

# 音声対話システムの移植性に関する考察 -観光案内システムとデータベース検索システム-

小暮 悟 伊藤 敏彦 中川 聖一

豊橋技術科学大学 情報工学系  
〒441-8580 豊橋市天伯町字雲雀ヶ丘 1-1

近年、音声認識技術や言語処理技術に関してはある程度確立しつつあり、実用化に向け開発が進んでいる。実用化などを考慮した場合、今までのような使い易さ、頑健性などに関する技術だけでは不十分であり、拡張性や移植性なども十分考慮する必要がある。そこで我々はこれまで開発を行ってきた富士山観光案内対話システムのタスクを富士山観光案内から三河観光案内に変更する作業を通して音声対話システムの移植性の問題点などを考察する。また、一般的なデータベースを用意するだけで音声によるデータベース検索ができる汎用性の高いデータベース検索システムを開発した。

## The study of portability for a spoken dialogue systems - a sightseeing guidance spoken dialogue system and a database retrieval system -

Satoru KOGURE, Toshihiko ITOH and Seiichi NAKAGAWA

Department of Information and Computer Sciences  
Toyohashi University of Technology, Tenpaku, Toyohashi, 441-8580, Japan

Recently the study of robustness and usability for speech recognition and language processing has been established, and speech recognition systems and dialogue systems have been developed to be practical use. But if these systems will become practical, it is important that not only those fundamental techniques but also the techniques of portability and expansibility should be developed.

Based on this consideration, we examined our system in portability by transferring the domain of the system from the Mt.Fuji sightseeing guidance to the Mikawa sightseeing guidance. Also we designed a domain independent platform of spoken dialogue system for database retrieval, and applied the platform to a literature retrieval system.

### 1 はじめに

近年、音声対話システムの研究が広く行なわれている [1]。最近では特に「ロバスト性」や「ユーザビリティ」に関する研究が、注目を集めている [2]。我々も、タッチスクリーンと音声による入力とグラフィカルな出力を実現した富士山観光案内システムを開発し、音声認識・言語理解・応答生成・マルチモーダルインターフェースのそれぞれについて、頑強で使い易いシステムを目指して、今まで開発を行ってきた [3][4][5][6][7]。しかし、これらのシステムはあるタスクやドメインに限定した上での開発であり、処理データにはタスクやドメインに依存している部分がかかなりあると考えられる。

現在では音声認識システムなどはかなり実用化に近いものも開発され、音声対話システムなどもこれから益々実用になると思われる。しかしながら、実際に新しい音声対話システムの構築には莫大なコストがかかり、そのため、これまで開発してきたシ

ステムを他のタスク用に変更したり、汎用的なシステムを開発することが重要になってくることは間違いない。実際、「移植性」や「拡張性」を重要視する研究も行なわれてきている [8][9]。

田中らは、「情報検索」にタスクを限定することで、データベース自体には依存しない汎用的な情報検索音声対話についての研究を報告している [10]。このシステムでは、アプリケーション開発者は、複雑な文法、語彙の設定を行なうことなく、手持ちのデータベースに対しての音声対話による情報検索を利用することが出来るが、ユーザの発話できる内容をメニューに提示することでユーザの発話を制限している。

秋葉らは、マルチモーダルインターフェースの汎用性に関する研究を報告している [11]。彼らは、MILES というマルチモーダル対話記述言語を開発し、ジャンケンや地下鉄乗り換え案内などいくつかの対話タスクを実際に試作している。

そこで、本稿では「音声対話システムにおける移

植性」に関する考察をシステム開発を通して行なった。最初に本研究で開発している既存の富士山観光案内音声対話システムを三河観光案内のタスクに変更する作業を通じて、音声対話システムの移植性の問題点などを考察する。結果としてタスクやドメインに依存したデータが多く存在し、比較的近いタスクへの変更は容易であるが、全く別のタスクへの変更には大変な労力がかかることが分かった。

