようとする場合などに、それらの係りから「属性 = 値」の表現を生成する。この際、修飾成分が被修飾概念のどの属性を限定しているかを決定する必要があるが、ここで複数の可能性が残ってしまう場合がある。このような場合には、可能な属性に対し、「属性 = 値」という形式で意味表現を生成し、そのすべてを候補として保持することとする。構文的な曖昧性をもつ文の場合、考えられるすべての依存構造を想定した場合の解釈が「属性 = 値」の形式の意味表現に変換され、可能な候補として蓄積されることとなる。

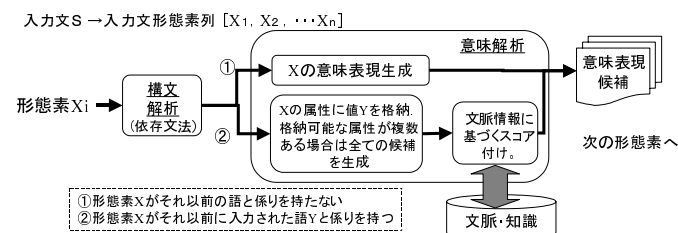


図 1: 構文・意味解析の処理の流れ

このようにして得られた複数の意味表現候補に対し、各々を文脈、知識と照らし合わせて、そのいずれかにマッチする「属性値」の組があるか否かを調べ、4.1 節で述べるスコアを求めておく。

このような意味表現を生成しながらの 2 語間の係り受け解析を、1 形態素ずつ取り込みながら文末にいたるまで行い、スコアの高いものを残す。処理過程で作成される入力文の意味表現候補数は、4.1 節のスコア付けに基づいて最大 6 4 個保持し、文脈は得点の高い方から 4 つを常に保持する。

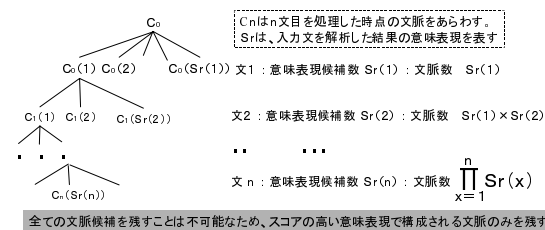


図 2: 入力文の解釈結果から文脈を構成する概念図

### 4. 曖昧性解消の方法

#### 4.1 スコア付けに基づく絞り込み

語と語の係りが決定する度に、「属性値」の意味表現を生成する。このとき、係る語が受ける語の属性に属性値として格納される度に次のような項目で得点付けを行い曖昧性の解消を行う。得点は一律一点とする。

- (1) 概念が先行文脈に既出
- (2) 固有名詞を除く実体概念について、先行文脈の実体概念と同定が可能
- (3) 「属性値」で重なるものが先行文脈に存在

(1) は、同じ見出しで、語義の曖昧性がある場合には、全ての語義にあたる概念を選択し、候補を生成するが、既に文脈に既出の概念については、加点をを行う。(2) は、例えば「ホテル」のような語が出現した場合に、先行文脈で「A ホテル」という具体的なホテルが同定出来た場合に、加点をを行う。(3) では、「属性値」で文脈を参照することで、先行文脈との意味内容の比較を行い、重なるものがあつた場合には、加点をする。文脈の参照範囲は、次のように提題の「は」で取り立てられている範囲で行う。ただし、それより前に取り立てられた語が、直前の取り立てられた語と全体 - 部分の関係にあれば、さらに遡る。一文の解析が終了した時点で、得点の高い候補を選択する。

#### 4.2 構文的曖昧性の解消例

「浜松の温泉がついているホテルがいい」

上記の文では、「浜松の」の係りが曖昧となっている。ここでは、先行文脈として、既に「浜松にあるホテルに泊まりたい」という文が存在しているとす。「浜松の温泉がついているホテルがいい」の形態素を前から順に解析すると、「浜松の」が「温泉」に係る場合と「ホテル」に係る候補が生成される。それぞれ温泉の属性である「存在場所 浜松」とホテルの属性である「存在場所 浜松」という意味表現を生成する。この時点では、どちらか一方を選択する手がかりが存在しない。次に、文脈情報に基づくスコア付けを行う。ここで文脈の参照を行い、先行文脈のホテルの「存在場所 浜松」と、生成した「存在場所 浜松」同士が重なる。そこで、(3) の項目から「浜松の」が「ホテル」に係る候補の得点上がり、最終的に候補として残る（図 3 参照）。

もしも、先行文脈に「存在場所 浜松」のような有効な手がかりとなる意味表現が存在しない場合には、「浜松の」が「温泉に」に係る候補と「ホテル」に係る候補が同じ重みの候補として、文脈に保持される。

この例では人間が考えた場合には、常識的に「浜松の」の係り先は「ホテル」と分かる。しかし、「浜松の一温泉」と「浜松の一ホテル」はどちらも十分に係り得る候補である。一般的な近いところにかかるという規則を優先させるが、文脈によっては、遠い係りを優先すべき場合が存在する。

