

学位論文

1993年  
262

音声応答生成のための対話処理と  
音声合成に関する研究

1992年12月21日提出

指導教官 広瀬啓吉助教授

工学系研究科 電子工学専攻 07084

浅野 康治

# 目次

1	序論	1
1.1	本研究の意義	2
1.2	本研究の背景	4
1.3	本研究の目的	6
1.4	本論文の構成	8
2	音声応答システム	10
2.1	音声応答システムの種類	11
2.2	音声応答システムの構成	12
2.3	データベース	15
2.4	音声認識処理	15
3	音声対話処理	19
3.1	はじめに	20
3.2	音声対話資料の分析	20
3.2.1	音声対話資料	20
3.2.2	音声対話の特徴	21
3.3	処理の対象となる対話	25
3.4	システムの構成	27
3.5	システムの特徴	29
3.6	深層意味表現	32
3.7	まとめ	33
4	対話履歴管理	37
4.1	はじめに	38

4.2 ユーザ入力発話の形態の分類	38
4.2.1 分類カテゴリ	39
4.2.2 分類アルゴリズム	41
4.2.3 評価	43
4.3 対話の流れの表現方法	43
4.4 対話の流れの管理	44
4.5 対話履歴管理に付随する機能	48
4.5.1 焦点の管理	48
4.5.2 システム応答における省略・照応表現の使用	49
4.6 まとめ	50
5 応答内容立案	52
5.1 はじめに	53
5.2 話題に関連する知識の構成	53
5.2.1 単語辞書	54
5.2.2 話題依存規則	57
5.3 主題の推定	63
5.3.1 主題推定処理の概要	63
5.3.2 主題推定アルゴリズム	66
5.4 応答内容立案の手順	69
5.4.1 質問文入力の場合	70
5.4.2 要求文入力の場合	72
5.4.3 応答文入力の場合	73
5.5 知的システムへのアクセスとタスク依存規則	73
5.6 対話処理システムの動作例	74
5.7 まとめ	76
6 音声合成処理	80
6.1 はじめに	81
6.2 音声合成処理の概要	81
6.2.1 音声応答システムにおける音声合成	81
6.2.2 文生成処理	84

6.2.3	音響処理	84
6.3	対話調の発話の韻律的特徴に関する分析と検討	85
6.3.1	分析資料	85
6.3.2	発話速度に関する分析と考察	88
6.3.3	発話のパワーに関する分析と考察	91
6.3.4	基本周波数に関する分析と考察	91
6.4	対話調の音声合成のための音韻規則	102
6.4.1	音韻規則化	102
6.4.2	音声合成実験	104
6.5	まとめ	106
7	結論	110
	謝辞	114
	参考文献	115
	発表文献	119
A	話題に依存する知識の作成手順	121



# 目 次

1.1 プランの階層構造によって対話を表示する例 . . . . .	5
2.1 音声応答システムの全体構成の概略 . . . . .	13
3.1 模擬対話テキストの例 (書き起こし) . . . . .	22
3.2 模擬対話テキストの例 (不要語などを除去) . . . . .	23
3.3 模擬対話テキストの例 (文法的整合性をとったもの) . . . . .	24
3.4 音声対話処理システムの構成と情報の流れ . . . . .	28
3.5 深層意味表現の例 . . . . .	34
4.1 基本ルーチンの入れ子構造によって表現される対話の例 . . . . .	45
4.2 主張文が応答文として機能する例 . . . . .	51
5.1 単語辞書における単語の記述例 . . . . .	58
5.2 話題依存規則の具体例 . . . . .	64
5.3 質問文入力時の応答内容生成の流れ . . . . .	71
5.4 対話処理システムによる対話の例 . . . . .	78
5.5 対話処理システムにおける動作の概略の例 . . . . .	79
6.1 音声合成部の構成 . . . . .	83
6.2 直列、並列型回路併用型のターミナルアナログ型音声合成器の構成 . . . . .	86
6.3 マルチカスケード型音声合成器の構成 . . . . .	87
6.4 発話環境の違いによる発話速度の変化 (男声の例) . . . . .	89
6.5 発話環境の違いによる発話速度の変化 (女声の例) . . . . .	90
6.6 質問文の発話におけるパワーの時間変化 . . . . .	92
6.7 応答文の発話におけるパワーの時間変化 . . . . .	93
6.8 焦点が存在する発話のパワーの時間変化 . . . . .	94

6.9 基本周波数パターン生成過程のモデル . . . . .	96
6.10 無声化のため基本周波数の抽出が困難である例 . . . . .	98
6.11 質問文の文末において基本周波数の上昇がみられた例 . . . . .	99
6.12 応答文の基本周波数パターンの例 . . . . .	100
6.13 発話中に焦点がある場合の基本周波数の時間パターンの例 . . . . .	101
6.14 応答合成用音韻規則とテキスト合成用音韻規則による合成音声の基本 周波数 . . . . .	107
6.15 焦点情報の表出に関する音韻規則を用いた場合と用いない場合の合成 音声の基本周波数 . . . . .	108
A.1 意味マーカの一般的な例 . . . . .	123

# 表 目 次

2.1	本システムで使用するデータベースに記述されている情報の例 . . . . .	18
3.1	本システムにおいて用いる意味付与 . . . . .	36
3.2	本システムにおいて用いる深層格 . . . . .	36
4.1	発話形態分類アルゴリズムの評価 . . . . .	45
5.1	対話処理で用いる品詞の種類 . . . . .	56
5.2	意味マーカの例 . . . . .	56
5.3	主題の例 . . . . .	56
5.4	データ形式と疑問文の形態の組合せ . . . . .	59
6.1	テキストからの音声合成における韻律記号の種類と対応する指令の大きさ . . . . .	105
6.2	応答用音声合成におけるアクセント記号に対応する指令の大きさ . . . . .	105

# 第 1 章

## 序論

## 1.1 本研究の意義

現代社会が情報化社会と呼ばれるようになって久しい。その間に社会生活の様々な場面にコンピュータを用いたシステムが用いられるようになり、今やコンピュータは社会生活においてなくてはならない存在となるに至ったといえる。

このように社会においてコンピュータが浸透し、コンピュータに関する専門的知識をほとんどもたない人々が、一般の生活を営む上でコンピュータと接する機会が増えるにつれ、コンピュータと人間との間の情報伝達のぎこちなさが大きく取り上げられるようになってきた。そして、コンピュータと人間の間の情報伝達を人間にとって容易なものにしたいという要望が高まっている。

このような社会的状況から、人間とコンピュータの間の情報伝達手段、いわゆるマンマシンインターフェース(またはヒューマンインターフェース)に関する研究が、工学をはじめ、情報科学、心理学、医学などの幅広い分野で盛んに研究されるようになってきた。そしてその最も効果的な手段の1つとして、マンマシンインターフェースに音声を利用しようという試みが行なわれている。

音声は、言語的情報のみならず、発話者の意志や感情などいわゆるパラ言語・非言語情報をも同時に伝達することができる、人間のもつもっとも根源的な情報伝達手段である。このことは、文字言語をもたない民族は地球上に存在しても、音声言語をもたないものは存在しないということからもうかがい知ることができる。

マンマシンインターフェースに音声を用いることを目的として、これまでに音声認識、音声合成を中心とした研究が数多く行なわれてきている。音声認識の研究は、大語彙不特定話者連続音声認識をその究極の目的として、音声の変動の多い音響的特徴から、いかに多くの言語・パラ言語・非言語情報を取り出すかに研究の主眼がおかれている。一方音声合成の研究は、人間にとって明瞭で、かつイントネーション、アクセント、ポーズなど韻律的な情報が的確に付与された理解しやすい音声を合成することを目的としている。そして入力情報としてあらかじめ記述されたテキストを用い、それを適切に読み上げることを目的とする研究が主として行なわれてきている。

しかし、音声言語による円滑なコミュニケーションを目的とした場合、コンピュータにとっての情報の入出力技術である音声認識・合成の研究だけでは不十分で、コンピュータが音声による人間との対話を適切に処理できるようになることが必要不可欠である。しかし音声による対話は文字による対話と共通する点もあるが、音声

に固有な特徴も多く、単純に文字対話における手法を音声対話に適用することは困難である。例えば、音声対話処理の対象となる話し言葉においては、言い直しや言いよどみが起ったり、また主語や述語などの文の構成要素が省略されるなど、文字によって記述された文に比べてはるかに非文法的である。さらに日本語の場合には、文における語順の制約が緩く、格助詞などの省略も行なわれる。

このような特徴をもつ音声対話を処理するにあたっては、従来のテキストを対象とする自然言語処理でよく用いられるような入力文を形態素解析、統語解析して、さらに意味や意図を解析するというような処理を単純に適用することは困難である。このような背景から、音声対話における特徴を十分に考慮した対話処理の重要性が最近になって認識されはじめ、研究が行なわれるようになってきた。

人間とコンピュータとの間で行なわれる情報伝達の形態としては、人間がある目的をもってコンピュータにアクセスし、その目的を達成するために必要な処理をコンピュータに要求し、コンピュータがその結果について応答するというものが最も一般的であると考えられる。したがって、ユーザからの入力に応じて適切な応答を生成することを目的とする対話処理に関する研究は、音声対話処理における1つの重要な課題である。そしてこのためには、その時点における対話の状況を把握し、その状況に応じてユーザである人間に対して適切な応答を生成する機能を対話処理に持たせることが必要不可欠である。

一方、対話処理の結果をユーザに対して音声で応答する際には音声合成の技術が用いられる。これまでの音声合成の研究は、文字で記述されたテキストを入力として、それを読み上げるような音声の合成が主として研究されており、音声応答の研究としても、応答内容をいったんテキストの形で生成し、この技術をそのまま用いて音声応答を生成している研究も存在する。

しかし、一般に対話において発声される音声は、その韻律的特徴がテキストを読み上げる場合と異なり、また応答内容をいったんテキストで表現し、それに基づいて音声の合成を行なう方式では、対話処理の過程で得られるテキストには表現されない様々な深層レベルの情報を適切に合成音声に反映させることができない。そこで対話処理の結果として得られる情報を最大限応答音声の合成に反映させ、特に韻律の面でユーザにとって理解しやすいような応答音声を合成する方式を検討することも、非常に重要であるといえる。

## 1.2 本研究の背景

本節では、本研究の対象となる対話処理と音声合成に関連して、これまでに行なわれてきた研究の概略について述べる。

コンピュータがユーザに対して音声で応答する音声応答システムを考えると、音声合成の技術が必要不可欠である。これまでの音声合成に関する研究の多くは、人間によってあらかじめ作成されたテキストをよりよく読み上げることに重点がおかれている。しかし、音声応答システムに用いる音声合成としてはよりコンピュータとのインターフェースの取りやすい別のアプローチが考えられる。

そのアプローチの1つとして、コンピュータ上での情報の表現形態に近い知識表現をもとにして音声を合成するという研究が [Young et al. 79] 以来いくつか行なわれている。しかしこれらの研究は知識表現の音声言語への変換に主眼をおいたもので、音声応答システムを構築するためには、ユーザとコンピュータの間の対話の履歴を保存し、対話の流れを制御しながらその時点での対話の状況に応じた適切な応答を生成するような音声対話処理機能が不可欠である。

対話処理に関する研究としては、人間同士の対話をコンピュータに理解させるという観点からの研究として、[Litman et al. 87, 飯田他 90] などの例がある。これらの研究では、プランの概念をもとに primitive な人間の情報伝達行為を表わすプランと対話の domain に依存した対話の展開を表わすプランを階層的に用いることによって、人間同士の対話の理解を行なっている (図 1.1 参照)。しかしこれらの研究は、対話の理解という点に重点がおかれていて、適切な応答を生成するという点からは力点が異なっている。

応答を生成するという観点からの対話処理の研究もいくつか行なわれている。例えば、[Takebayashi et al. 92] では、まず対話の話題をファーストフードショップにおける注文に限定し、それに依存した形でシステム側とユーザ側にそれぞれいくつかの対話状態を設定している。そして、どのような入力があったときにどのような対話状態に移動し、システムはどのような応答をするかというように、対話の進行の状態遷移図を作成しておき、応答を生成するという手法を用いている。また [田路他 91] では、対話の話題を電話取次対話とし、対話の話題に依存する形で対話の進行をプラン形式で記述して対話処理を行なっている。またプラン形式による記述を用いて対話処理を行なう別の例としては、[Yamamoto et al. 91] などもある。しかしこれらの研究においては、対話処理の構成が対象とする話題に依存している部

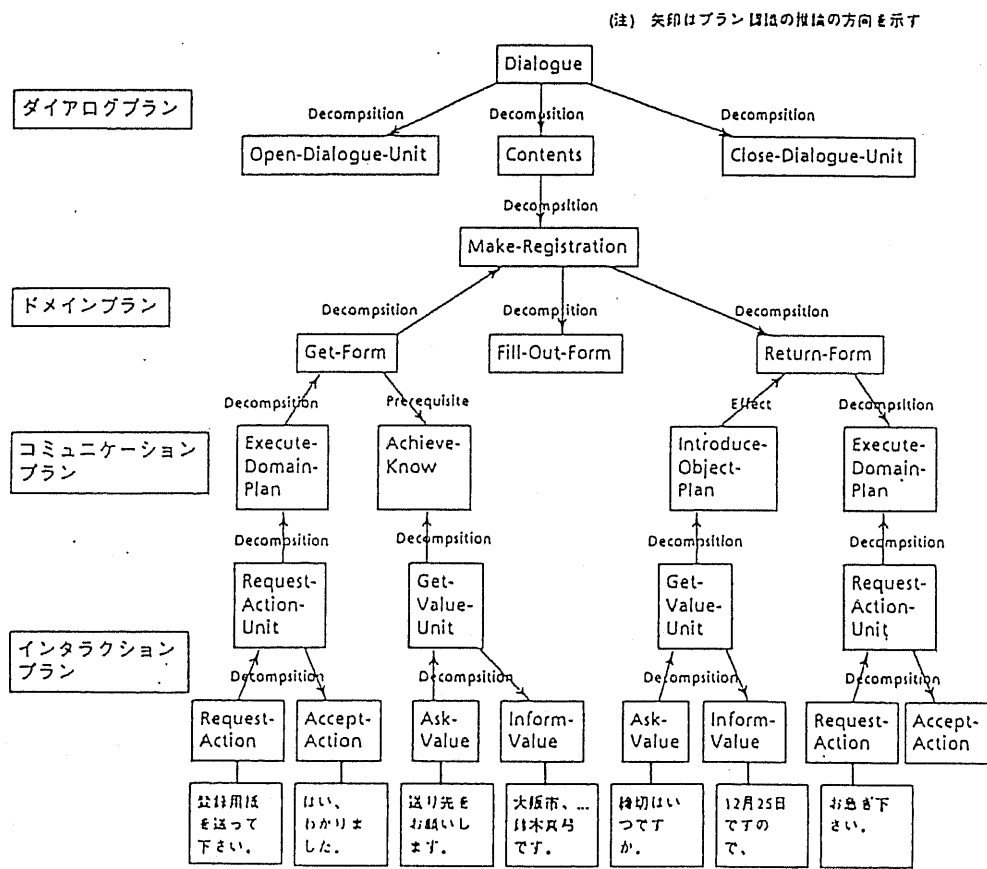


図 1.1. プランの階層構造によって対話を表現する例 (文献 [飯田他 90] より転載)



分が少なくなく、対話において扱う話題の範囲を拡大したり、別の話題に関する対話処理を新たに構築しようとするのが困難であると思われ、システムの汎用性という点はあまり考慮されているとはいえない。

一方音声合成に関して、テキストを入力とする音声合成方式の多くは、(1)入力されたテキストを解析する言語処理部、(2)言語処理部での解析結果に基づいて実際に合成される音声の分節的・韻律的特徴を記号レベルで記述した単音・韻律記号列を生成する音韻処理部、(3)単音・韻律記号列に基づいて音声合成器を駆動し音声波形の合成を行なう音響処理部の3つの処理部で構成される [Fujisaki et al. 90a]。

音声応答システムにおける音声合成を考えた場合、これらの処理部のうち、(3)音響処理部に関しては、そのまま適用することができるが、(1)言語処理部と(2)音韻処理部に関しては、対話処理によって応答内容として生成される情報によって、応答音声の単音・韻律記号列を生成する処理に置き換える必要がある。これらの処理に関連する研究としては、意味的記述から単音記号列を生成するルールに関して [Sato 90] の研究があるが、韻律的な検討に関してはほとんど行なわれていない。

### 1.3 本研究の目的

本研究は、音声応答を行なうシステムにおける対話処理と応答音声の合成処理を、実用性という点も考慮した形で実現することを目的とする。具体的には以下の4点に主眼をおく。

第1に、対話処理への入力情報として単語列を想定し、ある程度限定された情報から駆動することができる対話処理を実現することである。一般に対話処理への入力となる情報は、音声認識の結果として得られるものであるが、音声応答システムの場合に音声認識の対象となる発話は話し言葉である。音声は、読み上げ発話においても個人ごと、発声ごとの音響的特徴の変動が大きく、認識が困難である上に、話し言葉は読み上げ発話に比べると非文法的な発話が多く、そこから多くの情報を安定して引き出すのはかなり困難である。そこで本研究においては、対話処理への入力情報として、発話中で意味的に重要で音声認識において比較的安定して認識しやすいと思われる自立語を中心とする単語列を用いる。そして各単語の意味的、機能的な特徴に注目して、ユーザから入力された発話の機能的な形態の分類や、発話において言及されている主題の推定を行なうアルゴリズムを提案し、この情報に基づいて対話処理を進める方式を提案する。

第 2 に、対話処理に不可欠な対話の話題に依存する知識を局在化させ、さらにその記述を明確に規定することによって、対話処理機能を比較的容易に他の話題に対して適用させることができるような構成にすることである。本研究では、対話処理に必要な規則を、(1) 対話の話題に関係なく汎用的に用いることができる一般規則、(2) 対話の話題に依存して記述される話題依存規則 (T<sub>OPIC</sub> Dependent Rules: 以下 TODRs<sup>1</sup>と記述する)、(3) 対話を通じて音声応答システムがどのようなタスクを行なうかに関連して記述されるタスク依存規則 (T<sub>ASK</sub> Dependent Rules: 以下 TADRsと記述する) の 3 種類に分類して記述している。そして、これらの規則を用いる対話処理部本体は対話の話題に依存しない形で動作するように構築し、TODRs と TADRs を新たな話題に適合した形で構築することによって、比較的容易に他の話題に関する対話の処理を行えるような構成とする。さらにこのうち TODRs に関しては、従来よく用いられている plan-based な記述ではなく、実際の対話中において生じる事例に基づいて (case-based) 記述することとし、対話の話題に依存する規則を、具体的に記述しやすいようなものとする。

第 3 に、対話処理において応答音声の合成の際に韻律的特徴を制御する際に有効な情報を生成することである。自然言語インターフェースにおける対話処理としては、従来は応答する内容を表層レベルで表現したテキストを生成する点にのみ重点がおかれている。そして音声応答システムにおいても、対話処理の結果としては応答テキストを生成し、それをもとにテキストを入力情報とする音声合成システムで応答音声を生成するというものもある。しかしこのような方式では、対話処理の過程で得られる情報を十分に利用して、応答音声の合成においてユーザにとって理解しやすい音声を合成することは不可能である。本研究では、対話処理の過程で、応答音声の合成において韻律的特徴の制御に有効と思われる情報を生成し、それを用いた音声の合成を行なうこととする。

第 4 に、応答音声の合成において、対話中における発話の韻律的特徴の分析を行ない、話し言葉による自然な応答音声の合成を行なうことである。従来の音声合成での韻律的特徴の生成に関する規則は、アナウンサーなどがテキストを読み上げた発話の分析に基づいて作成されたものが多く、合成される音声は読み上げ調の合成音声となる。しかし、音声応答システムにおいては合成される音声はユーザとの間で

<sup>1</sup>後述するように、話題依存規則は複数の規則が個々に独立して記述されている。よって記述されている規則の 1 つ 1 つを指す場合は TODR、個々の規則の集合としての全体を指す場合は TODRs と記述する。

対話を構成するものであり、対話中の音声の特徴に基づいた形で音声を合成するのが望ましい。この点について、本研究では特に対話処理で得られる情報を基に、どのように韻律的特徴の制御を行なうかについて、音韻規則化を行なうこととする。

## 1.4 本論文の構成

本論文は全 7 章よりなっている。

第 2 章では、ユーザとの間で音声による情報伝達を行なうシステムである音声応答システムについて説明する。まず音声応答システムの概念を説明した後、本研究で想定している音声応答システムの具体的な構成を提示し、各構成要素の機能の概略を説明する。さらに、これらの構成要素のうち、本研究の直接の研究対象ではないデータベースと音声認識処理に関して、本研究としてどのようなものを想定しているかについて説明する。

第 3 章では、音声応答システムにおいて対話処理を行なう音声対話処理システムについて概説する。まず、音声対話資料の分析について説明し、音声の対話処理を行なうにあたって考慮する必要がある音声による対話の特徴について考察する。そしてこの結果を受けて、対話処理においてどのような対話を処理対象として扱うかについて考察、規定する。それに続いて、規定された対話を処理する対話処理システムの構成とその特徴を説明する。さらにこのシステムにおいて対話の履歴の保存や出力情報として用いられる深層意味表現について定義する。

第 4 章では、対話処理システムのうち、対話の状況の把握や対話履歴の保存、対話の流れの制御などを司る対話履歴管理機能について詳説する。まず、対話状況の把握や対話の流れの制御を行なうのに必要不可欠な情報である、ユーザから入力される発話の形態について、本研究における分類カテゴリと、その分類を行なうアルゴリズムを提案し、さらにそのアルゴリズムの評価を行ない有効性を検証する。次に、対話の流れを対話履歴として保存する際にどのように表現するかという表現方法と、それに基づいて対話の流れの制御を行なう汎用的な規則を提案する。最後に、音声合成において韻律的特徴を制御するのに重要な役割を果たす焦点情報の設定と、システム応答における省略・照応表現の利用について言及する。

第 5 章では、対話処理システムのうち、ユーザに対して対話状況に応じた適切な応答を生成する応答内容立案機能について詳説する。まず、ユーザの発話内容を把握し、適切な応答内容を生成するために必要不可欠である対話の話題に依存する知

識について、単語辞書と TODRs による記述を提案し、さらに TODRs の記述として、対話中に現れる事例に基づいて記述する方式を提案する。次に、ユーザ入力発話で言及されている主題を、発話中の単語から単語辞書の記述に基づいて推定する処理を提案する。そして、TODRs を用いてシステムがユーザに対する応答を生成する手順を、ユーザ入力発話の形態ごとに説明する。さらに応答を生成するために必要となる情報を得るために行なう知的システムへのアクセスと、アクセスの過程で対話処理に関連する処理を行なうための規則である TADRs について言及する。最後に、実際にユーザとの間で行なわれた対話を例に挙げて、対話処理システム全体の動作を説明し、それが有効に機能していることを示す。

第 6 章では、対話処理の結果として得られるユーザに対する応答内容を記述した深層意味表現に基づいて、音声を合成する処理について説明する。まず音声合成処理の概要として、音声応答システムにおいて求められる音声合成処理について、テキストを入力とする音声合成と比較して説明する。そして深層意味表現から表層レベルの情報を生成する文生成処理と、音声波形を生成する音響処理において用いられる音声合成器の構成について説明する。次に対話処理で得られる情報を合成音声の韻律的特徴に反映させるため、対話中の発話や同じものを読み上げた発話、さらに焦点を付与した発話の発話速度、発話のパワー、基本周波数パターンの分析を行ない、それに基づいて、音韻処理において韻律記号列を作成する音韻規則として規則化する。そして、その音韻規則に基づいて音声の合成実験を行ない、その有効性を検証する。

最後に第 7 章では、本論文の内容のまとめを行なう。

## 第 2 章

# 音声応答システム

## 2.1 音声応答システムの概念

コンピュータを用いたシステムと、そのユーザである人間との間で行なわれる情報伝達について考えてみると、ユーザはある目的をもってシステムにアクセスし、その目的を達成するために必要となる情報をお互いの間でやりとりするという形態が大部分であると思われる。

例えばプッシュホンによる列車の指定席予約システムの場合は、ユーザは座席の予約を行なうという目的をもってシステムに電話でアクセスし、両者の間ではシステム側の質問にしたがって、日時や列車番号などの情報をユーザがプッシュボタンを押して応答するというものである。また別の例として文献データベース検索を考えてみると、ユーザはある文献を探すという目的でデータベースにアクセスし、必要な検索条件の情報をシステムに与え、システムはその条件に適合した文献を応答としてユーザに提示する。

しかし、現在このようなシステムを利用しようとした場合、システムを利用しようとするユーザは、列車番号や駅名のコード番号を覚えたり、またデータベースアクセスのためのコマンドを覚えなければならない。すなわち両者の間の情報伝達(コミュニケーション)は、システムを中心として考えられており、ユーザはシステムの規定する機械的な規則にしたがうことを義務づけられることになる。しかもこの手順を誤ったり、システムの処理に必要な情報が抜けている場合などは、システムとしてはエラーであることを指摘する程度で、システムの方から不足情報を求めるような機能が備わっていないものが大半である。これらの点が、便利な機能をもつシステムの利用をごく一部の人に留めてしまう最大の原因となっていると思われる。

このようなシステム中心の情報伝達の形態を、ユーザ中心のものにするための手法としてはいくつかのものが考えられているが、そのもっとも有力なもの1つは、人間が普段の情報伝達に最も頻繁に用いる手段である音声言語をシステムとユーザの間の情報伝達に用いるというものである。

この目的に基づき、人間とコンピュータとの間で音声による双方向の情報伝達を実現するシステムが、本論文における音声応答システムである。すなわち音声応答システムとは、一般的にいつて次のような機能をもつシステムであると考えられることができる。

- システムとその利用者であるユーザとの間の情報伝達は音声言語の形で行なわれる。
- ユーザからシステムに対して発話が行なされると、システムはその発話内容と、その時点までにユーザとの間で交わされた対話の履歴を参照して、ユーザに対して適切な応答を返す。
- システムが処理を行なうのに必要な情報がユーザからの入力に存在しない場合は、システムの方から積極的にユーザに働きかけて情報を得る。
- ユーザからの入力とシステムの応答内容にもとづいて、システムは対話の履歴の更新を行ない、対話処理に必要な情報を内部に保存する。

このような機能をもつ音声応答システムを実現するためには、システムとしては音声認識、音声合成に加えて、音声対話処理という、大きく分けて3つの処理を行なうことが必要不可欠である。

## 2.2 音声応答システムの構成

前節で述べたような機能をもつ音声応答システムの構成としては、図 2.1 のようなものが考えられる。

全体的な処理の流れは、次のようなサイクルで行なわれる。

1. ユーザからの入力は、まず音声認識部によって処理され、その結果が対話処理部に送られる。
2. 対話処理部では、この入力情報と対話履歴ベースに蓄積されているこれまでの対話状況から、必要に応じて知的システムにアクセスし、ユーザに対する適切な応答を深層意味表現の形で生成する。
3. 音声合成部で、この深層意味表現をもとに合成音声の生成が行なわれ、システムの応答が音声の形でユーザに提示される。
4. 応答音声を聞いたユーザは次の発話を行ない、音声応答システムの処理はまた1に戻る。

各処理部における具体的な処理の内容は、だいたい次のようなものである。

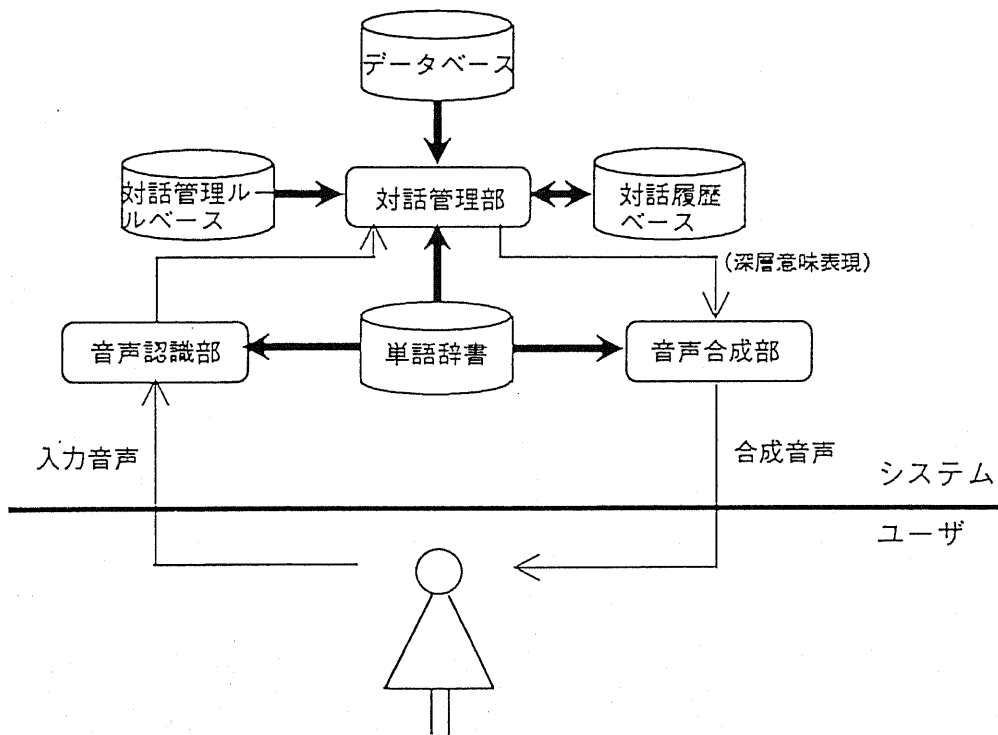


図 2.1. 音声応答システムの全体構成の概略



音声認識部 ここでは、ユーザからシステムに対して入力される音声波形を処理し、音響的な特徴を言語・知識処理に適した情報形態へ変換する。音声認識の方式としては様々なものが提案されており、認識結果としてどの程度の情報が得られるかも様々であるが、本研究では3.4節で述べるように、認識処理の結果として得られる情報として、ユーザからの入力音声に含まれる自立語を中心とする単語列を想定する。

