

協調的応答を備えた音声対話システムとその評価

伊藤 敏彦[†] 小暮 悟[†] 中川 聖一[†]

自然言語による音声対話システムにおいては、システムがユーザと協調的に対話を進めていくことは重要である。この考えを基に我々は音声対話システムにおける協調的応答生成システムを開発した。応答生成システムは対話制御部、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、対話システムの意味理解システムによって生成された意味表現を受け取り、可能な限り協調的応答をする。たとえば、ユーザの質問文に検索に必要な情報を含まれていなかったり、検索結果の数が多い場合などはユーザへの質問を行う。また、ユーザの望む検索結果が得られなかった場合、それに代わる代案を提案する。本論文では音声対話システムの評価実験であげられたいくつかの応答生成システムの問題点を改良し、ユーザの対話の焦点を抽出し、協調的な応答を行う応答機能を持った応答生成システムについて述べる。また、「システムの使い勝手の良さ」が協調的応答生成の導入によってどのように向上するのかに着目して行った評価実験について述べる。

A Spoken Dialogue System with Cooperative Response and Evaluation for the System

TOSHIHIKO ITOH,[†] SATORU KOGURE[†] and SEIICHI NAKAGAWA[†]

We have developed a robust dialogue system which aids users in information retrieval through spontaneous speech. Dialog system through natural language must be designed so that it can cooperatively response to users. Based on this consideration, we developed a cooperative response generator in the dialogue system. The response generator is composed of dialog manager, problem solver, knowledge databases, and response sentence generator. The response generator receives a semantic representation (that is, semantic network) which the interpreter builds for the user's utterance and generates as cooperative response sentences as possible. For example, if a user's query doesn't have enough conditions/information to answer the question by the system, and if there are many information retrieval candidates from the knowledge database for user's question, the dialog manager queries the user to get necessary conditions and to select the information. Further, if the system can't retrieve any information related to the user's question, the generator proposes an alternative plan. And evaluation experiments are described how the above improvement increases "convenience of the system".

1. はじめに

近年、音声対話システムにおいて中心となっている研究はユーザに自然な発話を許すロバストな音声対話システムの構築である。自然な発話において生じる現象や音声認識部で生じた認識間違いに対するいくつかの意味理解法が報告されている。また我々も人間の訂正ストラテジーを用いたロバストな音声対話システムを構築し、被験者を用いた評価実験を行った¹⁾。しかし、我々が構築した音声対話システムはQ&Aシステムに非常に近いもので、質問文以外の発話文に対する処理や協調的な応答に関して問題点があった。自然

言語によるタスク指向の対話においては、システムがユーザと協調的に対話を進めていくことは重要であり多くの研究が行われている^{2),3)}。

発話内容を決定する方法としては、談話の結束性に注目し、修飾構造、談話の焦点などの情報を利用し発話内容を決定するアプローチ^{4),5)}や、談話をある目的のためのプランとして考え、システムがユーザの質問意図としての談話ゴールを推論し、そのゴールの達成に必要な内容を協調的発話として生成するアプローチ^{6),7)}がある。しかしながら、プランに基づく対話理解モデルは、タスク構造より導出されるプランによって対話全体が規定されるという性格上、話者が明確な意図を持ち、システムの持つプランに沿った無駄のない受け答えをすることが前提となっている。しかし、実際の人間同士の対話での受け答えは、試行錯誤過程

[†] 豊橋技術科学大学
Department of Information and Computer Sciences,
Toyohashi University of Technology

を含み対話固有の柔軟な運用方式に沿って行われている部分が多い。そのため、高野ら⁸⁾は、局所的なやりとりの多様性を結束性認識によって吸収しながら対話文脈を制御するために、対話における質問発話と応答発話の意味的結束性のタイプ分類を通して、応答生成を1つの推論過程としてモデル化し、質問者の意図を考慮することによって協調的な応答をする方法を提案している。

山田ら⁹⁾は、文献 6) や 7) のように対話者双方の信念を考慮することなくユーザの意図を推論することによってより多くの協調的応答を行う応答生成モデルを提案している。また、より知的な対話システム構築のためには、ユーザの質問に対してシステムが適切かつ協調的に応答するだけでは不十分であり、ユーザの意図をうまく聞き出すためにシステムからどのような質問をすればよいのか、システムがユーザに対してどのような質問をすれば効率良く対話が進むのか考慮する必要がある¹⁰⁾。観光案内、道案内などのデータベース検索における協調的応答生成に関しては質問の答以外に付加的な情報を与えたり、失敗した質問に対する理由や代案を提示するものが多い³⁾。

Dybkjær ら¹¹⁾は、WOZ を用いて集められたコーパスを分析することによって、人間と機械の音声対話において協調的に対話を進めていくための原則を提案している。それらの原理は Grice の理論に比べてより一般的なものとなっている。同様に堂坂ら¹²⁾も話し言葉対話コーパスにおいて、発話単位、対話相手からの応答、応答に対処するための戦略などを分析することによって、話し言葉生成システムが対話に参加するための遵守すべき協調的対話原則を示している。また、協調的な応答戦略を自動的に獲得する方法に関しては、高間ら¹³⁾によって擬人化エージェントの応答が、ユーザとの対話を経験していくうちに協調的なものになっていく応答戦略の自動学習対話管理機構を開発している。これらの研究は、音声認識部が完全に動作する、すなわちテキスト入力を対象としており、音声認識に誤りがある場合は想定していない。誤りがある場合を想定すると、上記の方法はそのまま使用することはできない。

