

## 対話システムのための概念素理解方式による対話音声理解

永井 明人、石川 泰

{nagai, yasushi}@media.isl.melco.co.jp

三菱電機株式会社 情報技術総合研究所

〒247 神奈川県鎌倉市大船 5-1-1

我々は、自由発話理解を目指して、概念素に基づく意味理解方式を継続して検討しており、一般性を失わずに言語制約力を高めるために概念依存関係を利用した音声理解手法を既に提案した。しかし、本手法は、対話システムに対する質問文としての文意妥当性が未検証であり、理解性能が不十分であった。そこで、対話の意味的機能である発話タイプに着目して、複数の観点から文意妥当性を評価する制約関係知識を検討し、これらを各文意候補の言語尤度を算定する基準として用いた。音声対話実験システム上で収集した自由発話音声で評価した結果、理解率が第一位で70.8%から80.0%へ向上し、また、第十位までの累積では88.5%から94.5%へと誤り率が半減した。これより制約関係知識の有効性を確認した。

## Speech Understanding by Concept-Driven Semantic Interpretation for a Spoken Dialog System

NAGAI Akito and ISHIKAWA Yasushi

MITSUBISHI Electric Corporation

Information Technology R & D Center

5-1-1 Ofuna, Kamakura, Kanagawa 247, JAPAN

We have been carrying on research on spontaneous speech understanding, and have proposed a method of concept-driven semantic interpretation based on conceptual dependency to strengthen linguistic constraint without losing generality of the method. Its performance, however, was not sufficient because semantic validity as a query to a dialog system was not verified. In this paper, constraint knowledges for verifying the semantic validity in plural viewpoints focusing on utterance types are discussed, and they are used to estimate linguistic likelihood of meaning hypotheses. Experimental results of spontaneous speech understanding showed that the introduction of the constraint knowledge improved system performance dramatically.

## 1 はじめに

一般の利用者の多様な発話表現を頑健に理解するために、自由発話の音声理解技術が必要である。英語の音声理解では、文全体の解析に失敗した場合、統語的制約力を弛緩させる方法 [1] や、部分的な構文木を利用する **partial parsing** [2, 3] が提案されている。一方、日本語では文節の並びが比較的自由と考えられ、発話文中の長い範囲で構文規則に良く従っている英語と異なる。このため、我々は、日本語では構文規則によって発話文を規定できないと考え、意味主導的なアプローチで言語的頑健さを得る概念素に基づく意味理解方式 [4, 5] の検討を進めている。

本方式の実用化への課題は、理解性能の向上とタスクポータビリティの実現である。音声言語理解では、音響処理から得られる複数の対立候補を言語的に検証して、意味的に妥当でない候補を棄却する。従って、理解精度向上のために制約力を強めることが必須である。しかし、制約知識を詳細化する上で、タスク知識の領域まで記述することは、容易に制約力を強められる反面、制約力の一般性がなくなる。

このため、我々は、本方式の一般化の枠組みとして、概念依存関係を利用した音声理解手法 [6, 7, 8] を既に提案した。これは、一般性を失わずに制約力を強めることを目的として、タスク知識を分離した後、一般的な言語知識である概念依存関係知識を導入し、概念素間の部分依存構造を構築していく概念素の統合手法である。しかし、本手法は、概念素から文意を構成する段階での文意妥当性に対する言語制約がなく、また、目的を達成するために発話が行なわれる対話システムについて考える場合、質問文としての妥当性が検証されていなかったため、性能が不十分であった。

そこで、理解性能向上のために、(1) 一般的な文意妥当性を観点とした言語制約と、(2) 対話質問文としての妥当性を観点とした言語制約、に関する制約関係知識を検討した。本稿では、この制約関係知識を言語尤度として適用し、自由発話音声データに対して性能評価した結果を報告する。

## 2 言語制約の適用方法

システムとの対話における意味的機能(発話タイプ)に着目して、複数の観点から文意妥当性を評価する制約関係知識を検討した。言語制約の適用では、受理/棄却の二者択一的な判定として用いると、言語的頑健さが失われる。従って、言語制約は、文意妥当性を言語尤度として評価するための基準として用いるべきである。ここでは、文意妥当性を侵害する言語的特徴にペナルティを与える規則を制約関係知識として設定し、各文意候補に与えられたペナルティの総和を言語尤度とする。以下では、まず、制約関係知識の検討のために再整理した文意の定義を述べ、次に、定義された文意の構成要素の意味的関連を検証する制約関係知識について述べる。