対話処理部 音声認識部で得られた情報に基づき、まず発話中に含まれるユーザの意図や取り上げている内容(主題)の推定を行なう。そして対話管理規則に基づいて、ユーザとシステムとの間の対話の履歴を保存・管理し、対話の流れを制御する。そしてその制御に基づいて必要に応じてユーザに対するシステムの応答内容を深層意味表現の形で生成する。

音声合成部 対話管理部において生成された応答内容の深層意味表現をもとに、まず深層意味表現をもとに文法規則に基づいて統語情報の作成や述語の活用を行ない表層文を生成する。次にこの表層文や統語情報などをもとに、音韻規則に従い、異音化処理を施したりイントネーション・アクセントなどの韻律的情報を生成することによって、単音・韻律記号列を生成する。そして最後にこの単音・韻律記号列をもとに音響処理部において音声合成器を駆動し、合成音声を生じユーザに提示する。

また前節で規定したような応答システムを構築することを考えた場合、応答システムには対話の対象となる話題に関する知識を提供するエキスパートシステムやデータベースなどの知的システムが必要となる。これは同じような対話を人間が行なう場合を考えてみても、その話題に関する知識を全く持たない人は対話を行なうことが困難であるという点から考えても当然のことである。そして知らないことに関しては応答することができないように、応答システムの場合も知的システムによって提供される知識に関することだけしか応答することはできず、必然的に応答システムで扱うことのできる話題は限定される。

さらに音声認識、対話処理、音声合成の処理を行なうには、使用する語彙を登録した単語辞書が必要となる。しかし、対話中で使用される単語は、対話において言及される話題によって当然変わってくる。そこで単語辞書は、利用する知的システムによって提供される知識の分野に応じて用意すればよい。

ここで、本論文で用いる「話題」と「主題」という単語の違いについて説明しておく。話題とは、対話（人間同士の間と音声応答システムとユーザとの間の両方を含む）において対話中を通じて一貫して取り扱われている内容である。例えば、スキー場の案内や航空券の予約などである。これに対して主題とは、ある話題に関する対話の過程で、対話の主導権<sup>1</sup>をもっている対話参加者がその発話によって設定する細かい内容を指すものである。例えばスキー場の案内の場合はリフトの本数や料金、車での行き方などがこれにあたり、航空券の予約では出発時刻の問い合わせや運賃などがこれにあたる。

### 2.3 データベース

前節で述べたように、応答システムがユーザに対して適切な応答を返すためには、特定の話題に関する知識を提供できる知的システムの存在が必要不可欠である。

本研究では、市販されているスキー場に関するデータベースに基づいて知的システムの構築を行なった。そして応答システムで扱う対話の話題としては、スキー場の紹介及び案内とした。このデータベースを用いることとしたのは、次のような理由による。

- 対象分野に関する対話において用いられる語彙が数百語程度であり、比較的小規模の単語辞書を作成することによってシステムの性能評価が可能となる。
- データベースに記録されているデータの形式がテキストデータ、数値データ、2値データのすべてを網羅しており、データ形式の違いによる応答システムでの処理の違いなども検討できる。

このスキー場データベースに登録されているデータのうち、応答システムで利用しているものを表 2.1 に示す。

### 2.4 音声認識処理

音声応答システムにおいて、対話処理に必要なユーザからの入力発話に関する情報を得るための処理が音声認識処理である。

音声認識の方式としては、最近では Hidden Markov Model (HMM) や Neural Network (NN) など統計的手法を用いた認識手法が精力的に研究されている [中川 88]。さ

<sup>1</sup>対話の主導権に関しては、3.5節を参照。

らにこれらの音響的特徴のマッチング処理に加え、LR パーザや言語的統計モデルによる制約 (N-gram) を用いて認識率の向上をはかる方法が数多く提案されている [Sawai 90, Maltese 92, 山口他 92]。

しかし、現在の音声認識において処理の対象となっている発話としては、テキストの読み上げなどの比較的文法的整合性をともなった発話が多い。これに対して、音声応答システムにおける音声認識の対象となる発話は、システムとユーザとの間で行なわれる対話中の発話であり、これは読み上げ発話などとは異なる、いわゆる話し言葉による発話である。具体的には、ユーザはシステムの発話を受けてから次の発話を考えて発話することになるので、話し言葉に特徴的な言いよどみや言い直しが含まれたり、発話が非文法的なものとなる場合も存在する。そしてこのような発話を対象とした音声認識の研究は、まだその緒についたばかりである。

話し言葉のように非文法的な要素を含む発話を解析するのに、記述されたテキストを対象とする自然言語理解の処理でよく用いられるように、形態素解析、統語解析などを段階的に施して、徐々に高次の情報を求めるという方式をそのまま適用するには困難であろうということは想像に難くない。したがって、このような発話から対話処理を行なうのに必要な情報を安定して得るためには、意味的に重要な部分を中心として音声認識を行なう方法が有効であると考えられる。

以上のような考察に基づき、本研究では、音声対話処理を行なうのに必要な情報として自立語を中心とする単語列を想定し、音声認識処理においてこの程度の認識が安定して行えれば対話処理を行なうことができるような対話処理方式を検討することとした。自立語は発話中において意味的に重要な場合が多く、そのため発声においても比較的是っきりと発声されやすく、安定に認識されやすいと考えられる。

このように発話中から単語列を得るための音声認識手法の 1 つにワードスポッティング法がある。これはあらかじめ単語辞書中に登録されている単語を発話中から抽出して認識する手法で、現在の不特定話者連続音声認識の中で、比較的高い認識率を得ている [Wilpon et al. 90]。またこの方式には、原理上不要語<sup>2</sup>の除去が容易に行えるという利点もある。

またこのように単語列を対話処理の入力として想定することによって、将来入力情報を単語ラティスに拡張することが容易であるという利点もある。単語ラティスとは、多くの音声認識処理方式において、その処理の中間過程で生成されるもので、

<sup>2</sup>不要語については 3.2.1 節参照

認識処理によって認識対象である発話中に含まれていると推定された単語の候補を、発話の文頭から文末に向けてその確からしさの順番も含めて並べたものである。

後述するように、本研究における対話処理では、入力情報である単語の意味的な情報に基づいて対話処理を行ない、その処理の過程で意味的な整合性をチェックするような機能が含まれている。したがって対話処理の入力情報として単語ラティスを用いた場合には、この機能に関する情報を用いて、意味的に望ましいと思われる単語を単語ラティス中から選択することが可能となる。すなわち対話処理において、音声認識に有用な情報を提供することが可能となり、音声認識の認識率の向上に寄与することができる。

表 2.1. 本システムで使用するデータベースに記述されている情報の例

数値データ	2 値データ	テキストデータ
リフト料金	ナイターあり	所在地
リフト数	ポール練習可	問い合わせ先
駐車場料金	託児施設あり	
駐車台数	温泉あり	
標高	初滑り可	
コース数	春スキー可	
最大滑走距離	ポール練習可	
最大斜度	スノーボード可	
コース難易度割合		

## 第 3 章

# 音声対話処理

### 3.1 はじめに

音声応答システムにおける対話処理部の機能を実現するものが、音声対話処理システムである。

対話処理システムの目的は、ユーザから発話が入力された時、その内容とこれまでのユーザと応答システムの間で行なわれた対話の内容(対話履歴)に基づいて、適切な応答を深層意味表現の形で生成することである。このような処理を実現するためには、対話処理システムとしては次の3つの機能を備えていることが必要不可欠である。

- ユーザから入力された発話内容から、ユーザがどのような主題に関して、どのような意図をもって発話を行なったのかを、対話履歴も参照することによって推定する。
- 前項で推定された結果に基づき、そのユーザ発話が全体の対話中で果たす役割を対話履歴に基づき推定し、対話の流れを制御する。同時に、このユーザ発話を新たに対話履歴の中に登録し、対話履歴の更新を行なう。
- 前項の対話の流れの制御に基づいて、ユーザに対して適切な応答内容を生成する。同時にこの応答内容に基づいて対話履歴を更新する。

本章では、以上のような機能をもつように構成された音声対話処理システムについて、その概要について言及する。まず人間同士が行なった音声対話資料を収集し、それを分析して音声対話の特徴について検討する。さらにこの検討に基づいて、本研究における音声対話処理システムの処理対象となる対話や目的を明確にし、実際にシステムとして構築するために対話処理システムに導入される制約について説明する。そして、これらの条件のもとに構築された対話処理システムの特徴と、システムの全体構成について説明する。最後に、この対話処理システムで用いられる深層意味表現について言及する。

## 3.2 音声対話資料の分析

### 3.2.1 音声対話資料

音声による対話における個々の発話は、1.1節で述べたように文字によって記述されたテキストとは異なる点が多い。したがって、音声対話処理を考えるにあたって

は、音声による対話が文字による対話(例えばUNIXのtalkなど)とどの様な点で異なり、それを扱うためにはどの様な処理が必要となるのかという方針を立てる必要がある。そこでこのような観点から、人間同士が行った音声による対話の資料を収集し、その特徴を分析し、音声対話の特徴に関する考察を行なった。

音声対話の特徴の分析を行なうにあたって、収集した音声対話資料は次の2種類である。

- 日本音響学会連続音声データベース調査委員会の模擬対話テキスト[板橋他 92]  
これは、観光案内や大学案内などのあらかじめ決められた話題に対して、2人の人間が行なった対話を、次の3通りの基準にしたがってそれぞれテキスト化したものである。なおここでいう不要語とは、発話中における「あー」、「えー」などの言語的な情報の伝達とは直接関係のない発話のことをさす。

基準1 不要語や言いよどみ・言い直しなども含めてテキスト化(例:図 3.1)

基準2 不要語や言いよどみ・言い直しなどを除去してテキスト化(例:図 3.2)

基準3 語順などを調整し、文法的整合性をとった形でテキスト化(例:図 3.3)

この模擬対話テキストのうち、実際に対話資料として用いたのは、標準語(東京方言)で行なわれた対話のみとした。総対話数は27対話で、対話中での総発話数は1700発話<sup>1</sup>である。

- 独自に収録したスキー場を話題とした対話  
これは、数人の成人男女を2人1組にして、一方を質問者、他方を応答者というように役割を与え、その役割に応じて自由にスキー場を話題とする対話を行なってもらったものを収録したものである。さらにこれを前項の基準1でテキスト化した。総対話数は10対話、総発話数は318発話である。

### 3.2.2 音声対話の特徴

前節で説明した音声対話資料を分析した結果、音声対話に固有の特徴としては以下のものが挙げられる。

1. 「あー」、「えー」などの直接対話の流れに関係しない発声(不要語)が発話中に挿入される。

<sup>1</sup>発話数の定義としては、主として発話している人が交代するまでの発話のまとまりを1発話と数えた。例えば、図 3.1における発話数は11である。



Q: もしもし、浅野ですが、御無沙汰しております。

お元気でしょうか。

A: [あ、] 浅野君ですか。

どうしてます。

お変わりありませんか。

お久しぶりですねー。

Q: あいかわらず元気にやっています。

[えー、] ところで、あしたそちらの電子工学科の会議室で会議があるんですけども。

どうやっていったらいいでしょうか。

A: どっから来ますか。

Q: [えーと、] 新宿から行きます。

A: 新宿からだ、いちばんいい方法が、[えー、] (お茶の水、) [えー、] [えーと、] (何だ、山の手線じゃないや、中央線か、) 中央線のお茶の水で降りて、それから、

学バスに乗って来るのがいちばん近いかな。

Q: 学バスって何ですか。

A: 学バスは、あの、東大構内行の都営バスのことです。

Q: [おー、] そのバスに乗ってどうしたらいいんでしょうか。

A: [えー、] まず終点で降りて、[えー、] (お茶、) お茶の水から乗って終点で降りて、 [えー、] 降りたら歩いて、まっすぐ歩いて、[えー、] すぐ左に曲がって工学部 3 号館が有ります。

[もー、] 歩くのは 2、3 分。(うん)

その 1 階に、[えー、] 入口はいった右っかわに、[えー、] 会議室があります。

{はー}

そこだからすぐわかると思いますよ。

Q: はい。

でも、バスだとちょっと渋滞したりして時間をわかんなくなるんでー、バス使わない方法もありますか。

図 3.1. 模擬対話テキストの例 (書き起こし)

- Q: もしもし、浅野ですが、御無沙汰しております。  
お元気でしょうか。
- A: 浅野君ですか。  
どうしてます。  
お変わりありませんか。  
お久しぶりですねー。
- Q: あいかわらず元気にやっています。  
ところで、あしたそちらの電子工学科の会議室で会議があるんですけども。  
どうやっていったらいいでしょうか。
- A: どっから来ますか。
- Q: 新宿から行きます。
- A: 新宿からだど、いちばんいい方法が、中央線のお茶の水で降りて、それから、  
学バスに乗って来るのがいちばん近いかな。
- Q: 学バスって何ですか。
- A: 学バスは、あの、東大構内行の都営バスのことです。
- Q: そのバスに乗ってどうしたらいいんでしょうか。
- A: まず終点で降りて、お茶の水から乗って終点で降りて、降りたら歩いて、  
まっすぐ歩いて、すぐ左に曲がって工学部 3 号館が有ります。  
歩くのは 2、3 分。  
その 1 階に、入口はいった右っかわに、会議室があります。  
そこだからすぐわかると思いますよ。
- Q: はい。  
でも、バスだとちょっと渋滞したりして時間をわかんなくなるんでー、  
バス使わない方法もありますか。

図 3.2. 模擬対話テキストの例 (不要語などを除去)

- Q: もしもし、浅野ですが、御無沙汰しております。  
お元気でしょうか。
- A: 浅野君ですか。  
どうしてます。  
お変わりありませんか。  
お久しぶりですね。
- Q: あいかわらず元気にやっています。  
ところで、あしたそちらの電子工学科の会議室で会議があるんですが、  
どうやっていったらいいでしょうか。
- A: どこから来ますか。
- Q: 新宿から行きます。
- A: 新宿からだ、中央線のお茶の水で降りて、それから、学バスに乗って来るのが  
一番近くて良いでしょう。
- Q: 学バスって何ですか。
- A: 学バスは、東大構内行の都営バスのことです。
- Q: そのバスに乗ってどうしたらいいんでしょうか。
- A: まず、お茶の水から乗って終点で降りて、降りたら次にまっすぐ歩いて、すぐ  
左に曲がれば工学部 3 号館が有ります。  
歩くのは 2、3 分です。  
その 1 階の入口を入った右側に会議室があります。  
そこだからすぐにわかると思いますよ。
- Q: はい。  
でも、バスだとちょっと渋滞したりして時間が分からなくなるので、  
バスを使わない方法もありますか。

図 3.3. 模擬対話テキストの例 (文法的整合性をとったもの)

2. 言い直しや言いよどみが発話中に起こる。
3. 個々の発話は比較的短く、対話の途中で相手の発話に対して割り込み的に質問などが発生する場合がある。そのため、一時的に質問と応答の立場が容易に入れかわり、対話の主導権の移動が頻繁に発生する。
4. 聞き取りミスによる誤解等が生じやすい。
5. 対話内容の保持が人間の脳によるため（テキストとして残らない）、過去の内容との重複や、再確認などが生じる。

対話の途中で起こる割り込みは、一般的に協調的なものと競合的なものの2種類に分類することができる [飯田 88]。

協調的な割り込みとは、相手の発話内容の一部が理解できずに質問する場合など、現在の対話の流れに沿った上での割り込みで、この場合は割り込みで要求された処理を実行した後に、対話の流れをもとの状態に戻すことができる。これに対して競合的な割り込みとは、一方の発話目的が達成されないうちに、他方から現在の対話の流れに無関係な別の主題に関する発話が行われるというものである。そして実際の対話に見られる割り込みのほとんどは対話の進行に沿って行なわれる協調的なものである。

以上であげた音声対話における特徴のうち、1と2については、人間は発話内容を考えながら発声するためにこのような現象が生じるので、システムが音声で応答する場合にはこの現象を実現する必要はないと考えられる。これに対し3～5については、対話の流れに直接関係する問題で、音声応答を目的とする対話処理においてはこれらに対処する必要がある。

### 3.3 処理の対象となる対話

本節では、音声対話処理システムを構築する場合、その処理対象となる対話として、どのようなものを設定すべきかについて考察する。

人間がコンピュータシステムに対してアクセスする場合を考えると、通常はユーザである人間はコンピュータシステムに対してある目的をもって(列車の指定席を予約する、データベースを検索して必要な情報を得るなど)アクセスを行なう。そして、両者の間でその目的を達成するのに必要な情報のやりとりが行なわれる。

このようなユーザとシステムとの間で行なわれる目的指向の情報伝達を、そのまま音声言語を用いた両者の間の対話によって行なうことを考えた場合、この対話の大筋は広い意味で質問応答型の対話であるとみなすことができる。質問応答型対話とは、対話参加者のうちの一方がある知りたい情報に関して、他方に対して質問をすると、質問された側はそれに対する答えを応答するという形の対話である。

例えば、列車の指定席の予約の場合は、システムの側が日時、列車名、乗降駅などの情報をユーザに質問し、ユーザがそれに応答することで予約が行なわれる。一方データベース検索の場合は、逆にユーザがシステムに対して得たい情報に関する質問を行ない、システム側がそれに対して応答を行なうと考えることができる。

またこのような情報伝達は、たいていの場合はシステムとユーザの間で1対1で行なわれると考えられる。

以上の点を考慮して、実用的な観点から、本研究では音声対話処理システムの処理対象となる対話を、目的指向の対話の典型である質問応答型対話にしぼり、システムは1人のユーザとの間で1対1で対話を行なうものとしてシステムの構築を行なうこととした<sup>2</sup>。

さらに、3.2.2節でも述べたように、通常の問題型対話の音声対話においては、協調的な割り込みが生起して、発話の主導権が移動することは生起するが、競合的な割り込みはほとんど生起しない。したがって、対話処理システムの処理対象とする対話は協調的に行なわれるものとし、割り込みに関しては協調的なものだけに対処することとする。

実際にどのような場合に協調的な割り込みが行なわれるかということ、相手の発話中にわからない点があって聞き返す場合や、相手からの質問に、その質問に答えるのに十分な情報が含まれていないときに、相手に対して不足する情報を問い合わせるような場合が考えられる。

2.2節で述べたように、応答システムとしては、付随する知的システムによって与えられる範囲のことにしか応答することはできない。また言語処理的な面から考えても、汎用的な知識だけを用いて、あらゆる話題の対話に対してすべて処理できるようなシステムを構成することは不可能である。したがって対話処理システムとしては、扱うことのできる対話は必然的にある範囲の話題に限定される。

<sup>2</sup>特殊な事例を対象としたもので、航空管制官のトレーニングシミュレータ用として、複数の航空機のパイロットの応答を行なうという同時に複数の対話を行なうようなシステムの研究 [Howes 89] も行なわれている。

しかし対話で扱う話題が限定されているからといって、特定の話題に依存した処理アルゴリズムで対話処理を行なうのではなく、処理アルゴリズムとしてはなるべく一般性を持たせ、話題に依存する知識を必要に応じて記述することによって目的とする対話処理システムを構築することとする。

以上をまとめると、本研究における対話処理システムの処理対象とする対話は、ユーザとシステムの間で協調的に行なわれる質問応答型の対話で、付随する知的システムによって提供される情報により対話の話題が限定されるものとする。

### 3.4 システムの構成

音声対話処理システムは、対話履歴管理部と応答内容立案部の2つに大きく分けて構成されている。システムの構成とシステム内での情報の流れを図 3.4 に示す。

対話処理システムへの入力情報としてどのようなものにするかの決定には、音声認識処理の性能が大きく影響を与える。3.2.2節で述べたように、音声対話における個々の発話は、不要語の挿入や主語・述語の省略などが頻繁に生じ、非文法的な発話が数多く存在する。そのため音声認識においてこのような発話の一字一句を認識することは、テキストの読み上げ文を認識するのと比べてはるかに困難である。そこで対話処理システムへの入力情報としては、発話中における意味的重要度が高く、比較的安定して認識されやすい自立語を中心とした単語列を用いることとした(2.4節参照)。

一方、対話処理システムの出力情報としては、ユーザに対して応答すべき意味内容を規定した深層意味表現を用いる。この深層意味表現については、詳しくは3.6節で述べるが、この記述内容に基づいて処理が行なわれる音声合成処理において、特に合成音声の韻律的特徴を制御するのに必要となる情報も記述できるようにしている。

対話履歴管理部の機能は、まずユーザから入力された発話を解析し、後述するアルゴリズム(4.2.2節、5.3節)にしたがって、その発話の形態の分類と発話で言及されている主題の推定を行なう。そして推定されたユーザ発話の形態と、これまでの対話の履歴をもとに、この発話の対話中での役割を推定し、それにそった形で対話の流れの制御を行なう。同時に、この発話も含めたかたちで対話履歴を更新する。さらにシステムの応答音声を合成する際に重要な情報となる焦点位置の制御や、システムからユーザへの応答における省略・照応表現の処理も、対話管理部で取り扱う。

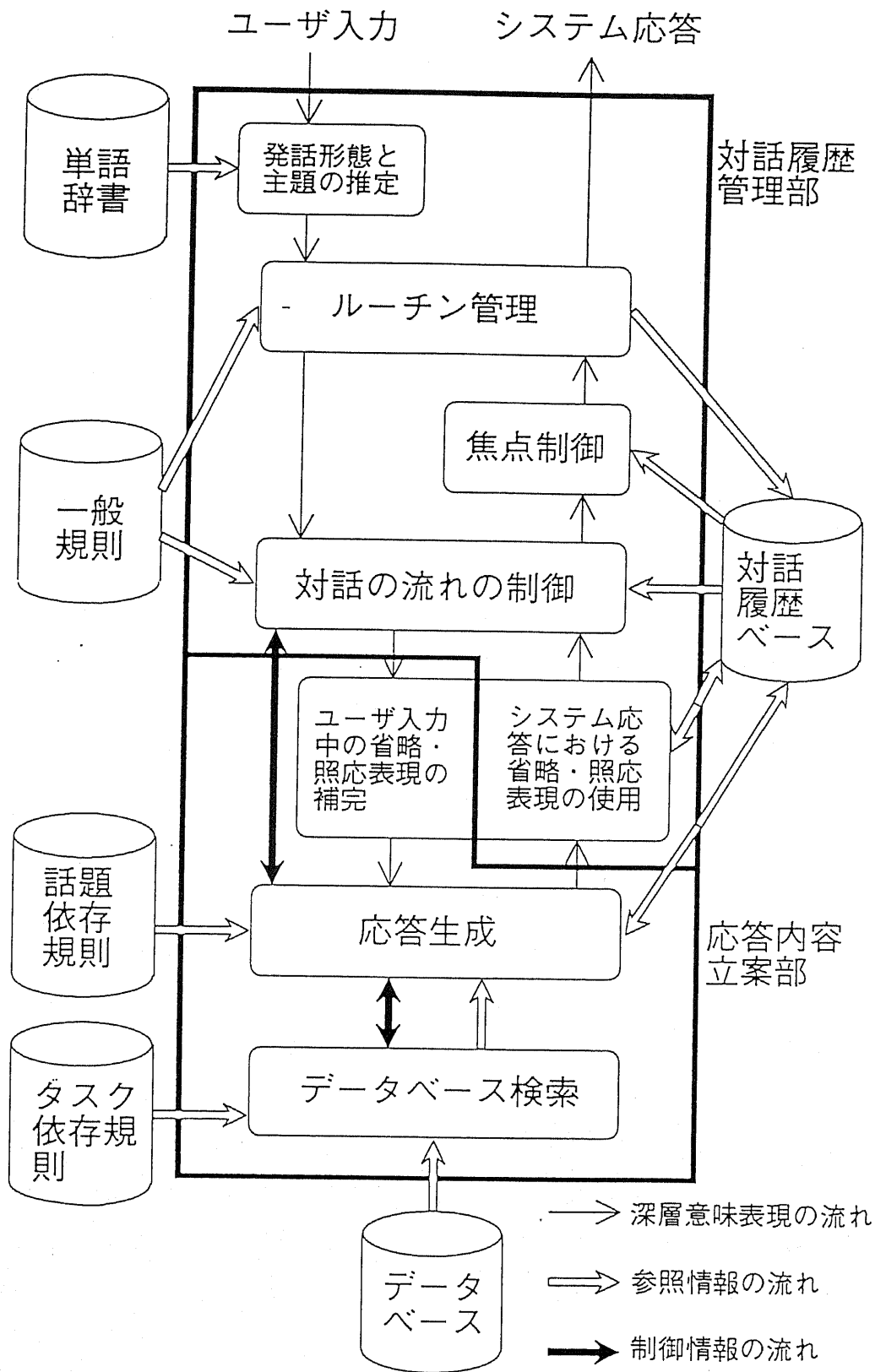


図 3.4. 音声対話処理システムの構成と情報の流れ

一方応答内容立案部では、対話履歴管理部での対話の流れの制御にしたがって、システムが発話する必要がある対話状態においてユーザに対する応答内容の立案を行なう。具体的には、まず対話履歴部で推定されたユーザ発話の形態分類と主題の推定結果からユーザがその発話で要求している情報を判断する。そしてユーザ発話中で省略されている情報を対話履歴から補完し、可能であれば知的システムにアクセスして、その結果を基に応答内容を生成する。また情報不足でアクセスができない場合は、その不足情報をユーザに対して質問する発話を生成することによってアクセスを可能にするための情報収集を行なう。

### 3.5 システムの特徴

以上が本研究において構成した音声対話処理システムの概要であるが、個々の処理の詳細については次章以降で説明することとし、本節ではこのシステムの特徴について、箇条書形式でまとめてみる。

#### ・自立語を中心とする単語列を入力とする処理

音声対話における発話には、不要語の挿入や主語・述語の省略など文字言語としてみた場合には非文法的と思われる文が頻繁に発生する。このような文を処理するのに、従来の文字言語で記述された文を扱う際によく用いられる形態素解析、統語解析などを適用して実効性をあげるのは困難である。

また音声認識の面から考えても、自立語などの意味的に重要な部分については、比較的認識がしやすいが、短い助詞や文末の助動詞などを正確に認識する事はまだ難しい状態である。

これらの点を考慮して、対話処理システムに対する入力情報としては、発話を構成する自立語を中心とした単語列を用いることとし、個々の単語の意味に重点をおいた処理を行なうこととした。具体的には、ユーザから入力される発話中の単語から、その発話の形態と言及されている主題を推定するアルゴリズムを提案し、この推定された情報に基づいて対話処理を行なう方式を提案した。

ただし、入力情報として自立語のみの情報を用いるため、次のように意味的に同じ属性をもつものが異なった項目に対して用いられる対話に対しては、対処することは不可能である。

「東京から大阪までの列車」



## 「東京までの大阪からの列車」

この2つの発話の場合は、自立語だけを取り出すと「東京」、「大阪」、「列車」の情報しか得られないので、この2つの発話の違いを検出することができない。

このような発話に対処するためには、「から」、「まで」などの助詞も正確に認識することが必要不可欠である<sup>3</sup>。本システムの構成としては、音声認識の結果としてこのように助詞などの認識も可能な場合も想定して、後から処理を追加することによってこれらの発話に対処できるような拡張性を持たせている。

## ・対話の基本ルーチンによる対話履歴の管理及び対話の流れの制御

対話履歴の管理と対話の流れの制御を行なうに際して、対話状態を記述する手法として、質問・要求系の発話とそれに対する応答の発話を対にして扱う対話の基本ルーチンという概念を導入した。そしてこの概念に基づいて、発話ごとに対話の履歴を管理し、さらに対話の流れを制御する手法を提案し、対話履歴管理部に実装した。