協調的な応答をする音声対話システムに関しては、Allen ら¹⁴⁾がユーザと協調してタスクを実行する音声対話システムを実現し、音声対話システムの実用可能性に関する評価を行っている。その対話システムでは協調的な応答戦略に関して、リアルタイム性重視のために対話制御に計算コストのかかるプランニングは使用せず、入力からロバストパーシングによって得られ

た発話行為のシーケンスと音声対話の談話構造を用いて対話制御を行っている。対話システムで処理できない不確かさ・曖昧さが生じた場合、それらを解消するための対話を起動する。

協調的応答の1つである不明瞭な情報に関する明確化は我々の以前のシステム¹⁵⁾でも採用されていた。これは対話の流れを記述する対話ルールに明確化の命令を記述することによって実現していた。

我々が開発した協調的な応答機能を持った応答生成システムも、複雑な計算を必要とするプランニングは用いず、ユーザの発話から予測された発話目的や対話の焦点に注目し協調的な応答を行う。この発話目的や対話の焦点の予測はシステムへの質問だけに限定されず、願望やシステムの質問に対する応答からも予測される。また音声対話システムにおいて問題となる誤認識による誤動作を考慮して、複雑で精密な協調的応答生成モデルは構築せず、対話を最低限うまく終了させるために必要な目的や焦点の予測や協調的応答に限定している。つまり、テキスト入力対話と違い音声対話では、認識結果やその意味理解結果が正しいかどうかシステム側では確実に判断できないので、現在のところ文脈知識を強く用いることを避けている。

本論文では、構築した協調的な応答機能を持った応答生成システムについて述べ、その生成システムを組み込んだ富士山観光案内音声対話システムに関して、「システムの使い勝手の良さ」、「協調的応答」に着目して行った評価実験について述べる。

2. 従来の「富士山観光案内日本語音声対話システム」

「富士山観光案内日本語音声対話システム」^{15),16)}は、富士山周辺の観光案内をタスクとしており、ユーザの発声する音声を入力とし、その発話内容に対する観光案内を合成音声で応答する。現在のシステムでは、ふだん、我々が使用しているような話し言葉に近い『自然な発話 (Spontaneous Speech)』を理解することが可能になっている。ここでいう『自然な発話』とは、発話文中に「関投詞」、「未知語」、「助詞落ち」、「言い淀み・言い直し」、「倒置」、「多様な言い回し」といった話し言葉特有の現象を含んだ発話のことである。

従来の「富士山観光案内音声対話システム」は、「入力音声認識部」、「対話理解・管理部」および「応答音声合成部」の3つの部分から構成されている。

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。さらに、HMM は話者適応化を行って、認識率の向上をはかっている。音節 HMM, 文脈自由文法の構文解析法, 音声の「取込み」「分析」「認識」を同時に行う並列化アルゴリズム, および, 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, ユーザの発話を認識する「入力音声認識部」は構成されている。さらに「問投詞」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¹⁷⁾。文脈自由文法は自然対話音声を認識するために, 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。

ユーザの発話は「入力音声認識部」で認識され, 5-best の認識結果が, 「対話理解・管理部」に送られる。このうち第 1 位の認識結果のみ (もちろん, 第 5 位の認識結果まで用いれば, この「対話理解・管理部」の性能は向上する¹⁸⁾) 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。

続いて, 富士山周辺の観光地データベースの検索を行うことによって, 応答文の文字列を生成する。

「構文解析」および「意味解析」においては, 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ックスを用いて解析を行っている¹⁹⁾。

「応答音声合成部」は, 対話システムが生成する応答文の文字列をワークステーション上で動作する音声合成サーバに送り, 音声を出力している。

3. 協調的応答生成システム

3.1 従来のシステムの問題点と改良点

従来のシステム^{15), 16)}ではシステムがまだ不完全なために, 以下のような問題点があった。

- (1) システムが Q&A システムに近いシステムであり, 質問文以外の文をほとんど処理できない。
- (2) 使用できる文型に制限がある。
- (3) ユーザの提示した検索条件が少ない場合など, システムからの応答文は多くの内容を含んで長くなり, 応答内容の一部を聞き逃す可能性がある。
- (4) 音声認識部の処理時間が実時間の数倍程度で, 自然な対話とは言い難く, 十分な評価実験ができない。

このように発話できる文に制限があることや, 対話によって得られる情報の一部が欠落してしまうこと, 対話状況が不透明であることは, ユーザに不安, もしくは, 負担を与えることになりかねない。そこで, ユー

