

Voice Q-A System の言語処理について

鹿野 清宏

(日本電信電話公社 武蔵野電気通信研究所)

あらまし 会話音声で動作する質問回答システム (Voice Q-A System) の言語処理の構成と性能について述べる。システムの対象は列車の座席予約である。言語処理では、音響処理から音韻系列を受けとり、単語を認識し、構文解析して意味内容を把握し、推論を行い、会話モデルに従って応答文を作成する。言語処理は、top-down的に構成され、left-to-rightの手順でマッチングを行って会話音声を認識する。

計算機室で文節ごとに区切って発声した文章で認識実験を行った。すべての音韻を学習した発声者に対しては、95.7%の文節認識率(予約項目の認識率)が得られた。母音は学習し、子音を特定の1人に固定した男性8名の平均の文節認識率は86.0%であった。さらに、男性8名に対して会話モデルを用いて、オンラインの質問回答実験を行い、予約達成率99.1%を得た。質問回答の回数は、最初の入力を除くと、平均3.21回であり、少ない質問回答の回数で予約が達成できるシステムが作成できた。

1. まえがき

音声認識の研究は、最初、音声タイプライタを目ざして行われた。しかし、音韻を正しく切り出して認識することが、きわめて困難であることがわかり、次に、孤立して発声された単語音声全体を一つのパターンと見なし、パターン・マッチングで単語音声を認識しようとする単語音声の認識が試みられた。最近では、様々な単語音声認識装置が実用化され始めている。単語音声認識では、単語辞書により音韻系列を制限し、その冗長性によって音韻認識のあいまいさを解消していると考えることができる。会話音声認識の研究は、単語辞書の情報の他に、構文・意味情報などの冗長性も利用して、連続音声の中の音韻認識の誤りをおさず、会話音声を認識しようとする試みである。音声認識で利用する情報および代表的な手法を図1に示す。

我々は、会話音声認識の対象(task)として、新幹線の座席予約をとりあげ、会話音声認識システム (Voice Question-Answering System) を作成した。^{(1)~(4)} 入力

音声は、文節(予約項目に相当)ごとに区切って発声した座席予約の文章であり、予約項目の発声順序は自由である。座席予約の発声における利用情報の冗長性を概算した。その結果を表1に示す。情報量、冗長度の計算方法は文献(1)を参照されたい。我々の対象では、構文情報による冗長度が大きいことがわかる。これらの単語辞書、構文、意味内容、時刻表による情報を

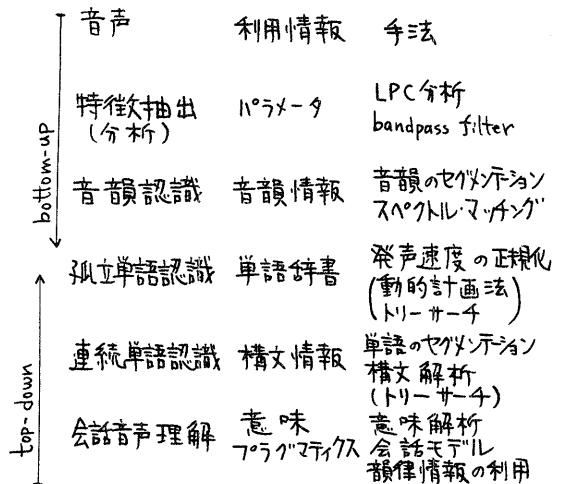


図1. 音声認識で用いる情報

利用して、音韻認識のあいまいさを解消する手法を見い出すことが、会話音声認識における言語処理の主たる問題となる。

表1 言語情報による冗長度

	冗長度	情報量 (bits)
音韻空間	—	52.5
単語辞書	1.42	36.9
構文情報	3.10	11.9
意味内容	1.57	7.6
時刻表	1.60	4.8
全体	11.1	—

我々は、通常行われているように、会話音声認識システムに、次のような制約を設けた。

- (1) 会話の対象を限定する。
- (2) 発声内容を一語一句完全に正しく認識できなくても、発声者が意図した内容が理解できればよい。
- (3) 質向回答形式で発声者との対話を行う。

(1)は入力音声の内容を限定し、構文・意味・単語などの言語情報を利用しやすくするために必要である。(2)は、出力の認識結果に対する要求をゆるやかにしており、(3)は対話を通じて人間の助けも利用できるようにしており、目標を、現実に可能な、実用に近いところに設定している。

我々は、現在の音声分析、音韻認識の技術レベルを考慮し、あまりにも高度な認識対象をとることは得策ではないと判断して、アメリカのARPAの研究グループの目標より冗長性のある認識対象として、列車の座席予約サービスをとりあげ、1974年に研究を開始し、オンラインで動作するシステムを1976年末に作成した。このシステムでは、連続音声で計算機と会話して、新幹線の座席予約のシミュレーションを行うことができる。

本稿では、2節でVoice Q-A Systemの概要を述べた後、3節で、言語処理について簡単に説明する。4節でVoice

Q-A Systemの性能について、認識実験と質向回答実験に基づいて述べる。入力音声に発声の制約を設けた場合の認識実験の結果についても示す。5節で、今後の問題点について述べる。