## ・対話主導権の動的な遷移(協調的割り込みへの対応)

対話主導権とは、対話の流れの中で、次にどのような主題に関して話を進めるかを決定することができる権利のことである。そして、この主導権をユーザとシステムにどのように割り当てるかによって、ユーザとシステムの間での情報伝達の形態が決ってくる。例えば、プッシュホンによる列車の指定席の予約システムの場合は、システム側が次にユーザが入力すべき情報を常に指定し、それに対してユーザはただ応答するだけで、対話主導権は常にシステム側に存在する(システム主導型対話)。これに対してデータベース検索の場合は、ユーザが項目を問い合わせ、それに対してシステムが情報を提供するので、対話主導権はユーザ側にあるといえる(ユーザ主導型対話)。

3.2.2節で述べたように、音声対話においては協調的な対話への割り込みが頻繁に生起し、音声対話処理システムとしてはこれに対処する必要がある。この協調的割り込みを対話の主導権という観点から考えてみると、協調的割り込みに対処するためには、対話の主導権がユーザとシステムの間で動的に遷移するような対話処理方式を用いることが必要であるといえる。

本システムでは、先の対話の基本ルーチンに基づいて、これを入れ子状に適用す

<sup>3</sup>このような場合の「から」「まで」などはこの助詞自身が意味的に非常に重要であるため、はっきり発声される傾向にあり、比較的認識は行ないやすい。

ることによって、協調的な割り込みに対して対処できる対話の流れの制御方式を用いている。

#### ・対話処理規則のうち、対話の話題に依存する部分の分離

対話処理を行なうために必要となる対話処理規則としては、次の3種類に分類して、システムに実装している。

**一般規則** 対話履歴の管理や対話の流れの制御に関する規則で、本システムが対象としている協調的に行なわれる質問応答型対話全般に対して、汎用的に用いることができ、対話において取り扱う話題には依存しないで記述することができる。この規則は対話履歴管理部において手順として記述されている。

**TODRs** 質問応答型の対話では、応答するために必要となる情報や、知的システムへのアクセスの形態などが、対話で取り扱う話題によって変わってくる。このような話題に固有の対話進行上の知識を記述したものが TODRs である。

**TADRs** システムがユーザとの間の対話を通じて、どのようなタスクを実行するかという点に依存する規則である。本システムでは対話を通じてデータベース検索を行なうことを目的としており、それに基づいて規則が記述されている。

このうち TODRs に関しては、対話処理システムを他の話題に対して適用することを考えたとき、システム管理者の規則記述の負担を低減するために、比較的単純で具体的な記述を用いることを目的とし、処理対象となる話題の対話中において、実際に生起する対話事例に基づき、それをフレーム形式で規則として記述する方式を提案した。これは人工知能の分野における事例に基づく推論方式 [Kolodner et al. 85, 小林 92] を対話処理に対して応用したと解釈することもできる。

#### ・応答音声合成用の情報の生成

ユーザにとって理解しやすい音声を合成するためには、読みにあたる音声の分節的特徴が十分な明瞭度をもっているのはもちろん、イントネーションやアクセントなどにあたる韻律的特徴が適切に制御されている必要がある。そしてこの韻律的特徴の制御には、文中のどこに焦点がおかれているか、この発話がどのような意味情報を表出しようとしているかといった情報が重要な役割を果たしている。

本対話処理システムでは、この焦点位置に関する情報や意味情報を、対話処理の過程で生成し、音声合成処理においてこの情報を利用して適切な韻律的特徴を付与

された音声の合成を行えるようにしている。

### 3.6 深層意味表現

深層意味表現とは、表層の文が包含している意味内容に関する情報だけを抽出して表現したものを一般的には指す。したがって、統語構造などの文の表層構造に関する情報は含まれず、また理想的には個々の表層言語(英語や日本語などの個々の言語)にも依存しない抽象度の高い形態で表現される。

深層意味表現には一階述語論理や R.C.Shank による CD 表現 [Schank et al. 81] などいくつかの表現方法が提案されているが [西田 88]、本システムにおいては研究の目的上、対象となる言語は日本語だけであるので、中間言語表現を用いた (pivot 方式の) 機械翻訳システムなどのように言語間の差異を吸収するために深層意味表現の抽象度を高める必要がないことから、深層意味表現としては Fillmore によって提唱された格構造の概念 [Fillmore 68] を用い、個々の深層格を占める要素に直接表層文の構成要素(文を構成する単語)を用いることとした。

音声認識の認識率が向上し、認識結果としてユーザ発話の正確な深層意味表現が得られるようになれば、音声認識部と対話処理部間の情報伝達の形態に深層意味表現を用いるのが理想的である。しかし現状の音声認識のレベルではこれは困難であり、本システムでは両者間の情報伝達の形態を単語列で行なっている。したがって本システムでは、深層意味表現は対話処理部内における対話履歴の保存と対話処理部と音声合成部間の情報の伝達に用いられている。

本来の格構造表現には、意味の中心となる述語と、その他の文の構成要素が述語に対して意味的に果たす機能(深層格)とともに記述されている。しかし、これを対話処理部と音声合成部間の情報伝達の形態として用いることを考えた場合には、これだけの情報記述では不十分で、音声合成を行なうのに必要となる情報を表現するのに必要十分な拡張を行なう必要がある。

具体的な格構造としては、次のようにリスト形式で記述することとした。

[述語, [意味付加], [文脈情報], [[深層格 1, 構成要素 1],[深層格 2, 構成要素 2], … ]]

この記述のうち、述語は発話中の意味的中心となる述語で、これに対する深層格とこの格を占める発話の構成要素の組をリストにして表現している。これらが発話文の意味的内容の骨格をなす部分である。

意味付加とは、表層文において述語の活用と助詞、助動詞の付加によって表現される意味情報を表わし、本システムでは表 3.1 に示すものを用意した。同時に対話の状況から焦点がおかれている深層格をここに記述し、音声合成の際の韻律的特徴の制御にこの情報を利用する。

また文脈情報とはこれまでの対話状況から得られる情報に基づいて、この格構造から表層の発話文を生成するのに関連する情報を記述するもので、具体的には、対話処理の結果から、表層の発話文においては省略・照応表現を用いるべき深層格をこのリストに記述する。

この格構造において用いる深層格としてどのようなものを用意するかについては、機械翻訳システムの開発を目指した Mu プロジェクトにおいて用いられている深層格を参考にして決定した [辻井 91]。Mu プロジェクトにおいては動詞・名詞間の深層格として全部で 34 個が用意されているが、このうち本研究で対話処理の話題としたスキー場の案内に必要であると考えられる 15 個を深層格として用いることとした。本システムにおいて用いる深層格の一覧を表 3.2 に示す。

この用いる深層格を選択する際には、一般的によく用いられるものについてはだいたい網羅しているが、本システムを他の話題に適用するには若干深層格の種類を追加する必要がある場合も考えられる。これに対しては、後述する単語辞書と TODRs の記述の際に利用したい深層格を記述すればよく、対話処理のアルゴリズムとしては特定の深層格の種類に依存していない。したがって、深層格の種類を追加、変更は非常に容易である。

最後に深層意味表現の例として、リスト形式で表現された深層意味表現と、これを意味ネットワーク形式で表現したもの、及びこれによって表現される表層文の一例を図 3.5 に示す。

### 3.7 まとめ

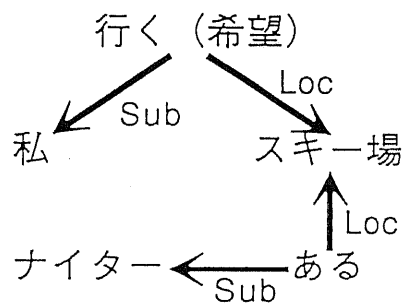
本章では、音声応答システムにおいて対話処理を行なう音声対話処理システムについて、その概要を説明した。

はじめに音声対話処理を考えるにあたって、音声対話資料を収集し、その分析を行なって、音声対話に特有の特徴について考察した。

次に分析結果に基づき、ユーザとコンピュータを用いたシステムとの間で行なわれる情報伝達の形態について考察し、音声対話処理システムの対象とする対話とし

(行く (希望) ( $\phi$ )  
( (主体 私) (場所 スキー場[1]) ) )  
(ある ( $\phi$ ) (限定)  
( (主体 ナイター) (場所 スキー場[1]) ) )  
[1]は同じものを指していることを示す添字

リスト形式による格構造表現



対応する表層文の 1 例

私はナイターのある  
スキー場に行きたい。

ネットワーク形式による表現

図 3.5. 深層意味表現の例

て、協調的に行なわれる質問応答型対話を設定した。

そして、対話処理システムの構成として、入力情報としては自立語を中心とした単語列とし、出力情報としては深層意味表現を用いることとした。そして、この間の処理で求められる機能を大きく対話履歴管理と応答内容立案に分け、その概要を説明し、さらに対話処理における本研究の特徴についても言及した。

最後に、システムにおいて用いる深層意味表現として、Fillmore の格構造に基づき、これを音声合成部へ伝達すべき情報も記述する形で拡張したものを定義し、本研究で用いる深層格の種類についても言及した。

表 3.1. 本システムにおいて用いる意味付与

否定 (Negative)	疑問 (Question)
丁寧 (Polite)	完了 (Perfect)
希望 (Desire)	

表 3.2. 本システムにおいて用いる深層格

主体 (Subject)	時・終点 (Time-To)	場所・経過 (Space-Through)
対象 (Object)	時間 (Duration)	原因・理由 (Cause)
相手 1- (Partner)	場所 (Space)	手段・道具 (Tool)
時 (Time)	場所・始点 (Space-From)	目的 (Purpose)
時・始点 (Time-From)	場所・終点 (Space-To)	度合 (Degree)

## 第 4 章

# 対話履歴管理



## 4.1 はじめに

対話履歴の管理において、最も重要な機能は、ユーザとシステムの間で行われる対話の流れ(進行)を適切な形式で表現して保存し、それに基づいて対話の流れの管理・制御を行なうことである。

対話の流れを的確につかむためには、ユーザからシステムに入力される発話が、どのようなユーザの意図のもとに発話され、対話の中においてどのような役割を果たすかを推定する必要がある。

本章では、まずユーザからシステムに入力される発話を本システムではどのように分類して処理を進めるかについて説明した後、3.2.2節での音声対話の特徴に関する考察結果に基づいて、対話の基本ルーチンという概念を導入し、それに基づく対話の流れの表現方法および対話の流れの管理について説明する。そして最後に応答音声の合成において韻律的特徴を制御するために重要な情報である焦点位置の制御と、システム応答における省略・照応表現の利用について言及する。

## 4.2 ユーザ入力発話の形態の分類

システムが、ユーザから入力された発話に対して適切な応答内容を決定するという対話処理を行なうにあたっては、ユーザから入力された発話について、ユーザがどのような意図からその発話を行なったのかを的確に把握することが、必要不可欠である。

このような目的を達成するための1つの方法として、対話処理で扱える話題を限定して、その話題に固有の方法で行なうことが考えられる。例えば[田路他 91]では、対話処理の対象となる話題を電話取次対話の一部という狭い範囲に限定し、ユーザの発話意図を「発呼者の名前の伝達」など電話取次対話という話題に依存した形式で13種類に分類している。そして、ユーザからの発話がこれらのどの発話意図を持つかを、電話取次対話に用いられる95種類の言い回しに基づいて推定するという方法を提案している。

しかしこのような方法では、ユーザの発話意図とその分類に用いられる表現が対話の話題に依存しており、そのため対話の話題を変更するごとに、その話題に関する対話例を収集し、言い回しなどを分析することが必要となる。

本研究では、このように話題に依存してユーザの発話意図を細かく分類するよう

な方式は用いずに、ユーザからシステムへ入力される発話を、システムが処理対象とする協調的な質問応答型の対話に対して汎用的に適用できる程度の大まかきで分類し、そのかわり後述する発話の主題に関する情報も用いることによって、ユーザに対して適切な応答を生成するような方式とした。これによって、ユーザ入力の分類は、対話の話題に依存しない形で実現することができ、システムを適用する話題を変更する場合にも、新たに対話例を分類して分類アルゴリズムを生成する必要はなくなる。

以下では、質問応答型対話における大まかなユーザ発話の分類カテゴリとその分類を行なうためのアルゴリズムについて説明する。

#### 4.2.1 分類カテゴリ

質問応答型の対話とは、対話の話題に関する情報の授受を目的とした対話である。従って、情報の授受という観点からは、対話中において現れる発話は以下の質問文、要求文、応答文の3つに大きく分類することができる。

##### 質問文:

対話参加者のうち、一方が知りたいと思った情報について他方に対して直接的に疑問表現を用いて問い合わせるのが質問文である。質問文はその求めている情報の種類に応じて、さらに以下のように分類することができる。

- 単純質問文 (YN 型)

質問文で問い合わせている情報に対して、はい (Yes)、またはいいえ (No) で答えることができるもの。

例: 「温泉はありますか。」

- 深層格の内容に関する質問文 (内容型)

質問文で問い合わせている情報が、深層意味表現で表わした場合、ある深層格を占める内容にあたるもの。

例: 「リフトは何本ありますか。」

- Which 文・比較文 (比較型)

同じ意味マーカ―をもち、同じ深層格の要素であるいくつかの候補のうち、最も条件を満たしたものを問い合わせる質問文である。

例：「苗場と草津のどちらがリフトが多いですか。」

#### 要求文:

対話の相手に対して、ある発話行為を行なうことを求めたり、こちらの望む条件を伝達したりする発話をまとめたものが要求文である。これは以下のように分類することができる。

- 質問型要求文

質問文の形態をとってはいないが、対話の相手に対して特定の情報を要求している発話。この発話は、応答内容の生成という観点からは、質問文の場合と同様に3種類(YN型、内容型、比較型)に分けて処理することができる。

例：「リフトが何本あるか教えてください。」

- 主張文

対話の相手に対して、こちらの希望や意向などを伝達する発話である。データベース検索における条件の追加などによく用いられる表現である。

例：「温泉のあるところがいいです。」

- 挨拶

挨拶に対しては、儀礼的に挨拶を返すのが一般的である。そこで対話のはじめとおわりに発話される挨拶は、対話の相手に対して挨拶を返すという発話行為を要求すると解釈し、要求文のカテゴリに分類する。

#### 応答文:

前2項の質問文、要求文において、対話の相手から求められた発話行為を行なう発話である。すなわち質問文において問い合わせられた情報に関して処理を行なった結果を返すのはもちろん、挨拶を返したりするような発話もこれに含まれる。

以上の表現は、発話者の意図が直接的に表層の発話文の形態に反映されている直接話法(Direct Speech Act)である。これに対して実際の発話者の意図は、発話文の形態とは異なった形で表現される間接話法(Indirect Speech Act)という表現形態も、音声対話中に限らずよく用いられる。例えば、

「リフトが何本あるか知っていますか。」

という発話は、表層の形態からは知っているか、知らないかを問う YN 型の質問文と分類することができる。しかし、この発話文においてユーザが真に意図しているのは、応答者がその情報を保持しているかどうかではなく、情報内容そのもの、すなわちこの場合はリフトが何本あるかということで、発話者の意図という観点からは、この発話は次の内容型の直接話法の発話と同義である。

「リフトは何本ありますか。」

従って、このような間接話法の発話がユーザから入力された場合には、システムは、発話者の意図という観点からその発話と同義である直接話法の発話に対するのと同じものとして分類を行ない、処理を行なうこととした。

#### 4.2.2 分類アルゴリズム

前節での発話形態の分類を行なうにあたっては、本研究では対話処理に入力されるユーザ発話の情報として自立語列を想定しているため、この情報に基づいて分類処理を行なう必要がある。この分類アルゴリズムは、3.2.1節での対話資料に基づいて検討を行ない、以下のようなものとした。

1. ユーザ発話中に「知っている」、「わかる」などの表現が含まれている場合、この発話は間接話法による表現であると判断される。この場合システムはユーザ発話が間接話法であるというフラグを立てた後、この発話を質問文に分類する。そしてその他の部分(従属節の部分)の内容に基づいて、後述する一般の質問文の場合と同様にさらに3種類に分類される。
2. ユーザ発話中に「教えてください」などの表現が存在する場合、この発話は質問型要求文に分類される。質問型要求文は発話の形態的には要求文であるが、以下の処理では質問文と同様に処理される。
3. ユーザ発話の中に疑問詞が含まれている場合、その発話は質問文に分類される。さらにその疑問詞が深層格の内容に関する情報を求めるものである場合、この発話は内容型に分類される。このような疑問詞には以下のようなものがある。(内容型疑問詞)

例:「いくら」、「どこ」、「どれくらい」

一方、疑問詞がいくつかの候補からの選択を要求するものである場合、この

発話は比較型に分類される。このような疑問詞には以下のようなものがある。

(比較型疑問詞)

例:「どちら」、「どれ」

4. ユーザ発話中にユーザの希望や意向などを表現すると考えられる語が存在する場合、この発話は主張文に分類される。このような語には次のようなものがある。

例:「好き」、「嫌い」、「いい」、「いや」

5. 以上の処理の結果、どれにも属さない発話としては次のものが考えられる。

- 応答文
- YN 型質問文
- 内容型・比較型質問文のうち疑問詞が省略されているもの

このうちまず応答文と質問文の分類に関しては、対話が協調的であるという前提に基づき、後述する対話の基本ルーチンによって表現される対話の対応関係によって、応答文が生成されてよい対話状況である場合は応答文として分類し、その他の場合は質問文であると判断することとした。この質問文と応答文の分類には、さらに発話の基本周波数パターンを利用する事も考えられるが、今回は検討対象から外した。

また3種類の質問文を分類する手法としては、次のような方法を用いることとした。

(a) 文末が格助詞「は」で終了している場合は、内容型か比較型の質問文であるとし、文末が終助詞「か」かまたはその他の終わり方をしている場合は、YN 型の質問文であると判断する。

(b) 内容型と比較型の質問文に関しては、発話文中に数値的大小に関連する形容詞・形容動詞（例えば「近い、遠い」、「多い、少ない」など）が含まれている場合は比較型の質問文とし、含まれていない場合は内容型の疑問詞であるとした。

なお、疑問詞やユーザの意向を表わす単語などこれらの分類の基になる単語に関しては、単語辞書中にそれぞれの属性を(疑問詞の場合はさらに内容型か比較型か

という情報も)あらかじめ記述しておき、分類の際に参照する(5.2.1節参照)。

### 4.2.3 評価

前節で提案した入力発話の形態の分類アルゴリズムの有効性を検討するために、3.2.1節で説明した対話資料の一部を用いて、本アルゴリズムの評価を行なった。

具体的には、対話中の各発話に対して、質問側と応答側のそれぞれの立場から前節の分類アルゴリズムを適用し、相手側の発話の形態の分類を行なった。そして、その結果を人間が正解であると判断した発話形態と比較し、その結果を正誤それぞれ合わせたものを評価結果とした。結果を表 4.1 に示す。なお評価に際しては、3.3 節で述べた音声対話処理システムにおいて処理の対象となる発話の部分のみに対して行ない、相手が発話しているの途中で他方が割り込んで発話を行なっている部分などは評価の対象から除外した。

評価の結果として誤って分類された例は、質問文の中での 3 種類の発話間での分類を誤った例で、そのうちの 1 つは YN 型の質問文を誤って内容型と推定しているが、内容型と解釈して応答することも可能である発話であった。

この結果から、自立語を中心とした単語列だけを入力情報として用いて、ユーザからの入力発話を約 98% 以上の高い正答率で分類できることが確認され、提案した分類アルゴリズムの有効性を示すことができた。

## 4.3 対話の流れの表現方法

本研究が対象としている質問応答型対話においては、通常対話参加者双方による以下の 2 種類の発話が 1 組となって 1 つのルーチンを形成し、これが組み合わされて全体の対話が構成されていると考えることができる。

主題の設定及び質問・要求 対話に新たな主題を持ち出し、それに関して相手に対して情報の提供や処理を求める発話である。ただし前回の主題に関連して、引続き質問・要求の発話が行なわれる場合は、主題の設定は省略されて、質問・要求だけの発話となる。

質問・要求に対する応答 上の発話で求められた情報の提供や処理を行なった結果を相手に返す発話である。

この1組の発話によって構成されるルーチンを対話の基本ルーチン(Fundamental Routine of Dialogue:以下FRDと記述する)とする。

また3.2.2節に記述したように、音声対話では文字による対話に比べて個々の発話が短く、またはるかに容易に相手の発話に対して割り込めるため、対話の途中で割り込みが数多く生じる。そしてその割り込みは協調的なものがほとんどである。

このような特徴を持つ音声対話の流れを表現するのに、本システムはFRDを用いている。すなわち、主題の設定及び質問・要求の発話によってFRDが生成され、その応答発話によってFRDは終了し、1つのFRDが完結する。そして協調的割り込みはこのFRDの入れ子で表現し、入れ子の内部のルーチンが終了するとその外側のルーチンにシステムの処理を移し、対話処理を進める。このFRDによって表現される対話の簡単な例を図4.1に示す。

このFRDという概念のように、対話中における発話を質問の発話とそれに対する応答の発話というように2つの発話を対にして考えるというやり方は、会話分析に用いられた adjacency pair の概念 [Levinson 83] と類似のもので、これを音声対話処理に適用した他の例としては、Q/A Cycle [Proctor et al. 89] や SR-Plan [Yamamoto et al. 91] などが挙げられる。しかしこれらの適用例では、Q/A Cycle や SR-Plan 自体に応答生成のために必要となるプランなどが記述されており、対話履歴管理と応答内容生成がまとめて行なわれているのに対して、FRDは純粋に対話履歴の管理にのみ用いられ、応答内容の生成は後述する TODRs を用いてこれとは独立して行なわれる。そのため、それぞれの処理プロセスを独立して改良することが容易で、システムの柔軟性が高められている。

#### 4.4 対話の流れの管理

対話履歴管理部における対話の流れの管理は、前節で導入したFRDに基づいて行なう。具体的には個々のFRDを対話履歴管理処理の処理対象とし、処理対象となるFRDを対話の進行に合わせて遷移させることによって、対話の流れの管理を行なう。

この対話の流れを管理、制御するための必要となる規則は、質問応答型対話に対して話題には無関係に記述することが可能である。そこでこの規則を質問応答型対話一般に関する規則(一般規則)としてシステムに組み込んだ。この規則には、ユーザ入力の分類に関する規則と対話の流れの制御に関する規則の2種類が存在する。

表 4.1. 発話形態分類アルゴリズムの評価

正誤	正	誤
発話数	247	3

U 1: 「A スキー場ではナイターはできますか。」 \_\_\_\_\_ FRD(1)  
 S 1: 「はい、できます。」 \_\_\_\_\_  
 U 2: 「そこまで時間はどのくらいかかりますか。」 \_\_\_\_\_  
 S 2: 「列車と車のどちらで行きますか。」 \_\_\_\_\_ FRD(3)  
 U 3: 「車です。」 \_\_\_\_\_ FRD(2)  
 S 3: 「それなら約 4 時間かかります。」 \_\_\_\_\_

FRD(1),(2) はユーザ発話 (U1,U2) で生成

FRD(3) はシステム発話 (S2) で生成

図 4.1. 基本ルーチンの入れ子構造によって表現される対話の例



### ユーザ入力のカテゴリに関する規則

ユーザ入力のカテゴリに関する規則とは、4.2.2節で記述したユーザ入力発話の形態分類アルゴリズムにおいて、FRDによって表現された発話の対応関係を用いる部分に関する規則と、対話の流れの制御における主張文の扱いに関する規則である。

まず前者の規則については、これは具体的には質問文と応答文の分類を行なう際の規則である。4.2.2節で述べたように、ユーザからの入力発話の形態の分類を行なう際、質問文と応答文は表層的な特徴だけでは厳密に分類することが困難である。しかしFRDの概念から考えると、応答文が入力されるのは、いちばん内側の（現在処理対象の）FRDがシステムの発話によって生成され、しかもそのFRDがまだ終了していないときである。

従って4.2.2節のアルゴリズムで、ユーザから入力された単語列の表層レベルの情報だけでこの2種類の発話が明確に分類できない場合には、システムからユーザへの発話によって生成されたFRDが現在システムの処理の対象となっている場合、すなわちユーザは現在の主題に関してシステムから情報の提供を求められている場合は、その発話は応答文であるとする。そしてそれ以外の場合、すなわちユーザが自由に主題を設定して質問できる状態である場合は、その発話は質問文であると推定する。ただし、5.4.3節で述べるように、応答文と推定された場合でも、発話内容から考えて応答文であるとは判断できない場合には、質問文と解釈しなおして処理を進める。

一方後者の主張文の扱いに関する規則は、主張文は発話されたときの対話状況によって、対話の流れの制御の点から扱いを変える必要があり、この点に関する規則である。すなわち、FRDの定義からは主張文などの要求文は通常FRDを生成する機能をもつ。しかし、例えば図4.2の場合は、ユーザからの入力U2は主張文であるが、システムからの質問に対する応答として機能しているので、対話履歴管理上は応答文として取扱い、FRDを終了させる働きを持たせる必要がある。

そこで主張文に関しては、システムの発話によって生成されたFRDがシステムの処理対象となっている場合は、対話履歴管理上はこの主張文を応答文として扱うこととした。

### 対話の流れの制御に関する規則

対話の流れの制御とは、具体的にはユーザからの入力発話のカテゴリに応じてFRDの生成・終了を行ない、またそのFRDの状況から必要に応じて応答内容立

案部に応答内容の立案処理を指示する。これらの処理は以下の規則にしたがって行なわれる。

### 1. 質問・要求文の入力時

履歴管理部は新たに FRD を生成するとともに、質問・要求内容を応答内容立案部に送り、応答内容の立案を指示する (図 4.1 の U1、U2)。応答内容立案部での処理の結果として質問に対する回答が得られた場合はその深層意味表現を出力して FRD を終了し (図 4.1 の S1)、1 つ外側の FRD に処理対象を移す。一方、回答を得るのに十分な情報が得られていない場合は応答内容立案部がユーザに対して不足情報に関する質問の発話を生成するので (協調的割り込み)、それに対応する新たな FRD を生成する (図 4.1 の S2)。

### 2. 応答文の入力時

この場合、まずユーザの応答内容がシステムの質問した内容に対する応答として適切であるかどうかを判断するために、対話履歴管理部は応答内容を応答内容立案部に送る。応答内容立案部での処理の結果 (詳しくは 5.4.3 節参照)、ユーザの応答内容が適切であれば、対話履歴管理部は現在処理対象となっている FRD を終了させ、1 つ外側の FRD に処理対象を移す (図 4.1 の U3)。一方、応答の内容が不適切な場合は、発話の形態における推定が誤りであったと判断し、その発話を質問文であると推定しなおして、処理を行なう。

### 3. 外側の FRD に処理対象が移ったときの処理

以上の処理のなかで FRD が終了し 1 つ外側の FRD に処理対象が移ったときの処理は、次の 2 つの場合に分けて行われる。

- 新たに処理対象となった FRD がユーザの発話によって生成されているとき

新たに処理対象となった FRD が生成されたときのユーザ発話の内容に基づき、その発話に対する応答の生成を応答内容立案部に指示する。この時、応答内容立案部でユーザに回答を提示するのに十分な情報が得られていないと判断された場合は、再びユーザに対する質問がシステム発話として生成され、対話履歴管理部は再び入れ子状に FRD の生成を行なう。もしすでに十分な情報が得られたと応答内容立案部が判断した場合は、回答を生成し、この FRD を終了させる (図 4.1 の S3)。

- システムの発話によって生成されているとき

システムはユーザからの入力待ちの状態になるが、一定時間待ってもユーザから入力がない場合は、システムからユーザに対して入力を催促する発話を行なう。

以上の規則に従って FRD の生成、終了を行ない、さらに状況に応じて応答内容立案部を呼び出すことによって、対話中の個々の発話の対応関係を記述できると同時に、状況に応じた適切な対話進行をはかることができる。

## 4.5 対話履歴管理に付随する機能

### 4.5.1 焦点の管理

音声応答システムがユーザにとって理解しやすいような音声を合成するためには、合成音声の分節的特徴の品質が高く、十分な了解度が得られることはもちろん、システムがユーザに伝達したい部分を強調するなどの適切な韻律的特徴を合成音声に付与することも不可欠である。そしてそのためには対話処理システムは焦点の管理を行なうことが必要となる。これは文字対話の処理にはない機能である。

焦点とは、対話のある時点（ある対話の状況下）において、1つの発話によって相手に伝達される情報の中心となるもの、すなわち発話者が相手に最も把握してもらいたいと考えている情報である。この管理を対話処理において行なうことによって、音声合成処理では指定された焦点位置の情報から、焦点がある部分が強調されるような形で韻律的特徴を制御し、ユーザに対してより理解しやすい合成音声を生成することが可能となる。またこのように韻律的情報を制御することによって、重要な情報の聞き取りミスの減少が期待できる。

具体的にどの情報に焦点をおくかは、3.2.1節で用いた対話資料を参考にして、最も典型的と思われるものに対して、以下のように焦点の設定に関する規則化を行なった。

- 内容型質問文に対する応答の場合

ユーザから入力された発話が内容型質問文の場合は、まず疑問詞によって求められている情報の意味マーカをこの質問文に対する FRD のシリアル番号とともに対話履歴ベース中に保存する。そしてこの質問文に対する応答が応答内容立案処理によって得られたときに、応答深層意味表現中から保存されてい