そこで、上述の問題点を考慮し、一般的なデータベースを用意すれば音声対話によるデータベース検索ができる汎用性の高いデータベース検索音声対話システムの枠組を開発し、文献検索というタスクに適用した。

## 2 他の観光地案内音声対話システムへの移植

### -東三河観光案内音声対話システム-

本章では、既存の富士山観光案内対話システムをベースとして、タスクを富士山周辺の観光案内から東三河周辺の観光案内に変更することを行なった。この移植作業から、この富士山観光案内音声対話システムの移植性に関する問題や一般的な問題点を考察する。

### 2.1 知識データベース

東三河観光案内音声対話システムの開発において、最も重要なデータはもちろん東三河に関する知識である。この知識データを基に音声認識部や言語処理部の処理データが作られる。東三河に関するデータは、駅のインフォメーションやホームページから収集し、意味ネットワークのデータ形式で入力した。東三河に関する地名、観光地や宿泊施設などの98個の固有データが入力されている。時間的な労力としては、情報の収集に1週間(1人)、データの入力と修正に2日間(1人)を費やした。

### 2.2 音声認識部

東三河観光案内音声対話システムの音声認識部には、富士山観光案内システムで用いている連続音声認識システム SPOJUS-SYNO[5]を用いた。この音声認識システムは、文脈自由文法と辞書を変更すればどのようなタスクやドメインにも対応できるため、その汎用性は非常に高い。

ユーザの発話を認識する音声認識部は、音節 HMM、文脈自由文法の構文解析法、音声の「取り込み」「分析」「認識」を同時に行なう並列化アルゴリズム、及び、One Pass Viterbi サーチアルゴリズムに基づいたフレーム同期型の連続音声認識の統合アルゴリズムを基礎として構成されている。認識に用いる HMM (Hidden Markov Model) には、日本語の113音節を単位とする音節 HMM で、5状態4出力分布の単一連続分布を持つ、離散継続時間制御 HMM (DDCHMM) を用いている。更に、HMM は話者適応化を行なって、認識率の向上をはかっている。更に「間投詞」や「言い淀み・言い直し」の部分には、未知語処理に基づいた処理を施している[5]。

言語制約としてバイグラムなどの言語モデルを用いるバージョンも存在するが、東三河観光案内に関する発話コーパスなどが存在しない理由から、今

回は言語制約に文脈自由文法を用いている。最初に、知識データベースに存在する全ての固有名詞と名詞を抜きだし、音声認識用文法と辞書を作成した。辞書の方は富士山観光案内で使用していた辞書を基本に富士山周辺の固有名詞を除き、東三河観光案内に必要な固有名詞と名詞を辞書を追加した。音声認識用文法の方は、富士山観光案内対話システムで受理できる文型を基本とし、東三河観光案内では出現しない文型を取り除き、発話する可能性のある文型を追加して作成した。この文脈自由文法は自然な対話音声を確認するために、助詞落ちや倒置を含む文も受理できるように記述している。変更後、単語数433、規則数499となった。東三河観光案内システムに入力できる主要な25文でのテストセットパープレキシティは148.8となった。なお、この音声認識用文法作成に2日(1人)を費やした。

### 2.3 言語理解部

言語理解部は、音声認識部で認識された発話文字列から意味表現である意味ネットワークへの変換を行なう。ユーザの発話は音声認識部で認識され、5-best の認識結果が、言語理解部に送られる。この内、第1位の認識結果のみを文字列に変換し、変換した文字列に対して「形態素解析」「文節解析」「構文解析」「意味解析」「文脈解析」を行う。構文解析・意味理解部は、文節解析を行なった結果の係り受けに基づく文節間依存関係を解析する。解析の途中結果はチャートデータベースに格納され、一度行った部分解析結果を保存するようになっている。またこの時、助詞落ち、助詞誤り、倒置に対応するためにいくつかのヒューリスティクスを用いている。係り受け解析が成功した場合は、再帰的に文節の意味表現を組み合わせて文の意味を作る。意味表現は文脈的、背景的な観点からおかしなところがないかどうかチェックされる。これをフィルタリングと呼ぶ。さらに、ボトムアップの意味表現を得ることができない場合は、トップダウン的にキーワードによる意味の抽出を行う。パターンに記述された制約に適合する単語を探すことにより、全体の意味表現を得る[3]。

