人間の理解手法を用いたロバストな音声対話システム

山本幹雄[†] 伊藤敏彦[††]
肥田野勝[††] 中川聖一[††]

現在の音声認識技術では，音声対話システムの品質を向上させるためには，間投詞，助詞落ち，言い直し，倒置等を含む文の理解はもちろん，誤認識文からの発話文の復元も必要不可欠である．そこで復元ストラテジー開発のため，人間は誤認識された音声認識結果をどのようにして，またどのくらい原文と意味的に正しい文に訂正できるかという実験を行った．単語数241，パープレキシティ74の文法で，音声認識部だけでの平均文認識率は57.4%であったが，認識結果からのエキスパートの復元訂正によって文理解率は87%に向上した．この人間の訂正ストラテジーを用いた自然発話の意味理解システムの開発を行い，その評価実験を行った．語彙数275，パープレキシティ102の文法を用いて，最終的にシステムの利用に関して助詞落ちや言い直し，間投詞を含む初心者の発話した自然言語に対して，文音声認識率52%，意味理解率72%を得た．

A Robust Spoken Dialogue System Based on Understanding Mechanism of Human Being

MIKIO YAMAMOTO,[†] TOSHIHIKO ITOH,[††] MASARU HIDANO[††]
and SEIICHI NAKAGAWA[††]

In a current speech recognition technology, an interpreter that receives the recognized sentences must be developed so as to cope not only with spontaneous sentences but also with illegal sentences with recognition errors to improve a spoken dialogue system property. Therefore, we carried out experiments to investigate how humans modify or correct the recognized sentences which might include errors. Although 43% of the sentecnes were the results of misrecognition, the results showed that the subjects who were familiar with the system could correctly interpret 87% of all the sentences. And several findings were obtained on human strategies for robust interpretation. We developed a robust interpreter by integrating some of the above findings. By using the interpreter efficiently, we constructed a robust spoken dialogue sytem. Experiments were conducted to evaluate the overall performance of the system. The results indicated that the system could correctly interpret 72% of all the sentences uttered by naive (unfamiliar) users for utterances with the recognition accuracy of 52% (the vocabulary was 275 words and the perplexity was 102).

1. はじめに

ユーザに自然な発話を許す対話システムは，これまで音声認識で評価用に用いられてきた朗読文などの発話に比べてバリエーションの大きな発話を扱わなければならない．話し言葉の文法は書き言葉に比べてかなり緩くなり，間投詞，言い直し，曖昧な発話などの現象も多く生じてくる．制約の多くを文法的制約に頼る音声認識システムではパープレキシティが増大し，認識率が下がる．また，間投詞や言い直し，未知語などの問題によって認識率はさらに下がり，音声認識結果には単語の置換，挿入，欠落などが増大する．受理可能な文を多くすることと認識率はトレードオフであるため[1]，どこかで妥協するしかない．スポッティング認識とキーワードを用いた意味理解などの方法も考えられるが，自然な発話を認識，意味理解するには，文節ラティスを介する方法よりもOne-pass型のような最適な照合方法を用いることが重要であるという結果が出ている[8]．このため，現状の音声対話システムでは誤りを含んだ認識結果を解析しなければならない．すなわち，音声対話システムにおいて自然な発話を理解するためには間投詞，助詞落ち，言い直し，倒置な

† 筑波大学電子・情報工学系
Institute of Information Sciences and Electronics, Tukuba University
†† 豊橋技術科学大学情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology

どを含む文の理解はもちろん，誤認識文からの発話文の復元も対話システムの品質を向上させるために必要不可欠である．

いくつかのシステムでは，これらの自然な発話文や誤認識を含む文を扱うためにロバストなマッチングを使用している．SRIのTemplate Matcher（TM）は，発話での最適な単語やフレーズをスロットに埋めようとするテンプレートを使用している[3]．最もスコアの高いテンプレートが意味表現を生む．Carnegie Mellon UniversityのPhoenixは，フレーム構造のスロットとして現れるいくつかのsemantic tokenに一致するフレーズパターンをRTN（Recursive Transition Network）で表現している[4]．システムは，dynamic programming beam searchの方法を使い，並列に異なったフレームのスロットを埋めていく．フレームのスコアは，埋められた入力単語の数である．SRIのTMやCMUのPhoenixは，非文法的な文を意味理解するために，ロバストなマッチングを使用している．つまり，文全体を正しくパージングすることを目的とするのではなく，重要な単語やフレーズをマッチングによって獲得していくことでその文のパーズを獲得し文の意味を得る．そのためのマッチングパターンとして，TMはテンプレートを用いたマッチング，PhoenixはRTNを用いたマッチングを用いている．また，R. KuhnとR.De Moriは意味理解用のルールをトレーニングデータから自動的に学習させる新しい方法を提案している[2]．そのルールはSemantic Classification Trees（SCTs）と呼ばれる木の集まりからエンコードされ，その方法によって学習されたルールは，非文法的な文や入力文の誤認識に対してロバストになる．新しいデータ構造のSCT（Semantic Classification Tree）はロバストマッチング用のパターンになることができるし，SCTによって得られた意味理解用ルールは発話の少数の単語を頼りとして意味理解しているので，音声認識部の誤認識を含んだ文や非文法的な発話に対してロバストである．