2. Voice Q-A Systemの概要

2.1. タスク

会話の対象として、現在の技術レベルや質向回答形式で行うことを考慮すると、入力の語い数が数百以下であること、文章には意味上の冗長性があること、会話のやりとりがある程度継続することが望ましいと考えられる。実用的ないくつかの候補をあげて、これらの事柄について比較検討した結果、座席予約サービスが適当であると判断した。⁽⁵⁾

具体的な対象として新幹線の座席予約サービスをとりあげた。システムは、文節ごとに区切って発声した入力音声を認識して、予約内容を把握し、発声者に音声で応答する。この座席予約のタスクの概要を表2に示す。入力音声中に含まれる単語は、表3に示す112個である。

ARPAの研究グループでは、タスクの複雑さの指標として、branching factorを用いている。branching factorは、次のように定義されている。⁽⁶⁾

"The average branching factor is defined to be the average number of words that would have to be considered at each point along the correct left-to-right path through the syntactic production rules during the processing of a typical utterance."

このタスクのbranching factorは約20であり、ARPAの代表的なシステムであるCMUの

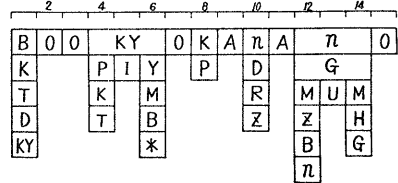
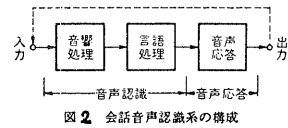
単語数	列車の座席予約(新幹線)
予約項目	112個(駅名28個、列車181本)
発声の制限(発声例)	日付、発駅、着駅、発時刻、列車名、等、枚数 文節ごとに0.5秒以上のポーズ 明日の・ひかり191号で・東京から・博多まで・指定席を・4枚・予約します。

表3 認識対象の単語

日付 16	TSUITACHI FUTSUKA MIKKA YOKKA ITSUKA MUIKA NANOKA YOKA KOKONOKA TOKA HATSUKA KYO ASU ASHITA ASATTE HONJITSU
数字 21	1 2 3 5 6 7 8 9 0 10 100 SHICHI KYU YO I RO HA ZYU ZI RE
駅名 28	TOKYO SHINYOKOHAMA ODAWARA ATAMI MISHIMA SHIZUOKA HAMAMATSU TOYOHASHI NAGOYA GIFUHASHIMA MAIBARA KYOTO SHINOSAKA SHINKOBE NISHIAKASHI HIMEJI AIOI OKAYAMA SHINKURASHIKI FUKUYAMA MIHARA HIROSHIMA SHINIWAKUNI TOKUYAMA OGORI SHINSHIMONOSEKI KOKURA HAKATA
その他の名詞 15	EKI NICHII JI FUN PUN PPUN HIKARI KODAMA GO GREEN FUTSU SHITEI KEN SEKI MAI
助詞 15	KARA HATSU YORI MADE YUKI IKI E NO O DE NOO NODE WA GA NOWA
動詞 12	YOYAKU ONEGAI ITA SHI MASU ARI MASEN KA DESU DES MAS MOSHIKOMI
その他 5	HAI IIE CHIGAI SO (PAUSE)

計 112 個 (音便変化した単語も含む)

2.2. システム構成
 質問回答システムは、
 図2に示すように、音響処理⁽¹⁾、言語処理⁽²⁾、音声応答⁽³⁾の3つの部分からなる。音響処理は、音韻を認識し、音韻ラティスの形で、音韻認識の結果と音韻の継続時間的情報を言語処理に送る。音韻ラティスの例を図3に示す。この図に示されるように、音韻ラティスは、音韻のセグメンテーションのあいまいさと識別のあいまいさを表す音韻系列の一表現形式である。言語処理では、この音韻ラティスをもとに、単語を認識し、構文を解析して意味内容を把握し、さらに、時刻表を利用して予約項目間の推論を行い、会話モデルに基づいて、発声者への応答文を作成する。音声応答では、この応答文に従って音声を合成し、発声者に音声で応答する。



(T ō KY ō K A R A ŋ ō)

図3. 音韻ラティスの例。

Voice Q-A System は、2台のミニコン (図4) の上に作成されており、2台のミニコンはチャンネル結合されている。Neac 3200/70 上に音響処理と音声応答が、PFU-400 上に言語処理がつくられている。2台のミニコンは並列に動作し、チャンネルを介して音韻ラティスや応答文を受け渡す。音響処理は実時間の3.4倍で動作し、言語処理は実時間の2.0倍の処理時間で動作する。システム全体では、実時間の約5倍の処理時間で動作する。処理の流れの例を、処理時間ととみに、図5に示す。

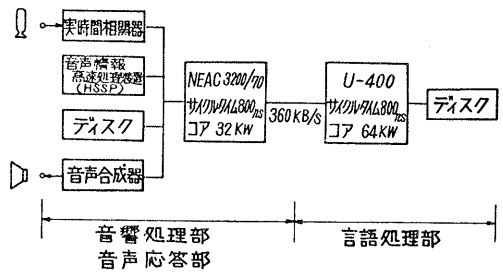


図4. 計算機システム

2.3. 質問回答の例

Voice Q-A System での、システムと発声者との質問回答のやりとりの例を示す。最初の例は、典型的な質問回答のやりとりである。多くの場合、この例のような単純な質問回答によって、新幹線の座席を予約することができると。次の例は、認識誤りが生じたときの質問回答のやりとりである。誤りが生じたとき、誤りを訂正するための