ザに余計な不安や負担を与えないようにすることを目的として, 従来のシステムの改良や協調的応答生成のできる応答生成システムの開発を行った。

3.1.1 言語処理 (理解) 部の改良

これまでの言語処理 (理解) 部は, 正しく意味表現に変換処理できる文型にかなりの制限があった。システムがもとと Q&A システムに近いシステムであったため, 疑問文以外の文はほとんど正しく意味理解できなかった。また副詞や形容詞を含んだ文の発話も多くは許していなかった。さらに指示詞などの処理も完全に行ってなかったため, 文を正しい意味表現に変換することは難しかった。

しかし, 言語理解部で使用する単語辞書や意味表現変換ルールデータベースへのデータの追加により, 疑問文以外の願望・依頼といった発話や副詞や形容詞を含んだ文も正しく意味理解できるようになった。さらに指示詞データベースを使用し, 以前の発話で使用された名詞などを登録することによって, 指示詞や省略の補完もある程度正しく処理できるようになった。

3.1.2 応答生成部の改良

応答生成部においても, ほとんど質問文以外の文は受け付けることができなかった。これは, 従来の応答生成システムはユーザの対話の意図 (焦点) をユーザの発話から得る機構を持っておらず, 質問文以外の文に対しては IF 文に似たルールで対応しているからである。そのため, そのルールにない発話の場合はシステムはどう対処すればよい判断できない。これは応答生成システムを作成したときの仕様によるものであるが, 実際の音声対話システムでは非常に不便である。これには, ユーザの質問文以外の発話からもユーザの対話の意図 (焦点) を抽出し, システムがユーザとの対話を制御できる機構が必要と思われる。

また, ユーザの発話文から得られた情報が少ない場合, 知識データベースの検索に失敗したり, 逆に大量の検索結果データを応答として出力したりすることがあった。これはユーザの発話文から得られた検索条件だけで, 知識データベースを検索し, さらに知識データベースから得られた情報をすべてユーザに提示しようとするからである。これらの点を改良するためには, システムが知識データベースを検索するための情報が足りない場合や検索結果データが多く得られた場合に, システムがユーザへ質問を行う機構を組み込む必要がある。

同様に, 知識データベースを検索した結果, データが見つからなかった場合, 知識データベースにデータがなかったことをユーザに示すだけであった。それで

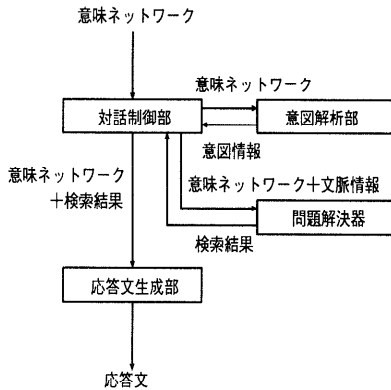


図1 応答生成部の構成図
Fig.1 Response generator.

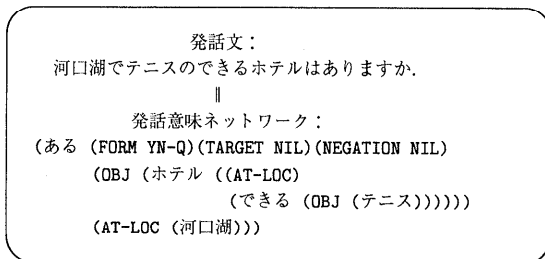


図2 発話意味ネットワーク例
Fig.2 An example of semantic network.

はタスクを達成するための発話数は非常に多くなる。このような場合には、ユーザの要求条件に近い検索結果データを代案として出すことが必要である。

このような点を改良するために、システムがユーザに対して協調的な応答を行い、ユーザのタスク達成のための労力を減少させる協調的応答生成システムを構築した。

3.2 改良した応答生成部の概要

構築した応答生成部の構成図を図1に示す。ユーザの発話した発話文は、意味理解部によって図2のような等価な意味ネットワーク（意味表現）に変換され、最初に対話制御部に入力される。

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。ここでの対話の意図とは、speech actのような抽象化された意図ではなく、その時点の話題に相当するものである。現在は観光案内において出現する可能性のある話題のみを意図と

して用いている。それらの意図は、その他のタスクにも適応できる一般性のある意図とドメインに依存するヒューリスティックな意図に分類されるが、観光案内というタスクの性格上、ドメインに依存した話題が多くなっている。しかしながら、話題は対話の流れを表現できる一般的なものと考えられる。また、より細かい発話意図と談話文法のようなもので対話の流れを制御することも考えられるが、認識間違いや談話文法の想定ミスによって対話の失敗を引き起こす可能性がある。つまり、テキスト入力対話と違い音声対話では、認識結果やその意味理解結果が正しいかどうかシステム側では確実に判断できないので、文脈知識を強く用いることは難しい。

そのため、大まかな対話の流れの制御においては、話題程度の粒度が適しており、細かい対話の制御が必要な場合には speech act を用いたプランベースの方法が適していると考えている。対話制御部は獲得したユーザの対話の意図（焦点）から対話の流れを決定し、以前の発話からの文脈的な情報があれば、それを文脈情報として利用する。

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。このとき、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。そのときの質問はユーザの対話の意図（焦点）に沿ったまだ獲得されていない検索条件に関してである。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よいかを決定し、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。