ドメインやタスクを変更する時、この言語理解部に必要な変更は、形態素解析部の形態素辞書の変更、構文解析部の係り受けルールと名詞の意味素性の変更、意味理解部のフィルタリング(常識)処理のルールの変更、キーワード処理のルールの変更である。

形態素解析には富士山観光案内音声対話システムと同様に京都大学長尾研究室で開発された JUMAN を利用した。富士山観光案内音声対話システムでは、JUMAN の接続ルール(接続可能性規則)を自然な発話及び誤認識結果にも対応するように変更し、辞書としては富士山観光案内で使用される固有名詞や単語を追加したものを使用していた。東三河観光案内では接続ルールの変更を行わず、富士山周辺の固有名詞を除き、東三河周辺の固有名詞と発話される可能性のある単語とその属性を追加した。

構文・意味解析部の係り受けルールに関しては、入力できる文型が富士山観光案内音声対話システムとほぼ同様のため今回の移植では変更を行っていない。しかしながら、構文解析部で使用する各単語の意味素性は、新たに形態素解析部で追加した単語に関して追加してある。意味解析部のフィルタリング処理は、意味解析部によって生成された意味ネットワークが、タスク・ドメイン的に許されない形である場合にリジェクトしたり、修正したりする処理

である。処理データは基本的に許されない意味ネットワークの型とマッチした場合の処理で構成される。キーワード解析処理は、構文・意味解析処理によって意味ネットワークが生成できなかった場合、入力された発話単語列中に存在するキーワードから文全体の意味を予測し意味ネットワークを生成する処理である。処理データは、キーワードとなりえる単語の意味素性と意味ネットワークの雛型から構成されている。この2つの処理に関しては、今回の移植がタスク的に観光案内という同タスクのため、データには全く変更を行っていない。

言語理解部の各データの修正・追加に1日(1人)を費やした。

## 2.4 応答生成部

応答生成部は対話制御部、問題解決器、知識データベース、応答文生成部から構成され、意味理解部によって生成された意味表現を受けとり、可能な限り協調的応答をする。

対話制御部は対話の流れ、文脈情報の管理、必要な情報の質問などを行なっている。対話制御部は、対話の流れを決定するためにユーザの対話の意図(焦点)を抽出する意図解析部に、ユーザの発話意味ネットワークを送る。意図解析部は、入力された発話意味ネットワークからユーザの対話の意図(焦点)と、ユーザの提示した検索条件を抽出する。次に知識データベースを検索する問題解決器に、ユーザの発話意味ネットワークと先ほど述べた文脈情報を入力する。問題解決器はそれらのデータをもとに知識データベースを検索する。この時、問題解決器はユーザの発話意味ネットワークに含まれている検索条件と文脈情報を活用して検索するが、もし検索結果が得られなかった場合は、検索条件の一部を同じ概念の検索条件に変更し検索をやり直す。もし再検索後、検索結果が得られたなら、代案としてその検索結果を対話制御部に送る。

対話制御部は獲得された検索結果を調べ、検索結果が多い場合にはユーザにデータの選択のための質問をする。検索結果や代案のデータ数が適当であるなら、検索された情報をユーザへ提示するために対話制御部は、応答文生成部へユーザの発話意味ネットワークと検索結果(代案結果も含む)を送る。

応答文生成部では、入力された発話意味ネットワークと検索結果からどのような形で応答すれば良いかを決定し、それに従い応答文意味ネットワークを形成する。それから応答文意味ネットワークを通常の文字列に変換し、ユーザへの応答として出力する。