音響処理部 (Neac 3200/70)		音声応答部	
1日の新大阪から博多までの6時22分の	1/19号	1日の新大阪から博多までの6時22分の	1/19号
4.8 s	3.8	5.9	7.5
	音韻ラティス		応答文
	27s	38	20
言語処理部 (U-400)		(推論)	
		4.9	14
処理速度			
システム全体	実時間の約5倍		
音響処理	実時間の3.4倍		
言語処理	実時間の2.0倍		

図5. Voice Q-A System のタイムチャート

質問回答の方法は、音声認識での会話モデルの作成における重要な問題である。

(例1) こちらは新幹線の座席予約センターです。ご希望をおっしゃい
 計算機： てください。
 発声者： 姫路より、小郡駅までの、グリーン券で、5
 日の、こだま217号を、予約します。
 計算機： 3日ですか？
 発声者： いいえ
 計算機： 何枚ですか？
 発声者： 5枚
 計算機： あなたの予約は、5日の、こだま217号で、姫
 路駅、8時24分発、小郡駅までの、グリー
 ン券を、5枚、ですね？
 発声者： はい、そうです。
 計算機： ご希望の指定券はとれました。予約番号は4番です。最寄
 りの緑の窓口でお受取り下さい。

この例では、日付が、3日と5日のど
 ちらかであること認識されたが、3日だ
 ないことを確認して、5日に決まった。
 そのあと、まだ認識していない枚数の
 向合せが行われた。また、発時刻は発
 駅と列車名による推論から定められた。

(例2)
 計算機： (予約開始の案内)
 発声者： 新横浜から、浜松への、28日の、こだま215
 号の、グリーンを、1枚。
 計算機： 普通席、グリーン席の、どちらですか？
 発声者： グリーン席。
 計算機： 御希望の予約は、こだま215号、新横浜駅、
 9時2分発、浜松駅まで、と了解しましたが、
 該当する列車がございません。当方の聞き違い
 と思いますので訂正してください。
 発声者： 8時34分発
 計算機： あなたの予約は、28日の、こだま215号で新
 横浜駅、8時34分発、浜松駅までの、グリー
 ン券を、1枚、ですね？
 発声者： そうです。
 計算機： (予約終了の案内)

この例では、最初に入力した「グリーンを」
 を「9時2分」に誤認識した。そのため、
 普通席かグリーン席かの向合せが行われ、
 全部の予約項目が決まったところで、
 予約項目間の矛盾の訂正が求められた。

3. 言語処理の構成

言語処理は、単語認識部、構文解析部、
 推論部からなる認識部と、会話モデルと
 からなっている。この構成を図

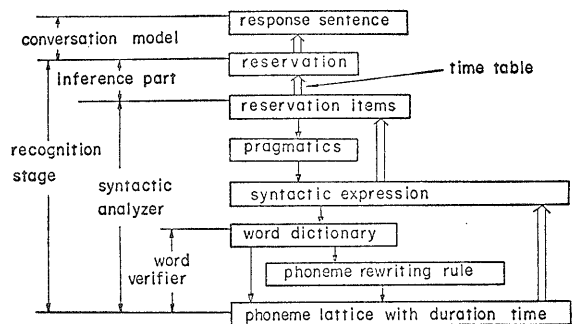


図6. 言語処理の構成

6に示す。

認識部は、推論部を除くと、top-down
 的な方向でのマッチングを用いて構成
 されており、単語認識部は構文解析部
 の一部に属している。処理の流れ(図
 6の太い矢印)は下位の概念(音韻う
 ティス)から上位の概念(構文表現や深
 層構造)に向っているが、認識の手順
 は上位の概念を仮定し、下位の情報に
 より、その仮定を検証するという、上
 位から下位へのマッチングの流れ(図
 6の細い矢印)に従っている。

マッチングは、left-to-rightの順序で
 行われる。left-to-rightのマッチング
 をとる理由は、入力者声を文節ごとに
 区切って発声する場合、強く明確に発
 声されるキーワードが文節の先頭にく
 ることが多いこと、処理の制御が簡単
 になることによる。

推論部では時刻表を用いて予約項目
 間の推論を行う。この部分はbottom-up
 的な構成をとっている。これは、認識
 においてある程度の誤りが生ずるのは
 さけられなく、認識結果と時刻表を利
 用して構文解析での単語を予測するよ
 うなtop-down的な構成にした場合、
 誤った認識結果が生じたとき、その誤
 りと時刻表による予測がシステムの動
 作に悪い影響を与えることがあると考
 えられるためである。