「大仏に行きたいんですが、長谷の駅からどうやって行けば良いんですか」

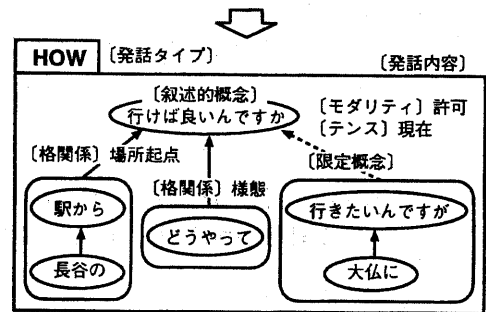


図 1: 文意の定義

## 3 文意の定義

文意表現を以下のように定義する(図1参照)。

文意：発話タイプ+発話内容

発話内容：叙述的概念+格要素概念+限定概念

叙述的概念：格構造+モダリティ+テンス

ここで、発話タイプは表1に示す分類を用いた。叙述的概念は、発話文の主述語であり、(1) 概念依存関係により表現された格構造と、(2) 主述語の文末表現で表わされる話者の判断・態度(表2)であるモダリティと、(3) 時制を表わすテンスの情報を持つ。格要素概念は、叙述的概念の格構造

に統合される概念素である。限定概念は、限定条件(順接、逆接、目的、条件、理由など)を表現する連用節の概念素である。

表 1: 発話タイプの分類

|      |             |         |
|------|-------------|---------|
| 要求表現 | 質問表現 (返答要求) | 真偽質問    |
|      |             | 確認質問    |
|      |             | 各 WH 質問 |
|      | 命令表現 (行為要求) | 命令文     |
| 伝達表現 | 事物宣言        | 名詞文     |
|      | 判断行為表明      | 述語文     |
|      | 応答宣言        | 肯定、否定   |

表 2: モダリティの分類

|   |
|---|
| 試行、依頼、命令、意思、勧誘、推量、経験、忠告、希望、許可、義務、禁止、不必要、可能、決心、伝聞、様態、受身、使役、やりもらい |
|---|

## 4 制約関係知識

妥当な文意をなすための言語制約として、以下に例示する制約関係知識を用いた。

### 4.1 一般的な文意妥当性の観点

【発話タイプ  $\leftrightarrow$  叙述的概念】叙述的概念を、(1) 動作 / 状態、(2) 継続 / 瞬間、の観点で分類することで、発話タイプとの関係を制約できる。

- 例：発話タイプ HOW (焦点は手段) の場合  
動作述語「探す」「予約する」「行く」→妥当  
状態述語「ある」「空いている」→ペナルティ

【発話タイプ  $\leftrightarrow$  格構造】WH 質問では質問の焦点が明確であるため、格構造への一般的な制約が可能である。焦点となる概念素が、(1) 移動性(場所の時間変化性)、(2) 変化性(概念素自身の時間変化性)、の素性を持つか、持たないかに応じて、次の制約が可能である。

- 例：発話タイプ WHERE (焦点は場所) で、焦点概念素に移動性がない場合、時間格にペナルティ(「そのホテル、明日どこにありますか」)。

- 例：発話タイプ WHAT (焦点は対象物) で、焦点概念素に変化性がない場合、時間格にペナルティ(「明日、玄って何屋ですか」)。

【発話タイプ  $\leftrightarrow$  モダリティ】発話タイプと、叙述的概念のモダリティの間には一般的な共起関係がある。例えば、質問的表現の HOW 質問では、許可、可能などのモダリティが主に用いられるので、それら以外に対してペナルティを与える。

【限定概念  $\leftrightarrow$  叙述的概念】叙述的概念が動作、状態のどちらを表現するかによって、限定概念が表現する限定条件との制約関係がある。例えば、動作動詞が接続助詞「て」によって連用節になった場合は、主述語が動作動詞であることが必要であり、それ以外の場合にペナルティを与える。