応答生成部を他タスクへ変更する場合、変更する必要があるデータは、知識データ、意図解析用データ、知識データ検索用データ、応答文生成用データの4種類である。知識データは先に述べたように最初に構築する必要があるため、すでに構築済みである。

意図解析用データはその発話がどのような意図で発話された可能性があるかを示すデータベースであるが、今回の東三河観光案内では富士山観光案内との類似から変更は行っていない。

知識データ検索用データは、他システムへの移植において最も手間のかかるデータであり、知識データの構造を示すデータである。つまり、ホテルの概念から実際のホテル名を抽出するための方法(意味ネットワークの遷移順序)や、ある土地に存在する全ての物体を抽出するための方法などを記述したものである。これは知識データベースの表現法に非

常に依存しており、知識データベースの構造に変更があった場合、このデータは全て変更する必要がある。今回、以前の富士山観光案内の知識データベースで存在していた不満点を改良するために知識データベースの構造を少し変更したため、この部分を大きく修正する必要があった。

応答文生成用データは、応答文意味ネットワークと知識データベースから応答文に必要な助詞などを決定するためのデータである。知識データベースの構造に若干依存しているため、知識データベースの構造を変更によりこのデータも幾らか修正した。

知識データ検索用データ変更のために1週間(1人)を、その他のデータ変更に1日(1人)を費やした。

## 2.5 マルチモーダルインターフェイス部

マルチモーダルインターフェイス部は、「地図・写真などの表示」、「システムとの対話の途中経過表示」や「メニューへのタッチ入力、及び、地図などへの指示詞や指示代名詞を含んだユーザ発話」といった、マルチモーダルインタフェースを実現している。

マルチモーダルインタフェース、特に画面構成や使用モーダルなどは基本的にタスクやドメインに非常に依存しているため、他タスクなどへの流用は難しい。しかしながら、今回の移植に関してはタスク的に同一の観光案内であることから、使用モーダルや画面構成を変更する必要がないため、富士山観光案内音声対話システムのインターフェイス部は全く変更せず、メニューや地図・写真などの変更を行った。

地図、写真の準備とデータの変更に1週間(1人)を費やした。

## 2.6 まとめ

音声認識部は、文法と辞書を用意すればどんなタスクやドメインでも対応することができる。今回の移植においては、文脈自由文法を言語制御として用いた。しかし、自然な発話を受理するために非常に緩い言語制御となり、テストセットパープレキシティは約150にもなってしまった。富士山観光案内音声対話システムにおいて言語制御としてプログラムを用いた場合、パープレキシティが約10程度であった(文脈自由文法では約120)ことを考えると言語制御の能力として非常に低いものとなった。また、今回の移植においては移植元ドメインと非常に近いため、元の文法のほとんどを使用することができた。しかし、全く別タスクへの移植では、発話される内容が大きく変化し、文法の作成は非常に労力のかかるものとなることが容易に予想される。今後は、移植タスク用の発話コーパスからの言語モデル作成や、非常に一般的な大語彙言語モデルから特定のタスク用への適応(小語彙化)などが移植性を考慮した場合、現実的であると思われる。

形態素解析にはJUMANを用いており、形態素辞書の登録により他タスクやドメインへの変更は容易である。しかし、構文・意味解析に使用している係り受けルールと意味素性はドメインに依存したものである。そのため、タスクなどを変更する場合、係り受けルールと意味素性は全て付け直す必要がある。これは、意味素性に一般的な概念などを使用することである程度、汎用性を高めることができると思われる。フィルタリング(常識)処理やキーワー

ド処理は非常にドメインに依存したデータを使用しており、ドメインが変更になればそのほとんどは使用できない。自動的これらの処理データを獲得できる方法などを開発する必要がある。