る意味マーカ―をもつ単語を求め、その単語が応答深層意味表現中で占めている深層格に焦点を設定する。

- 比較型質問文に対する応答の場合

ユーザから入力された発話が比較型質問文の場合、まず同一の意味マーカ―をもつ複数の単語を探索することによって、比較されている対象を推定し、この意味マーカ―をこの質問文に対する FRD のシリアル番号とともに対話履歴ベース中に保存する。そしてこの質問文に対する応答が応答内容立案処理によって得られたときには、前項と同様にその意味マーカ―をもつ単語を求め、その単語が占める深層格に焦点を設定する。

- システムがユーザに質問する場合

システムが質問を生成する際に応答内容処理において利用した TODR において、質問によってユーザから得ようとする情報(目的情報)に対応する疑問詞に焦点がおかれるので、その疑問詞が深層意味表現中で占める深層格に焦点を設定する。

#### 4.5.2 システム応答における省略・照応表現の使用

ユーザに対してシステムが応答する場合、応答発話にユーザにとってわかりきっている情報が含まれていると、ユーザにとってはくどいものとなる。そしてシステムとユーザの間の自然なコミュニケーションをはかる場合には、このようなユーザの不快は極力取り除くべきで、そのためには、対話処理システムはユーザに対する応答を提示する場合に、省略・照応表現を適切に利用することが必要である。ユーザに対する応答内容の立案については、次章で詳しく説明するので、本節では、応答内容立案部で生成されたユーザに対する応答内容に対応する深層意味表現から、どの部分に省略・照応表現を用いるかを決定するアルゴリズムについて説明する。

省略・照応表現を用いるためには、その時点でどのような単語に対して省略・照応表現を用いることができるのかを、対話履歴管理機能の一部として把握しておく必要がある。そのため、対話履歴管理部では、ユーザから入力された発話中の各単語と、対話処理システムがユーザに対して応答内容として提示した深層意味表現中の各単語について、品詞(名詞の場合は意味マーカ―も)を単語辞書の記述をもとに求め、それぞれの品詞(意味マーカ―)ごとに対話履歴ベース中に使われた単語を保存する。そしてこの保存内容に対話の進行とともに更新し、各品詞(意味マーカ―)

ごとに対話中で用いられた最新の単語を常に対話履歴ベース中に保持する。

そしてこの情報に基づいて、システム応答として得られる深層意味表現中に含まれている単語に対して、その単語の品詞(意味マーカ)を調べ、それに対して対話履歴ベース中に保存されている各品詞(意味マーカ)の対話中で用いられた最新の単語と比較し、同じ単語である場合に、その部分に対して省略・照応表現を用いることとした。なお、その単語のしめる深層格が述語に対して必須格である場合は省略表現を、任意格である場合は照応表現を用いることとした。

#### 4.6 まとめ

本章では、音声対話処理システムのうち、対話履歴管理部について詳説した。

はじめに、処理対象となる対話が質問応答型対話であるという点に基づき、ユーザからシステムに対して入力される発話文を、質問文、要求文、応答文に大きく分類し、質問文についてはさらにYN型、内容型、比較型の3種類に分類して扱うこととした。そしてこの分類を入力される単語列の情報に基づいて行なうアルゴリズムを提案し、実際の音声対話資料にそのアルゴリズムを適用して評価し、その有効性を確かめた。

次に、質問応答型対話の対話の状況を表示する手段として、対話の基本ルーチン(FRD)という概念を導入し、これに基づいて協調的に行なわれる質問応答型対話を表現する方法を提案した。

さらに、対話の基本ルーチンによって表現された対話の状況に基づいて対話の流れを制御する規則を、対話の話題に依存しない形で記述できることを示し、これを対話履歴管理部中に実行手順の形で実装した。

最後に、音声合成の際に韻律的特徴を制御する重要な情報となるシステム応答中の焦点位置の設定と、システムの応答における省略・照応表現の利用について規則化を行なった。

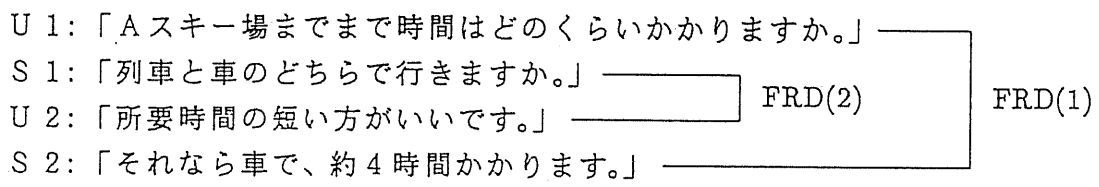


図 4.2. 主張文が応答文として機能する例

## 第 5 章

# 応答内容立案

## 5.1 はじめに

システムが適切な応答内容を立案するには、対話の話題に関連した知識を用いる必要があることは明らかである。しかし対話処理システムとしての汎用性を高めるためには、知識の記述を明確に規定し、対話処理プロセスからは分離して実現する必要がある。

本システムでは話題に関連する知識を単語辞書と TODRs という形で記述し、これを利用する処理プロセス自体は話題には依存せず、質問応答対話に汎用に適用できるような構成とした。

また本研究では、対話処理システムへの入力は一語列を用いているので、応答内容の立案処理は入力される個々の単語の意味を主体とした意味主導の処理によって行なわれることとなる。

本章では、まず 5.2 節で単語辞書と TODRs の 2 つの話題に関連する知識の記述について述べる。次に 5.3 節で、単語辞書中の記述に基づいて、入力された発話でどのような主題について言及されているのかを推定する処理についてその概要とアルゴリズムについて説明する。5.4 節では、実際にシステム応答を立案する手順をユーザ入力発話の形態ごとに説明し、5.5 節で、ユーザからの質問に対する答えを得るために行なわれる知的システムへのアクセスとそれに関連して TADRs について言及する。そして最後に、5.6 節で対話処理システム全体の動作を解説するために、ユーザと対話処理システムの間で行なわれた対話の例をあげ、それに沿う形で対話履歴管理処理、応答内容立案処理をあわせた対話処理システム全体の流れを解説し、システムが有効に動作していることを示す。

## 5.2 話題に関連する知識の構成

本システムを構成するにあたっては、研究の目的上なるべく汎用性を高めるために対話の話題に関連する知識を明確にし、知識を利用する対話処理プロセス自体とは区別して実現することを方針とした。そして具体的には話題に関連する知識を単語辞書と TODRs という形で実現した。

単語辞書は、対話処理の対象となっている話題において、対話中で発話される単語をその意味情報などとともに記述したものである。一方 TODRs とは、質問が与えられたときにそれに応答するにはどのような情報が必要になるかというような、対



話の話題に固有な対話の進行に関係している情報をまとめて表現したものである。

この2つの知識に関しては、記述すべき情報と記述形式が明確に規定されている。よってこの規定にしたがって2つの知識を書き換えることによって、システム管理者は対話処理システムを他の話題の対話に比較的容易に適用することが可能となる(知識の構築手順について詳しくは付録 A を参照)。

以下にこの2つの知識の記述内容について説明する。

### 5.2.1 単語辞書

単語辞書は対話において使用される単語を、以下のような情報とともに記述したものである。各単語は単語辞書中に以下のような形式で個々に記述されている。

dic(単語表記, 品詞, 意味情報, 音声情報).

それぞれの項目の記述内容は以下の通りである。

- 単語表記

各単語に対する文字表記で、深層意味表現中ではこれが述部や深層格の filler として使用される。また、応答出力として表層のテキストも出力ことができるように考慮して、表記としては漢字かな混じりを用いている。

- 品詞

各単語の品詞である。品詞の種類としては表 5.1 のものを用意した。このうち特殊述語とは、4.2.2 節におけるユーザ入力発話の形態分類の際に、間接話法や主張文を分類する際にキーとなる単語である。また、その他の品詞は助詞や助動詞など本システムでユーザ入力を解析する際に用いられない単語である。しかし、3.5 節のように音声認識によってこれらの単語が利用できるような状況になれば、新たに品詞を追加して、対話処理に用いるようにシステムを改良することは容易に行なえる。

- 意味情報

この部分の記述はそれぞれの品詞ごとに、次のような情報がリスト形式で記述されている。

- 名詞

名詞の意味記述には、この単語の意味マーカーと、この単語が発話中に現

れた場合に、関連している可能性がある主題のリストが記述されている。意味マーカ―とはその名詞のもつ意味属性を表わすもので、ユーザ入力中の名詞と後述する TODR の各情報項目とを対応づけるために用いる。また主題の記述は、5.4.1 節で述べる手順にしたがって、応答生成に適切な TODR を選択するためにのみ用いられる。従って意味マーカ―と主題は、単語辞書と TODRs の 2 つの話題に依存する知識の間で統一がとれていればよく、意味マーカ―、主題としては一般的な意味概念などに固執することなく、知識作成者が対象とする話題に特化した形で決めることが可能である。例として、スキー場に関する話題で用いた意味マーカ―の例を表 5.2 に、主題の例を表 5.3 に示す。

－ 動詞

動詞の活用形と主題記述、さらにその動詞がとり得る深層レベルでの必須格、任意格が、その格の filler となりうる単語のもつべき意味マーカ―と表層レベルで用いられる代表的な助詞とともに記述される。

－ 形容詞・形容動詞

単語が数値の大小に関連するものである場合は、その単語が数値の大小のどちらと関連するのかを+-で記述する。例えば「多い」という単語は、数値的に大きいことを表現するので、+の記号が記述される。この記述によって、比較の質問文に対して容易に対処できるようになっている。

また、動詞と同様に主題記述と深層レベルでの必須格、任意格が動詞の場合と同様に記述される。

－ 疑問詞

内容型疑問詞の場合は、その疑問詞によって求められる情報の意味マーカ―と関連する主題記述が記述される。一方、比較型の疑問詞の場合は比較の対象となる意味マーカ―を特定することはできないので、意味マーカ―と関連する主題の部分には何も記述されない。

－ 特殊述語

特殊述語のうち、間接話法と質問型要求文に関する述語には意味情報は何も記述されない。主張文に関するものには、その述語によって希望しているのか希望していないのかを yes,no で意味情報の中に記述する。例えば「いい」という単語はそのようなものを希望していると考えられるの

表 5.1. 対話処理で用いる品詞の種類

名詞	動詞
形容詞	形容動詞
内容型疑問詞	比較型疑問詞
特殊述語	その他

表 5.2. 意味マーカの例

意味マーカ	意味概念
act	行動
inst	施設
hum	人間
sp	場所、地点
tp	時点
dura	時間
tl	手段、道具
cost	金額

表 5.3. 主題の例

リフト数	リフト料金
滑走距離	コース数
駐車台数	駐車料金
所要時間	所在地

で意味情報の中に yes と記述しておく。

－ その他

意味情報には何も記述されない。

- 音声情報

音声情報の部分には、音声合成を行なう際に必要となるその単語の音素記号とアクセント型が記述される。

単語によっては、使われる文脈によってその意味マーカ―や、深層格として取る単語の意味マーカ―が異なる場合がある。例えば、「リフト」という単語は、「リフトは何本ありますか。」という発話中で使われれば、施設としてのリフトの意味であるし、「リフトはいくらですか。」という発話中では、リフトの料金という意味に用いられる。また、「高い」という形容詞も料金が高いという場合と標高が高いという場合で、深層格としてとる単語の意味マーカ―が異なる。

このように意味的に複数に用いられる単語に関しては、本システムでは単語辞書中で別々の辞書項目とし、意味マーカ―や主題記述などの情報はそれぞれに記述することとした。そして、実際のユーザ入力発話の解析においてどの辞書項目に記述されている情報を利用するかは、5.3.2節で述べる処理の結果として推定される発話における主題をもつ辞書項目を利用することとした。

最後に、単語辞書における各単語の記述例を図 5.1 に示す。

### 5.2.2 話題依存規則

話題依存規則 (TODRs) は話題に固有の対話進行上の知識を規則化したもので、質問応答対話における様々な質問に答えるために必要となる情報などを、関連する主題ごとに個々に独立して記述する。

ユーザ入力の発話形態が質問文である場合は、4.2.1節で述べた質問文の形態ごとにその応答の仕方やデータベースへのアクセスの仕方が変わってくるために、同じ主題に関する TODR も質問文の形式に応じて YN 型、内容型、比較型の 3 種類に分類して登録されている。但し、ある主題に関連する情報がデータベース中でどのようなデータ形式で表現されているかにより、存在する質問文の形態は表 5.4 の組合せようになり、TODR としては存在する組合せだけを用意すればよい。

TODR には、対話処理システムがユーザに対する応答内容を生成するのに必要と

## %名詞

dic(リフト, n, [inst, [リフト数, 施設存在確認, 輸送力]], 'ri\$futo').

dic(リフト, n, [cost, [リフト料金]], 'ri\$futo').

dic(料金, n, [cost, [リフト料金, 駐車料金]], 'rjoH\$kiN').

dic(温泉, n, [inst, [施設存在確認]], 'oN\$seN').

## %動詞

dic(する, v, [さ変, [], [[sub, hum, が], [obj, act, を]], [[time, tp, に], [space, sp, で]], 'suru\$').

dic(行く, v, [か 5, [行き方], [[sub, hum, が]], [[obj, act, に], [s-to, sp, に], [tool, tl, で], [partner, hum, と], [time, tp, に], [dura, dura, で]] ], 'iku\$').

## %形容詞

dic(高い, a, [+ , [リフト料金, 駐車料金], [[sub, cost, が]], [[space, sp, で]], 'taka\$i').

dic(高い, a, [+ , [標高], [[sub, inst, が]], [[space, sp, で]], 'taka\$i').

dic(多い, a, [+ , [リフト数, コース数, 駐車台数, 積雪量], [[sub, inst, が], [[space, sp, に]], 'oH\$i').

dic(安い, a, [- , [リフト料金, 駐車料金], [[sub, cost, が]], [[space, sp, で]], 'jasu\$i').

## %疑問詞

dic(何本, value, [num, [リフト数, コース数]], 'na\$NboN').

dic(何時間, value, [dura, [所要時間]], 'na\$NzikaN').

dic(いくら, value, [cost, [リフト料金, 駐車料金]], 'i\$kura').

dic(どれ, comp, [[], []], 'do\$re').

図 5.1. 単語辞書における単語の記述例

表 5.4. データ形式と疑問文の形態の組合せ

データ形式	疑問文の形態		
	YN	内容型	比較型
2値データ	○	×	×
数値データ	△	○	○
テキストデータ	○	○	×

○：組合せが存在する

×：組合せが存在しない

△：存在する場合としない場合がある

例えば、個数の場合は存在する。

距離の場合は存在しない。

なる知識がフレーム形式で記述されている。TODR の記述フォーマットは次のようになっている。

発話文形態 (主題,[目的情報リスト],[アクセス必須情報リスト, …],[アクセス任意情報リスト, …],アクセス形態,単位,大小基準,応答意味表現の枠組み).

目的・アクセス必須・アクセス任意情報リストの個々の要素の記述形式  
[情報項目名,意味マーカー,検索情報,複数指定フラグ]

個々の項目における記述内容は以下の通りである。

- 発話文形態

4.2.1における3種類の質問文の形態(YN型、内容型、比較型)と要求のいずれか1つが記述される。

- 主題

このTODRが対話処理で選択されるときに対話における主題を規定したもので、これは単語辞書中の主題記述部分との関連が保たれていれば、知識作成者が自由に設定することができる。

- 目的・アクセス必須・アクセス任意情報リスト

このTODRに記述されている中でもっとも重要な項目である。目的情報リストとは、このTODRが選択されたときにユーザまたはシステムが相手から得ようとしている情報に関する記述である。アクセス必須情報リストは、知的システムにアクセスし、目的情報リストの内容を決定するために必ず必要な情報に関する記述である。またアクセス任意情報リストとは、このアクセスを行なう際にユーザが付加的に任意に指定することができる情報に関する記述である。

これらのリストを構成する要素は、以下の4つの項目で構成される。目的情報リストに関しては、要素は1つしか許されないが、アクセス必須・アクセス任意情報リストは必要なだけ複数個要素を取ることができる。

- 情報項目名

個々の情報項目につけられる名称で、知識作成者が自由に設定することができる。この名称は検索情報や応答意味表現の枠組みに実際の目的・ア

クセス必須・アクセス任意情報をあてはめる際や、ユーザ入力発話中で省略された情報を前の TODR から継承して補完する際 (5.4.1 節参照) に利用される。

－ 意味マーカ－

この情報項目にあてはまる単語のもつべき意味マーカ－を記述する。この記述に基づいて、ユーザ発話中から適切な単語がこの情報項目にあてはめられる。

－ 検索情報

目的情報リストにおいては、システムで扱う対話の話題に関する具体的な知識を提供する知的システム (データベースなど) へ、実際にアクセスするための書式が記述される。

またアクセス必須・アクセス任意情報リストにおいては、同義語の参照テーブル名を記述して、様々な表現をここで吸収して知的システムへアクセスする際の名称に統一する処理を行なう (5.4.1 参照)。

－ 複数指定フラグ

これは必須・任意情報の場合に用いられるもので、この情報項目名が複数存在する必要がある場合は、この部分に pl が記述される。具体的には比較を行なう場合に、比較対象が複数必要になるという場合に用いられている。

これらの情報リストに情報が登録される手順は、詳しくは 5.4.1 節で記述されるが、アクセス必須・アクセス任意情報に関しては、ユーザ入力発話中から適切な単語があてはめられ、目的情報に関しては知的システムへのアクセスの結果またはシステムからの質問に対するユーザからの応答として得られる。

● アクセス形態

知的システムに情報を得るためにアクセスする際、どのような形態でアクセスに行くかを記述した情報である。本研究で構築した対話処理システムの場合、知的システムとしてはデータベースを用いているので、このアクセス形態としては以下のものを用意した。知識作成者はこの 3 種類の中からこの規則で目的としている要求情報をデータベースから得るのに最適なものを 1 つ記述する。



## - simple

データベースに記述されている情報を単純に求める場合である。これは目的とする情報のデータベース中でのデータ形式に関係なく利用できる。

発話例:「苗場にはリフトが何本ありますか。」

## - comp

いくつかの候補に関するある情報項目の値をある条件で比較する場合である。これは比較する情報項目のデータ形式が数値型である場合に限って利用できる。また、これを選択する場合は、アクセス必須情報として必ず「条件」という情報項目名をもつものを用意しなければならない。

発話例:「苗場と草津温泉ではどちらがリフトが多いですか。」

## - yncomp

ある情報項目の値をある基準と比較した結果を yes、no で答える場合である。これは比較する情報項目のデータ形式が数値型である場合に限って利用できる。また、これを選択する場合は、アクセス必須情報として必ず「条件」と「基準」という情報項目名をもつものを用意しなければならない。また、ユーザ入力発話で基準が省略された場合は、大小基準のスロットに記述されている値を基準として比較が行なわれる。

発話例 1:「苗場はリフトが 15 本以上ありますか。」

発話例 2:「苗場はリフトが多いですか。」(基準が省略された例)

## ● 単位

この TODR が数値データを扱うものである場合、その数値に用いられる単位を記述する。

## ● 大小基準

この TODR が数値データを扱うものである場合、その数値が大きいか小さいかを判断するための基準として、これ以上だったら大きいと判断する値 A とこれ以下だったら小さいと判断する値 B を [A,B] の形で記述しておく。

## ● 応答意味表現の枠組み

知的システムへのアクセス結果として得られた情報をユーザに応答する際の応答深層意味表現の枠組みを記述する。

また TODR には、ユーザからの質問に答えるために用意されるものの他に、シス

テムがユーザに対して質問するために用意されるものもある。そしてシステムが質問を行なう場合には、システムがユーザに対して質問したい情報項目名を目的情報リストの情報項目名にもつ TODR を用い、その応答意味表現の枠組みを用いてユーザへの質問内容を生成する。

話題依存規則の具体的な記述例を図 5.2 に示す。

TODRs をこのような方式で記述する最大の利点は、個々の TODR が独立しているため、データベースのデータ項目の追加や削除に対して TODR の追加、削除によって柔軟に対応できるという点である。またシステム管理者が独自に TODR を追加することによってシステムを容易に Tuning する事が可能であるという利点もある。

この TODR の記述方法は、対話中で起りうる個々の事例を規則として記述するもので、人工知能の観点からみると、この応答生成の方法は事例に基づく推論 (Case-Based Reasoning) [Kolodner et al. 85, 小林 92] の一例で、適用する事例の選択を個々の単語の意味的記述 (意味マーカー、主題記述) に基づいて行なったものと解釈することができる。

plan に基づく対話処理規則の記述は、対話中での情報伝達行為を抽象的に表現したものであるため、その記述には対話の深い分析が必要となる。また対話中に対応できない事例が生じた場合や、システムを他の話題を扱うように変更する場合に対処するのが困難で、あまり実用的であるとはいえない。それに対して、本研究における事例に基づいた記述を用いる方式は、plan などの抽象的な記述に比べて記述が具体的で容易であり、そのために規則の追加・変更によるシステムの拡張性やシステムの管理者が独自に規則を作成することによるシステムの他の話題への適応性が高く、対話処理システムとして実用性の高いものとなっているといえる。

## 5.3 主題の推定

### 5.3.1 主題推定処理の概要

入力された発話の中で、ユーザがどのような主題に対して情報を求めているかをシステムが推定することは、適切な応答をユーザに返すためには必要不可欠な機能である。本節では、このユーザから入力された発話における主題を推定するアルゴリズムについて具体的な発話例を交えて説明する。なお、この処理は入力された発話を構成する単語の辞書引きと同時に、4.2.2 節で説明した発話形態の分類処理に引き続いて行なわれる。

yn( リフト数, [[リフト存在確認,yn,[リフト種, スキー場名,X],-]], [[スキー場名,sp, スキー場名,-],[リフト種,inst, リフト種,-],[条件,pred,-,-]], [], yncomp, 本, [15,5], yn).  
value( リフト数, [[リフト数,num,[リフト種, スキー場名,X],-]], [[スキー場名,sp, スキー場名,-],[リフト種,inst, リフト種,-]], [], simple, 本, [15,5], [ある,[pol],[],[[space, スキー場名],[deg,Y],[sub, リフト種]]]).  
comp( リフト数, [[スキー場名,sp,[リフト種, スキー場名,X],-]], [[スキー場名,sp, スキー場名,p],[リフト種,inst, リフト種,-],[条件,pred,-,-]], [], comp, [], [], [条件,[pol],[],[[space,Y],[sub, リフト種]]]).

図 5.2. 話題依存規則の具体例

主題の推定にあたっては、発話の内容だけでは決定することが困難で、その発話以前の対話状況を考慮して推定することが要求される場合がある。このような場合のために、このアルゴリズムではそれ以前の発話状況を表わす情報として、前回主題と前回主題候補という2つの情報を対話履歴ベースの中に保存しておき、主題推定の際に必要な応じて利用する。

前回主題とは、前回の発話がどのような主題に関しての発話であったのかに関する情報で、この発話が行なわれる前に行なわれた最新の主題推定処理の結果として得られた主題が1つ設定されている。一方、前回主題候補とは、前回の主題推定処理において主題が選ばれる際に、その処理の過程において候補としてあがった主題である。すなわち、前回の主題推定処理の対象となった発話に含まれていた主題記述をもつすべての単語の主題記述に含まれるすべての主題から、前回主題となったものを除いたものである。そして、この前回主題候補は複数存在することが可能である。

具体例として、次のような発話に対して主題推定処理が行なわれたとする。

「リフトは何本ありますか。」

この主題の推定処理は後述する推定アルゴリズムの1.(a).ii.Aによって主題「リフト数」が推定される。このとき、発話を構成する単語のうち主題記述をもつものとその主題記述は以下のようなになる(5.1参照)。

「リフト」の意味マーカー inst に対する主題記述:

[リフト数]

「リフト」の意味マーカー cost に対する主題記述:

[リフト料金]

「何本」の主題記述: [リフト数, コース数]

この主題推定処理の終了後に対話履歴ベースに登録される前回主題と前回主題候補は次のようになる。

前回主題 この主題推定処理によって推定された主題である「リフト数」

前回主題候補 2つの「リフト」に関する辞書記述中の主題記述に含まれる「リフト数」、「リフト料金」と、何本に含まれる「リフト数」、「コース数」のなかから、主題と推定された「リフト数」を除いたもの、すなわち「リフト料金」と「コース数」が前回主題候補となる。

## 5.3.2 主題推定アルゴリズム

以下に主題推定アルゴリズムを示す。

## 1. ユーザ入力発話中に主題記述をもつ名詞がある場合

## (a) 主題記述をもつ名詞が1つしか存在しない場合

## i. その名詞に対する辞書記述が1つしかない場合

A. その名詞の主題記述が1つしかない場合は、それを主題とする。

例：「温泉はありますか。」

「温泉」の主題記述: [施設存在確認] ⇒ 主題

B. 主題記述が複数ある場合は、発話を構成する他の品詞(疑問詞、述語)の主題記述と共通するものがあれば、それを主題とする。なお、この処理においては疑問詞の方を述語より優先させて適用する。

例：「コースは何本ありますか。」

「コース」の主題記述: [コース数, 滑走距離, 最大斜度]

「何本」の主題記述: [リフト数, コース数]

両者に共通する主題記述: [コース数] ⇒ 主題

C. 他の単語と共通するものが複数ある場合は、前回主題、前回主題候補と共通するものを選択する。

例：「苗場と草津温泉のどちらがリフトが多いですか。」

「苗場です。」

「料金はいくらですか。」

前回主題: [リフト数]

前回主題候補: [リフト料金]

「料金」の主題記述: [リフト料金, 駐車場料金]

「いくら」の主題記述: [リフト料金, 駐車場料金]

前回主題候補と共通する主題記述: [リフト料金]

⇒ 主題

## ii. その名詞に対する辞書記述が複数ある場合

A. 他の品詞(疑問詞、述語)と共通する主題を主題記述にもつものがあれば、その辞書記述を用い、共通するものを主題とする。なお、

この処理においては疑問詞の方を述語に優先させて適用する。

例 1: 「リフトは何本ありますか。」

「リフト」の意味マーカー inst に対する主題記述:

[リフト数]

「リフト」の意味マーカー cost に対する主題記述:

[リフト料金]

「何本」の主題記述: [リフト数, コース数]

共通する主題記述: [リフト数] ⇒ 主題

例 2: 「リフトはいくらですか。」

「いくら」の主題記述: [リフト料金, 駐車場料金]

共通する主題記述: [リフト料金] ⇒ 主題

B. 他の単語と共通する主題記述がなければ、前回主題、前回主題候補と共通する主題記述を選択する。

例: 「コースは何本ありますか。」

「15本です。」

「では、リフトは。」

「リフト」の意味マーカー inst に対する主題記述:

[リフト数]

「リフト」の意味マーカー cost に対する主題記述:

[リフト料金]

前回主題: [コース数]

前回主題候補: [滑走距離, 最大斜度, リフト数]

前回主題候補と共通する主題記述: [リフト数] ⇒ 主題

(b) 主題記述をもつ名詞が複数存在する場合

i. 複数の名詞のそれぞれの主題記述中に共通する主題がある場合、それを主題とする。

例: 「リフトの料金はいくらですか。」

「リフト」の意味マーカー inst に対する主題記述:

[リフト数]

「リフト」の意味マーカー cost に対する主題記述:

[リフト料金]

「料金」の主題記述: [リフト料金, 駐車場料金]

共通する主題記述: [リフト料金] ⇒ 主題

2. 主題記述をもつ名詞がなく疑問詞がある場合、疑問詞の主題記述に基づいて、主題の推定を行なう。

(a) 疑問詞が1つのみ存在する場合

- i. 述語の主題記述との間で共通する主題がある場合は、それを主題とする

例: 「どうやって行くのですか。」

「どうやって」の主題記述: [行き方]

「行く」の主題記述: [行き方, 交通手段]

共通する主題記述: [行き方] ⇒ 主題

- ii. 述語と共通する主題記述がない場合は、前回主題と疑問詞の主題記述のリストで共通するものを主題とする。

例: 「苗場と草津温泉のどちらがリフトが多いですか。」

「苗場です。」

「何本ですか。」

前回主題: [リフト数]

「何本」の主題記述: [リフト数, コース数]

共通する主題記述: [リフト数] ⇒ 主題

- iii. 前回主題との間にも共通する主題がない場合は、前回主題候補と共通する主題記述を主題とする。

例: 「苗場と草津温泉のどちらがリフトが多いですか。」

「苗場です。」

「いくらですか。」

前回主題: [リフト数]

前回主題候補: [リフト料金, コース数, 駐車台数, 積雪量]

「いくら」の主題記述: [リフト料金, 駐車料金]