【概念素  $\leftrightarrow$  テンス】文意候補内にテンスの不一致がある場合、ペナルティを与える(「明日、行った」)。

【その他の一般的な制約関係知識】以下の場合にペナルティを与える。

- 一文一格の原則：格の重複がある場合  
例：「五時 から、駅 から、行けますか」
- 時間的な位置：話題格が文頭以外に存在する場合  
例：「良いお寺を、鎌倉は、探しています」
- 助詞の省略：助詞のない概念素が複数連鎖している場合  
例：「鎌倉、小松荘、明日」
- 意味限定を要する抽象的概念：依存概念素がない場合  
例：「鎌倉で、ところはありますか」

### 4.2 対話質問文としての妥当性の観点

【発話タイプ  $\leftrightarrow$  限定概念】対話システムに対する発話では、命令的表現や宣言的表現などにおいて、条件を指定する発話は生じにくいと考えられる。この場合に、条件節の存在自体にペナルティを与える。

【限定概念 ↔ モダリティ】情報を提供するシステムに対して、限定概念は情報を得るための意味限定の役割を持ち、主述語では話者の情報獲得に関するモダリティ(依頼、希望など)が用いられると考えられる。従って、限定概念が、相手(システム)への働きかけである命令、忠告、禁止、勧誘などのモダリティと共起した場合にペナルティを与える。

## 5 理解性能評価

以上検討した制約関係知識の有効性を評価するために、音声理解実験を行なった。

### 5.1 評価用音声データ収集

人-機械間の自由発話音声により評価するために、鎌倉観光案内の音声対話実験システム(図2)を構築した。実験システムは、(1)観光名所案内、(2)レストラン案内・予約、(3)ホテル案内・予約、について、各検索対象物の関連情報をテキストと静画像で提供する。主なシステム構成は、音響分析部、音声認識(文節スポッティング)部、意味理解部、タスクフィルタ(タスク知識による制約とシステム命令変換)部、対話管理部[10]である。認識語彙は1005単語であり、ワークステーション上で準実時間で動作する。

データ収集時の音声理解部は、後述の表4中の、ベースラインとなる(1)の理解方式を動作させたものである。従って、評価文には、誤認識時のリカバーのための繰り返し発話などが含まれる。

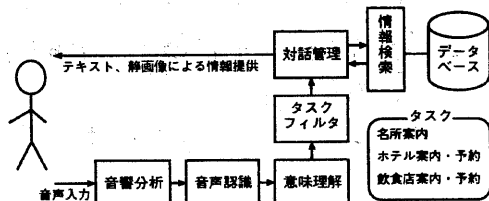


図2: 音声対話実験システム

男性6名による対話の収録をオフィス環境で行なった。被験者への指示は、画面表示内容の説明、音声入力時のタイミングなどにとどめ、発話内容に関する制約は一切せずに、「このシステムは、

鎌倉観光案内システムです。自由に質問して下さい。」とだけ指示した。従って、対話の目的設定も自由である。また、対話の終了も各話者の判断にまかせた。

得られた発話文を分析した結果を表3に示す。全般的に発声は丁寧で不要語発声はなく、発話文の言い回しも比較的平易であった。これらの収集データのうち、表中の92文を除いた260文を、意味理解部の評価音声データとして用いた。

表3: 発話の分類

| 全発話数         | 352文        |
|--------------|-------------|
| 未知語、文節内文法の違反 | 66文 (18.8%) |
| 始末端検出誤り      | 12文 (3.4%)  |
| 正解文節の脱落誤り    | 9文 (2.6%)   |
| 言い直し、発話中断    | 5文 (1.4%)   |
| 合計           | 92文 (26.1%) |

### 5.2 評価条件

音声認識部の文節スポッティングは、文献[9]の未知語処理なしの方式を用いた。認識に用いた文節内ネットワークは不特定話者の音節HMMで構成され、ネットワークの規模は語彙1005単語(自立語848、付属語134、不要語23)である。意味理解部では、概念素を102、概念依存関係を13種類用いた。制約関係知識は、40のペナルティ規則を用い、ペナルティ規則の違反の程度が軽微か重大かに応じて、1~3点をペナルティとして設定した。文意候補の言語的特徴がこの規則に違反するごとに、ペナルティを言語尤度として加算していき、音響尤度との線形和を文意候補の総合尤度とした。言語尤度を加算する比率は事後的に決定した。

