

音声対話システムにおける協調的応答

伊藤敏彦 中川聖一

¹豊橋技術科学大学 情報工学系

〒441 豊橋市天伯町字雲雀ヶ丘1-1

我々はユーザが自然言語を用いてシステムと対話することによって情報を検索するロバスタ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なった。その評価実験において、応答生成部に関するいくつかの改良点が挙げられた。それは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あり、質問文以外の発話文に対する処理やユーザへの協調的な応答に関してであった。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とり、可能なかぎり協調的応答をする。例えば、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なう。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。

本稿では音声対話システムの評価実験で挙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、ユーザの簡単な対話の焦点を抽出し、協調的な応答を行なう応答機能をもった応答生成システムについて述べる。

Cooperative Response in Spoken Dialogue System

Toshihiko ITOH and Seiichi NAKAGAWA

¹ Department of Information and Computer Sciences

Toyohashi University of Technology, Tenpaku, Toyohashi, 441, Japan

We have developed a robust dialogue system which aid 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 considerations, we deloped 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 be answer the question by sysytem, and if there are many retrieved information 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.

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タ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タな音声対話システムを構築し、被験者を用いた評価実験も行なった[10]。その評価実験において、応答生成部に関するいくつかの改良点が挙げられた。それは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あり、質問文以外の発話文に対する処理やユーザへの協調的な応答に関してであった。

自然言語による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り

多くの研究が行なわれている [1], [2]。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 [3], [4] 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 [5], [6] がある。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 [2]。高野ら [7] は、質問に対する協調的応答生成モデルを使用し、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 [8] は、[5], [6] 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なう応答生成モデ

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 [9]。

本稿では評価実験で挙げられたいくつかの応答生成システムの問題点を改良するために、構築した協調的な応答機能をもった応答生成システムについて述べる。

2 富士山観光案内システム

本節では、我々が開発した富士山観光案内音声対話システムについて簡単に述べる。この音声対話システムはユーザに自然な発話を許すため、音声認識システムで一部誤認識された認識結果にもある程度対応することができる意味理解システムを用いている [10]。

ユーザの発話は音声認識システム [11] で認識され文字列に変換された後、意味理解システムへ送られる。意味理解システムでは認識結果を形態素解析、文節解析、構文解析、意味解析、省略された動詞の補完と代名詞の補完の処理である文脈解析を行ないユーザの発話した内容と等価である意味表現 (意味ネットワーク) に変換する。

また、自然な発話や音声認識結果の一部に誤認識が含まれている認識結果に対応するために構文解析・意味解析では、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。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。意味表現はドメインやタスクといった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、ボトムアップに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 [12]。

応答生成システムは意味理解システムが生成した意味表現 (意味ネットワーク) を受け取り、必要な情報を知識データベースから検索して、ユーザへの応答を文字列として生成する。生成された応答文は音声合成され、ユーザに音声で応答する。それぞれの部分はシーケンシャルに処理がなされる。

3 音声対話システムの評価

我々はまた前節で述べたような富士山観光案内システムを使用し、音声対話システムの性能を評価するた

めに評価実験を行なった [10]。被験者は音声対話システムに対して全く知識のない学生 7 名である。被験者には富士山周辺へ 1 泊 2 日の旅行を計画する必要があると想定してもらった。それから富士山周辺の観光、つまり観光地や宿泊施設の情報について質問する事ができる富士山観光案内音声対話システムを使用してもらい、研究室旅行の計画を決定するために必要な、いくつかの項目 (1、2 日目の目的地とそこでのプラン、宿泊施設の場所、種類、料金、名前の計 8 項目) を決めてもらうタスクを課した。

結果として、全発話 232 文に対して音声認識システムでの文認識率が 52.1% であり、この結果を対話システムに入力すると 69.8% が正しく意味理解できた。また被験者の発話を書き起こし、言い直し、間投詞などを除いたものを対話システムに入力した (テキスト入力) 場合には 78.4% が正しく意味理解できた。これを見ると音声入力の場合とテキスト入力の場合の差はそれ程大きくはない。テキスト入力でも正しく意味理解できなかった理由として、約半分が意味表現生成用規則がないためであり、残りは意味理解部の解析用語彙への単語の未登録のためであった (登録単語のうち自立語は 284)。

4 協調的応答生成システム

以上に述べてきたように、実際に富士山観光案内音声対話システムを構築し、その性能を評価した。音声対話システムを構築した時点でも、いくつか応答生成システムの改良すべき点は挙げられていたが、この被験者を用いた音声対話システムの評価実験で被験者とシステムの対話のやりとり、更に実験終了後のユーザの感想を聞き、応答生成システムの応答に対する改良点が浮き彫りになってきた。以下にその改良点を挙げる。