会話モデルは、発声者とシステムと
 が自然な質問回答をくりかえして、最
 終的には、システムが正しい予約内容

を把握するように、作成される必要がある。本システムでは、7つの会話状態を設け、各状態間の状態推移で会話モデルを表現した。

構文解析部、単語認識部は、top-down的に、left-to-rightのマッチングで構成されている。その制御手法として、tree searchのdepth-first methodを用いた。会話音声認識システムで用いられているtree searchの手法としては、次の5つの方法があげられる。

- (1) depth-first method (1)
 - (2) breadth-first method (7)
 - (3) best-first method (11)(12)
 - (4) blackboard method (9)
 - (5) beam search method (10)(8)
- において信頼度の高い一定数の候補だけを残して絞りこむ方法)

いずれの手法も、tree searchの方法であり、探索の順序が異なるだけで、探索能力には、本質的な差がないといえる。問題は、探索のコストである。よって、これらの手法をそのまま用いることは少なく、常にコストの低減を考え、探索を能率化する手法を用いる必要がある。よって、探索の能率化の手法を抜きにして、各手法の評価は不可能であると考えられる。我々は、depth-first methodを用いることにし、depth-first methodにおける探索の能率化について検討した。能率化の手法は3.1で述べる。depth-first methodは、他の手法に比べて、探索の手法の制御構造が簡単で、push-down stackを用いればよく、かつ、記憶容量も少なくて済むという利点を持っており、我々の計算機システムに適していると考えられる。

言語処理は、ミニコンPFU-400上に、著者が作成したFORTRANによるリスト処理システム⁽⁴⁾を用いて作成されている。

3.1. 単語認識部

単語認識部では、構文解析部から提示された単語辞書と音韻ラティスとのマッチングが行われる。連続音声の中の単語

を認識するためには、単語のセグメンテーションと単語の識別という2つの問題を解決する必要がある。この2つの問題のからみ合い、音韻ラティスのありまじ、種々の音韻変形規則の適用などが、連続音声の中の単語のマッチングのプロセスを非常に複雑にしている。よって、これらの複雑さを秩序づけて整理し、強力かつ、能率のよい単語認識のアルゴリズムを得ることが必要となる。

3.1.1. 音韻変形規則

音韻の性質を音韻変形規則として、書換え規則の形で表わす。各々の書換え規則は、規則の適用条件として、音韻ラティスの音韻の継続時間も利用できる。また、各々の規則にはペナルティがつけられている。音韻変形規則は、母音に書換える規則(6種の型)、子音に書換える規則(6種の型)、音韻の読みとはしの規則、および母音の無声化の訂正規則からなる。各々の書換え規則は、単語辞書の音韻をキーとして連想的に検索されるようにリスト形式で蓄えられており、単語認識アルゴリズムにおいて、能率よく検索されるデータ構造にまつている。

3.1.2. 単語辞書

単語辞書は、通常の音素表記で、リストで表わされている。単語辞書のリストには、単語名と同じ名前がつけられており、構文表現から直接に検索される。

3.1.3. 単語認識アルゴリズム

単語辞書と音韻ラティスのマッチングを、tree searchのdepth-first methodで実行する。tree searchによるマッチングは、動的計画法によるマッチングよりも、音韻の様々な性質を利用することができるといえる。しかしながら、音韻の様々な性質を利用するために、計算量が増大しやすい。よって、tree searchによるマッチ

ングにおいては、ステップ数の削減が重要な問題となる。

単語認識部では、構文解析で予測された単語の単語辞書と音韻うティス上のスタートセグメントのリストを受けとり、マッチングを left-to-right の順序で行う。マッチングは、すべてのスタートセグメントから試みられ、マッチングが可能なラティス上のエンドアドレスをすべて求める(図7)。マッチングで音韻変形規則を適用するごとに、各々の規則のペナルティが加えられ、このペナルティの総和でマッチングの度合を表わす。単語認識の結果は、リストで表わされ、ペナルティの総和とともに構文解析部に返される。

言語処理での処理時間の大部分は、単語認識部で費やされており、処理を高速化するには、マッチングに要する計算量を減らす必要がある。そのために、次に示す4つの手法を併用した。

- (I) ペナルティの総和に制限を設け、tree search の枝刈りを行う。
- (II) tree search のプロセスを記憶し、それを参照することによって、同じプロセスをくり返さないようにする。
- (III) 単語の境界をありのまま、まとめて処理し、単語認識の回数を少なくする。
- (IV) 単語認識の結果を記憶する。

手法(I)のペナルティの総和に対する制限は、depth-first method でのバックトラックが必ず生じるようにするためにも不可欠である。

手法(II)は、音声認識でよく利用されている動的計画法において、途中の段までの最適値のみを求めて記憶し、後段でその最適値を利用して全体の最適値を求めるという考えの拡張とみなすことができる。これにより、処理速度が4倍向上した。⁽¹³⁾

次に手法(III)について説明する。マッチングにおいて、単語の境界は、通常、一意にはきまらな。ここで、複数の

単語境界の候補について、別々に、次に続く単語とのマッチングを行

うと、単語認識の回数が増大する。これをさけるために、すべての単語境界の候補から単語辞書とのマッチングを同時に行うようにした。これにより、処理速度が約1.2倍向上した。⁽¹³⁾