応答生成部では、基本となる知識データベースの構造が非常にドメインに依存して記述され、また一般的知識とドメイン知識も分離されていない。そのため、知識データベースの構成(構造)が変わるとほとんどの処理データを変更する必要がある。他ドメインに移植する場合、知識データベースの変更に伴いほとんど全てのデータを新たに書き直す必要がある、その移植性は非常に低い。移植性を高めるためには、知識データベースを汎用的なデータとドメインに依存したデータに分離することが重要であろう [12]。さらに、知識データベースが意味ネットワークで表現されているため、知識データベースを作るのに専門的な知識が必要となり、知識のない人間には作成が難しい。汎用性を高めるためには、一般的なデータベースで使われている表やフレーム形式の知識表現が良いと考えられる。

マルチモーダルインターフェイスでは、ドメインに依存した部分が多量に多く、画面構成や使用するモーダルなどはタスクごとに考慮し直す必要がある。そのため、簡単に画面構成や使用するモーダルを変更できる汎用的なマルチモーダルインターフェイスの開発が重要であると考えられる。

今回の移植は富士山観光案内対話システムから東三河観光案内対話システムへの移植と言う類似部分の多い変更であったため、それほど大きな労力ではなかったが(合計で約30人日)、全く別のタスクへの移植の際は、ドメインに依存している処理データが多いため、変更するデータの量が多くかなりの労力になると考えられる。

### 3 移植性の高いデータベース検索システム

前章で富士山間港案内を三河観光案内に変更してタスク適応の問題点などをあげた。本章では前章の問題点を考慮して、移植性の高いデータベース検索システムの構築について検討した。

#### 3.1 使用データ

タスクを適用するのに必要なものを以下に示す。

- データベース
- データベースのフォーマット情報
- ユーザ発話文コーパス(対話例)

データベース検索システムのため、データベースがまず必要となる。また、データベースのフォーマット情報に従って、用意したデータベースのアクセスを行なう。さらに、データベース検索システムのタスク適応のために、ユーザ発話例の文集合が必要となる。システムの構築後に新しいデータを追加することは可能である。

#### 3.2 辞書の生成

ユーザ発話文を解析する際に必要な辞書は、形態素解析辞書と意味辞書である。

前述の富士山観光案内では、形態素解析に京都大学で開発されている形態素解析アプリケーション JUMAN 3.2 [13] を利用している。データベース検索案内システムでも同様に JUMAN3.2 を利用する。

形態素解析辞書のうち普通名詞辞書と固有名詞辞書は、データベースの項目や、ユーザ発話文コーパスから作成する。データベースのフォーマット情報には、データベースの各フィールドをどの辞書に登録するか情報が書かれており、そのデータに従って、辞書を作成する。また、助詞辞書には、「に関する」「のような」といった話し言葉で頻発する格助詞相当句を格助詞として前もって登録しておく。その他の辞書については、デフォルトのものを使う。

次に、意味解析時に使用する意味辞書は、ユーザ発話文コーパスに存在する動詞と形容詞を格とした格フレーム情報と、データベースとユーザ発話文コーパスに存在する名詞の概念情報から構成されている。格フレーム作成時に必要となる格と名詞間の接続関係、名詞の概念情報は、EDRの日本語動詞共起パターン副辞書、日本語共起辞書、概念辞書等 [14] から抽出する。

#### 3.3 意味表現の生成

前節により生成された形態素解析辞書と、意味辞書を用いて意味解析を行なう。前述の富士山観光案内で使用している意味理解部では、意味表現の統一化や省略語補完など、「富士山観光案内」というタスクに依存した処理が含まれているため削除し、それ以外の構文解析、意味解析、指示詞処理についてはそのまま使用する。

#### 3.4 文献データベース検索システムへの適用

実際のタスク適応の大まかな流れを図1に示す。アプリケーション開発者はデータベース、ユーザ発話の文集合を用意し、辞書生成ツールを実行することによって、システムへのタスク適応を行なう。ここからは、「文献検索」を例にして説明していく。