このような流れで、協調的応答生成システムは対話の制御を行っている。次節からシステムを構成している各部について詳しく述べる。

3.3 対話制御部

対話制御部は前節で述べたような対話の流れの管理、文脈情報の管理、必要な情報に対する質問などの制御を行っている。以下にその処理の流れを示す。

- (1) ユーザの発話意味ネットワークを意図解析部に送り、ユーザの対話の意図（焦点）やユーザの提示した条件など（意図情報）を抽出。
- (2) 直前の意図情報を検索し、今回の意図情報へ移行できる情報（文脈情報）を移行。
- (3) 問題解決器にもユーザの発話意味ネットワークを送り、データの検索。
 - (a) 検索結果がある一定数以下なら、検索結果をユーザに提示。
 - (b) 検索結果が一定数より多い場合、ユーザに情報の選択をしてもらうため、検索結果の取捨選択の質問を提示。その質問は意図情報を利用し選択。
 - (c) ユーザに提示する質問がない場合、情報のある一定数提示。
 - (d) 検索部の検索結果が代案であったのなら、代案についてユーザに提案。
 - (e) 代案用の検索データを用いて検索しても検索結果が出なかったら、データベースに希望の検索結果や代案がないことを応答。

3.4 意図解析部

意図解析部は入力されたユーザの発話意味ネットワークから、ユーザの対話の意図（焦点）や検索条件を抽出する。その対話の意図（焦点）の抽出方法は、次のような方法である。

- (1) まず発話意味ネットワークのルートである主動詞の部分（ある、できる、宿泊するなど）が対話の意図的な情報を持っているか調べ、持っていればそれを仮の対話の意図とする。
- (2) 次に発話意味ネットワーク中に含まれている意図情報や検索条件を抽出するために、サブネットワーク単位で調べていく。サブネットワークは、ルートである主動詞にアークによって接続されている各々の部分ネットワークである。つまり、文における主部・述部などに相当するものである。このときの探索順序は、我々が主観で決めた主部となりやすいサブネットワーク（その順序は、OBJ, INSTANCE, ACT, AT-LOC, COST, TIME, HOW-LONG, THEME, AMONG）からである。
- (3) サブネットワークの上位部分から下位部分に探索は進んでいき、探索位置がサブネットワーク

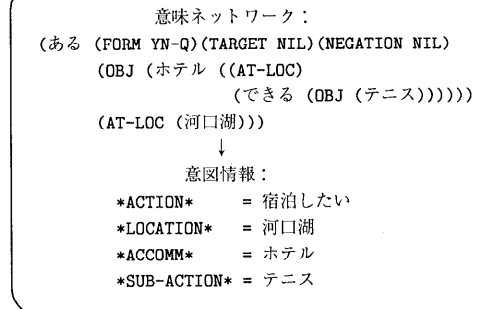


図 3 抽出された意図情報例

Fig. 3 An example of extracted intention.

の各名詞単語の部分に来たときに情報の抽出を試みる。

- (4) まずその名詞単語の意味クラスを知識データベースから検索する。
- (5) 次にその名詞の意味クラス、その名詞が掛かっている動詞、サブネットワークの種類を検索条件として意図データベースを検索する。そのデータベースにはその単語が問題解決器の検索条件として意味を持つか、また意味を持つならばどういった条件としてか、さらに意図に関するものは含まれているかといったことが書かれている。それらのデータを用いて最も中心となる対話の意図（焦点）と検索条件を抽出していく。
- (6) もし、同じ種類の検索条件が複数抽出された場合は、検索条件として制約が強いものが優先される。

図 3 に実際に抽出された意図情報の例を示す。「河口湖でテニスのできるホテルはありますか。」という文が意味理解システムに入力されると、図 3 に示したような発話文と等価な意味ネットワークを出力する。この意味ネットワークと発話文には情報の欠落はないと考えている。

この意味ネットワークが入力として意図解析部に入力される。この意味ネットワークの「ある」という主動詞からだけでは対話の意図は分らない。そこで次に主語になりやすいサブネットワークの順序に従って、最初に主動詞と OBJ のアークによって結ばれているサブネットワークを探索する。そのサブネットワークの最初の名詞「ホテル」の意味クラス「宿泊施設」、YN 型の質問文であることを示す「(FORM YN-Q)」, その名詞の掛かっている動詞「ある」といった検索条件を用いてデータベースを検索し、この対話の意図（焦点）は「どこかに宿泊したい」とであると仮決定す

る（これより優先度の高い意図が他の部分で抽出されれば、意図は変更される）。これは意図の抽出には現在のところ文脈的な情報を用いていないために（認識結果の誤りも想定されるので、文脈の利用法には検討を要する）、発話文のみから最も使用範囲の広い意図を抽出するためである。つまり、「情報を聞きたいだけ」という意図は「どこかに宿泊したい（だから情報を聞きたい）」という意図に含まれている。現在の対話制御では、上に述べたような2つの意図の違いを分ける必要がない（必要に応じて意図の確認や変更ができる）ため、これ以上の細かい決定は行っていない。しかし、対話制御部の対話の意図履歴からこれが本当に宿泊したいと思っているか否か判定可能だと考えている。さらに意図決定のときにユーザの望んでいる宿泊施設の種類は「ホテル」であることも同時に抽出する。そして探索は下位のネットワークへ進んでいき、「テニス」で次の情報抽出を行う。このときも同様な方法で抽出する。