前回主題候補と共通する主題記述: [リフト料金] ⇒ 主題

3. 発話中に主題記述のある名詞、疑問詞がなく、述語のみの場合

- (a) 形容詞がある場合は、その形容詞に基づく比較型の質問文であると思われる。この場合はこの形容詞の主題記述が1つしかない場合は、それを主

題として用いる。

例：「苗場と草津温泉で近いのは。」

「近い」の主題記述: [所要時間] ⇒ 主題

(b) 形容詞がなく、動詞しかない場合は、その動詞だけで主題が決まる場合はそれを主題として用いる。

例：「もうやっていますか。」

「やっている」の主題記述: [営業期間] ⇒ 主題

以上のアルゴリズムで推定処理を行なった結果、主題記述が1つも残らなかった場合には、協調的割り込みによってユーザに対して何に関する発話であるかを問い合わせる。同様に、推定処理の結果として主題記述が複数残った場合には、その残った主題をユーザに提示して、このうちのどれについての発話であるかをユーザに対して質問する。

また発話中に複数の疑問詞が存在する場合は、複数の主題に対する問い合わせとなるので、「1つずつ質問して下さい。」という発話をユーザに対して提示し、単独の主題に対する質問を改めて入力してもらうようにユーザの発話を誘導する。

この主題推定方式の利点としては、次のような点があげられる。

- 前回主題、前回主題候補を推定処理に導入のすることによって、単独の発話では主題を確定するのが困難な発話に対しても、的確な主題の推定が行なえる。さらに本方式では、単語辞書中の主題記述だけで主題の推定が行なえるため、主題階層などの別の知識を利用して主題の管理を行なうような方式[新美 87]のように、新たに別の知識を構築する必要がない。
- 主題を1つに推定するまでに至らない場合でも、推定処理によって絞りこんだ主題候補をユーザに提示して、その中からユーザに選択してもらうことが可能である。したがって、ユーザの次の発話に対して再び初めから推定処理を行なう必要がなくなる。

#### 5.4 応答内容立案の手順

本節では、5.2.2節で説明した TODRs を用いて応答内容を立案する手順について説明する。処理手順は対話履歴管理部から応答内容立案部に送られてくるユーザ入力発話の形態によって異なる。以下に、質問文、要求文、応答文が入力された場合



に、それぞれどのような処理手順でユーザに対する応答内容の立案が行なわれるかを示す。

#### 5.4.1 質問文入力の場合

ユーザからの入力発話の形態が質問文であった場合、それに対する応答内容の立案は以下の1から4の手順にしたがって行なわれる。このうち1から3までの処理の流れの概略を図5.3に示す。

##### 1. TODRsからの適切な規則の選択

4.2.2節と5.3.2節で説明したアルゴリズムにしたがって推定されるユーザ入力発話の形態と主題に関する情報に基づき、TODRsの中からこの質問文を処理するのに最適なTODRを選択する。

具体的には、ユーザ入力発話から得られた形態と主題の情報と、個々のTODRに記述されている発話文形態と主題情報とのマッチングをとることによって、適切なTODRの選択を行なう。

##### 2. アクセス必須情報、任意情報への情報の登録

発話を構成する個々の単語に対して、辞書引きの結果として得られる品詞情報と、さらに名詞の場合は単語辞書中に記述されている意味マーカーの情報に基づき、1で選択されたTODR中のアクセス必須情報、アクセス任意情報に該当する単語を登録する。

各アクセス必須・任意情報には、そこに登録されるべき情報の意味マーカー(または品詞)があらかじめ記述されており、その情報と各単語の意味マーカー(または品詞)とを一致させることによって、この登録処理は行なわれる。

またこの登録処理と同時に、名詞の同義語の処理も行なわれる。すなわち、各アクセス必須・任意情報に対して、そのスロットに登録されうる単語に関して、知的システムにアクセスする際に利用する名称とその同義語が記述された参照テーブルを用意しておき、同じ概念を表わすのにいくつかの表現方法がある単語が入力された場合は、このテーブルを参照して知的システムへのアクセス用の名称に統一され、登録される。

##### 3. 前に使用したTODRからのアクセス必須情報の継承

2での処理において、該当する情報がなかったアクセス必須情報に関しては、

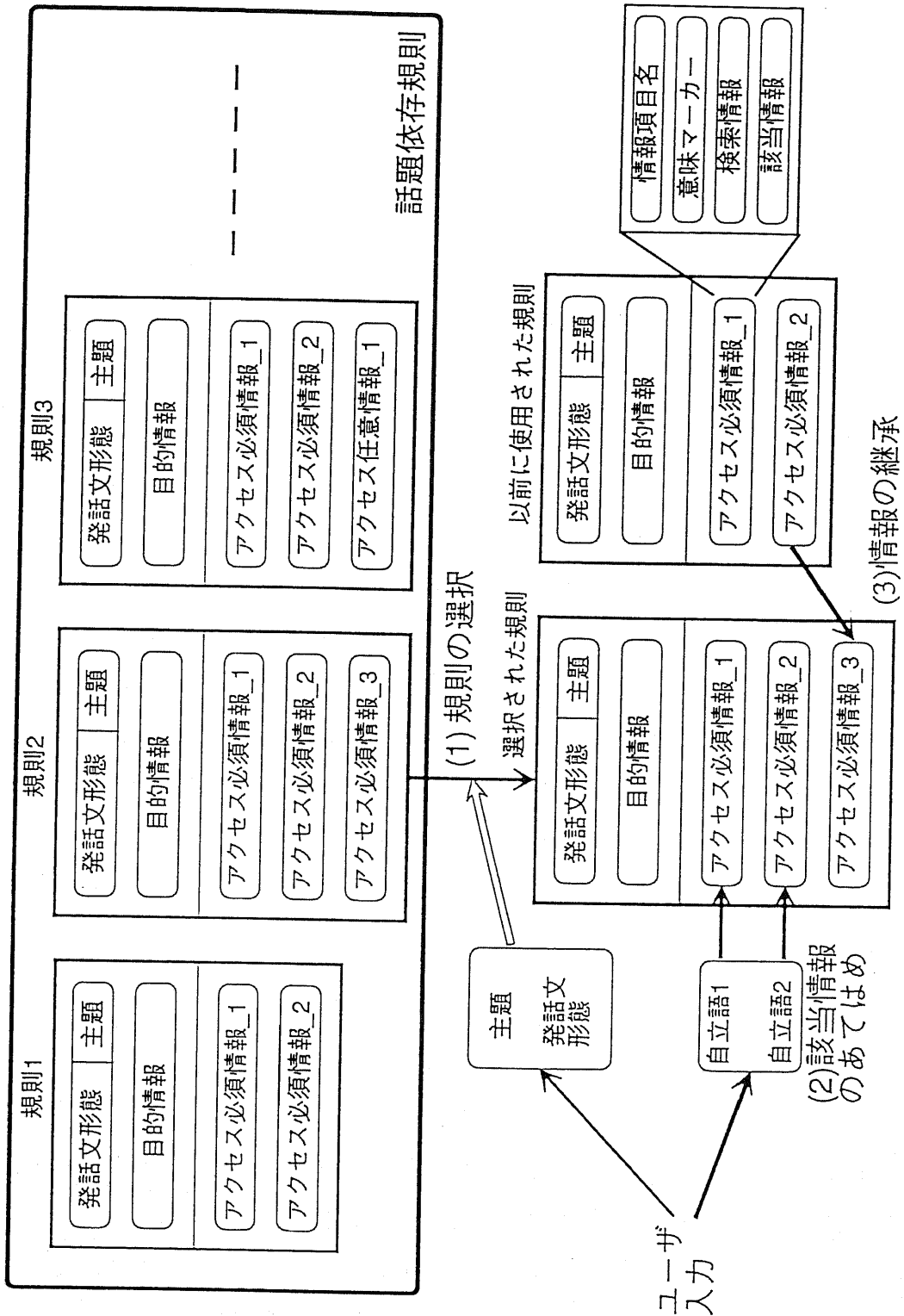


図 5.3. 質問文入力時の応答内容生成の流れ

対話処理の過程で1つ前に用いられた TODR から同じ情報項目名を持つものを検索し、その情報をアクセス必須情報の該当情報として利用する。この処理によって、ユーザ入力において省略・照応表現のために陽に記述されていない情報の補完が行なわれる。また1つ前の TODR からしか情報の継承を行なわないが、その情報が対話処理において必要である限り継承が繰り返されるので、実際にはある程度前に発話された情報でも、対話において暗黙の前提として両者で利用することができる。

また継承される情報項目の選択にあたっては、目的情報、アクセス必須情報、アクセス任意情報の順に情報項目名の検索を行ない優先的に情報の継承が行なわれる。

#### 4. 応答内容生成

前項までの処理で、TODR に記述されているすべてのアクセス必須情報が求まっていれば、知的システムへアクセスの必要となる情報はすべて得られているので、知的システムへのアクセスを行ない、そのアクセス結果をもとに、ユーザに対する応答内容を生成する。

一方、まだ求まっていないアクセス必須情報がある場合には、未登録のアクセス必須情報と現在処理対象となっている TODR をこれまでの処理で登録されている情報とともに対話履歴ベースに保存する。そしてシステムの方から協調的割り込みを行ない、ユーザに対してその不足情報を質問する。

#### 5.4.2 要求文入力の場合

4.2.2節のアルゴリズムによって要求文に分類される発話のうち、質問型疑問文に関しては、応答内容立案処理は質問文が入力された場合と全く同様に行なわれる。

主張文の場合には、FRD の対応関係によって応答文として扱う場合は次節の応答文入力の場合の手順で処理が行なわれる。要求文のまま扱う場合は、現在処理対象となっている TODR のアクセス任意情報にその内容を登録し、新たな TODR の記述にそって知的システムにアクセスに行き、その結果をユーザに応答する。

最後に挨拶が入力された場合は、同じように挨拶を返す TODR を用意しておき対処する。

### 5.4.3 応答文入力の場合

ユーザからの入力発話の形態が応答文である場合は、応答内容立案処理は以下の手順で行なわれる。

#### 1. 未登録のアクセス必須情報への単語の登録

この応答文に対応するシステムからの質問文が生成されたときに、対話履歴ベース中に保存されたアクセス必須情報に対して、この応答文で得られる情報を登録する。具体的にはアクセス必須情報に記述されている登録されるべき情報の意味マーカー（または品詞）と、応答文中の単語のそれを一致させることによって行なわれる。

この処理において、もし意味マーカー（または品詞）が一致する単語が応答文中にない場合は、この発話は応答文としては不適切であると判断され、4.4節で述べたように、この発話は質問文であると判断され、処理される。

#### 2. 対話履歴管理部への処理の移行

アクセス必須情報への登録処理が終了すると、処理は対話管理部に戻り、4.4節での記述に基づいて次の処理を行なう。

この場合は、FRDの対応関係から次に発話を行なうのはシステムの側となるので、対話履歴管理部から再び応答内容立案処理が呼ばれ、5.4.1節の4の処理を行なう。そして現在処理中のTODR中のすべてのアクセス必須情報が得られたかを調べ、得られていれば知的システムにアクセスにいき、その結果をもとに応答内容を生成する。まだ未登録のアクセス必須情報がある場合には、その情報に関してユーザに質問する。

## 5.5 知的システムへのアクセスとタスク依存規則

対話処理に用いられているTODRに記述されているアクセス必須情報がすべて確定すると、知的システムにアクセスすることによってユーザの求めている情報を得ることができる。このアクセスのための処理は知的システムの構成に依存するのでそれに適した方式で実現する必要がある。この実現の際に対話処理に関連してくる規則として、対話処理システムにTADR<sub>s</sub>を設けている。

本研究で構築した対話処理システムにおいては、知的システムとしてデータベースを用いているので、データベース検索用に次のようなTADR<sub>s</sub>を作成した。まず、

データベース検索の結果として該当する項目が多い場合、該当項目を全部ユーザに提示するか、該当個数だけ提示するか、ユーザが追加条件を設定するかのどれがいいかをユーザに質問して選択してもらうこととした。また、ユーザの要求する情報がデータベース中に保存されていない場合は、お詫びの発話を生成するようにした。

## 5.6 対話処理システムの動作例

以上が本研究で構築した音声対話処理システムの処理の詳細である。本節ではこのシステムがどの様に動作してユーザに対する応答を生成するかを図 5.4 に示す対話例をもとに図 5.5 にそって具体的に説明する。なお対話処理システムへの入力情報は単語列を、出力は深層意味表現を用いているので、図 5.4 の対話例はユーザ入力に相当する単語列と、それを入力することによってシステムで生成される応答意味表現をそれぞれ見やすくするために人間が表層文に変換したものである。

まずユーザからの「苗場と草津温泉のどちらがリフトが多いですか」という入力に対応する [苗場, 草津温泉, どちら, リフト, 多い, か] という単語列の情報が対話処理システムに入力される。対話履歴管理部ではこの入力に基づき、4.2.2 節、5.3.2 節におけるアルゴリズムにしたがい、この入力発話の形態の分類と主題の推定を行なう。形態の分類は、「どちら」という単語をもとに比較型の質問文と推定され、主題は「リフト」と「多い」の単語の主題記述から「リフト数」と推定される(5.1 節参照)。このとき「リフト」に関しては、意味マーカ―として施設と料金の 2 種類の辞書記述が存在するが、「多い」の主題記述と共通する主題「リフト数」をもつ施設の方の辞書記述が用いられる。

さらに対話管理部は、この推定結果から 4.4 節の手順にしたがい、この入力発話に対する FRD(1) を生成し、応答内容立案部に対して応答内容の立案処理を要求する。応答内容立案部では、推定された発話形態と主題の情報から、まず TODRs の中から TODR(1) が選択される。次に各単語の意味マーカ―と選択された TODR(1) のアクセス必須情報のそれをマッチングさせることによって、図 5.5 のようにアクセス必須情報を埋める。その結果、すべてのアクセス必須情報が得られており、データベースアクセスが可能であるので、その結果である「苗場です」を応答内容として生成する。そしてそれを受けて、対話履歴管理部では FRD(1) を終了させる。

次にユーザから「ナイターはありますか」という入力発話があると、同様の手順

対話履歴管理部はまず発話の形態と主題の推定を行ない、発話の形態として YN 型の質問文、主題として「施設存在確認」という結果を得、それに基づいて FRD(2) を生成する。さらに応答内容立案部では、この 2 つの推定された情報から TODR(2) が選択され、そのアクセス必須情報が埋められる。しかしこの場合はアクセス必須情報のうち「スキー場名」がユーザからの入力には省略されている。そこで前に使われた TODR(1) に「スキー場名」という同じ情報名をもつ項目を捜しにいき、その中で最も優先度の高い目的情報の項目から「苗場」という情報を継承する。その結果すべてのアクセス必須情報が得られたので、データベースにアクセスする。そしてその結果である「はい、あります」を応答内容として生成し、それを受けて対話履歴管理部では FRD(2) を終了させる。

次にユーザから「行くのにどのくらい時間がかかりますか」という入力があると、対話管理部ではこれまでと同様な過程で、発話形態として内容型の質問文、主題として「所要時間」という推定結果を得、それに基づいて FRD(3) が生成される。そして応答内容立案部では TODR(3) の選択とアクセス必須情報の登録が行なわれ、「スキー場名」が「苗場」であるという情報が前の TODR(2) から継承される。しかしもう 1 つのアクセス必須情報である「交通手段」については、以前の TODR 中には記述されていないので、応答内容立案部は「車と列車のどちらでいきますか。」という質問を生成し、この情報をユーザに対して問い合わせる。そしてこれを受けて、対話履歴管理部はこの質問に対応する FRD(4) を生成する。

システムからの質問に対して、「車でいきます」とユーザが応答した場合、4.4 節におけるユーザ入力の分類に関する規則に基づき、現在処理の対象となっている FRD(4) がシステム発話によって生成されているので、この入力発話の形態を応答文と推定する。そして引き続いて応答内容立案部で、この応答文が内容的に適切であるかを判断する。具体的には、応答文に含まれる単語「車」、「行く」の意味マーカと、質問の対象になっている情報項目(交通手段)の意味マーカを比較する。この場合は、「車」と「交通手段」の意味マーカがともに tl(tool:手段・道具)で一致し、応答として適切であると判断され、「交通手段」の情報項目に登録される。その結果を受けて、対話履歴管理部は FRD(4) の応答が得られたとして FRD(4) を終了させる。そして外側の FRD(3) がユーザの入力によって生成されているので、応答内容立案部に応答生成を要求する。応答内容立案部では、TODR(3) のアクセス必須情報が既にすべてもたまっているので、データベースアクセスを行ない、その結果であ

る「2時間26分です」をユーザに提示し、対話履歴管理部ではFRD(3)が終了する。

このように対話処理システム全体として、対話履歴管理部による対話の流れの制御に従って、必要に応じて適切な応答生成が行なわれている。また、表層のテキストには表現されないが、焦点位置に関する情報や応答発話のもつ意味情報も深層意味表現中に適切に記述されており、これらが音声合成処理において韻律情報の制御に用いられる。なお、焦点位置については図 5.4、図 5.5において下線で示した。

これらの処理は、ユーザ入力発話中の単語列から、単語辞書の記述内容に基づいて推定される発話の形態と主題に基づいて行なわれる。さらに TODR の選択、情報項目の埋め込み、必要情報の前の規則からの継承などが行なわれ、適切な応答が生成されていることがわかる。

## 5.7 まとめ

本章では、音声対話処理システムにおける応答内容立案処理について詳説した。はじめに、応答内容を立案する際に必要不可欠である対話の話題に依存した知識に関して、単語辞書と TODRs という 2 種類の知識を用意し、それぞれで記述する内容を規定した。単語辞書には、対話中に現れる単語に関して、単語表記、品詞、音声情報、さらに品詞によって異なる意味情報を記述することとした。一方 TODRs には、対話処理の過程で応答を生成するのに最適な TODR を選択するために用いられる発話の形態と主題に関する記述、さらにこの TODR が選択されている対話状況において対話の目的とされている情報である目的情報、目的情報を知的システムにアクセスして得るのに関連するアクセス必須情報・アクセス任意情報などをフレーム形式で記述することとした。

次に、応答を生成する際に最適な TODR を選択するのに必要となる、ユーザ入力発話で言及されている主題の推定に関して、その概要と推定アルゴリズムを説明した。主題推定アルゴリズムとしては、発話中の単語のうち、名詞、疑問詞、形容詞、動詞の順に優先順位を設け、単語辞書中の主題記述に基づいて、ユーザからの入力発話中の単語列の情報だけから主題を推定するアルゴリズムを提案した。

さらに、応答内容の立案処理の流れについて、ユーザ入力発話の形態が質問文、要求文、応答文である場合にわけ、それぞれの場合について手順を追って説明した。

また、ユーザから質問された項目に関する実際の情報を入手するために行なわれる知的システムへのアクセスについて概説し、アクセスの過程で対話処理に関連す

る処理を行なうための規則として TADR<sub>s</sub> を導入し、その具体例について言及した。

最後に、対話処理システム全体の動作を解説するために、ユーザと対話処理システムの間で行なわれた対話の例をあげ、それに沿う形でシステムの対話履歴管理処理、応答内容立案処理の流れを解説し、システムが有効に動作していることを示した。



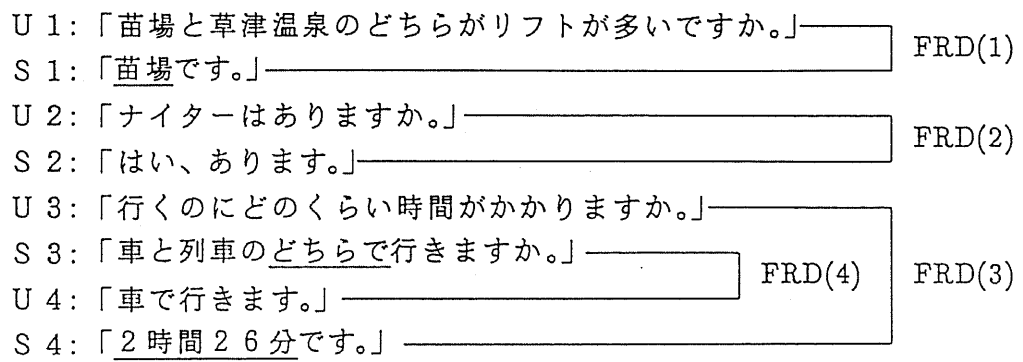


図 5.4. 対話処理システムによる対話の例

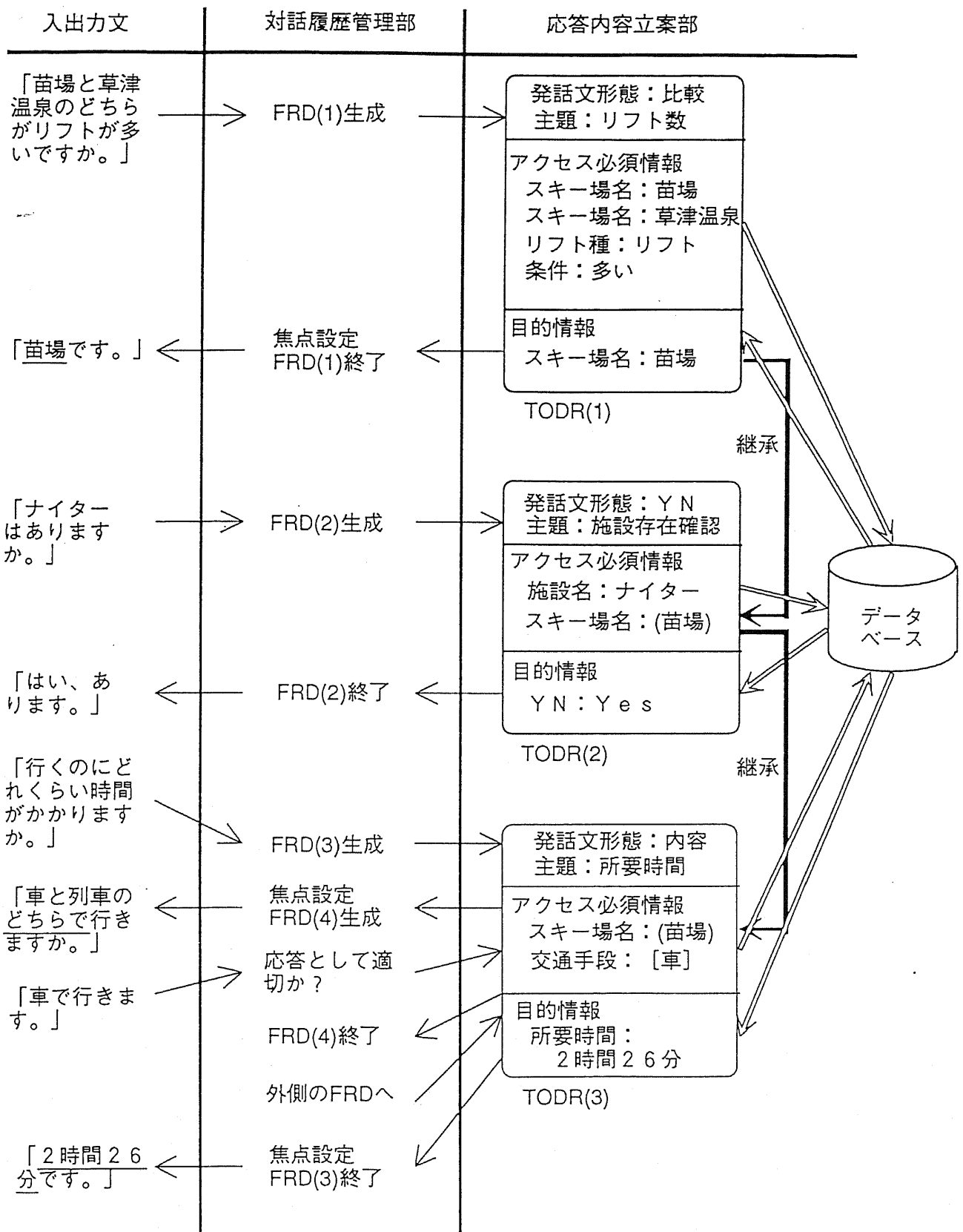


図 5.5. 対話処理システムにおける動作の概略の例

## 第 6 章

# 音声合成処理

## 6.1 はじめに

音声合成処理は、深層意味表現から表層レベルの音素記号列を生成する文生成処理と、音韻規則に基づいて、どのような音声を合成するかを記号レベルで作成する音韻処理、音韻処理で作成された記号列に基づいて音声合成器を駆動し、音声波形を生成する音響処理に大きく分けることができる。このうち音響処理に関しては、記号レベルでの記述に基づいて音声波形を生成するだけであるので、音声応答システムにおける処理もテキストを入力とした音声合成システムにおける処理もその処理内容は同じである。

一方音韻処理に関しては、テキストを入力とした音声合成に対しては、既にその音韻規則に関して検討が行なわれている [匂坂他 83, 広瀬他 89]。しかし、これを音声応答システムというユーザと音声対話を行なうことを前提としたシステムの音声合成に用いるには、対話処理の過程で生成された焦点や発話文の形態というような情報を反映でき、またテキストの読み上げ調ではなく、対話調の音声を合成できるように音韻規則の一部に修正を加える必要がある。

本章では、まず音声応答システムにおける音声合成処理とテキストを入力とする音声合成処理との違いと、それに対処するために音声応答システムにおける合成処理をどのように行なう必要があるかについて検討する。また、音声合成処理のうち、文生成処理の概要と両方の合成処理に共通する音響処理において用いられる音声合成器の構成について説明する。次に、対話処理において生成される韻律的特徴の制御に重要な役割を果たす情報を音声合成において反映させるため、対話中の発声に関して、発話文の対話の中での役割(発話の意味情報)と焦点との関連から、発話速度と発話のパワー、基本周波数の時系列パターンの分析を行ない、その分析結果を検討し、音声応答のための音韻規則の改良を行なう。そして作成した音韻規則に基づいて合成実験を行ない、その規則の有効性を検討する。

## 6.2 音声合成処理の概要

### 6.2.1 音声応答システムにおける音声合成

音声合成処理に関して、音声応答システムの中には、単に表層文のテキストを生成することを対話処理の目的とし、その生成されたテキストを市販のテキストを入力とする音声合成器に入力することによって応答音声を合成している例も存在する

[Zue et al. 91]。このようなシステム構成で音声応答システムを構築した場合には、次のような問題点が存在する。

- 対話処理部での深層レベルの処理結果をいったん表層レベルのテキストに変換し、音声合成部ではテキストとして入力される表層文を解析して、再び深層レベルの処理を行なって音声を合成するため、処理に無駄が多い。
- 表層のテキストの解析では、対話処理で得られる深層レベルの情報のすべてを抽出することは非常に困難で、その情報をすべて韻律的特徴の制御に反映させることは不可能である。
- 音声合成を行なうためのテキスト解析処理において、現状の言語処理のレベルでは解析誤りの発生は避けられない。
- テキストを入力とする音声合成では、テキストを読み上げるような感じの音声の合成を行なうが、対話中における発話はテキストの読み上げとは韻律的な点でかなり異なる音声となっている。

このような問題点を解決するために、本研究では対話処理の結果として得られる情報を深層意味表現中に反映させ、その記述に基づいて、表層レベルのテキストではなく合成する音声の分節的・韻律的特徴を直接的に制御できる単音・韻律記号列を生成し、適切な音声の合成を行えるような方式にした。これによって、合成音声に対して対話処理の過程で得られる焦点などの情報を適切に反映させることができる上に、合成に際して表層のテキストを解析する必要がなくなり、処理の無駄や解析誤りを取り除くことができる。さらに合成音声の韻律的特徴に関しては、対話中における発話の韻律的特徴についての分析を行ない、対話中の発話に近い音声を合成することとした。

この音声合成部の構成を図 6.1 に示す。音声合成部は文生成処理、音韻処理、音響処理の 3 つの処理によって構成されている。文生成部では合成される音声の読みに大まかに対応する音素記号列が生成され、さらにそれに対して音韻処理部において異音化処理が施されて、単音記号列が生成される。また音韻処理部では、音声のイントネーションやアクセント、抑揚などに対応する韻律記号列が生成される。そして音響処理部では、これらの処理の結果として得られる単音・韻律記号列に基づいて音声合成器を駆動し、音声波形が合成される。