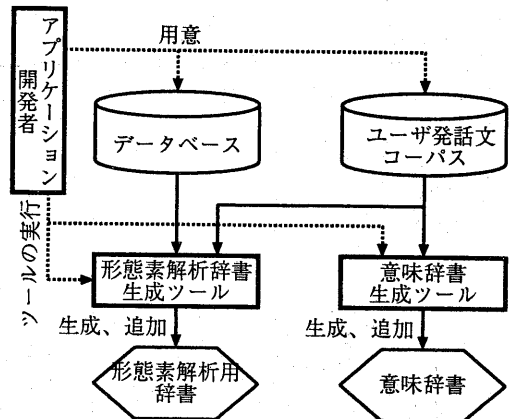


図1: タスク適応

TITL: 計算場における分散プロセスの準最適配置

AUTH: 上原, 稔 / 所, 真理雄

AUTE: Uehara, Minoru / Tokoro, Mario

CTIN: 情報処理学会論文誌

ABST: 「論文を執筆できる」ということは、大学教育における大きな到達目標のひとつであるに違いない。にもかかわらず、大學生にとって論文を書くことがどのような意味をもつか、また大學生は學生に対して論文執筆をどのように教育したらよいのか、といった基本的な問いに対しては、従来、正面から答えられることが少なかった。本稿では、論文執筆の原点は、他者との交流を通じて自らの主張を節度をもって根拠づけることであるとの認識から出発し、レポートと論文の相違点、論ずるに足る主題とは何か、論拠の示し方などについて述べ、さらに、執筆上のトラブルや悩みの対処の仕方などを論じた。

KYWD: 大学教育 / 論文執筆の意義 / 「論文が書けない症候群」 / 「知の技法」 / 「知の論理」 / 論文執筆のトラブル / 主題と論拠 / よい論文と悪い論文 / 論文審査

図 2: 文献データベース

### 3.4.1 データベースの準備

学術情報センターの文献検索システムの対話データ [15] から、電子情報通信学会誌、情報処理学会誌、日本音響学会誌、人工知能学会誌の日本語の論文情報 (3092 件) を実際に使用した。

次に、データベースのフィールド情報 (図 3) を用意した。このファイルは、データベースの各項目のデータの形式と、そのデータの処理の方法を定義している。例えば、図 3 の CTIN というフィールドを例にとる。これは、データベース (図 2) の、[CTIN:] という場所に学会の情報が書かれてあり、固有名詞の辞書に登録する必要があるということを示している。

### 3.4.2 形態素解析用辞書生成ツールの実行

データベース、ユーザ発話コーパス、フィールド情報を元に、形態素解析用の辞書を生成する。

著者についての例を図 4 に示す。同じように、キーワード、組織名についても辞書を生成していく。この例では、3092 件の論文の各データから、キーワード (5503 単語) とタイトルから抽出された名詞 (4008 単語) を普通名詞辞書に、名前 (姓 + 名、姓、6755 単語)、組織名 (4 単語) を固有名詞辞書にそれぞれ登録した。

```
S|TITL: |$|tit_j
R|AUTH: |,|/|$|aut_j1,aut_j2
R|AUTE: |,|/|$|aut_e1,aut_e2
S|CTIN: |$|ctin
R|KYWD: |,|/|$|keyword_japanese
$tit_j, SENTENSE
$aut_j1, WORD_REPEAT, 著者, 固有名詞
$aut_j1.$aut_j2, WORD_REPEAT, 著者, 固有名詞
$ctin, WORD, 学会, 固有名詞
$kywd, WORD_REPEAT, キーワード, 普通名詞
```

図 3: フィールド情報の例

### 3.4.3 意味辞書生成ツールの実行

次に、意味辞書を生成する。対話の文集合は、情報検索のユーザ発話文集合 (ユーザ発話 365 文) [15] を使用した。文集合に含まれる、動詞、形容詞を格とする格フレームの情報を EDR の辞書から抽出する。例えば、図 5 のような意味辞書を生成する。このような意味辞書 (及び形態素解析用の辞書) が生成されている状態で、「秋田成司さんの書いた論文はありますか」という入力に対して、図 6 のような意味理解結果 (意味表現) を得る。今回のユーザ発