そのサブネットワークの探索が終わると、それから、次の AT-LOC のアークで結ばれたサブネットワークも探索し全体の意図抽出を終了する。このような方法で獲得された意図情報が図3の意図情報の部分である。

3.5 問題解決器

問題解決器は、対話制御部から入力された意味ネットワークと文脈情報を用いて、応答に必要な情報を知識データベースから検索し出力する。また知識データベース検索の結果、直接に関係した情報が見つからなかった場合には、一部の検索条件を変更して代案の探索を行う。

探索は以下のような方法で、ユーザの望む情報または代案を検索している。

- (1) ユーザの発話意味ネットワークに含まれている文型に関する情報（主動詞、発話のタイプ、ターゲット、名詞の意味クラスなど）からユーザの望んでいる情報のタイプ（たとえばホテル名、場所など）を決定する。
- (2) 次に主動詞といくつかのアークによって接続されているサブネットワークを優先順位（主語になりやすい順番）に従って、サブネットワーク単位で、さらにサブネットワークの上位位置から検索を行う。
- (3) サブネットワークの上位からノードとアークを調べていき、ノードが名詞（たとえば、ホテル、水上スキー）であった場合に実際の情報の検索が行われる。
- (4) まずその名詞（たとえば、湖の名前）からユー

ザの望んでいる情報（たとえば、ホテル名）を検索するためには、どのような順序で知識データベースを検索していけばよいか記述されている検索プランを獲得する。1つの検索プランはいくつかの関数で構成されて、その関数は自由に定義することができる。そのため、複雑な検索も自由に設定することができる。検索プランは、その名詞が掛かっている動詞、その名詞が上位概念と結ばれているアーク名、名詞の意味クラスなどを利用し選択される。

- (5) 選択された検索プランに記述されている関数どおりに意味ネットワークで構成された知識データベースを検索していく。そのときに得られた結果がその名詞における検索結果となる。
- (6) 同様にして次の名詞においても検索結果が得られる。そこで得られた検索結果と先ほどの検索で得られた検索結果の AND をとり、残った結果を新たな検索結果とし、検索を続けていく。
- (7) 文脈情報として得られた情報に関しても同様に検索を行っていく。
- (8) 最終的にすべての情報に関して検索が終わったとき、残った検索結果が最終的な検索結果となる。
- (9) 途中、どこかの名詞で検索結果が得られなかった場合、または各検索結果の AND をとったときに検索結果が得られなかった場合、代案処理に移行する。
- (10) 代案処理では代案モデルに従って付けられた各名詞の代案概念があり、検索に失敗した名詞の代案概念クラスと同じ概念を持つ名詞を知識データベースより検索し、それを代わりの検索条件として用いて検索を続行する。代わりの検索条件が5個以下の場合、見つかったすべての代案を出力する。代わりの検索条件の数がそれ以上の場合には検索時間の関係から代案が1つ見つかった時点で検索を終了し検索結果を出力する。

図4に実際の動作例を示す。問題解決器に図4の意味ネットワークが入力されると、前述した検索方法によって検索結果が出力される。この検索結果はユーザの望む検索結果がなかった場合に出力される代案の例である。これは河口湖には水上スキーができるホテルは見つからなかったが、テニスのできるホテルとして、河口湖ホテルというホテルが見つかったことを示している。先ほど述べたように代案を出力するために再検索をするとき、検索に失敗した水上スキーの部分の検

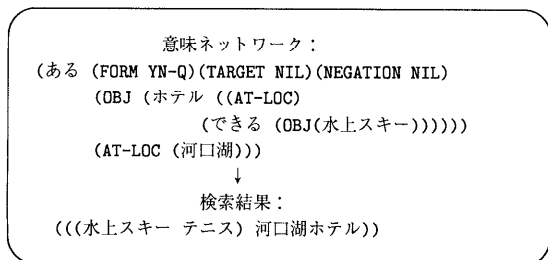


図4 検索結果例

Fig. 4 An example of information retrieval result.

索条件を変更する。水上スキーの上位概念はスポーツであり、スポーツの下位概念である「テニス・水上スキー・スケート・水泳」などを検索条件として再検索する。その結果、うまく検索結果が出力された検索条件のテニスが代案として出力される。

3.6 応答文生成部

応答文生成部はユーザの発話意味ネットワークと検索結果から、その2つを統合しユーザの質問や発話に対する応答文用の意味ネットワークを生成する。その後、生成された応答文意味ネットワークの各単語とそれらの単語が接続されているアークの種類によって必要な助詞などを選択し、実際の応答文を生成する。

最初に応答文生成部に図2のユーザの発話意味ネットワークと検索結果が入力される。図4の検索結果の形から、入力された検索結果はユーザが望む検索結果が見つからないための代案結果であることが分かる。そのため、応答文生成部はまずユーザが望む検索結果がなかったことを示す応答文意味ネットワークを作成する。次に代案を提案するための応答文意味ネットワークを作成する。作成された応答文意味ネットワークの各単語を並べ替え、単語の意味や概念から必要な助詞などを付与し、応答文として出力している。