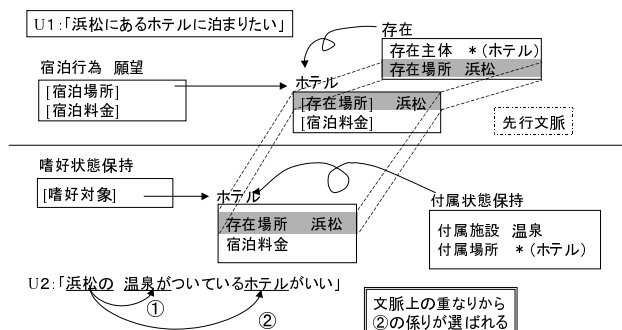


図 3: 構文的曖昧性を解消する例

### 4.3 意味的な曖昧性を解消例

「Aホテルは温泉がありますか」

この例では「ある」の語義の選択について述べる。「Aホテル」には「温泉」が付いているかどうかという知識が、「存在場所 浜松」、「付属施設 温泉」などの「属性値」の意味表現形式で書かれている。「ある」は「存在」と「付属状態保持」の語義が候補となるが、「温泉がある」は、それぞれ「存在主体 温泉」と「付属施設 温泉」という意味表現の候補が生成される(図4)。その結果、「付属施設 温泉」は、Aホテルのもつ知識と重なり、4.1の(3)の項目の加点がされ、「ある」の語義を決める。

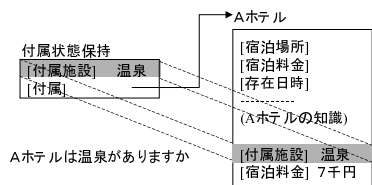


図 4: 語義の曖昧性を解消する例

次に、文脈を参照して文の意味的な曖昧性を解消する例として「いくらですか」という文について述べる。この文では、「いくら」がどのような料金の属性を限定しているのかが曖昧である。このとき、先行文脈で「宿泊料金」や「駐車料金」が直接言及されていれば、「駐車料金」あるいは「宿泊料金」を属性概念とする「属性値」の意味表現が文脈に存在するので、どのような料金の属性を限定しているのが限定できる。また、直接具体的な属性は限定されていなくても、現象が言及されていて(例えば、「宿泊現象」であれば「宿泊料金」属性が含まれているように)、該当する具体的な属性が含まれていれば、やはり曖昧性解消に利用することができる。

一方、「Aホテルに泊まりたい」のように、宿泊現象などが言及されている場合にも、Aホテルの知識中の宿泊料金属性と重なることにより、(3)の項目の加点がされ、何の料金であるかを定めることができる。

## 5. 実験

文脈情報を参照しながら構文解析・意味解析を行う効果を確かめる実験を行った。実験データには、情報学部関係の学生12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

曖昧な文の総数	54
人間がみれば、正しい解釈が可能な文	44
人間にも判断がつかない文	10

表 1: 実験データ

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20, ユーザ発話544)に対し、話し言葉を書き言葉に直し、言い淀みや言い直しは削除をした上で、1文内に構文的または意味的な曖昧性を含む54文を抽出した(表1)。ただし、この54文の中には、文脈が全く存在しない場面で入力された文は含まれていない。この中で人間が見て正しい解釈が可能な44文を構文・意味解析器に入力し、どれだけ曖昧性が解消出来たかを調べた。その結果、この44文中33文(75.0%)において、正しい解釈結果である意味表現を選択することができた。正しい意味表現が生成できなかった理由として「なるべく」や「ちょっと」等の副詞表現の係る先が曖昧な場合や、格助詞の係り先でも、文脈の情報と直接関連せず、その一文から決められない場合が挙げられる。文脈情報を利用しても尚、曖昧性が解消できない場合には、語と語の係りやすさや文の解釈がどのように行われやすいかを表す統計情報を利用することで、一層精度を高めることができる。

## 6. まとめ

構文解析と意味解析を融合させ、文脈情報文を利用して文の曖昧性を解消する枠組を検討した。意味表現で文脈、知識を構成することで、複雑な推論ルールを必要せずに一貫した方法で文脈及び知識と意味の比較ができるということが特徴である。今後、より多くの対話文から、曖昧性が生じる文のタイプ分類、対話文脈の場面などの分析を行い、対話文脈を有効に利用した構文・意味解析の手法を検討したい。

## 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, Vol.36, No.10, pp.2362-2370 (1995)

[福本 02] 福本文代: 語義の曖昧性解消のための最適な属性選択, 情報処理学会論文誌, Vol.43 No.1 pp.20-33 (2002)

[Noguchi 02] Y. Noguchi, Y. Ikegaya, A. Takagi, H. Nakashima, T. Konishi, T. Itoh, M. Kondo, Y. Itoh: "A Framework for Semantic Representations for a Natural pp Language Dialog System", *Proceeding of SNLP-Oriental COCOSDA*, pp.231-236 (2002)

[高木 02] 高木朗, 中島秀之, 伊東幸宏, 近藤真, 今仁生美, 三宅芳雄: 文脈への意味の位置付けを重視した対話意味表現, 言語・音声理解と対話処理研究会資料, SIG-SLUD-A202, pp.55-62 (2002)

[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)

[長尾 96] 長尾真: 自然言語処理, 岩波講座ソフトウェア科学, 第15巻, 岩波書店 (1996).