一方，我々のシステムは，文節解析結果の係り受けに基づく文節間依存関係を解析するという通常の構文解析・意味解析を基礎としている．このとき，助詞落ち，助詞誤り，倒置に対応するためにいくつかのヒューリスティックスを用いている．係り受け解析が成功した場合は，再帰的に文節の意味表現を組み合わせて文の意味を作る．意味表現はドメインやタスクといった背景的な観点からおかしなところがないかどうかチェックされる．さらに，ボトムアップに意味表現を得ることができない場合は，トップダウン的にキーワードによる意味の抽出を行う．このような方法を用い，非文法的な文や誤認識を含んだ文に対するロバスト性を向上させた．

本論文では，人間が誤認識を含んだ認識文をどのようにして復元しているか調査した実験結果を参考に考案した復元ストラテジーを報告する．また復元ストラテジーを用いたロバストな意味理解システムとシステム評価実験について報告する．

2. 誤認識結果の人間による復元実験

本章では，いろいろな種類の誤認識を含んだ音声認識結果を被験者に提示し，この認識結果から実際に発話された文と同じ意味（意味同等文）になるように訂正できるか調査した実験について報告する．

2.1 復元訂正実験

音声認識させるための原文として，富士山観光案内対話システムへの，観光地・宿泊施設に関する入力文115文を使用した（このうち，15文はシステムで受理不可能な文である）．それらは定型文（50文），倒置文（10文），間投詞を含む文（10文），助詞落ちを含む文（10文），言い直しを含む文（10文），以上の複合文（10文），未知語を含む文（5文），音声認識用生成規則のない文（生成規則で受理できない規則外の文）（5文），以上の複合文（5文）から構成され，まったく同じ文は含んでいない．未知語としては名詞，動詞，形容詞，助詞，終助詞を使用した．これらの文を話者2人に発話してもらい，音声認識システムで認識を行った．また誤認識のバリエーションを増やすため，認識条件を変化させ115文×3回（1: 不特定話者モデル・メルケプストラム・回帰係数，2: 話者適応モデル・メルケプストラム・継続時間制御，3: 話者適応モデル・メルケプストラム・回帰係数・継続時間制御）の認識を行った[8]．平均文認識率は3つの認識条件でそれぞれ40.5%，52.6%，57.4%であった．

こうして得られた1文あたり6つの認識結果から1つずつ実験で使用する認識結果を選択した．選択方法は音声認識システムでよく生じる間違い，未知語処理が行われたものを優先的に選択し，しかも全体的にはバラエティーに富んだものになるようにした．その結果115文中100文は，なんらかの誤認識を含んでいるが，これらの誤認識された文の中には人間から見て原文と意味的にまったく同等な文もある．

使用した認識結果の分類と，認識結果が原文と意味的に異なる文（受理可能文65文，受理不可能文15文）になった原因を分類した内訳を**表1**に示す（分類は著者らの主観で行った．また，1文に複数の原因を

表1 実験に使用した115文の認識結果の内訳
Table 1 Experimental recognition results for 115 sentences.

原文と認識結果の比較	受理可能文	受理不可能文
意味の同等な文に認識	35文	0文
意味の異なる文に認識	65文	15文

意味が異なる原因	受理可能文	受理不可能文
助詞の置換	14文	1文
混入	12文	2文
欠落	18文	6文
終助詞の置換	10文	1文
欠落	1文	2文
名詞の置換	9文	5文
混入	15文	2文
欠落	7文	4文
述部の置換	11文	3文
混入	2文	1文
欠落	0文	4文
疑問詞の置換	3文	0文
混入	1文	2文
欠落	6文	2文

原文1：
えーと，安い旅館はありますか．
原文1の認識結果：
|えと| 安い料金はありますか．

原文2：
河口湖にどんな，りょ…，民宿がありますか．
原文2の認識結果：
河口湖にはどんな宿 [mi syu ku ga] ありますか．

原文3：
サイクリングしたいんですが．
原文3の認識結果：
サイクリングしたいんですか．

図1 認識結果の例
Fig. 1 Examples of recognition results.