### 5.3 評価結果

表4に文意理解率を示す。話者によって理解率に差が認められるが、制約関係知識による言語尤度を導入することで、第一位平均で70.8%から80.0%へ向上し、第十位までの累積では88.5%から94.5%へと、誤り率が半減する効果が認められた。

また、制約関係知識の導入の効果として、意味的に妥当性のある文意候補が理解結果の上位となっていることが要求される。そこで、正解以外の文意候補の文意妥当性について調査した。(2)の場合について、正解が第一位に得られなかった評価文(50文)に関して、第一位の対立候補の文意妥当性に関してまとめた結果を表5に示す。この結果より、50文中の34文(68%)が意味的に妥当であり、上位の理解誤りは音響性能に起因するものが大部分を占めていたことがわかった。

さらに、文意が妥当でなかった理解誤り(50文中の16文)を詳細に検討した結果、これらの理解誤りは、制約関係知識の抜けによるものがほとんどであり、容易に対処可能であること確認した。以上より、本稿で検討した制約関係知識は、(1)理解性能の向上に有効であり、(2)言語的な観点から妥当な文意候補を上位に浮上させる効果を持つ、といえる。

表4: 文意理解率(%):表中、(1)制約関係知識なし、(2)制約関係知識あり、を示す。

| 話者  |      | A   | B  | C  | D  | E  | F   | 平均   |
|-----|------|-----|----|----|----|----|-----|------|
| (1) | 1位   | 68  | 83 | 64 | 78 | 65 | 42  | 70.8 |
|     | ≤3位  | 79  | 90 | 73 | 83 | 80 | 67  | 80.8 |
|     | ≤10位 | 91  | 95 | 82 | 88 | 85 | 75  | 88.5 |
| (2) | 1位   | 86  | 88 | 73 | 78 | 73 | 58  | 80.0 |
|     | ≤3位  | 97  | 91 | 88 | 88 | 88 | 83  | 91.2 |
|     | ≤10位 | 100 | 93 | 91 | 90 | 93 | 100 | 94.6 |

表5: 理解誤りの文意妥当性:正解が第一位に得られなかった50評価文(表中の理解誤り)について、第一位の対立候補の文意妥当性を示す。

| 話者      | A  | B | C | D  | E | F | 合計 |
|---------|----|---|---|----|---|---|----|
| 理解誤り(文) | 10 | 7 | 8 | 11 | 5 | 9 | 50 |
| 妥当(文)   | 9  | 7 | 6 | 5  | 4 | 3 | 34 |

#### 5.4 考察

上記以外の対立候補に関し、意味的に妥当でない場合について、ペナルティが与えられていない

ものを調査した。その結果、言語的制約に関しては、以下の三つが今後の主な課題である。

#### 1. 表層的な言語制約に関して:

制約関係知識は概念を単位とする制約であり、助詞や単語間の、共起・係り受けに対する詳細な制約を行なっていない。このため、表層的に妥当性でない以下の例があった。

例:助詞「リラを、玄って、教えて下さい」

例:係り受け「これって、すごく、いくらで」

例:並立格の単語共起「鎌倉と日数」

この制約のためにはペナルティ規則の詳細化が考えられるが、発見的知識の詳細化は、言語知識の一般性の観点から望ましくない。言語表現は統計的現象であると捉えて、対話音声データにより、助詞の連鎖の出現確率を利用した方が良いと考える。同様に、単語間の共起(並立格、助詞「の」による連体修飾関係)などに関しても、一般的な統計的制約を取り入れていくべきだと考える。

#### 2. 深層的な言語制約に関して:

(1) 格依存関係の成立に関する制約が弱く、時間格、場所格などの任意格は全ての叙述的概念で無条件に成立可能になっているため、次のような時間格に違和感を感じる理解誤りがあった。

例:「門から来たら、銭洗い弁天に、五月、行けば良んですか」

この質問の焦点は道順の確認であり、道順を尋ねる場合には「五月」といった大きな時間単位よりは、「五時に」のような限定された時間の方が通常用いられると考えられる。このような制約を実現するためには、a. 質問の焦点に応じた語用的知識によるペナルティ、または、b. 単語共起の統計量の利用、などが考えられる。