図5に実際の対話システムの対話例を示す。検索情報が多く検索された場合、システムはユーザへ情報の選択を行うための質問を行ったり、検索結果がなかった場合、それに近い代案を出力していることが分かる。図中のゴシック体の文は、表層的な文章に関する処理を行っていないため、まだ文法的に不自然である。

4. 音声対話システムに関する評価実験

4.1 実験の形式・条件

実験は、被験者らに「1泊2日の富士山周辺への研究室旅行」を、システムとの対話で得られる情報をもとに計画してもらうという形式で行った。決定してもらう内容は、1、2日目の目的地とそこでのプラン、および、宿泊先の施設の所在、種類、料金、宿泊施設名

```

SYS： 富士山観光案内システムです。ご利用をお願いします。
USR： 本栖湖で何ができますか。
SYS： どんな事をなさりたいのですか。
USR： テニスがしたいのですが。
SYS： テニスは本栖湖でできます。
USR： 本栖湖でボートに乗れますか。
SYS： ボートは本栖湖では乗れません。ボートは山中湖でのれます。ボートは河口湖でのれます。
USR： 水上スキーのできるホテルは河口湖にありますか。
SYS： 水上スキーのできるホテルは河口湖にありません。しかしテニスのできる河口湖ホテルが河口湖にはあります。

```

図5 音声対話例

Fig. 5 An example of spoken dialog.

の計8項目である。被験者は、音声対話システムに関してまったく知識のない本大学生（学部2～修士1年）10人であり、特にシステムの使用法は教示せず、音声認識モデルの話者適応化用に発声してもらったタスクに関する50文を発声例として示した。さらに別の評価実験で同一のシステムを1時間程度使用してもらった。

実験では以下の2種類のシステムで旅行計画を立ててもらった。

- (1) 従来の対話システム¹⁶⁾ (旧システム)
- (2) 言語処理部（意味理解部、応答生成部）改良後の対話システム（新システム）

新システムは旧システムと比べると、言語理解部で使用する単語辞書や、意味表現変換ルールデータベースへのデータの追加が行われている。そのため、疑問文以外の願望・依頼といった発話や副詞や形容詞を含んだ文も正しく意味理解できる。さらに指示詞データベースを使用し、以前の発話で使用された名詞などを登録することによって指示詞や省略の補完もある程度可能である。

音声認識部は、どちらのシステムも同一のものを使用している。実験に使用した認識条件は特徴パラメータとしてケプストラム、時間変化を表す回帰係数を使用し、ビームサーチのビーム幅は5である（認識を高速に行うためにビーム幅を狭くし、また音声の継続時間分布制御も使用していない）。またHMMには研究室内の雑音環境に適応化した不特定話者モデルを、実験前、被験者に発話してもらった富士山観光案内適応化文50文で話者適応化し、使用している。音声認識用の文脈自由文法は単語数275、規則数467、単語単位のテストセットパープレキシティは約100である。

自然な発話を受理するように文法にロバスト性を持たせたため、パープレキシティは非常に大きい。しかしながら、音声認識処理に要する時間は、発話時間を含め実時間の1.5~2倍程度(HP-C160使用)で、時間的には比較的スムーズな対話が続行できる。

実験の手順としては、最初に実験の内容と旅行プランを記入してもらい紙を渡し、熟読してもらった。その後、旧システムと新システムのそれぞれで(最初に対話してもらうシステムは被験者ごとに変更)、旅行計画を2案立ててもらった。実験終了後、使用感や応答に関するアンケートに答えてもらった。

4.2 実験の結果

まず、この10人の項目達成率(必要な8項目を決定できた割合)は、旧システムで平均96%、新システムで平均99%であった。旧システムの対話システムでは宿泊施設の料金に関する情報を聞き出すことができなかった人が多く、新システムでは1人だけ2日目の目的地を決定できなかった。全発話数に関しては旧システムで190文、新システムで240文となり、新システムの方が発話数は多くなっている。この理由として、全発話数には正しく認識できなかったときの再発声が含まれているため、認識率が悪かった新システムには再発声が多く含まれていること、またシステムからの質問(22回)に回答する発話も全発話数に含まれていることがあげられる。

表1にシステムの評価結果を示す。正解音声認識率は正しく認識された文の割合を示している。認識文理解率は、(終)助詞落ち・誤りなどを許した場合の音声認識率である(つまり人間が容易に理解できる割合)。正解意味理解率は言語理解部で正しい意味表現に変換できた割合である。情報提供率はシステムが正しい情報を提供した割合である。システム質問率はシステムが質問をした割合、代案提供率はシステムが代案を提示した割合である。正解応答率は情報提供率、システム質問率、代案提供率の他に意味表現作成に失敗した割合(システムは発話理解に失敗したことをユーザに回答する)を加えたものである。情報検索失敗率は、データベースの不備などによって回答に失敗した割合である。

正解音声認識率、認識文理解率は発話された文型の豊富さの違いから、同じ音声認識部を使用しているにもかかわらずシステムによって大きな差がでている。つまり、新システムと旧システムではシステムが受理できる文型の種類が異なり、新システムの方が多種の文型を受理できる(システムの質問に対する応答も新しい文型)。旧システムはその文型の制限がきつく、被験

表1 評価結果

Table 1 Evaluation results.