手法(IV)では、単語認識の結果を、失敗した場合と成功した場合とに分けて記憶し、この結果を単語認識の直前に参照し、すでに認識が試みられていれば、その結果を構文解析部に返す。これによる処理速度の向上は約1.1倍であった。⁽¹³⁾

さらに、本システムでは、マッチングのプロセスや音韻変形規則の記憶形式として、リスト形式をとり、連想的な検索ができるようにデータが蓄えられている。このデータ構造の改良によって、処理速度が約6倍向上した。⁽¹⁴⁾

以上のステップ数削減の手法(II)~(IV)とデータ構造の改良によって、処理速度が約30倍に向上した。現在、言語処理は実時間の約2倍の処理時間で動作している。

3.2. 構文解析部

会話音声認識システムの多くは、構文情報をBNF(Bakus-Naur Form)で表現するか、あるいは、transition network grammar で表現している。transition network grammar は、文章の構文の表現に適し、かつ、その構文表現に従って単語を予測するのに適した方法であると思われる。我々は、この transition network grammar とほぼ等価な能力を持つ、リスト表現を用いた簡潔な構文の表現方法を提案する。

3.2.1. プラグマティクス

プラグマティクスでは、会話モデルの会

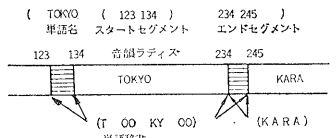


図7 単語認識の結果のリストと単語のセグメンテーションのあいまいさ

話状態, 予約項目の把握状態を参考にして, 構文解析をする文節の順序を指定する。プラグマティクスは, リスト表現で手続きとして, 図8のように表わされている。このリスト表現では, 次の演算子とセマンティクルーチンを用いている。

SEM (YES-NO) 会話状態が「はい, いいえ」で予約内容を確認する状態であるか。

SEM (VERB) 入力音声がある文章の最後の文節であるか。

SEM (LATTICE a b) 入力の音韻ラティスの長さが a 以上で b 以下であるか。

SEM (SYNTAX i) i 番目の予約項目が未定か。ただし, 会話状態が「矛盾の訂正の要求」のときには, 常に未定と見なされる。

以上のセマンティクルーチンでは, 条件が成立すれば, 次のリスト表現に, さもなければ, (#) の次のリスト表現にとぶ。

(# X) X という名前のリストで表わされる構文表現を構文解析せよ。

3.2.2. 構文解析

構文は, 7つの予約項目の文節, 「はい, いいえ」の文節, 重詞の文節ごとにリスト表現で表される。各文節のリストには名前がつけられており, プラグマティクスのリスト表現から検索される。更に, この構文表現から単語辞書を検索する。構文表現の一部を図9に示す。構文表現は, 次に示す演算子とセマンティクルーチンによって表される。

(AND X₁ ... X_n)

X₁, ..., X_n のリスト表現が, この順序で存在する。通常, AND は省略される。

(OR X₁ ... X_n)

X₁, ..., X_n のリスト表現のうちの一つが存在する。

(OPT X) リスト X は存在しなくてもよい。

SEM (x a₁ ... a_n) セマンティクルーチン X を動作させる。

(* X) X という名のリストを代入する。

上記以外の演算子は, 単語辞書とみなし, 単語認識のプログラムを動作させる。

各演算子は, スタックを用いた帰納的なプログラムで記述されている。入力音声全体が矛盾なく構文解析されると, そのキーワードが抽出され, 予約内容がシステムに把握される。さらに, その単語列を構成する単語のマッチングの度もシステムに記憶される。また, 一つの予約内容が認識されても, 同じ文節内の他の構文の解析も試みる。

3.3. 推論部

時刻表を用いて, 発駅, 発時刻, 列車名の予約項目間で推論を行う。推論では, 認識した予約内容のあいまいさを解消し, 未定の予約項目をすでに認識した予約項目から推定する。推論部は会話モデルとも密接に関連している。

(SEM (YES-NO) (# YES-NO) SEM (VERB) SEM (LATTICE 4 20) (# VERB) SEM (SYNTAX 1) SEM (LATTICE 4 24) (# DATE) SEM (SYNTAX 2) SEM (LATTICE 6 26) (# STARTING-STATION) SEM (SYNTAX 3) SEM (LATTICE 8 26) (# ARRIVING-STATION) SEM (SYNTAX 4) SEM (LATTICE 12 64) (# STARTING-TIME) SEM (SYNTAX 5) SEM (LATTICE 10 64) (# NAME-OF-TRAIN) SEM (SYNTAX 6) SEM (LATTICE 2 18) (# SEAT-CLASS) SEM (SYNTAX 7) SEM (LATTICE 4 14) (# NUMBER-OF-TICKETS))

図8 プラグマティクスのリスト表現

DATE: ((OR (* TSUITACHI) (* FUTSUKA) (* MIKKA) (* YOKKA) (* ITSUKA) (* MUIK A) (* NANOKA) (* YOKA) (* KOKONOKA) (* TOKA) (* HATSUKA) (* KYOASU) ((OPT (* 2)) (* 10) (OR ((OR (* 1) (* 2) (* 3) (* 5) (* 6) (* SHICHI) (* 7) (* 8) (* KU) ((* NICHU) (* YOKKA))) ((* 3) (* 10) (OPT (* 1)) (* NICHU))) (* JOSHI))