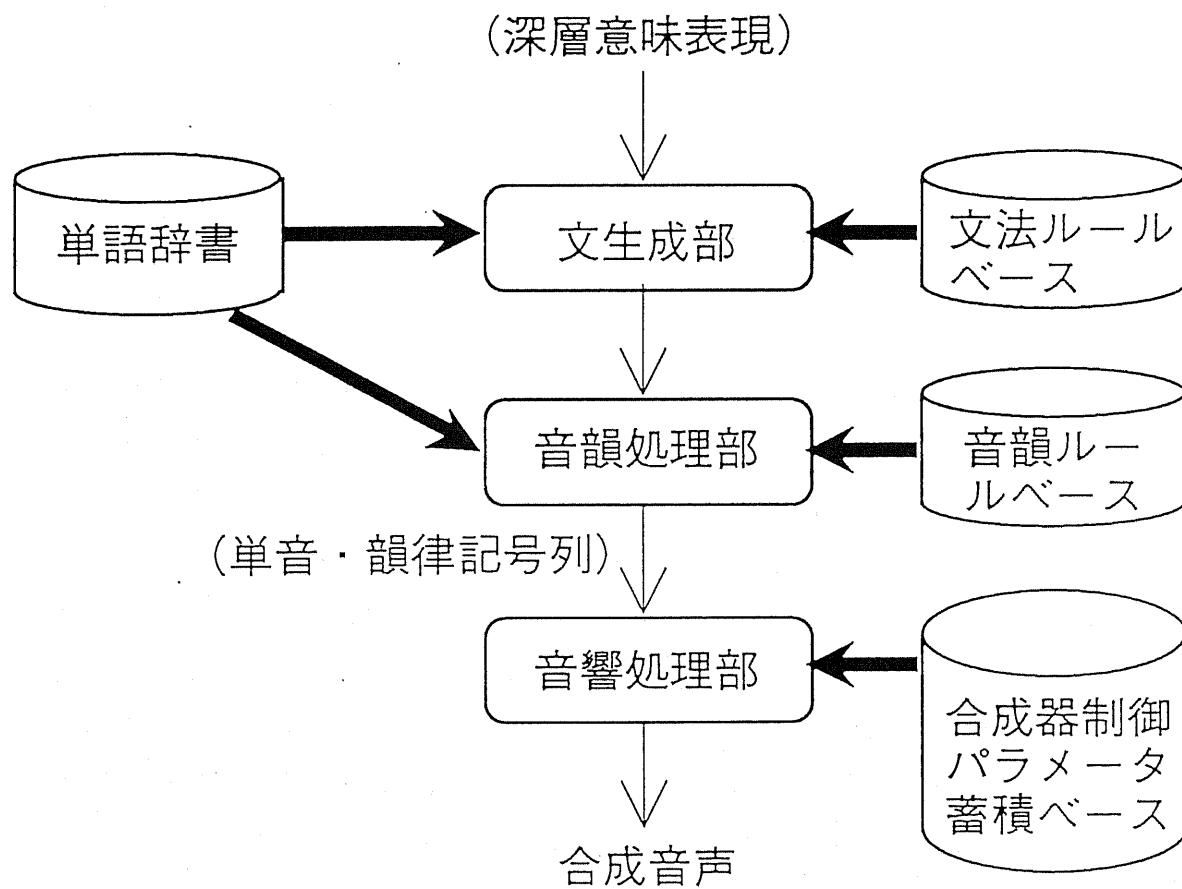


図 6.1. 音声合成部の構成

### 6.2.2 文生成処理

文生成処理においては、対話処理の結果として得られる深層意味表現中の記述に基づき、表層レベルの応答文の生成を行なう。この時、音声応答システムとしての最終目的は、応答内容をユーザに合成音声で提示することであるので、表層レベルの応答文の生成においては、合成する音声の読みに大まかに対応する音素記号列を生成する。ただし、5.2.1節で述べたように、単語辞書中の単語表記には漢字かな混じりを用い、音素記号列の生成と同時にテキストも生成できるような構成とした。

文生成処理の手順としては、まず深層意味表現の述語の単語辞書中での記述に基づき、深層意味表現中の各格要素に用いられる助詞を決定し、それを格要素の filler に結合し、表層文の文節を作成する。次に述語について、深層意味表現の意味付与における記述をもとに述語の活用と助詞・助動詞の付与を行なう。この処理に関しては、[草薙 88]の記述に基づいた。なお、システムの応答としては、文末に「です」、「ます」を用いる丁寧な発話とし、また発話がその意味情報として疑問をもつ場合は、文末に疑問の終助詞「か」を常に付与することとする。そして最後に各文節と述語部分を結合して表層文を作成する。

### 6.2.3 音響処理

音響処理では、まずあらかじめ蓄えられている個々の単音に対応する音声合成器の制御パラメータを単音記号列にしたがって接続する。さらに韻律記号列にしたがって、後述する基本周波数パターン生成過程のモデルに基づいて有声音源の基本周波数を求める。そして、これらの情報に基づいて音声合成器を駆動し、音声波形を生成する。この処理は音声応答システムにおける合成処理においても、テキストを入力とする音声合成方式の場合と全く同じ処理が行なわれる。

本研究では音声合成器として、音源特性と声道伝達特性が明確に分離されており、人間の音声生成過程のシミュレートが容易なターミナルアナログ型音声合成器を用いている。このターミナルアナログ型の音声合成器の方式としては、これまで主として次の3種類のものが用いられている。

- 周波数スペクトル上の極と零点にあたるフィルタを縦続接続した直列型回路ですべての音声を合成する方式 [Fant 53]。
- 極にあたるフィルタを並列接続した並列型回路ですべての音声を合成する方式 [Lawrence 53]。

- 従来の研究で声道伝達特性が比較的明確になっている母音などの音声に関しては直列型回路で、摩擦の音などあまり明確でないものは並列型回路で合成を行なうという直列、並列型回路を併用する方式 [Klatt 80, Klatt et al. 90](図 6.2 参照)。

これに対して本研究では、生成過程の異なる音声ごとに複数の直列型回路を構成し、これらを独立に制御して音声を合成するマルチカスケード型の音声合成器の構成を新たに提案し、これを用いて音声の合成を行なっている。この合成器の構成を図 6.3 に示す [Fujisaki et al. 90b]。

### 6.3 対話調の発話の韻律的特徴に関する分析と検討

前節で述べたように、テキストを入力とする音声合成は、読み上げ調の音声の合成を目的として音韻規則が記述されている。しかし音声対話中の発話(以下、対話調の発話と呼ぶ)は、読み上げ調の発話とは韻律的特徴の面で違いがみられる。そこで両者の韻律的特徴がどのように異なっているかを調べるために、韻律的特徴のうち、発話速度と発話のパワー、基本周波数の時系列パターンの分析を行なった。

#### 6.3.1 分析資料

韻律的特徴の分析の対象とする音声資料としては、3.2.1 節で述べた音声資料のうち、独自に収録したスキー場を話題とする対話中の発話を用いた。そして、対話中で他者の発話の割り込みや言い直しがなく、言いよどみが少ない発話を切り出して対話資料とした(以下対話中の発話と呼ぶ)。さらにそれをテキスト化し、そのテキストを対話での発声者と同じ人に対して、普通に読み上げるように指示した場合(以下通常の読み上げと呼ぶ)と、目の前に相手がいる、その相手に向かって発話しているような意識をもってという指示をして読み上げてもらった場合(以下意識した読み上げと呼ぶ)のそれぞれの読み上げについても収録した。

また、4.5.1 節での規則に基づき、対話の状況から発話中に焦点がおかれていると思われる発話を選び、焦点のある発話の韻律的特徴の分析に用いることとした。さらにこれとは別に、発声者にテキスト中のどこに焦点をおくかを指定して読み上げてもらった発話と普通に読み上げてもらった発話の収録も行ない、分析に利用した。

これらの収録はすべてデジタルオーディオテープ(DAT)を用いて、サンプリング周波数 48kHz、16bit 直線量子化で行なった。そして録音した DAT を再生して、遮



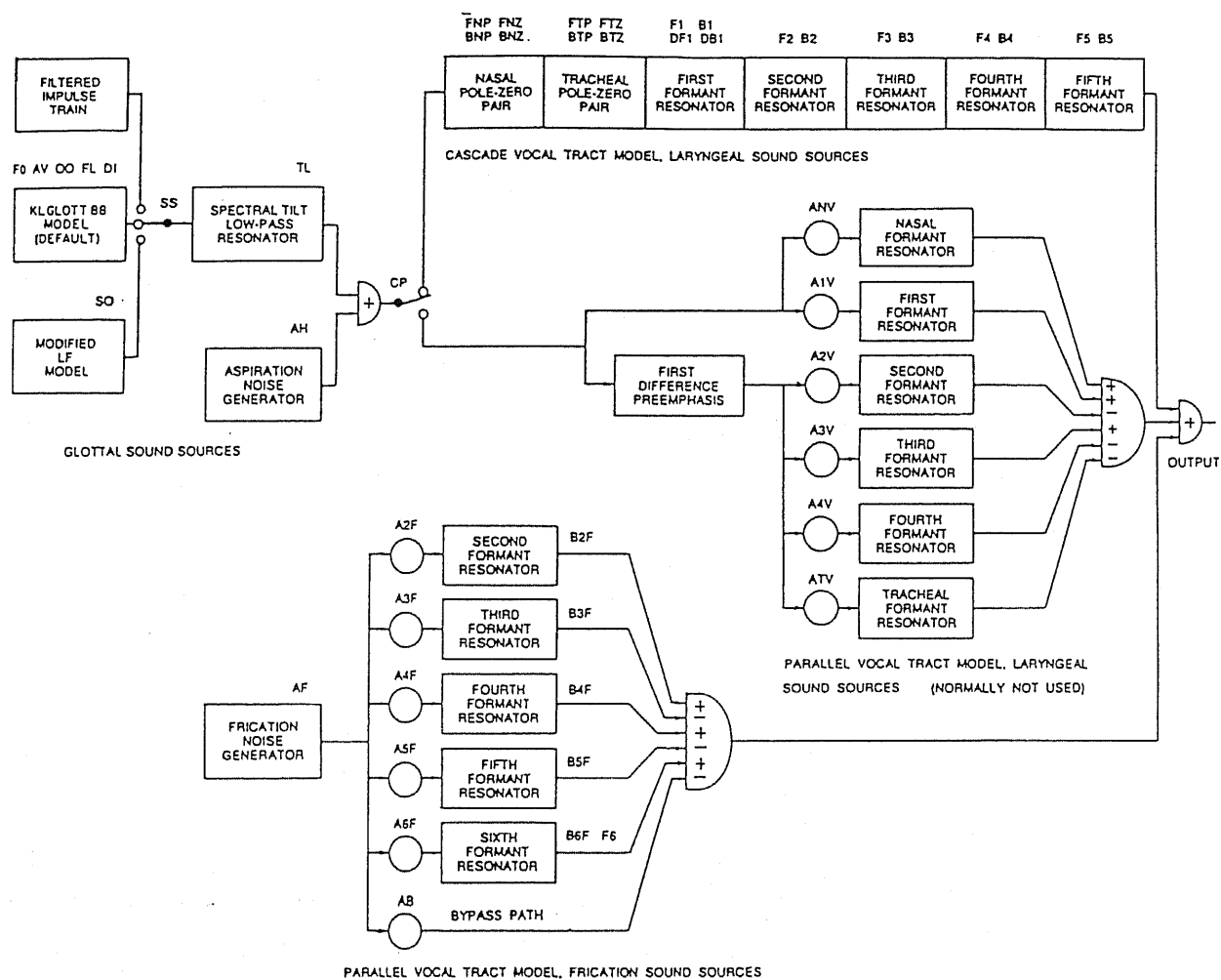


図 6.2. 直列、並列型回路併用型のターミナルアナログ型音声合成器の構成 (文献 [Klatt et al. 90] より転載)

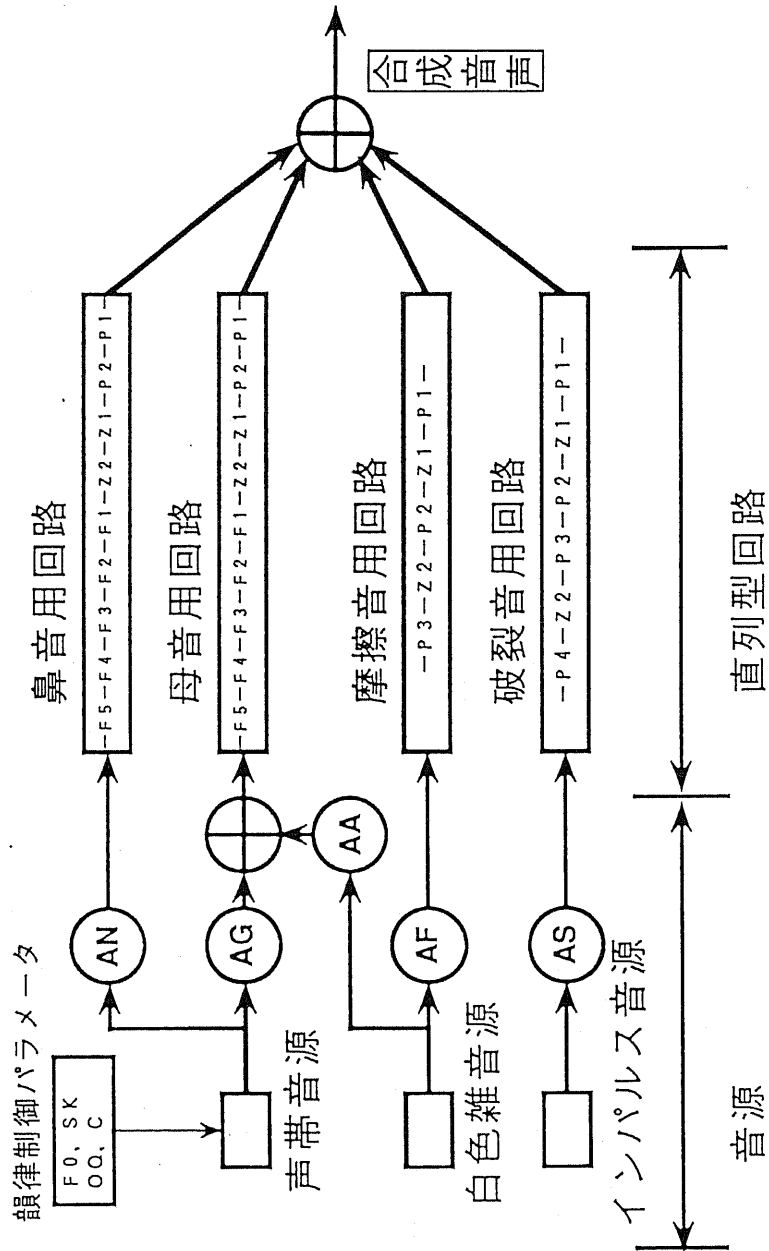


図 6.3. マルチカスケード型音声合成器の構成

断周波数 5Hz のハイパスフィルタ (HPF) と遮断周波数 4.5kHz のローパスフィルタ (LPF) を通した後に、サンプリング周波数 10kHz、12bit 直線量子化で AD 変換を行ない、PCM データとし、これを分析に用いた。

### 6.3.2 発話速度に関する分析と考察

発話速度は、分析対象区間において発声された mora 数を、その区間時間で割ることによって定義する。なお、区間時間の計測は、波形データに基づき視察によって行なった。

発話速度に関しては、発話者ごとに通常の発話自体の発話速度の個人差が大きいので、すべての分析結果をまとめることはできないので、発話者ごとにそれぞれ 3 種類の発話環境 (対話中の発話、通常の読み上げ、意識した読み上げ) での発話速度を集計した。

1 つの発話全体を分析対象区間とした発話速度の分析結果として、例として男女 1 人ずつの発話速度の分布を、それぞれの発話環境ごとに図 6.4、図 6.5 に示す。この分析から次のような結果が得られた。

- すべての発話者において、それぞれの発話速度の平均は通常の読み上げ、意識した読み上げ、対話中の発話の順で速くなり、通常の読み上げに比べて、発話者全体を通して、意識した発話で平均約 11%、対話中の発話で平均約 19% の発話速度の上昇がみられた。
- すべての発話者において、通常の読み上げや意識した読み上げに比べ、対話中の発声の発声速度はばらつきが大きい。これは短い言いよどみによる影響も多少あるが、それ以上に発話ごとに (1 つの発話内でも) 発話速度がダイナミックに変化しているためである。

さらに 1 つの発話内での発話速度の変化について分析してみると、次のような特徴がみられた。

- 文末にかけて発話速度が上昇する傾向にある発話が多く、発話によっては文末において速いところで 15mora/sec 程度にまで発話速度の上昇がみられる。さらに発話速度が速くなる部分においては、同時に発声のパワーの減少もみられる (次節参照)。また、文末で発話が速くなるものとしては、「と思います」、「あります」など意味的にあまり重要でない場合に多くみられる。

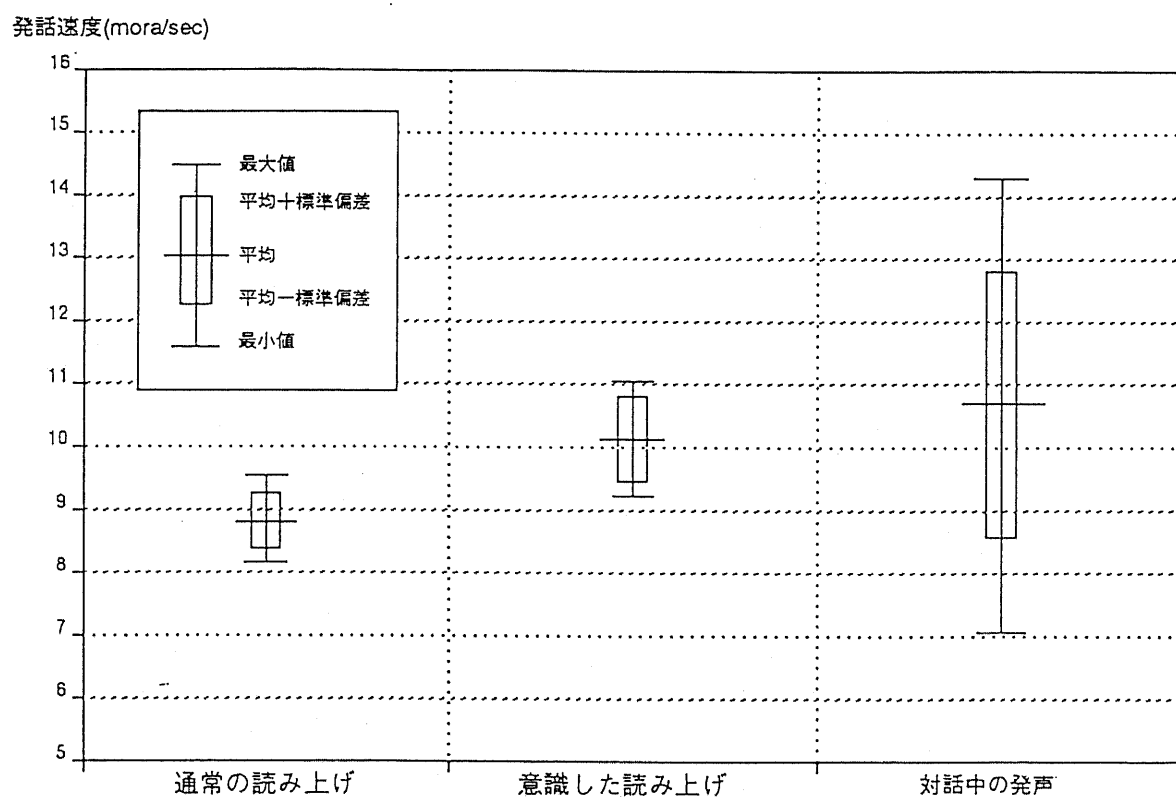


図 6.4. 発話環境の違いによる発話速度の変化(男声の例)

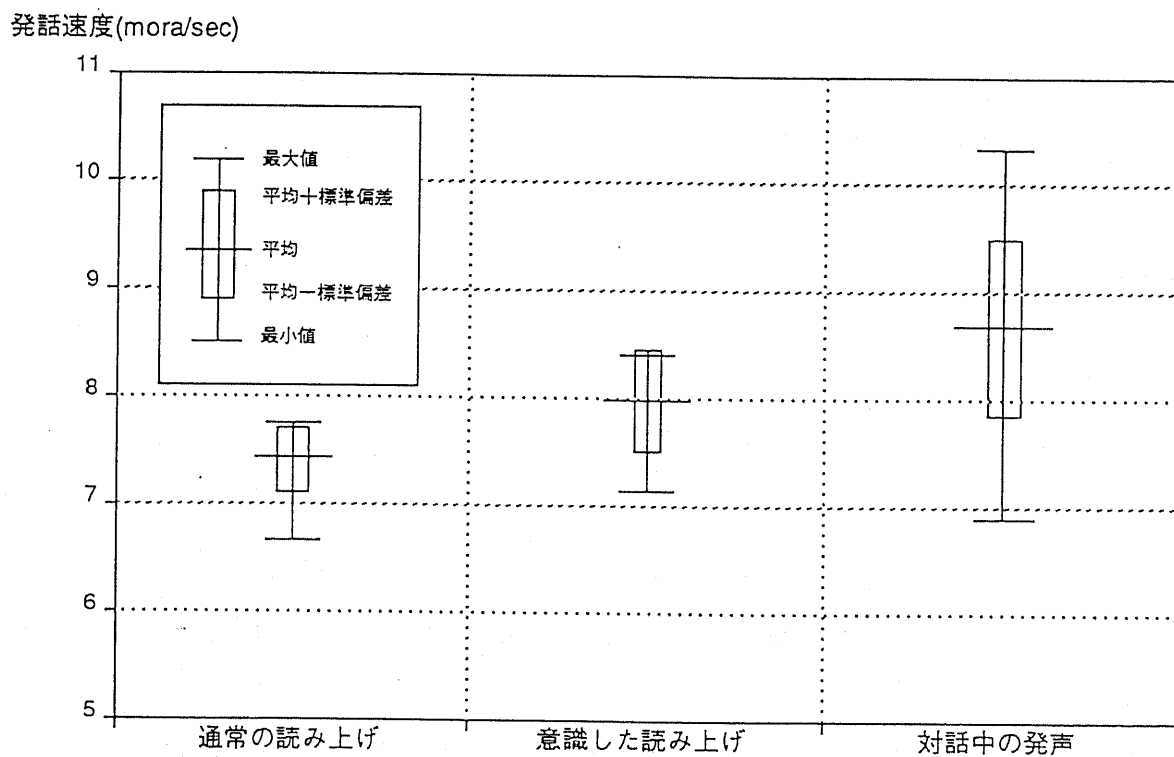


図 6.5. 発話環境の違いによる発話速度の変化(女声の例)

- 焦点の存在による発話速度の変化については、焦点のおかれている部分の発話速度とおかれていない部分の発話速度を比べてみると、有意な差はみられない。すなわち、焦点がおかれている部分で発話速度が遅くなるということはない。このことから、焦点の表出にはむしろ基本周波数パターンや発話のパワーが大きく関係していると考えられる。

### 6.3.3 発話のパワーに関する分析と考察

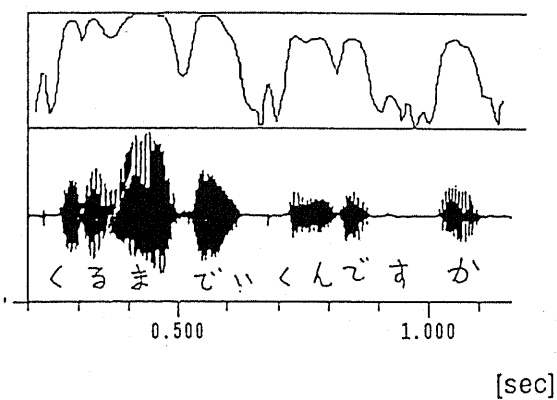
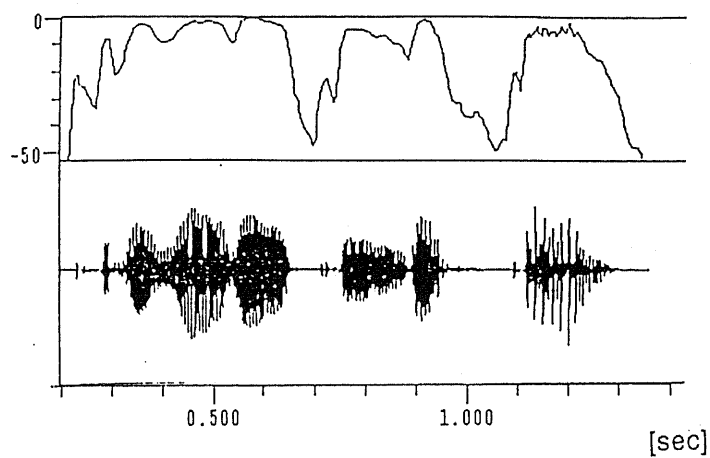
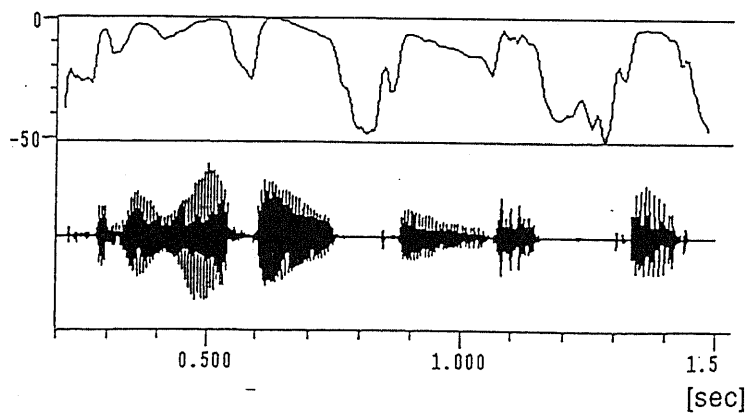
次に3種類の発話について、その発話のパワーの時間変化に関する分析を行なった。分析の方法としては、フレーム長 25.6ms のハミング窓をかけ、フレーム内の FFT パワーを求め、さらにそのフレームを 5ms ずつずらして発話全体のパワーの変化を分析した。

分析の結果として次のような特徴がみられた。

- すべての種類の発話において、文末にかけてパワーの減少傾向がみられる。これは呼気の問題などから生じる現象であると考えられるが、この減少の大きさは、通常の読み上げ、意識した読み上げ、対話中の発話の順に大きくなる傾向がある。
- 発話に疑問の意図が含まれる場合(質問文の場合)は、文末において基本周波数が上昇する場合が多く(次節参照)、その場合にはパワーにも上昇がみられる(例:図 6.6)。それに対して応答文の場合は、文末のパワーの低下が顕著に現れ、また読み上げの場合に比べて、対話中の発話においてはパワーの低下の開始時点が早い(例:図 6.7)。
- 焦点のある部分(及び焦点を意識して発声した発話における焦点の部分)では、他の部分に比べてパワーの増加がみられた(例:図 6.8)。

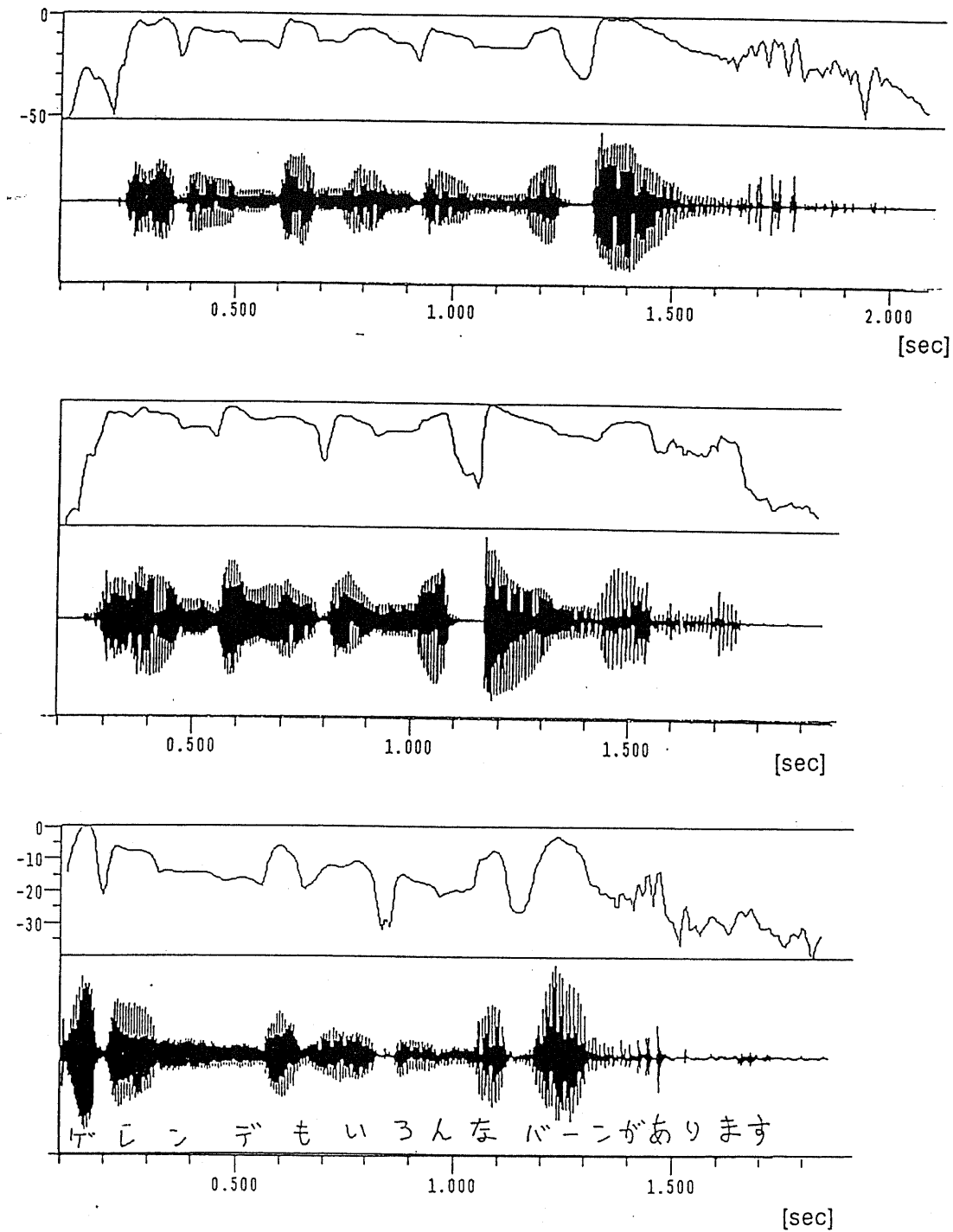
### 6.3.4 基本周波数に関する分析と考察

基本周波数の時系列パターンの抽出には、遅れ時間に比例した窓長で自己相関係数を用いるアルゴリズム [Hirose et al. 92] を採用した。そしてこのアルゴリズムに基づいて抽出された基本周波数パターンを視察によって修正して、基本周波数の次系列パターンを求めた。