データベース:

AUTH: 上原, 稔 / 所, 真理雄

AUTE: Uehara, Minoru / Tokoro, Mario

形態素解析用辞書:

(名詞 (固有名詞 ((見出し語 上原) (読み Uehara))))

(名詞 (固有名詞 ((見出し語 上原稔) (読み UeharaMinoru))))

(名詞 (固有名詞 ((見出し語 所) (読み Tokoro))))

(名詞 (固有名詞 ((見出し語 所真理雄) (読み TokoroMario))))

図 4: データベースからの形態素解析用辞書の生成

語文からは、形容詞 10 単語、動詞 33 単語について、それぞれを格とする格フレーム情報を抽出した。

```
(ある ((frame (object ((が) G-RONBUN) )))
(書く ((frame (object ((を) G-RONBUN)
                    (agent ((が) G-TYOSYA) )))
(論文 ((sem-features G-RONBUN))
(秋田成司 ((sem-features G-TYOSYA)))
```

図 5: 意味辞書の例

入力文: 「秋田成司さんの書いた論文はありますか」

意味解析結果:

(ある (FORM YN-Q) (NEGATION NIL)

(OBJECT (論文 ((OBJECT)

(書く (AGENT (秋田成司))))))

図 6: 意味理解の例

次に、この意味表現から「著者が秋田成司さんの論文が知りたい」というユーザの意図を抽出し (抽出する方法はデータベース検索というタスクに依存)、図 7 のような SQL 記述言語を出力する予定である。

## 3.5 その他のコンポーネント

ここまで意味理解部についての説明をしてきた。意味理解部についてはデータベース検索に関連したタスクについて汎用性を高くし、辞書にのみタスク依存の情報を持たせることによって汎用性の高いシステムを実現させることが出来ると考えている。ま

```
意味解析結果：
(ある (FORM YN-Q) (NEGATION NIL)
 (OBJECT (論文 ((OBJECT)
 (書く (AGENT (秋田成司)))))))
```

```
SQL 表現：
retrieve {
  SELECT: 著者, タイトル, 学会誌名
  WHERE: 著者 = "秋田成司"
}
```

図 7: SQL 記述言語

た、未知語の処理についても同時に検討する必要がある。

今後、音声対話システムを構築していく上で、音声認識部や応答生成部についても汎用性のあるコンポーネントを目標に構築する。

### 3.5.1 音声認識部

数千文のユーザ対話集合があれば、人手で作成した文脈自由文法よりもバイグラム言語モデルを文法とする方が認識率が高いということが分かっている [16]。しかしながら、新しいタスクに関する対話文集合を数千文集めることは困難であると考えられる。

そこで、「データベース検索」というタスクにおける文集合をあらかじめ数千文用意してデフォルトのバイグラムを求めておき、タスクを適用する時に、徐々にバイグラムをユーザ発話で再学習していく方法などが考えられる。

伊藤らは大量の既存データが存在する場合に、ある特定タスクの少量のデータがその大量の既存データと共通性があれば、その少量のデータを用いて N-gram の性能を改善できると報告している [17]。

### 3.5.2 対話制御・応答生成部

SQL による意味表現とデータベースが存在すれば、データベース検索部は特別なことをすることもなく実現できる。SQL を用いたデータベース検索のライブラリは多く一般に普及しており、また無料で公開されているものも多い。

ユーザが発話した検索条件が曖昧である場合にその曖昧性をどのように除去すべきか、またユーザの検索によって複数の検索結果が存在する場合にそのようにデータをユーザに提示するか、さらにその複数の検索結果をどのようにユーザに絞ってもらうかといった対話制御の問題を現在考察中である。