(2) 表層表現が異なる同一の意味表現が複数あり、効率的な表現でない。これに対しては、最終的な意味表現として、同一意味表現は最大尤度を持つものだけを残すという、理解結果のバックギンク処理を行なうことで、N-best 解の質を向上させることができる。さらに、探索の観点からは、助詞や言い回しの表層的な相違はあっても、深層的な格が同じ概念素であれば、文意全体での深層

的な構成要素として一つに統合していく方法が、効率的な意味表現方法として有望である。

### 3. システムとの対話の観点に関して：

省略された格の補完については、以下のような制約が可能な場合があり、一般性を持った言語制約として導入できるかは今後の検討課題である。

×「レストランを(あなたが)探していますね」

○「レストランを(私が)探しています」

これに対しては、発話文が仮定している主体と、述語のモダリティとの関連に対して、妥当性を評価する一般的な制約関係知識を設定する余地があると考えられる。

さらに、本評価で対象外とした、未知語や言い淀み、言い直しを含む発話文に対する頑健さを実現することも、今後の課題である。未知語については、音声認識結果としての未知語の区間候補情報を利用し、未知語以外の区間を部分的に理解する方法が考えられる。言い淀み、言い直しについては、言い淀んでいる発話区間や、言い直す前の不完全な発話部分を飛び越えて理解する枠組みが必要である。

## 6 おわりに

文意の妥当性を評価する一般的な制約関係知識を検討して、言語尤度として導入した結果、語彙1000単語規模の自由発話理解で良好な理解性能が得られ、制約知識の有効性を確認した。今後は、さらに語彙を拡張し、一般的な統計的手法の導入と、自由発話文の表現を広く扱える理解の枠組みを開発し、本方式の理解性能の向上と、言語知識の一般化を図っていく。

謝辞：研究の機会を与えて下さった、三菱電機株式会社 情報技術総合研究所 音声・言語インタフェース技術部 部長 中島邦男氏に感謝致します。

## 参考文献

- [1] Stephanie Seneff, "A Relaxation Method for Understanding Spontaneous Speech Utterances," Proc. DARPA Speech and Natural Language Workshop, pp. 299-304, (Feb. 1992).
- [2] David Stallard and Robert Bobrow, "Fragment Processing in the DELPHI System," Proc. DARPA Speech and Natural Language Workshop, pp. 305-310, (Feb. 1992).
- [3] Paolo Baggia and Claudio Rullent, "Partial Parsing as Robust Parsing Strategy," Proc. ICASSP 93, pp. II-123-126, (Apr. 1993).
- [4] 永井明人, 石川泰, 中島邦男 "概念素に基づく意味理解における音声認識への統合手法," 情報処理学会 音声言語情報処理研究会, SIG-SLP 7-4, pp. 23-28 (Jul. 1995).
- [5] A. Nagai, Y. Ishikawa and K. Nakajima, "Integration of Concept-Driven Semantic Interpretation with Speech Recognition," Proc. ICASSP'96, Atlanta (U.S.A.), pp. 431-434, (May 1996).
- [6] 永井明人, 石川泰, 中島邦男 "概念素理解方式の一般化とタスク知識の適用手法," 日本音響学会 春季研究発表会 講演論文集, pp. 45-46 (Sep. 1996).
- [7] 永井明人, 石川泰 "概念依存関係による概念素統合に基づいた音声理解," 情報処理学会 音声言語情報処理研究会, SIG-SLP 15-22, pp. 125-130 (Feb. 1997).
- [8] A. Nagai and Y. Ishikawa, "Speech Understanding Based on Integrating Concepts by Conceptual Dependency," Proc. EuroSpeech'97, RHODES (GREECE), pp. 2747-2750, (Sep. 1997).
- [9] 花沢利行, 阿部芳春, 中島邦男; "文節スポッティングにおける未知語検出方式の改良," 日本音響学会 秋季研究発表会 講演論文集, pp. 179-180 (Sep. 1996).
- [10] 渡邊圭輔, 永井明人, 石川泰, "ユーザの行動目的の推定による協調的な音声対話制御," 情報処理学会 第54年全国大会講演論文集, pp. 2-233-2-234 (Mar. 1997).