評価	旧システム	新システム
全発話数	190 文	240 文
正解音声認識率	65 文 (34.2%)	59 文 (24.6%)
認識文理解率	145 文 (76.3%)	129 文 (53.8%)
正解意味理解率	127 文 (66.8%)	129 文 (53.8%)
正解応答率	132 文 (69.5%)	198 文 (79.2%)
情報提供率	113 文 (59.5%)	106 文 (44.2%)
システム質問率	0 文 (0%)	22 文 (9.2%)
代案提供率	0 文 (0%)	7 文 (2.9%)
情報検索失敗率	17 文 (9.0%)	11 文 (4.6%)

者は自分が発話した文型がシステムで受理できないと気付くと、以後それと同種の文型はほとんど発話しなくなる。そのため、言語理解部・応答部の改良を行った新システムの方が、ユーザの発声する発話の文型の種類が豊富になっている。そのうえ、新システムで拡張された文型は認識が難しいものが多い(文長も長い)ことも認識率を下げている要因である。

そこで、人間には理解できる認識文理解率(人間にも理解できない文は、システムにも理解は難しい)に対する正解意味理解率、全情報提供率(情報提供率+代案提供率+システム質問率)などを比べると、今回の改良によってかなり性能が向上していることが分かる。つまり、言語処理部の改良によって、人間が理解できる程度の誤認識文に対してはかなり正しく意味理解できている。

また、正しく意味理解された文に対して新システムは、ほぼ有用な情報を提供していることや、情報検索失敗率も減少していること、さらにユーザに間違った回答をしなかった正解応答率も高いことから、応答生成部の改良の効果も表れている。しかし、被験者によってタスクを達成するための発話数の差があまりに大きく、質問や代案といった対話の制御による発話数の減少といった効果は得られなかった。また、システムの質問に対するユーザの回答が誤認識されるとシステムは誤った情報を提供したり、発話の誤認識によって文脈知識データが誤って更新されたりすることがあった。これらの失敗に対する対応(ユーザへの確認など)が必要である。

音声認識率を100%として被験者の発話を書き起こしたものを入力(テキスト入力)とした場合の評価結果を表2に示す。新システムは発話された文の種類が豊富なにもかかわらず、正解意味理解率(旧システム:78.9%, 新システム:85.4%), 正解応答率(旧システム:73.2%, 新システム:92.5%)ともに旧システムより新システムの方が大きく上回っている。この

表2 評価結果 (入力: 書き起こしデータ)

Table 2 Evaluation results (input: transcription).

評価	旧システム	新システム
全発話数	190 文	240 文
正解意味理解率	150 文 (78.9%)	209 文 (87.1%)
正解応答率	139 文 (73.2%)	212 文 (88.3%)
情報提供率	132 文 (69.5%)	150 文 (62.5%)
システム質問率	0 文 (0%)	23 文 (9.6%)
代案提供率	0 文 (0%)	11 文 (4.6%)

結果からも言語処理部の改良によって意味理解の性能が向上していることがはっきり分かる。さらにシステムがユーザにタスク達成のために有益な何らかの情報を提供する割合が旧システムの 69.5% から新システムの 80.4% (= 65.4 + 10.4 + 4.6) と大きく上昇している。この割合の上昇は今回改良した言語処理部と協調的応答によるもので、タスク達成に費やす労力を大きく軽減できた。

4.3 使用感に関するアンケート

実験終了後、言語処理部に関するアンケートを被験者に書いてもらい使用感などを調査した。計画の立てやすさ、システムの便利さに関しては、質問や代案を提示したり、処理できる文の種類が豊富な新システムの評価が高い。応答の聞きやすさ、応答の自然さ、次発話のしやすさといった項目では新システムが少し良い程度であった。

評価実験では質問や代案提示の回数がかなり少ないため、あまり差がでなかったと思われる。

次にそれぞれのシステムで同じ計画を立てた対話例の書き起こしを示し、システムの便利さ、システムの自然さ、次発話のしやすさの項目を質問した場合は、どれも新システムの評価がかなり高かった。

次に対話制御に関するアンケートでは「検索結果を取捨するためのシステムからの質問は必要か」という質問に半分が必要、半分が不必要(検索結果を全部提示してほしい)と答えた。代案に関しては全員があると便利であると答え、計画がうまくいかないときのシステムからの提案に関しても全員が望んでいた。またシステム主導型のような必要なことを全部聞いてくれるシステムが便利であるという意見もあった²⁰⁾。

これらのアンケートの結果からも、今回の改良による労力軽減の効果はあったと考えられる。

5. おわりに

本研究では、ユーザに協調的に応答するために、言語処理部の改良と協調的応答生成システムの開発を行った。またシステムの評価実験においては、言語処理部の改良によってタスク達成への労力の軽減が見ら

れた。

今後の課題としては、実際には対話の流れによって代案の候補は変化するので代案検索に文脈的な対話の流れを考慮することが必要となる。また表層的な協調性、たとえば、冗長な文の削除や文の統合といったことを組み入れていく予定である。また、評価実験によって明らかになったシステム全体の不備な点も改良していく予定である。