STARTING-STATION: ((* EKIMEI) (OPT (* EKI)) (OR (* KARA) (* HATSU) (* YORI) (* JOSHI))

ARRIVING-STATION: ((* EKIMEI) (OPT (* EKI)) (OR (* MADE) (* YUKI) (* IKI) (SEM (LATTICE 3 5) (* E))) (* JOSHI))

STARTING-TIME: ((* SUJI6-22) (* JI) (OPT (* PAUSE)) (* SUJIO-59FUN) (OPT (* HATSU) (* JOSHI))

NAME-OF-TRAIN: ((OR ((* HIKARI) (ORT (* PAUSE)) (* SUJII-199)) ((* KODAMA) (OPT (* PAUSE)) (* SUJII200-299))) (* GO) (* JOSHI))

SEAT-CLASS: ((OR (* SHITEI) (* FUTSU) (* GREEN)) (OPT (OR (* KEN) (* SEKI))) (* JOSHI))

NUMBER-OF-TICKETS: ((* SUJII-9) (* MAI) (* JOSHI))

YES-NO: (OR (* HAI) (* IE) ((* SO) (* DESS)) ((* CHIGAI) (* MASS)))

VERB: (OR ((OR ((OR (* YOYAKU) (* ONEGAI)) (OPT (* ITA) (* SHI)) ((* MOSHIKO MD) (OPT ((* ITA) (* SHI)))) (* MASS)) ((* ARI) (OR (* MASU) (* MASEN)) (* KA)) (* DESS))

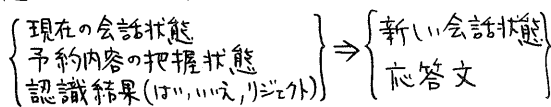
図9 構文表現 (リストの名前のうち, JOSHI, EKIMEI, SUJI6-22, SUJIO-59 FUN, SUJII-5, SUJII-199, SUJII200-299, KYOASU, MASS, DESS, NICHU は構文のリスト表現を, その他は単語辞書の音素表記したリスト表現を表わす)

3.4. 会話モデル

会話モデルは、次に示す7つの会話状態からなっている。

- (1) 予約開始の案内,
- (2) 未定の予約項目の向合せ,
- (3) あいまいな予約項目の確認,
- (4) 全予約項目の確認,
- (5) 予約項目間の矛盾の訂正の要求,
- (6) その他の訂正要求,
- (7) 予約完了の案内,

会話状態推移図を図10に示す。会話状態間の推移は、



の形の推移規則に従って行われる。

質問回答のやりとりができるだけ自然な形で行えるように、推移規則は、次の原則に従って作成されている。

- (1) 予約内容にあいまいな予約項目があるときには、まず、その予約項目の確認を行って、あいまいさをなくする。
- (2) 次に、まだ未定の予約項目がある場合には、その項目の向合せを行う。
- (3) 予約項目間の矛盾のチェックは、全部の予約項目が決まってから行い、そこで矛盾が見つければ訂正を求める。基本的な原則は以上であるが、質問回答のやりとりが自然に行われるように、次の考慮も払った。
- (4) 発時刻あるいは列車名の予約項目にあいまいさがあるときには、すぐには確認を行わず、次のようにする。まず、発駅の予約項目を認識し、

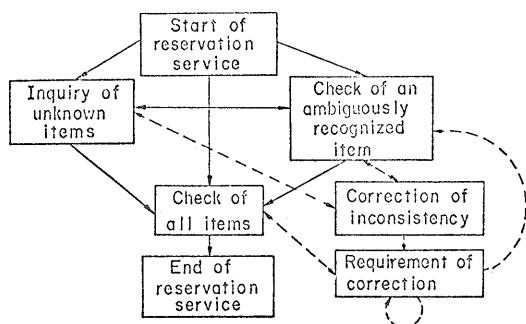


図10. 会話状態推移図

それに基づいて推論で、時刻表に合致しない発時刻や列車名を消して、まだあいまいさが残っていれば確認を行う。

(5) 通常は、推論を行いつつ、予約内容を認識する。しかしながら誤認識が生じたときには、誤った推論によって誤りが訂正できないことがありうる。このような事態をさけるために、推論によってある予約項目の認識結果が全部消された場合は、誤った認識結果に基づく推論が行われた可能性があるとして判断して、そのあと全予約項目の確認を行うまで、推論部の動作を一時止める。

4. システムの性能評価

騒音レベル69dB(A)の計算機室で、文節ごとに区切って発声した座席予約の文章を対象として、認識実験および質問回答の実験を行った。

4.1. 認識実験

発声者は男性8名で、各発声者は20文章を2回発声した。音響処理は文献(2)のシステムによる。音響処理では、母音の標準パターンは各自ごとに作成し、子音の標準パターンは、発声者RNのものに固定した。音韻ラティスに正しい音韻が含まれている割合は、平均で78.6% (母音85.9%, 子音71.4%)であった。音韻ラティスの情報量から推定した音韻認識率⁽¹⁵⁾は、平均で、58.5% (母音77.8%, 子音39.1%)であった。この音韻ラティスを入力として言語処理の性能を調べた。発声者ごとの音韻認識率と文節認識率を表4に示す。文節認識率は、文節中のキーワードの認識の正誤から求める。文節認識率には、複数個の認識結果が得られ、その中に正しい結果が含まれているものも含めた。すべての音韻の標準パターンを学習した発声者RNでは、95.7%の文節認識率が得られた。男性8名の平均の文節認識率は86.0%であった。