上段: 通常の読み上げ  
 中段: 意識した読み上げ  
 下段: 対話中の発話

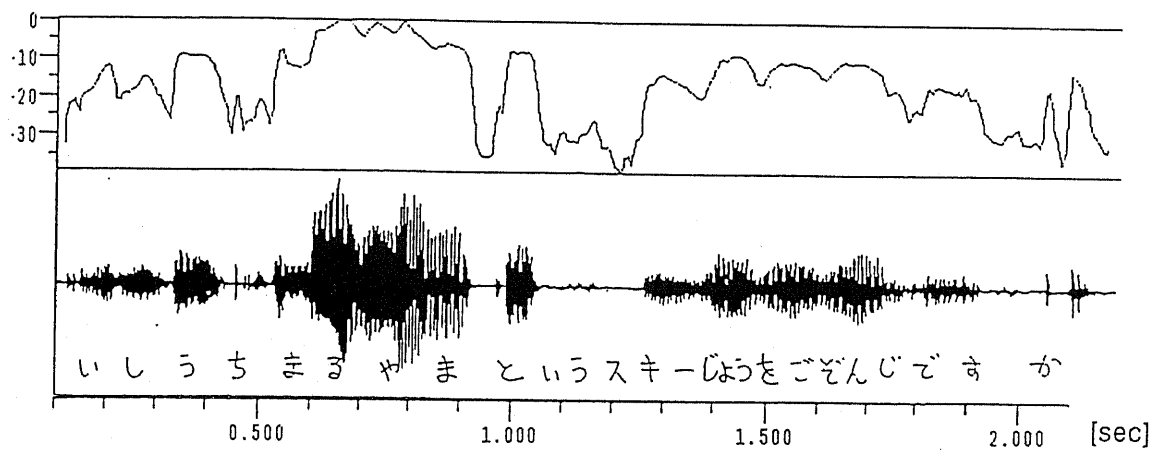
図 6.6. 質問文の発話におけるパワーの時間変化



上段: 通常の読み上げ  
 中段: 意識した読み上げ  
 下段: 対話中の発話

図 6.7. 応答文の発話におけるパワーの時間変化





「石打丸山」に焦点がおかれている。

図 6.8. 焦点が存在する発話のパワーの時間変化

抽出された基本周波数の時間パターンに対しては、さらに基本周波数パターンの生成過程のモデル [藤崎他 71, Fujisaki et al. 84] に基づいて分析を行なった。

基本周波数パターンの生成過程のモデルとは、基本周波数パターンを対数軸上で表現した場合の形状が、全体の声の高さによらず類似している点に着目し、基本周波数パターンを、句頭から句末にかけて緩やかに減少するフレーズ成分 (おおよそ発話のイントネーションに対応する) と、局所的な起伏のアクセント成分との和で表現する。そしてフレーズ成分をインパルス状のフレーズ指令、アクセント成分をステップ状のアクセント指令の臨界制動 2 次系の応答で近似するというモデルである。このモデルの概略を図 6.9 に示す。

数式では、基本周波数を時間  $t$  の関数として  $F_0(t)$  とすると、対数基本周波数  $\ln F_0(t)$  は次のように表わされる。

$$\ln F_0(t) = \ln F_{0\min} + \sum_{i=1}^I A_{pi} G_{pi}(t - T_{0i}) + \sum_{j=1}^J A_{aj} \{G_{aj}(t - T_{1j}) - G_{aj}(t - T_{2j})\}$$

ただし、 $G_{pi}(t)$ 、 $G_{aj}(t)$  はそれぞれフレーズ制御機構のインパルス応答関数、アクセント制御機構のステップ応答関数で、以下のように表わされる。

$$G_{pi}(t) = \begin{cases} \alpha_i^2 t \exp(-\alpha_i t) & (t \geq 0) \\ 0 & (t < 0) \end{cases}$$

$$G_{aj}(t) = \begin{cases} \min[1 - (1 + \beta_j t) \exp(-\beta_j t), \theta] & (t \geq 0) \\ 0 & (t < 0) \end{cases}$$

$\alpha_i$ 、 $\beta_j$  は、それぞれフレーズ制御機構、アクセント制御機構の固有角周波数である。 $A_{pi}$ 、 $A_{aj}$  は、それぞれフレーズ指令、アクセント指令の大きさを、 $T_{0i}$  はフレーズ指令の時点、 $T_{1j}$ 、 $T_{2j}$  はアクセント指令の始端と終端の時点である。また、自然音声の分析結果より、 $\alpha_i = 3.0 \text{ rad/sec}$ 、 $\beta_j = 20.0 \text{ rad/sec}$ 、 $\theta = 0.9$  に固定しうることが明らかになっている [広瀬他 82]。ただし  $F_{0\min}$  は個人によって異なる発声における最少基本周波数であり、特に男声と女声ではその値が大きく異なる。

またこのモデルは、その特徴としてフレーズ指令とアクセント指令が、統語情報などの言語情報と密接に関連するため、音声合成の際の基本周波数の制御にも有力な手法である。

このモデルに基づく分析としては、音声資料の発話から抽出された基本周波数パターンから、まずフレーズ指令とアクセント指令の位置と大きさを推定し、そして推定された指令の位置と大きさを視察によって修正した後、最後に合成に基づく分析法 (Analysis-by-Synthesis: A-b-S 法) によって、指令の正確な位置と大きさを求めた。

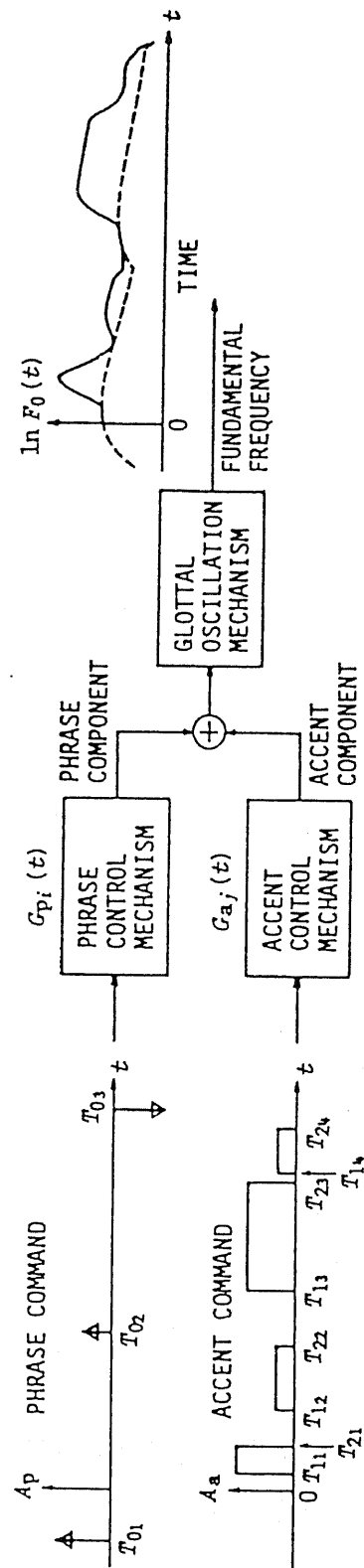
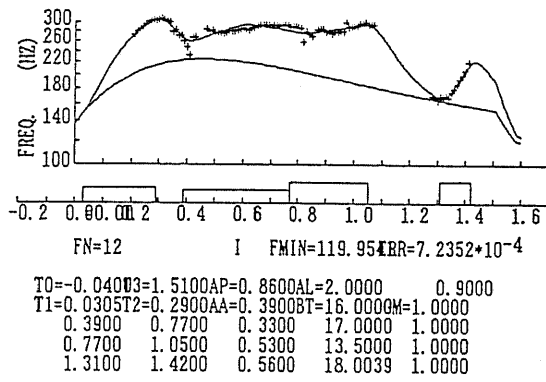
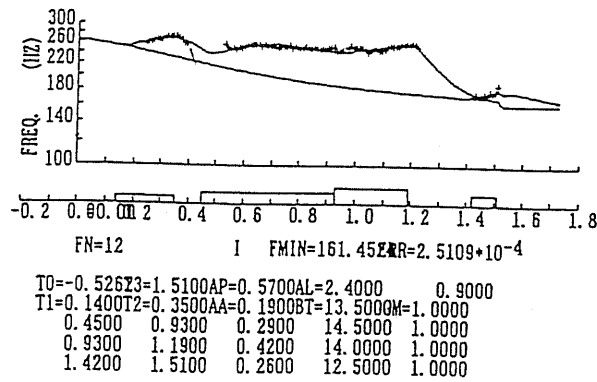


図 6.9. 基本周波数パターン生成過程のモデル

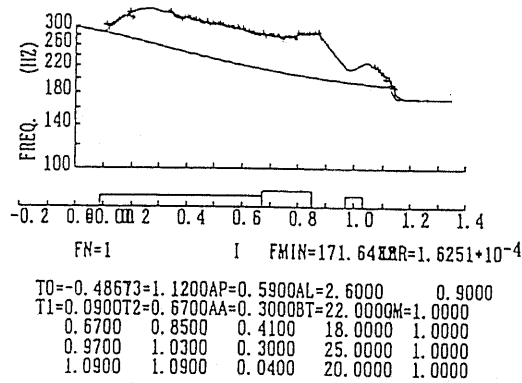
6.3.3節で述べたように、対話中の音声に関しては、声の大きさが発話中で大きく変わることがあり、特に文末にかけてが声が小さくなる場合が多い。さらにいわゆる「ですます調」の丁寧な発話においては、文末付近の「です」、「ます」における「す」の部分が無声化する傾向が顕著である(例:図 6.10)。これらの理由で、文末部分の基本周波数がうまく抽出できない発話も存在したため、丁寧な発話の他に、いわゆるくだけた感じ発話による対話も行なってもらい、分析に用いた。

基本周波数に関する分析の結果から、次のような特徴がみられた。

- くだけた感じの発話による対話の場合も含めて、対話中での発話の役割が質問である(意味情報として疑問をもつ)発話のうち、約 67%の発話において文末での基本周波数の上昇がみられた(例:図 6.11)。さらに、このうち文末に疑問の終助詞「か」が用いられている発話においては、約 88%のもので文末での基本周波数の上昇がみられた。
- 対話中での発話の役割が応答である発話においては、文末で発話のパワーが落ちている場合が多く、基本周波数がうまく抽出されない例があるが、おおむねフレーズ成分の低下にしたがった基本周波数の減少がみられる(例:図 6.12)。
- 通常の読み上げと意識した読み上げの2種類の発話を比較すると、質問の発話において、だいたい意識した読み上げ方が文末において顕著に基本周波数が上昇している。そして、対話中の発話で文末で基本周波数が上昇しているものに関しては、意識した読み上げとだいたい同じような基本周波数の時系列パターンとなっている。
- 焦点があると考えられる部分の多くでパワーの上昇とともに基本周波数の上昇もみられる(例:図 6.13)。またパワーのパターンと基本周波数のパターンとの間に高い相関がみられる。
- 読み上げの発話に比べて、対話中の発話においては、基本周波数の変化の幅が大きく、特にアクセント指令の大きさが大きくなる傾向にある。

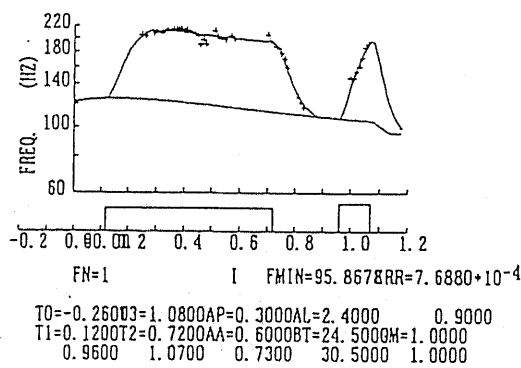
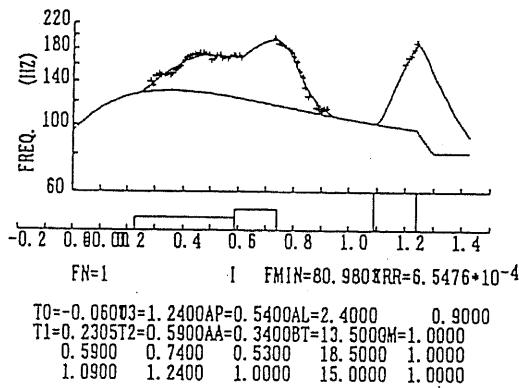
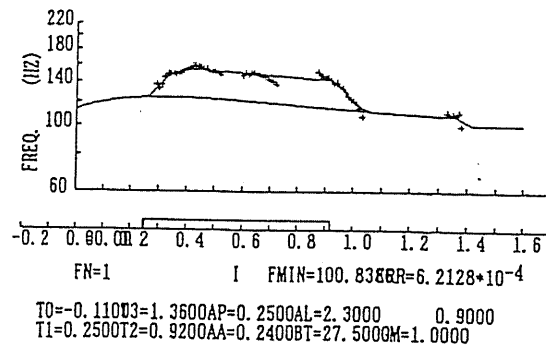


おんせんはありますか



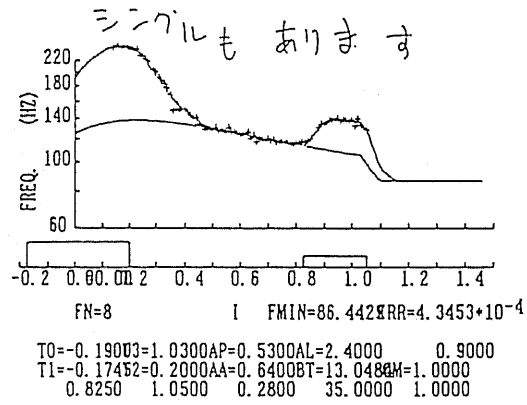
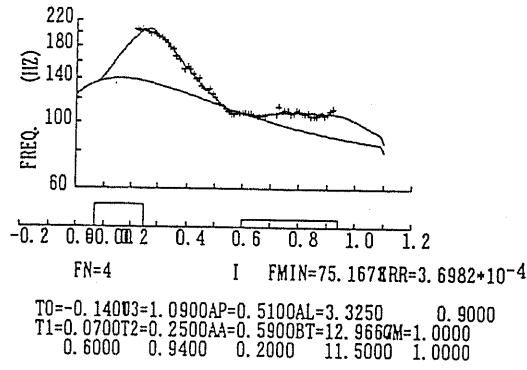
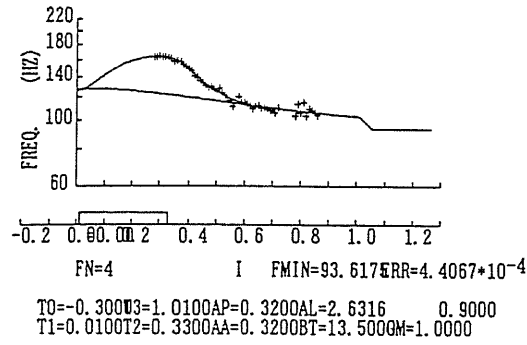
上段: 通常の読み上げ  
 中段: 意識した読み上げ  
 下段: 対話中の発話

図 6.10. 無声化のため基本周波数の抽出が困難である例



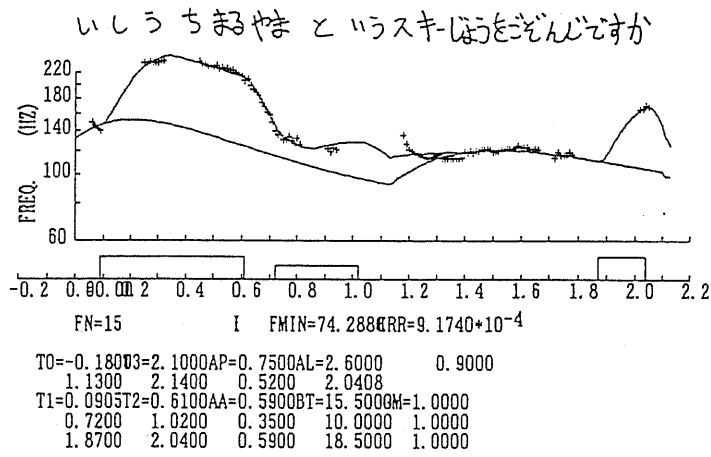
上段: 通常の読み上げ  
 中段: 意識した読み上げ  
 下段: 対話中の発話

図 6.11. 質問文の文末において基本周波数の上昇がみられた例



上段: 通常の読み上げ  
 中段: 意識した読み上げ  
 下段: 対話中の発話

図 6.12. 応答文の基本周波数パターンの例



「石打丸山」の部分に焦点がある。

図 6.13. 発話中に焦点がある場合の基本周波数の時間パターンの例



## 6.4 対話調の音声合成のための音韻規則

### 6.4.1 音韻規則化

以上の韻律的特徴に関する分析結果に基づき、音声応答のための音声合成を行なうという観点から、音韻規則をテキストを入力とする合成における音韻規則を基に、これを改良する形で作成した。なお、韻律的特徴のうち合成の際の基本周波数の制御には、前述の基本周波数パターン生成過程のモデルを用いた。

また前節で述べたように、発話のパワーと基本周波数の間には高い相関がみられる。さらに音声合成においてはその有声音源波形として多項式によって近似されたモデル [Fujisaki et al. 86] を用いているが、このモデルの特徴上、基本周波数が上がると、それにともなって発声のパワーも上昇する。これらの点から、音韻規則としては発話のパワーの制御は、文末の発話のパワーの低下の部分だけ独立に制御することとし、そのほかの部分では基本周波数の制御のみを行ない、パワーの制御は行わないこととした。

応答内容を合成音声でユーザに提示するという本研究の趣旨を考えた場合、応答システムが合成する発話を次のように規定することができる。(6.2.2節参照)

- 発話は丁寧なものとし、その文末は「です」、「ます」を用いる。
- 発話の意味情報と焦点位置の情報は対話処理システムで生成される深層意味表現に記述されている。
- 発話の形態が質問である発話には、文末に疑問の終助詞「か」をつける。

このように規定された発話を合成する際に韻律的特徴を生成する音韻規則として、前節での分析に基づいて以下のように規定した。

発話速度は、あまり速くし過ぎると音節間の接続が不自然になり明瞭性が低下するという音声合成器の特性も考慮して、多少はやい程度の発声となる 8mora/sec を標準の速度とした。さらに文末において、「です」、「ます」の 2mora 前の音節の持続時間を -2.5ms、直前の mora の音節の持続時間を -5ms、「です」、「ます」の 2mora の持続時間を -10ms だけ短縮する。質問文の場合はこのあとに疑問の終助詞「か」が付与されるが、これについては標準の持続時間とする。

発話のパワーについては、分析の結果としては文末で発話中の最大パワーから -20dB 程度低下している例も多かったが、音声合成器の明瞭度から考えて、大きく

パワーを落とすと明瞭度が極端に低下する恐れがある。したがって、「です」、「ます」の 2mora 前の音節のパワーを $-2.5\text{dB}$ 、直前の mora の音節のパワーを $-5\text{dB}$ 、「です」、「ます」の 2mora のパワーを $-10\text{dB}$  だけ低下させる。そして、「です」、「ます」における「す」は無声化させる。また、質問文の場合はこのあとに疑問の終助詞「か」が付与されるが、これについては標準のパワーとする。

基本周波数については、すべてフレーズ指令、アクセント指令に対応するフレーズ記号、アクセント記号の韻律記号で記述し、この記述に基づいて基本周波数パターン生成過程のモデルに基づいて実際の基本周波数を計算し、この値に基づいて合成音声を制御する。

テキストの場合の韻律記号列の生成規則については、すでに [河井 88] で検討が加えられている。この中でフレーズ記号は、文頭や句頭のフレーズ成分の立ち上がりに対して典型的な 3 種類の大きさに分け、これと文末の立ち下げ用の 1 種類の計 4 種類が用意されている。一方アクセント記号は、立ち上がりの大きさと時点を表わすのに起伏型と平板型のアクセントのそれぞれに対して 3 種類ずつ、そしてアクセントの立ち下がり時点を表わすための指令が両者に共通で 1 種類、全部で合わせて 7 種類が用意されている。そして、それぞれの指令の大きさは表 6.1 のように規定されている。また指令のおかれる時点としては、文頭におかれたフレーズ指令に関しては文頭の音節の母音開始時点の 210ms 前に、そのほかの位置のフレーズ指令に関しては韻律記号が記述されている次の音節の母音開始時点の 80ms 前、アクセント指令の場合は韻律記号が記述されている次の音節の母音開始時点の 70ms 前と規定されている。

対話調の音声の合成においては、分析結果に基づきこの規則を以下のように修正する。

まずフレーズ指令の設定に関しては、格構造による表現では、各深層格は述語を修飾するような形で記述されるので、表層構造として述語を文末にくるように統一すると、統語構造としては述語と各深層格の要素間の関係はすべて右枝分かれの構造となる。そして左枝分かれの構造は各深層格の内部にのみ存在する。したがって、フレーズ指令は右枝分かれの部分に対応するそれぞれの深層格の要素の間に設定し、それをテキストを入力とする音声合成用の音韻規則と同様に、フレーズ指令間のモーラ数に応じて修正することによって設定する。

次にアクセント指令に関しては、アクセント型やアクセント変形 (accent sandhi)

の現象はテキストからの合成の場合と同じであるので、文中における指令の設定に関しては、テキストからの合成の場合と同じである。ただし、分析の結果として、文末に終助詞「か」が用いられている発話の約 88%において、文末で基本周波数の上昇がみられることを考慮し、合成音声において疑問の発話意図を表出することを考え、文末の基本周波数を上昇させることを目的として、合成する発話が疑問の意味情報をもつ場合は、終助詞「か」の前に大きき DH のアクセント指令をおき、文末の基本周波数を上昇させる。

さらに対話調の発話においては、基本周波数の変化が通常の読み上げの場合に比べて大きいので、韻律記号のうち起伏型、平板型それぞれの 3 種類のアクセント記号に対応するアクセント指令の大ききを大きくし、を表 6.2 のように変更する。

最後に、焦点情報の表出に関しては、以下のように規則化する。

- 焦点がおかれた語のアクセント型が平板型である場合、そのアクセント指令を FH に変更する。
- 焦点がおかれた語のアクセント型が頭高・起伏型である場合、そのアクセント指令を DH に変更する。
- 焦点がおかれている語以降で、次のフレーズ指令の位置までに存在するアクセント指令を、平板型のものに対しては FL に、頭高・起伏型のものに対しては DL に変更する。

#### 6.4.2 音声合成実験

前節で作成した音韻規則を評価するために、次の 2 通りの実験を行なった。

(実験 1) 応答音声の合成用に本研究で作成した音韻規則 (以下、応答合成用の音韻規則と呼ぶ) と、テキストからの音声合成用に作成された音韻規則 (以下、テキスト合成用の音韻規則と呼ぶ) のそれぞれに基づいて音声を合成し、この 2 種類の音韻規則に基づく発声を 1 組にして被験者に提示し、どちらが話し言葉らしく聞こえるかという尺度で一対比較をしてもらい、どちらか一方を強制的に選んでもらうこととした。被験者は成人男女 7 人で、各被験者に 12 組の発声を提示した。

(実験 2) 応答音声用の音韻規則において、焦点情報の表出に関する音韻規則を用いた場合と用いない場合のそれぞれに基づいて音声を合成し、この 2 種類の音韻

表 6.1. テキストからの音声合成における韻律記号の種類と対応する指令の大きさ

フレーズ記号		アクセント記号	
記号	値	記号	値
P1	0.35	DH	0.50
P2-	0.25	DM	0.35
P3	0.15	DL	0.15
P0	-0.50	FH	0.50
		FM	0.25
		FL	0.10
		A0	

DH,DM,DLは起伏型用のアクセント記号  
 FH,FM,FLは平板型用のアクセント記号  
 A0はアクセント指令の立ち下げ時点を示す記号

表 6.2. 応答用音声合成におけるアクセント記号に対応する指令の大きさ

記号	値
DH	0.65
DM	0.50
DL	0.25
FH	0.60
FM	0.40
FL	0.20

規則に基づく発声を1組にして被験者に提示し、回答用紙に記入してある発話に対する応答としてどちらが文中の焦点がはっきりしているかという尺度で対比較をしてもらい、どちらか一方を強制的に選んでもらうこととした。被験者は成人男女7人で、各被験者に12組の発話を提示した。

この実験の結果、実験1においては全体の約77.4%で応答合成用の音韻規則に基づく合成音声を選択され、実験2においては、約65.5%で焦点の表出に関する規則を用いた合成音声を選択された。この結果から、本研究で音声対話システム用に行なった音韻規則の改良の効果が確認された。

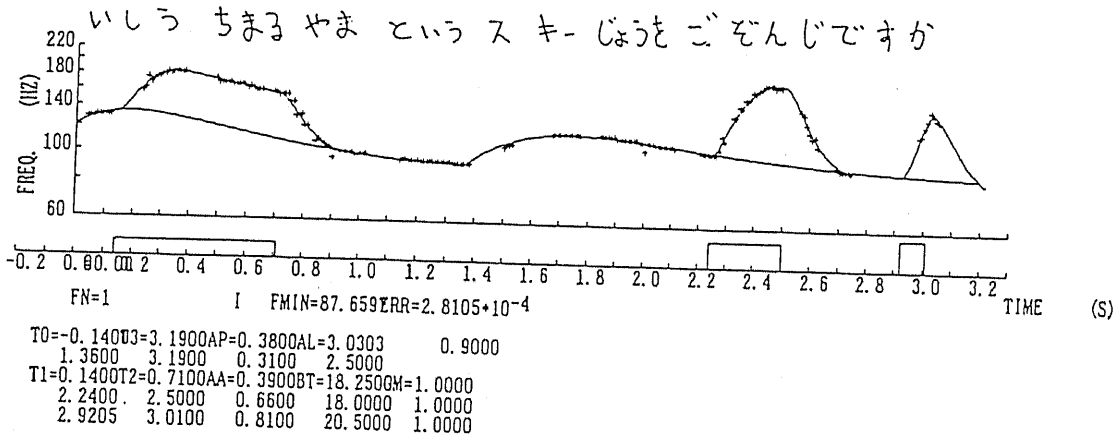
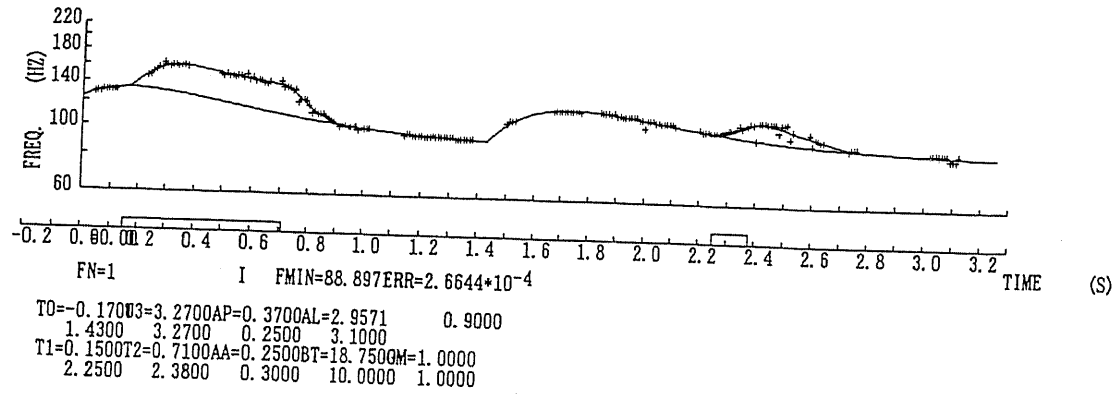
最後に、応答合成用の音韻規則とテキスト合成用の音韻規則で合成された合成音声の基本周波数パターンを図6.14に、焦点情報の表出に関する音韻規則を用いた場合と用いない場合の合成音声の基本周波数パターンを図6.15に示す。