参 考 文 献

- 1) 伊藤, 肥田野, 山本, 中川: ロバストな対話システム構築に関する一考察, 情報処理学会, 音声言語情報処理研究会, 95-SLP-7-22 (1995.7).
- 2) Kaplan, S.J.: Cooperative Responses from a Portable Natural Language Database Query System, Brady, M. and Berwick, R.C. (Eds.), *Computational Models of Discourse*, pp.167-208, MIT Press (1983).
- 3) Webber, B.: Question Answering, Shapiro, S.C. (Ed.), *Encyclopedia of Artificial Intelligence*, pp.814-822, New York, Wiley (1987).
- 4) McKeown, K.R.: Reasoning on a Highlighted User Model to Respond to Misconceptions, *American J. Computational Linguistics*, Vol.14, No.3, pp.52-63 (1988).
- 5) Moore, J.D. and Paris, C.L.: Planning Text for Advisory Dialogues, *Proc. 27th Annual Meeting of the Association for Computational Linguistics*, pp.203-211 (1989).
- 6) Cohen, P.R. and Perrault, C.R.: Elements of a Plan-Based Theory of Speech Acts, *Cognitive Science*, Vol.3, pp.177-212 (1979).
- 7) Allen, J.F. and Perrault, C.R.: Analyzing Intention in Utterances, *Artif. Intell.*, Vol.15, No.3, pp.143-178 (1980).
- 8) 高野, 平井, 北橋: 発話感の意味的結束性のモデル化—限定された領域独立知識を用いた応答タイプ絞り込み手法の提案, 電子情報通信学会論文誌, Vol.J79-D-II, No.12, pp.2146-2153 (1996).
- 9) 山田, 溝口, 原田: 質問応答システムにおけるユーザ発話モデルと協調的応答の生成, 情報処理学会論文誌, Vol.35, No.11, pp.2265-2275 (1994).
- 10) 石川, 加藤: 質問表現生成のための発話内容決定機構, 人工知能学会誌, Vol.10, No.6, pp.130-138 (1994).
- 11) Dybkjær, L., Bernsen, N.O. and Dybkjær, H.: Grice Incorporated Cooperativity in Spoken Dialogue, *Proc. COLING 96*, pp.328-333 (1996).
- 12) 堂坂, 島津: 話し言葉コーパスにおける協調的対話原則の分析, 信学技報, NLC97-1 (1997).
- 13) 高間, 土肥, 石塚: 擬人化エージェントにおけ

る音声対話を通じての協調的応答戦略の自動学習, 人工知能学会誌, Vol.12, No.3, pp.456-465 (1997).

- 14) James F.A., Bradford W.M., Eric K.R. and Sikorski, T.: A Robust System for Natural Spoken Dialogue, *Proc. 34th Annual Meeting of the Association for Computational Linguistics*, pp.62-70 (1996).
- 15) Yamamoto, M., Kobayashi, S., Moriya, Y. and Nakagawa, S.: A Spoken Dialog System with Verification and Clarification Queries, *IEICE Trans.*, Vol.E76-D, No.1, pp.84-94 (1993).
- 16) 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).
- 17) 甲斐充彦, 中川聖一: 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評価, 電子情報通信学会論文誌, Vol.J80-D-II, No.10, pp.2615-2625 (1997).
- 18) 肥田野勝, 中川聖一: 音声対話システムにおける N-best 文認識結果の一利用法, 第 52 回情報処理学会全国大会論文集 (2), 4D-2, pp.165-166 (1996).
- 19) 山本, 小林, 中川: 音声対話文における助詞落ち・倒置の分析と解析手法, 情報処理学会論文誌, Vol.33, No.11, pp.1322-1330 (1992).
- 20) 中川聖一, 山本誠治: 音声対話システムの構成法とユーザ発話の関係, 電子情報通信学会論文誌, Vol.J79-D-II, No.12, pp.2139-2145 (1996).

(平成 9 年 7 月 1 日受付)

(平成 9 年 12 月 1 日採録)



伊藤 敏彦

平成 6 年豊橋技術科学大学情報工学課程卒業。平成 8 年同大学大学院情報工学専攻修了。現在同大学大学院博士課程在学中。自然言語処理に関する研究に従事。



小暮 悟

平成 9 年豊橋技術科学大学情報工学課程卒業。現在同大学大学院情報工学専攻修士課程在学中。自然言語処理の研究に従事。



中川 聖一 (正会員)

昭和 51 年京都大学大学院博士課程修了。同年同大学情報工学科助手。昭和 55 年豊橋技術科学大学情報工学系講師。平成 2 年教授。昭和 60 ~ 61 年カーネギーメロン大学客員研究員。音声情報処理, 自然言語処理, 人工知能の研究に従事。工博。昭和 52 年電子情報通信学会論文賞, 1988 年度 IETE 最優秀論文賞受賞。著書「確率モデルによる音声認識」(電子情報通信処理学会編), 「音声・聴覚と神経回路網モデル」(共著, オーム社), 「情報理論の基礎と応用」(近代科学社) など。