- 現在のシステムは Q & A システムに近いものであり、あまり複雑な対話を行なうことができない。
 - ユーザの提示した情報が足りない場合にはデータ検索に失敗したり、又は大量の検索結果を出力したりする。
 - ユーザに親切な応答をすることができない。例えば、検索に合った結果がないならそれに近い結果を出力する等。
 - ユーザの意図をシステムが検出することができないので、システムが対話の流れを制御することができない。
- あまり複雑な応答文を生成することができない。

上に挙げたように現在の富士山観光案内音声対話システムは Q & A システムに近い状態にあり、ほとんど質問文以外の文は受け付けることができない。これは、現在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。そのため、そのルールにない発話の場合はシステム

はどう対処すれば良いか判断できない。これは応答生成システムを作成した時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。これはシステムがユーザが発話した文から得られた検索条件の情報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全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なう機構を組み込む必要がある。

さらに現在、もし知識データベースに検索した結果、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無かったことをユーザに示すだけである。それでは前節で述べた評価実験の様なタスクを達成しようとした場合に非常に対話数が多くなってしまいます。このような場合には、ユーザの条件に近い検索結果データを代案として出すことが望まれる。

そのために述べたように、システムがユーザに対して協調的な応答を行ない、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。

4.1 応答生成システムの概要

今回、構築した応答生成システムの構成図を図1に示す。ユーザの発話した発話文は意味理解システムによって、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)とユーザの提示した検索条件を抽出する。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、さらにこれまでの発話から獲得された情報で現在の発話にも移行できる情報(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ための問題解決器には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基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

脈情報を利用して検索するが、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対話制御部はユーザにデータの選択のための質問をする。その時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するものである。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。

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なっている。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。

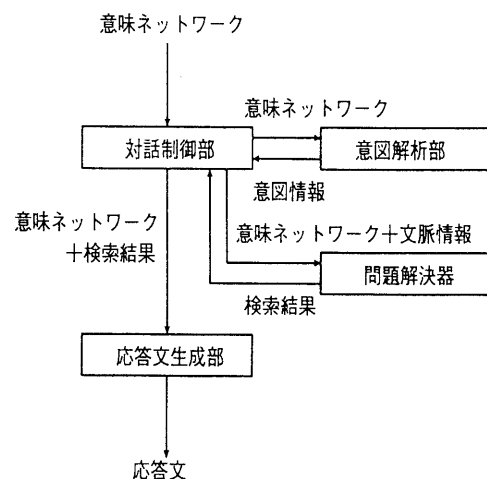


図1: 応答生成システムの構成図

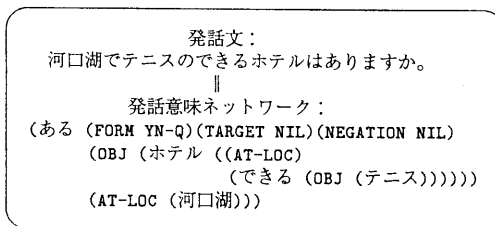


図2: 発話意味ネットワーク例

4.2 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なっている。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、データの検索を行なう。
 - (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する。
 - (b) 検索結果が一定数より上の場合ユーザに情報の選択をしてもらうため、検索結果を取捨選択するための質問を行なう。その質問は意図情報を利用し選択。
 - (c) ユーザへ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によって見つかったのなら、代案についてユーザに提案する。
 - (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、データベースを検索したが検索結果や代案がないことを応答する。

4.3 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、ユーザの対話の焦点(意図)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、次の様な方法である。

1. まず発話意味ネットワークの主動詞の部分(ある、できる、宿泊する等)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。この時の探索順序は現在は我々が主観で決めた主語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の各名詞単語の部分に来た時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。発話文として「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からない。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」とであると決定する。さらにその時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なう。この時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。そ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。

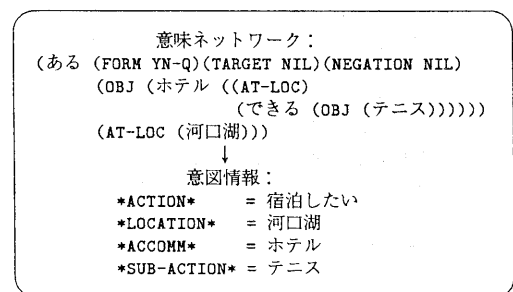


図3: 抽出された意図情報例

4.4 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、文脈情報からユーザの質問や発話に関して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は一部の検索条件を変更して代案の検索を行なう。

検索は以下のような方法で、ユーザの望む情報又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、名詞の意味クラス等)からユーザの望んでいる情報のタイプ(例えばホテル名、場所等)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、さらにサブネットワークの上位位置から検索を行なう。
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(例えば、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なわれる。
4. まずその名詞(例えば、湖の名前)からユーザの望んでいる情報(例えば、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ゆけば良いか記述されている検索プランを獲得する。一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通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。その時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なっていく。
8. 最終的に全ての情報に関して検索が終わった時、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、又は各検索結果のANDを取った時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. 代案処理では図4のような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も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、見つかった全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一つ見つかった時点で検索を終了し検索結果を出力する。