## 6.5 まとめ

本章では、音声応答システムにおいてユーザに対する応答の音声を合成する音声合成処理について説明した。

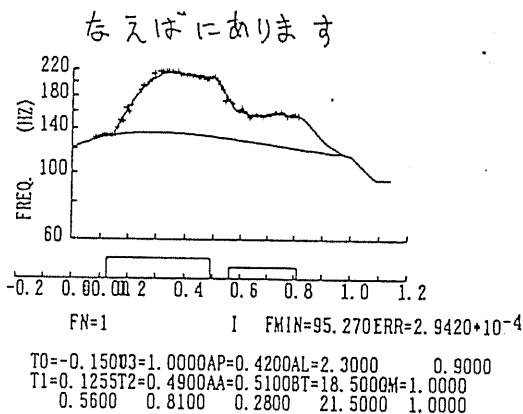
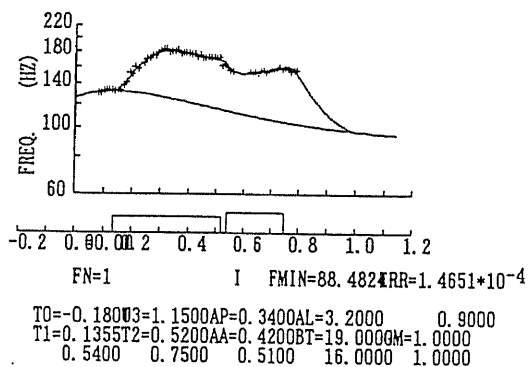
はじめに、音声応答システムにおける音声合成処理が、テキストを入力とする音声合成処理と比較して、どのような点が異なっているかを明らかにし、それに対処するために音声合成処理をどのように行なう必要があるかについてまとめた。そして、音声合成処理における文生成処理と音響処理について言及した。特に音響処理において用いる音声合成器の構成として、筆者らが構築した生成機構の異なる音声ごとに複数の直列型回路を持つターミナルアナログ型音声合成器の構成について詳説した。

次に、対話処理において生成された韻律的特徴の制御に重要な役割を果たす情報を、音声合成において反映させるため、対話中の発声に関して、対話中における発話の役割と焦点情報との関連から、発話速度と発話のパワー、基本周波数の時系列パターンの分析を行なった。さらに分析した発話と同じものをテキスト化し、同一発声者に普通に読み上げてもらったものと、人に話しかけるように読み上げてもらったものの分析もあわせて行い、これらの結果の比較を行なった。そしてこの結果に基づき、音声応答システムに用いることを考えて音韻規則の改良を行なった。さらに改良した音韻規則と従来のテキストを入力とする音声合成のための音韻規則とのそれぞれに基づく合成音声の比較と、焦点情報の表出に関する音韻規則を用い



上段:テキスト合成用音韻規則による合成  
 下段:応答合成用音韻規則による合成

図 6.14. 応答合成用音韻規則とテキスト合成用音韻規則による合成音声の基本周波数



上段:焦点情報表出に関する音韻規則を用いない合成  
 下段:焦点情報表出に関する音韻規則を用いた合成

図 6.15. 焦点情報の表出に関する音韻規則を用いた場合と用いない場合の合成音声の基本周波数

た場合と用いない場合のそれぞれに基づく合成音声の比較を行ない、改良を行なった音韻規則の効果を確認した。



## 第 7 章

### 結 論

本論文は音声応答における対話処理と音声合成に関して、実用的なシステムの構築を行なうことを目的とした。

第1章の序論に続いて、第2章では、ユーザとの間で音声による情報伝達を行なうシステムである音声応答システムについてその概略を説明した。まず音声応答システムの概念を説明した後、本研究で想定している音声応答システムの具体的な構成を提示し、各構成要素の機能の概略を説明した。さらに、これらの構成要素のうち、本研究の直接の研究対象ではないデータベースと音声認識処理に関して言及した。データベースとしては市販のスキー場に関するデータベースを用いることとし、また対話処理への入力となる音声認識の結果として得られる情報としては、比較的安定して認識結果が得られる自立語を中心とした単語列を想定することとした。

第3章では、音声応答システムにおいて対話処理を行なう音声対話処理部について概説した。まず、音声対話資料の分析について説明し、音声対話に固有な特徴に関する知見を得た。そしてこの結果を受けて、対話処理においてどのような対話を処理対象として扱うかについて考察し、協調的に行なわれる質問応答型対話とすることとした。それに続いて、規定された対話を処理する対話処理システムについて、対話処理への入出力情報を規定し、また内部的には対話履歴管理部と応答内容立案部の2つからなる構成を提案し、システムの特徴について概説した。さらに本研究で用いられる深層意味表現について、格構造に音声合成に関する情報の記述を拡張することによって定義した。

第4章では、対話処理システムのうち、対話の状況の把握や対話履歴の保存、対話の流れの制御などを司る対話履歴管理機能について詳説した。まず、対話状況の把握や対話の流れの制御を行なうのに必要不可欠な情報である、ユーザから入力される発話の形態について、質問文、要求文、応答文の3種類に大きく分類し、質問文と要求文に関してはさらに細かく分類する分類カテゴリを定義し、単語列の情報からその分類を行なうアルゴリズムを提案した。さらにこのアルゴリズムを対話資料を用いて評価し、%以上の正答率で分類が可能であることを示し、アルゴリズムの有効性を示した。次に、対話の流れを対話履歴として保存する際の表現方法として、質問・要求文とそれに対する応答文を対にして対話状況の把握を行なうFRDの概念を提案し、さらにこの概念に基づいて、発話の形態から対話の流れの制御を行なう汎用的な規則を提案した。最後に、音声合成において韻律的特徴を制御するのに重要な役割を果たす焦点情報の制御に関して、システムからユーザへの応答でど

の様な情報に焦点を設定するかを規定し、それを行なうアルゴリズムの提案を行なった。

第 5 章では、対話処理システムのうち、ユーザに対して対話状況に応じた適切な応答を生成する応答内容立案機能について詳説した。まず、対話処理を行なうのに必要不可欠である対話の話題に依存する知識について、本研究では単語辞書と TODRs によって記述し、さらに TODRs の記述としては、対話中に現れる事例に基づいてフレーム形式で記述する方式を提案した。次に、ユーザ入力発話で言及されている主題を、発話中の単語から単語辞書の記述に基づいて推定する処理を提案した。そして、これと関連してシステムからユーザへの応答に省略・照応表現の利用する手法についての検討も行なった。そして、TODRs を用いてシステムがユーザに対する応答を生成する手順を、質問文、要求文、応答文というユーザから入力される発話の形態ごとに説明した。さらに、応答を生成するために必要となる情報を得るために行なう知的システムへのアクセスと、アクセスの過程で対話処理に関連する処理を行なうための規則であるタスク依存規則について言及した。最後に、実際にユーザとの間で行なわれた対話を例に挙げて、対話処理の動作を説明し、それが有効に機能していることを示した。

第 6 章では、対話処理の結果として得られるユーザに対する応答内容に基づいて、音声合成する処理について説明した。まず、音声合成処理の概要として、本研究で音声合成に用いる音声合成器の構成について説明した。さらに対話処理で得られる情報を合成音声の韻律的特徴に反映させるため、対話中の発話及び同じものを読み上げた発話、焦点を付与した発話に対して、発話速度、発話のパワー、基本周波数パターン分析を行なった。そしてそれに基づいて、音韻処理において韻律記号列を作成する音韻規則の規則化を行ない、その音韻規則に基づく音声の合成実験を行ない、規則の有効性を確認した。

従来の音声応答システムに関する研究の多くは、音声の認識・理解の部分の処理に重点がおかれており、そこでいかに多くの情報を抽出するかという点に基づいて対話処理を行なっている。これに対して、本研究のアプローチは全く異なるものであり、現状の音声認識技術から安定して抽出することができる限られた情報から、対話処理においてあらかじめシステムに持たせている知識に基づいて、意味主導の形で発話の形態や主題の推定を行ない、より多くの情報をより適切な形態でユーザに伝達するということを、研究の最大の目的とした。さらにシステムにあらかじめ

持たせる知識として、対話の中の事例に基づく具体的な記述を用いることによって、知識の記述を容易なものとし、対話処理としての汎用性を高めることができた。この結果、本研究の目的の1つであるシステムの実用性という点も満足なものとすることができた。

本論文での説明には、システムとしての構築の際に用いたスキー場に関する紹介・案内を対話の話題としたが、付録 A での話題に依存する知識の作成手順を参考にして知識の記述を工夫することによって、かなり広い範囲の話題に対して本論文で述べたアルゴリズムを適用することが可能であろう。

## 謝 辞

本研究を進めるにあたって、大学院の5年間の長きに渡って親身に御指導いただいた指導教官の広瀬啓吉助教授に深謝致します。先生には、研究の方針や実施に対して、昼夜を問わず的確な御指摘、御助言を賜りました。

大学院修士課程の時の指導教官である藤崎博也東京大学名誉教授(現東京理科大学教授)には、特に音声合成器に関していろいろと御指導いただきました。

筆者の研究活動の場であった東京大学工学部電子工学科広瀬研究室の森川博由助手(現福井大学助教授)と高橋登技官には、本研究に対して貴重な御指摘やデータを提供していただきました。感謝致します。

筆者と同じ研究室に所属し、本研究に関して積極的に議論につき合って下さったり、研究の一部を手伝っていただいた大学院生、卒論生の皆様にも感謝致します。

最後に、音声対話資料の収集にあたって、スキーに対する話題に対して様々な対話を行ない、録音につき合って下さった友人達に感謝致します。

## 参考文献

- [Bobrow et al. 77] Bobrow D.G., Kaplan R.M., Kay M., Norman D.A., Thompson H. and Winograd T.: "GUS, A Frame-Driven Dialog System," *Readings in Natural Language Processing*, Morgan Kaufmann (1977).
- [Fant 53] Fant G: "Speech Communication Research," *Ing. Vetenskaps Akad. Stockholm*, Sweden 24, pp.331-337 (1953).
- [Fillmore 68] Fillmore C.J.: "The Case for Case," *Universals in Linguistic Theory*, edited by Bach E. and Harms R.T. (1968); 田中春美, 船城道雄訳: "格文法の原理," 三省堂 (1980).
- [Fujisaki et al. 84] Fujisaki H. and Hirose K.: "Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese," *J. Acoust. Soc. Jpn.*, Vol.5 No.4, pp.233-242 (1984).
- [Fujisaki et al. 86] Fujisaki H. and Ljungqvist M.: "Proposal and Evaluation of Models for the Glottal Source Waveform," *Proc. of ICASSP'86*, Vol.3, 31.2, pp.1605-1608 (1986).
- [Fujisaki et al. 90] Fujisaki H., Hirose K. and Takahashi N.: "Manifestation of Linguistic and Para-Linguistic Information in the Voice Fundamental Frequency Counters of Spoken Japanese," *Proc. of ICSLP'90*, Vol.1 12.1, pp.485-488 (1990).
- [Hirose et al. 92] Hirose K., Fujisaki H. and Seto S.: "A Scheme for Pitch Extraction of Speech using Autocorrelation Function with Frame Length Proportional to the Time Lag," *Proc. of ICASSP'92*, Vol.1, pp.149-152 (1992).

- [Howes 89] Howes J.R.: "Dialogue Supervision and Error Correction Strategy for a Spoken Human-Computer Interface," *The Structure of Multimodal Dialogue*, edited by Taylor M.M., Neel F. and Bouwhuis D.G., pp.375-383 (1989).
- [Klatt 80] Klatt D.H.: "Software for Cascade/Parallel Formant Synthesizer," *J. Acoust. Soc. Am.*, 67 (3), pp.971-995 (1980).
- [Klatt et al. 90] Klatt D.H. and Klatt L.C.: "Analysis, Synthesis and Perception of Voice Quality Variations among Female and Male Talker," *J. Acoust. Soc. Am.*, 87 (2), pp.820-857 (1990).
- [Kolodner et al. 85] Kolodner J.L., Simpson R.L. and Sycara K.: "A Process Model of Case-Based Reasoning in Problem Solving," *Proc. of IJCAI'85*, pp.284-290 (1985).
- [Lawrence 53] Lawrence W.: "The Synthesis of Speech from Signal which have a Low Information Rate," *Communication Theory*, edited by Jackson W., Butterworths, London, England, pp.460-469 (1953).
- [Levinson 83] Levinson S.C.: "Pragmatics," Cambridge University Press (1983).
- [Litman et al. 87] Litman D.J. and Allen J.F.: "A Plan Recognition Model for Subdialogues in Conversations," *Cognitive Science*, Vol.11, pp.163-200 (1987).
- [Maltese 92] Maltese G. and Mancini F.: "An Automatic Technique to Include Grammatical and Morphological Information in a Trigram-Based Statistical Language Model," *Proc. of ICASSP'92*, Vol. 1, pp.157-160 (1992).
- [Proctor et al. 89] Proctor C. and Young S.: "Dialogue Control in Conversational Speech Interfaces," *The Structure of Multimodal Dialogue*, edited by Taylor M.M., Neel F. and Bouwhuis D.G., pp. 385-398 (1989).
- [Sato 90] Sato S.: "Phonological rules for a Semantic-to-Speech System of Japanese: the segmental phase," *J. Acoust. Soc. Jpn.*, (E)11, 2, pp.83-94 (1990).
- [Sawai 90] Sawai H.: "The TDNN-LR Large-Vocabulary and Continuous Speech Recognition System," *Proc. of ICSLP'90*, 31.4, pp.1349-1352 (1990).

- [Schank et al. 81] Schank R.C. and Riesbeck C.K.: "Inside Computer Understanding: Five Programs Plus Miniatures," Lawrence Erlbaum Associates (1981); 石崎俊 監訳: "自然言語理解入門: LISP で書いた5つの知的プログラム," 総研出版 (1986).
- [Takebayashi et al. 92] Takebayashi Y., Tsuboi H., Sadamoto Y., Hashimoto H. and Shinchi H.: "A Real-Time Dialogue System Using Spontaneous Speech Understanding," *Proc. of ICSLP'92*, Vol.1, pp.651-654 (1992).
- [Wilpon et al. 90] Wilpon J.G., Rabiner L.R., Lee C.-H. and Goldman E.R.: "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models," *IEEE Trans. of ASSP*, Vol.38 No.11, pp.1870-1877 (1990).
- [Yamamoto et al. 91] Yamamoto Y., Ohta Y., Yamashita Y., Kakusho O. and Mizoguchi R.: "MASCOTS: Dialog Management System for Speech Understanding System," *IEICE Trans.*, Vol.E 74, No.7, pp.1881-1888 (1991).
- [Young et al. 79] Young S.J. and Fallside F.: "Speech Synthesis from Concept: A Method for Speech Output from Information Systems," *J. Acoust. Soc. Am.*, 66 (3), pp.685-695 (1979).
- [Zue et al. 91] Zue V., Glass J., Goodine D., Leung H., Phillips M., Polifroni J. and Seneff S.: "Integration of Speech Recognition and Natural Language Processing in the MIT VOYAGER System," *Proc. of ICASSP'91*, S10.7, pp.713-716 (1991).
- [飯田 88] 飯田 仁: "異言語間対話を目指す端末間通訳モデル," 認知科学の発展, Vol.1, pp.112-140 (1988).
- [飯田他 90] 飯田 仁, 有田英一: "4階層プラン認識モデルを使った対話の理解," 情報処理学会論文誌, Vol.31 No.6, pp.810-821 (1990).
- [板橋他 92] 板橋秀一, 速水悟, 小林哲則, 竹沢寿幸: "音響学会連続音声データベース," 日本音響学会春季研究発表会講演論文集, Vol.1, 2-2-17, pp.285-286 (1992).
- [河井 88] 河井 恒: "日本語テキストからの音声合成システム," 東京大学工学部学位論文 (1988).



- [草薙 88] 草薙裕: “形態素生成,” 同著: “LISP による自然言語処理,” pp.181-216 (1988).
- [小林 92] 小林重信: “事例ベース推論の現状と展望,” 人口知能学会誌, Vol.7, No.4, pp.559-566 (1992).
- [匂坂他 83] 匂坂芳典, 佐藤大和: “日本語単語連鎖のアクセント規則,” 電子通信学会論文誌, J66-D, No.7, pp.849-856 (1983).
- [辻井 91] 辻井潤一: “機械翻訳システムにおける辞書の構成,” 野村浩郷編: “言語処理と機械翻訳”, 講談社サイエンティフィック (1991).
- [田路他 91] 田路龍太郎, 吉田 孝: “電話取次対話における対話主導権の動的制御,” 電子情報通信学会論文誌, Vol.J74-D-II No.12 pp.1639-1650 (1991).
- [中川 88] 中川聖一: “確率モデルによる音声認識,” 電子情報通信学会編, (1988).
- [新美 87] 新美康永, 小林豊: “対話音声理解システムにおける対話管理とキーワード予測,” 電子情報通信学会技術研究報告, SP87-103, pp. (1987).
- [西田 88] 西田豊明: “意味の表現と処理,” 田中穂積, 辻井潤一共編: “自然言語理解”, オーム社 (1988).
- [広瀬他 82] 広瀬啓吉, 藤崎博也, 高橋登: “複文の基本周波数パタンの分析と合成,” 日本音響学会音声研究会資料, S82-40 (1982).
- [広瀬他 89] 広瀬啓吉, 藤崎博也, 河井恒, 山口幹雄: “基本周波数パターン生成過程モデルに基づく文章音声の合成,” 電子情報通信学会論文誌, Vol. J72-A, No.1, pp.32-40 (1989).
- [藤崎他 71] 藤崎博也, 須藤寛: “日本語単語アクセントの基本周波数パターンとその生成機構モデル,” 日本音響学会誌, Vol.27 No.9, pp.445-453 (1971).
- [山口他 92] 山口耕市, 他: “ATREUS: ATR における連続音声認識諸方式の比較,” 日本音響学会秋季研究発表会講演論文集, Vol.1, 2-Q-5, pp.181-182 (1992).
- [山下他 91] 山下洋一, 柴田宜宏, 溝口理一郎: “概念表現からの音声合成における文生成,” 日本音響学会秋季研究発表会講演論文集, Vol.1, 3-P-19, pp.329-330 (1991).

## 発表文献

- [藤崎他 89] 藤崎博也, 広瀬啓吉, 浅野康治: “高品質の音声合成に適したターミナルアナログ型音声合成器の構成,” 日本音響学会秋季研究発表会講演論文集, 3-P-2, pp.279-280 (1989-10).
- [藤崎他 90a] 藤崎博也, 広瀬啓吉, 浅野康治, “ターミナルアナログ型合成器による日本語音声の規則合成,” 日本音響学会春季研究発表会講演論文集, 1-4-6, pp.189-190 (1990-3).
- [Fujisaki et al. 90a] Fujisaki H., Hirose K., Kawai H. and Asano Y.: “A System for Synthesizing Japanese Speech from Orthographic Text,” *Proceedings of 1990 IEEE International Conference on Acoustics, Speech and Signal Processing*, S6a.5, pp.617-620 (1990-4).
- [藤崎他 90b] 藤崎博也, 広瀬啓吉, 浅野康治: “高品質音声合成のためのターミナルアナログ型音声合成器,” 電子情報通信学会技術研究報告, SP90-1, pp.1-8 (1990-5).
- [広瀬他 90] 広瀬啓吉, 藤崎博也, 浅野康治: “テキストからの日本語音声合成システム,” 電子情報通信学会技術研究報告, SP90-42, pp.23-30 (1990-8).
- [藤崎他 90c] 藤崎博也, 広瀬啓吉, 浅野康治: “知識表現からの文章音声合成システム,” 日本音響学会秋季研究発表会講演論文集, 2-6-6, pp.231-232 (1990-9).
- [Fujisaki et al. 90b] Fujisaki H., Hirose K. and Asano Y.: “Proposal and Evaluation of a New Type of Terminal Analog Speech Synthesizer,” *Proceedings of 1990 International Conference on Spoken Language Processing*, 19.20, pp.841-844 (1990-11).

- [浅野他 91a] 浅野康治, 広瀬啓吉, 藤崎博也, 服部雅一: “知識表現からの音声合成システムにおける対話管理部の役割と表層文の生成,” 日本音響学会春季研究発表会講演論文集, 1-6-19, pp.243-244 (1991-3).
- [浅野他 91b] 浅野康治, 広瀬啓吉, 藤崎博也: “音声応答システムにおける対話管理手法,” 日本音響学会秋季研究発表会講演論文集, 3-P-17, pp.325-326 (1991-10).
- [浅野他 92a] 浅野康治, 広瀬啓吉: “相互主導型対話機能を持つ音声応答システムとその対話管理手法,” 電子情報通信学会技術研究報告, SP91-110, pp.17-24 (1992-1).
- [浅野他 92b] 浅野康治, 広瀬啓吉, 藤崎博也: “音声応答システムにおける対話管理システムの実現,” 日本音響学会春季研究発表会講演論文集, 3-1-5, pp.79-80 (1992-3).
- [浅野他 92c] 浅野康治, 広瀬啓吉, 小杉康宏: “音声対話処理における話題に依存する知識の記述,” 日本音響学会秋季研究発表会講演論文集, 1-1-9, pp.17-18 (1992-10).
- [広瀬他 92] 広瀬啓吉, 浅野太郎, 浅野康治, 藤崎博也, 山口幹雄, John-Paul Hosom: “マルチ・カスケード回路構成のターミナルアナログ音声合成器を用いた規則合成,” 日本音響学会秋季研究発表会講演論文集, 1-P-9, pp.317-318 (1992-10).
- [Asano et al. 92] Asano Y., Hirose K. and Fujisaki H.: “A Method of Dialogue Management for the Speech Response System,” *Proceedings of 1992 International Conference on Spoken Language Processing*, pp.1367-1370 (1992-10).
- [Asano et al. 93] Asano Y. and Hirose K.: “A Dialogue Processing System for Speech Response with High Adaptability to Dialogue Topics,” *IEICE Transaction of Information and System*, Vol.E76-D, No.1, (1993-1).
- [浅野他 93] 浅野康治, 広瀬啓吉: “音声応答システムにおける音声の生成,” 日本音響学会春季研究発表会講演論文集, (1993-3 発表予定).
- [広瀬他 93] 広瀬啓吉, 浅野太郎, 浅野康治, 藤崎博也: “マルチ・カスケード型ターミナルアナログ合成器を用いたテキストからの音声合成,” 日本音響学会春季研究発表会講演論文集, (1993-3 発表予定).

## 第 A 章

# 話題に依存する知識の作成手順

本研究において構築した音声対話処理システムは、その処理プロセス部分是对話の話題に依存しないような形で、汎用性を考慮して構築されている(一部データベースアクセスに直接関連する部分は、データベースの構造に依存する)。そして対話の話題に依存する知識を、対話において使用される語彙を記述した単語辞書と、対話におけるシステムの応答内容の生成に利用される TODRs として実現し、システムに組み込んでいる。本節では、この2つの知識の作成手順について説明する。

#### 1. 対話の話題と使用する知的システムを決定する。

音声応答システムとして、どのような話題に関する対話を取り扱うのかを決め、その話題に関する情報を対話処理に対して提供する知的システムとして、何を用いるかを決定する。

#### 2. 対話中に現れる単語のリストアップ

1で決定した話題に関する対話において、どのような単語が現れるかを、対話資料などを参考にしてリストアップする。そしてこれらの単語を表5.1の分類に基づいて、それぞれの品詞ごとに分類する。

#### 3. 意味マーカの設定

2において品詞として名詞に分類された単語に対して、それぞれに意味マーカを設定する。意味マーカの種類として一般的なものとしては、機械翻訳システムの構築を目的とした Mu プロジェクトにおいて用いられた図 A.1 に示すような例 [辻井 91] があるが、本研究で構築した対話処理システムにおいては、対象となる対話の話題や知的システムに応じて知識作成者が任意に作成することができる。したがって、一般的な例などを参考にして、その話題において詳しく分類する必要がある部分に関してはより細かな意味マーカを用意するのがいいと思われる。

#### 4. 主題のリストアップ

決められた話題の対話において、現れ得る主題をリストアップする。この主題のリストアップについては、特に定められた方式があるわけではなく、知識作成者が対話の文例などを参照することによって任意に主題を設定することができる。もっとも、知的システムとしてデータベースを用い、対話を通じてデータベースアクセスを行なうことを想定している場合は、主題をデータベースの項目ごとに用意するのが最も簡単な方法であるといえる。しかし、規則の記

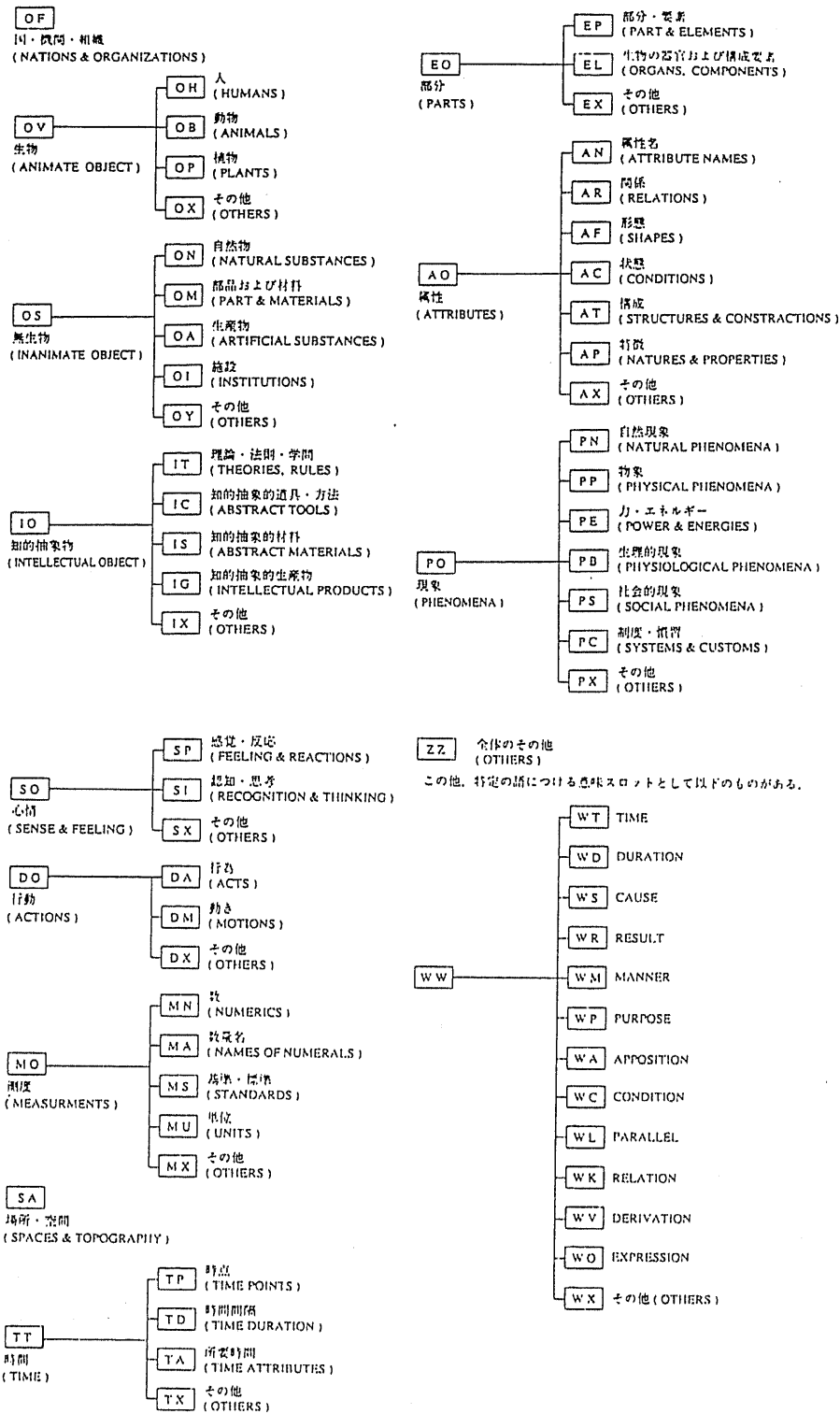


図 A.1. 意味マーカの一般的な例 (文献 [辻井 91] より転載)

述効率という点から考えると、まとめられるものに関してはいくつかの項目をまとめる方が望ましい。また同時にシステムがユーザに対して質問する場合に利用する TODR も用意する。

#### 5. 単語辞書の作成

2でリストアップされた単語を単語辞書に登録する。各単語に関して記述すべき情報に関しては5.2.1節で述べられている。

#### 6. TODRs の作成

4でリストアップされた主題が、それがどのような発話の形態 (YN 型、内容型、比較型、要求) で用いられるかを検討し、各主題に対してそれぞれ存在する発話形態との組合せで個々の TODR を記述する。各 TODR において記述すべき情報に関しては5.2.2節で述べられている。

#### 7. 実際に対話処理を行ない、知識の追加・変更を行なう。

構築した知識を用いて、実際に音声応答システムを運用する。そして運用の結果から、抜けている単語や対話の事例を検討し、それに基づいて新たに知識 (単語辞書、TODRs) の追加・変更を行なう。このように実際にシステムとして運用した後でも、その運用の状況に応じて知識の拡充を容易に行なうことができるのも、本方式の特徴の1つである。