## 4 まとめ

本稿ではまず、既存のシステムである富士山観光案内を三河観光案内に変更し、対話システムの各コンポーネントごとにその移植性についての検討を行なった。今回の移植のよきな類似部分の多い変更では、それほどの労力を費やす必要はなかったが、ある程度異なったタスクへの移植にはさまざまな問題点があることが分かった。

それらの問題点を考慮して、まずは、汎用性の高い音声対話データベース検索システムを構築するための前段階として、意味理解部に注目し、その実現性について考察を行なった。また、音声認識、

応答生成部についても、今後さらに細かな設計を行なっていく予定である。

システムが構築された後にはそのシステムの評価を必要とするが、移植性を評価する際には、いくつかのタスクを実際に適応してみても、適応にかかった時間、出来たシステムの使い易さなどを総合的に評価する必要がある。

さらに、ユーザビリティを更に向上させるために逐次発話単位での対話や認識エラー時のユーザの振舞いを考慮した対話などについても検討していくつもりである。

## 参考文献

- [1] Elisabeth Maier, Marion Mast, Susann LuperFoy: "Dialogue Processing in Spoken Language Systems", ECAI'96 Workshop, Revised Papers, 1996.8
- [2] 熊本, 伊藤: 「WWW を介した対話システムとの対話における混乱の分析」, 情報処理学会誌 126-18, 1998.7 pp.97-104, 1998
- [3] 山本, 伊藤, 肥田野, 中川: 「人間の理解手法を用いたロバストな音声対話システム」, 情報処理学会論文誌, Vol.37, No.4, pp.471-481 (1996.4).
- [4] 傳田, 伊藤, 中川: 「マルチモーダル観光案内対話システム」, 人工知能学会誌, Vol.13, No.2 (1998.3)
- [5] 甲斐, 中川: 「冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価」, 電子情報通信学会論文誌, Vol.80-DII, No.10, pp.2615-2625 (1997)
- [6] 伊藤, 小暮, 中川: 「協調的応答を備えた観光案内音声対話システムとその評価」, 情報処理学会論文誌 Vol.39 No.5(1998.5)
- [7] S.Kogure, T.Itoh, S.Nakagawa: "A Semantic Interpreter for a Robust Spoken Dialogue System", ICMF'99, Proceedings, pp.II-61—II-66, 1999.1
- [8] Kaspar.S, Hoffmann.A: Semi-automated incremental prototyping of spoken dialog systems, Proc. ICSLP 98, Vol.3, pp.859-862(1998)
- [9] Brondsted.T, Bai.B, Olsen.J: The REWARD Service Creation Environment. An Overview, Proc. ICSLP 98, Vol.4, pp.1175-1178(1998)
- [10] 田中, 川原, 堂下: 「汎用的な情報検索音声対話プラットフォーム」, 情報処理学会 音声言語処理研究会, SLP-24-14, pp.97-104, 1998
- [11] 秋葉, 伊藤: 「スクリプト言語を用いたマルチモーダル対話記述の試み」, 情報処理学会 音声言語情報処理研究会, 98-SLP-23, pp.1-6, 1998.10
- [12] Nils Dahlbäck, Arene Jönsson: Integrating Domain Specific Focusing in Dialogue Models, Proceeding of EuroSpeech97, pp.2215-2218(1997)
- [13] 長尾: 「日本語形態素解析システム JUMAN version 3.2」, 1997
- [14] 日本電子化辞書研究所: 「EDR 電子化辞書 version 1.5」, 1995
- [15] 藤崎, 大野, 飯島: 「音声対話の収録方法とその設定について」, 音声対話資料収録用参考資料, 1998.1
- [16] 中川, 大谷: 「Bigram の使用による話し言葉用確率文脈自由文法の自動学習」, 情報処理学会論文誌, Vol.39, No.3(1998.3)
- [17] 伊藤, 好田: 「対話音声認識のための事前タスク適応の検討」, 情報処理学会研究報告, 96-SLP-14, pp.91-98 (1996.12)