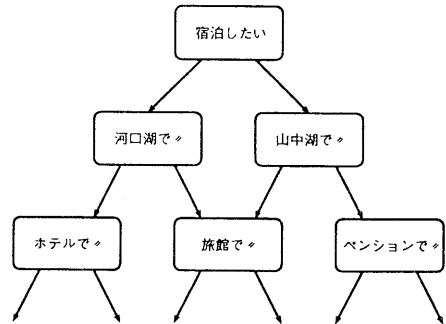


図4: 代案モデル(意図モデル)

図5に実際の動作例を示す。問題解決器に図5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され図5の検索結果に示されている様な結果が出力される。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はレイグランドホテルというホテルが見つかったことを示している。

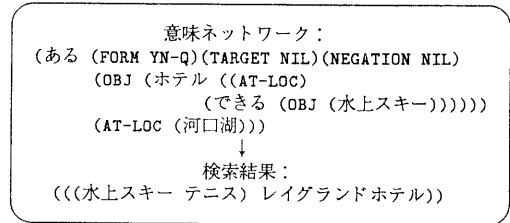


図5: 検索結果例

4.5 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、その二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。図6に実際の応答文生成例を示す。

最初に応答文生成部に図6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6の検索結果の形

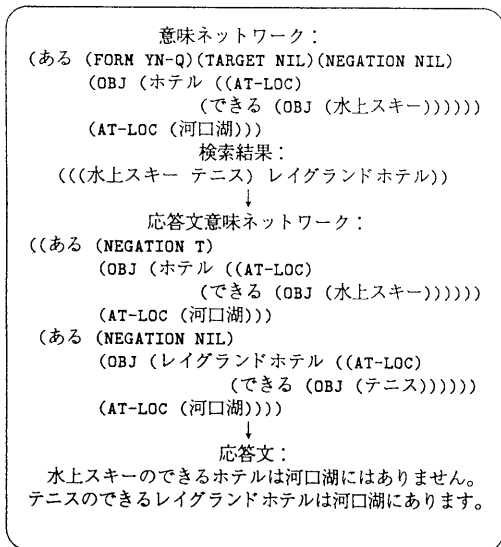


図 6: 生成例

から入力された検索結果は、ユーザが望む検索結果がないための代案結果であることが分かるため、応答文生成部はまずユーザにユーザが望む検索結果がなかったことを示す応答文を生成するための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ユーザの発話の形に合わせ並び換え、必要な助詞などを付与し応答文として出力している。

図7に実際の対話システムの対話例を示す。例では、検索情報が多く検索された場合にシステムはユーザへ情報の選択を行なうための質問を行ったり、さらに検索結果がなかった場合は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、まだ文章的に不自然である。

5 まとめ

実際に自然な発話や一部誤認識を含んだ認識文を意味理解できる音声対話システムを構築し、その評価を行なった。その評価でユーザに改良点として挙げられたユーザに対して不親切である応答生成システムをユーザにできるだけ協調的にそしてタスクを達成するためにできるだけ便利であるように改良した。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、例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。

SYS: 富士山観光案内システムです。ご用件をお願いします。

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レイグランドホテルが河口湖にはあります。

図 7: 対話例

参考文献

- [1] 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).
- [2] Webber, B.: Question Answering, Shapiro, S.C.(ed.), Encyclopedia of artificial intelligence, pp.814-822, New York: Wiley(1987)
- [3] 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)
- [4] Moore, J.D. and Paris, C.L.: Planning Text for Advisory Dialogues, Proc. Annual Meeting of the Association for Computational Linguistics, pp.203-211(1989)
- [5] Cohen, P.R. and Perrault, C.R.: Elements of a Plan-Based Theory of Speech Acts, Cognitive Science, Vol3, pp.177-212(1979)
- [6] Allen, J.F. and Perrault, C.R.: Analyzing Intention in Utterances., Artif. Intell., Vol.15, No.3, pp.143-178(1980)
- [7] 高野, 平井, 北橋: 「自然言語によるデータベース検索における協調的応答生成」, 電子情報通信学会, 信学技法, NLC95-4, pp.25-33(1995)
- [8] 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)
- [9] 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)
- [10] 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)
- [11] 甲斐, 間宮, 中川: 「自然発話の認識・理解のための解析・照合手法の比較」, 情報処理学会, 音声言語情報処理研究会報告, 94-SLP-2-12(1994.7)
- [12] 山本, 肥田野, 伊藤, 大谷, 中川: 「自然発話の意味理解と対話システム」, 情報処理学会, 音声言語情報処理研究会報告, 94-SLP-2-13, pp.91-98(1994.7)