音韻認識率と文節認識率との関係を

調べるために、各発声者の20文章ごとの音韻認識率と文節認識率をプロットしたのが、図11のグラフである。システムが非常に良好に動作する目安である、95%の文節認識率を達成するには、63%以上の音韻認識率を得る必要があることが推定できる。

システム全体でみた場合の誤りや拒絶の原因の大半は、次の4項目である。

- (1) 文頭の母音の誤り(21.4%)
- (2) 拗音、半母音の誤り(19.5%)
- (3) 有声子音や音韻の渡りの区間を母音とする誤り(10.7%)
- (4) 撥音/N/の誤り(10.7%)

認識システムの性能を向上させるためには、音響処理におけるこれらの誤りを少なくするとともに、言語処理もさらに精密化する必要がある。

4.2. 質内回答実験

質内回答システムとしての性能の評価を行うために、全部の予約項目が正しく認識できるまで、質内回答を繰返す実験を行った。発声者は認識実験と同じ8名で、各自40種類の予約を1回ずつ、延べ320回の予約について行った。

質内回答の実験では、入力ガリジェクトさせて、同じ内容の質内回答が5回連続した場合には、その予約を打切ることにした。その結果、途中で打切った予約は、320回のうち3回で、予約完了率99.1%であった。

この結果から、文節認識率が76~96%程度でも適切な会話モデルを用いること

によって、質内回答を通じて、非常に高い予約完了率を達成できることがわかった。

質内回答の様子を分析する

ために、途中で打切った回を除く317回の予約について、質内回答の回数を調べた。質内回答の種類は、(a)予約内容の発声、(b)あいまいな予約項目の確認に

対する「はい(b₁)、いいえ(b₂)」の発声、に分けられる。(a)は、さらに、各予約項目の最初の発声(a₁)、発声内容ガリジェクトされたための発声のしなおし(a₂)、誤認識を訂正するための発声(a₃)、に分けられる。全体での平均の質内回答数は、4.69回である。これらの回数を表5に示す。4.69回のうち、1.48回は予約が完了するために最小限必要な質内回答数(a₁)である。a₁が1より大きいのは、最初に、必ずしも、全予約項目を発声しないからである。よって、文節認識率が100%でないことにより生じる質内回答の回数は、3.21回である。発声者に負担に与える誤りの訂正(a₃)0.39回と「いいえ」(b₂)0.50回の回数は、「はい」(b₁)の回数1.71回に比べ少ない。この点でも本システムは、発声者にとって、比較的使いやすいシステムであると考えられる。

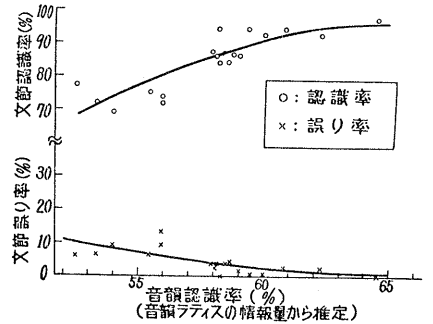


図11. 音韻認識率と文節認識率の関係

表4. 発声者ごとの文節認識率と音韻認識率

speaker		RN	HN	SF	KI	KS	SS	SA	MK	average
phrase recognition rate	correct	95.7	82.2	89.1	90.0	85.2	88.3	74.3	83.0	86.0
	reject	3.0	13.0	8.7	7.4	11.7	9.6	17.4	10.0	10.1
error		1.3	4.8	2.2	2.6	3.1	2.1	8.3	7.0	3.9
phoneme recognition rate		62.7	55.2	60.6	59.1	58.2	59.3	55.6	57.1	58.5

表5. 質内回答の回数

Speaker	RN	HN	SF	KI	KS	SS	SA	MK	Average	
Phrase recognition rate (%)	95.2	87.2	87.2	88.5	85.0	78.4	78.0	76.7	84.6	
Number of successful reservations	40	40	39	40	40	38	40	40	317 (99.1%)	
Question-Answering Cycles	First input (a1)	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	
	Check ambiguity (b)	YES (b1)	1.78	1.35	1.51	1.55	1.75	1.74	1.80	2.23
		NO (b2)	0.26	0.48	0.18	0.50	0.50	0.63	0.53	0.90
	Input repeat due to rejection (a2)	0.18	0.34	0.52	0.37	0.57	0.63	1.46	0.79	0.61
	Input repeat due to error (a3)	0.00	0.20	0.49	0.43	0.43	0.47	0.33	0.83	0.39
Total (b + a2 + a3)	2.22	2.37	2.70	2.85	3.25	3.47	4.12	4.75	3.21	

4.3. 発声の制約による効果⁽¹⁶⁾

これまで述べてきたシステムをSystem I とする。このSystem I のプログラムテキストを変更して、予約項目の発声順序を固定したのがSystem IIである。さらに、構文にも制約を設けたのをSystem IIIとする。各システムにおける発声の制約と、それに適合した発声例を表6に示す。System IIIでは、表6の下線で示された助詞をもった発声のみ受けつけるようになっている。

各システムのタスクの複雑さとして、文節あたりの、average branching factor (表6の発声例)、構文と単語辞書による冗長度、構文が表わしうる可能な単語列の数、意味内容の種類数を表7に示す。タスクの複雑さの指標としては、branching factorよりも、構文と単語辞書による冗長度の方が適切であると思われる。

各システムの制約に適合した発声 (data 1~data 3) を男性4名が、各々20文章ずつ発声した。このときの文節認識率を表8に示す。data 1~3による認識率の向上ほど、System I~IIIによる向上は大きくない。よって、システムを良

表6. 各認識システムの発声の制約とそれに適合した発声例

システム名(発声データ名)	発声の制約	発声例
System I (data 1)	文節の発声順序自由	ひかり15号で、名古屋から、相棒までの、 グリーン席を、1枚、おねがいします。
System II (data 2)	文節の発声順序固定	10日、名古屋から、相棒までの、8時27分発、 ひかり15号で、グリーン席を、1枚、あります。
System III (data 3)	文節の発声順序固定 構文制PR	10日の、名古屋から、相棒まで、8時27分、 ひかり15号の、グリーンを、1枚。

表7. 各システムの複雑さ

	A.B.F.	S.R.	N.P.W.	N.M.
System I	26	4.4	18151	1318
System II	9	5.6	16-12312	1-1020
System III	8	8.2	3-1368	2-1020

A.B.F.--average branching factor
S.R.--redundancy of syntax and word
N.P.W.--number of possible word
N.M.--number of meanings sequences

表8. 各システムの文節認識率

	System I	System II	System III
data 3	92.9	93.9	93.0
data 2	87.9	89.3	---
data 1	85.9	---	---

好に動作させるには、発声者に data 3 のように発声するようにしむけ、システムは自由な発声を受けつける System I のように作成するのがよいと考えられる。

5. 今後の問題

列車の座席予約サービスを対象にヒリあげた会話音声認識システム (Voice Q-A System) の言語処理について述べた。

今後の研究課題としては、次の専柄が考えられる。

- ① 発声者への適応方法。
 - ② 性能および処理速度の向上。
 - ③ 韻律情報の利用。
 - ④ 対象の複雑さの評価。
 - ⑤ 未登録単語の処理。
 - ⑥ 自然言語処理システムとの結合。
- ①および②については現在検討中である。③および④については検討を開始する予定である。⑤、⑥の問題についても、近い将来、検討したいと考えている。

謝辞 本研究を進めるにあたり、終始御指導を賜った斎藤特別研究室長ならびに好田調査役に深く感謝します。また、音響処理の立場から討
論してくださった中津調査員を始めとする斎藤特別研究室の皆様へ感謝します。

(参考文献)

- (1) 鹿野・好田：会話音声の機械認識における言語処理，
通研実報，27，No.6，1978.
- (2) 中津・好田：会話音声の機械認識における音響処理，
信学論 (D)，61-D，p. 261 (昭53-04)
- (3) 鹿野・好田：会話音声の機械認識における言語処理 信学論 (D)，61-D，
p. 253 (昭53-04)
- (4) 好田・中津・鹿野・伊藤：音声によるオンライン質問回答システム，音響学会誌，34，
No.3，1978.
- (5) 斎藤：最近の音声の研究について，信学技報，EA 75-26，1975.
- (6) D. W. Klatt: A Review of the ARPA Speech Understanding Project, JASA, 62, (1977-12)
- (7) 新美・小林・浅見・三木：「SPOKEN BASIC 1」の認識システム，情報処理，18，
No.5，p.453，1977.
- (8) S. Nakagawa: A Machine Understanding System for Spoken Japanese Sentences, 京都大学博士論文，1976.
- (9) V. R. Lesser, R. D. Fennel, L. D. Erman and D. R. Reddy: Organization of the HEARSAY II Speech Understanding System, IEEE trans. ASSP, 23, No.1, p.11, 1975.
- (10) B. T. Lowerre: "The Harpy speech recognition system", Ph.D. Dissertation, Carnegie-Mellon University (1976)
- (11) D. E. Walker: Speech Understanding Through Syntactic and Semantic Analysis, IEEE trans. C, 25, p.432, 1976.
- (12) W. A. Woods: Motivation and Overview of SPEECHLIS: An Experimental Prototype for Speech Understanding Research, IEEE trans. ASSP, 23, No.1, p.2, 1975.
- (13) 鹿野・好田 信学会会大，No.1212, 1976.
- (14) 鹿野・好田：オンライン会話音声認識系における言語処理部の構成，信学技報，PRL76-40.
- (15) 鹿野・好田：音韻特徴の評価システム，信学技報，PRL 78-42, 1978 1976.
- (16) 鹿野・好田：会話音声認識における発声の制約とその効果，信学会会大，No.1240, 1978.