含む場合もある）．表1の疑問詞とは，「どんな」・「どこ」といった疑問文を示す単語のことである．表の受理可能文とは，倒置文，間投詞，助詞落ち，言い直しに対処できる音声認識用文法（CFG文法）で受理可能な100文に対するものである．受理不可能文は，未知語を含む文，生成規則外の文で音声認識用文法で受理できない15文の内訳である．また実際の実験に使用した認識結果の例を**図1**（図2も参照）に示す．

音声認識システムから得られた情報はすべて被験者に提示した．その際，被験者に分かりやすいように，間投詞として認識されたものはその単語を ||，未知語として処理されたものはその音節列を [ ] で囲み区別した．また，音声認識システムが認識できる語彙を提示した場合と提示しない場合の訂正率の変化を調べた．さらに文脈の知識（対話の履歴）の有無による訂正率の変化も見るために，使用する115文の音声認識結果をまったくのランダムに並べたもの [文脈知識なし] と，音声認識結果とそれに対するシステムの応答を対話形式（20対話）で並べたもの [文脈知識あり] の2種類を用意した．被験者には，最初に文脈知識がない場合として，まったくランダムに115文を1文ずつ提示し，訂正を行ってもらった．文脈知識のない場合の訂正実験終了後，文脈知識のある場合として，同じ被験者にいくつかの対話例にそった順番で並んでいる認識結果を1文ずつ提示し，同じように訂正してもらった．このとき，1文の訂正終了ごとに，いま訂正してもらった文の正解（つまり，実際に発話された文）とそれに対する対話システムの応答文を提示することによって被験者に文脈知識を与え，次の新しい認識結果を訂正してもらった．

復元実験の被験者の人数は6人で，全員22～25歳の本大学の学生であり，出生地・成育地は静岡（2人）・愛知・岐阜・岩手・茨城である．さらに最低2年以上は本大学のある愛知県に居住しており，本実験内容に関しては方言等の言語生活の影響の差はない．またタスクに関する知識は一般的な知識を除いてほとんどない．その6人に，実験を行う前に実験の目的，訂正の方法，認識結果の見方，発話された文のタスクなどが書かれた紙を渡し読んでもらった．そのうち3人にはさらに音声認識システムに登録されている241単語の一覧表も渡した．実験終了後，被験者が訂正に成功した文に関して，その訂正戦略をインタビューした．ここでの訂正成功とは，我々が客観的に見て被験者が訂正した文の内部表現（意味表現）と，実際に発話された文の内部表現が一致した場合と定義している．すなわち，実際に発話された文とまったく同じ文になる必要はない．

2.2 実験結果

図2に，被験者が訂正に成功した文の例をいくつか示す．例1の原文は定型文である．認識結果には助詞の置換が生じているが，正しく復元訂正されている．例2の場合，原文は間投詞，助詞落ち，言い直しを含み，認識結果は述部の置換と助詞の混入によって原文とは異なる文になっている．この例の場合，被験者は最初，助詞が不自然なことに気づき「山中湖ホテルに近くしたいんですが」と考えた．次にその文では対話システムへの入力文としては不自然なので，述部を「近くしたい」から意味的に適切で音響的に似ている「宿泊したい」に変更し訂正に成功している．例文3は未知語「ペンションマリエ」を含んだ文で，文脈

例 1
原文：
鳴沢の氷穴はどこにありますか.

原文の認識結果：
鳴沢の氷穴をどこにありますか.

被験者の訂正結果：
鳴沢の氷穴はどこにありますか.

例 2
原文：
山中湖ホテル，その，とま…，宿泊したいんですが.

原文の認識結果：
山中湖ホテルは |その| |あ| 近くしたいんですが.

被験者の訂正結果：
山中湖ホテルに宿泊したいんですが.

例 3
システムの応答文：
ペンションマリエというペンションがあります.

原文：
ペンションマリエに食事はありますか.

原文の認識結果：
ペンション |あ| に [e ni] 食事はありますか.

被験者の訂正結果：
ペンションマリエに食事はありますか.

図 2 認識結果の訂正例
Fig. 2 Examples of human recovery from recognition errors.

の知識を使用して訂正に成功した例である．未知語の部分が間投詞，助詞，未知語として認識されている．この例の場合，被験者は前文のシステム応答からペンションマリエという単語が出現していること，間投詞，助詞，未知語の部分の音節系列が [a ni e ni] で，「マリエに」に似ていることから「ペンション」を「ペンションマリエ」に変更し訂正に成功している．

表 2 に訂正成功率を示す．表の「受理可」は，倒置文，間投詞，助詞落ち，言い直しを含むが音声認識用文法で受理可能な 100 文を 3 人が訂正した 300 文，「受理不可」は未知語を含む文と音声認識用文法で受理不可能な 15 文を 3 人が訂正した 45 文で評価した結果である．「語彙非提示」と「語彙提示」は，それぞれで認識システムの辞書を提示しなかった場合と提示した場合であり，被験者はそれぞれ別の 3 名である．

語彙の提示による訂正成功率の変化はあまり見られなかった．語彙の提示が通常役にたつ状況は，単語，特に名詞，述部，疑問詞を変更，挿入する場合である．しかしかしながら，今回の実験の結果では語彙を利用して修正を行ってもそのほとんどが不正解となっている．さらに，名詞，述部，疑問詞の訂正に成功した文を分析してみると，そのほとんどが認識結果自体，または文脈から変更，挿入されるべき単語が連想できるものが多い．そのため語彙の提示の有無による復元率の差が現れなかったと考えられる．他の理由としては，241 個の単語一覧表を修正を行うごとに参照することは非常に困難であるため，あまり活用されなかったとも考えられる．

表 2 文脈知識と語彙提示の有無による訂正成功文数
Table 2 The number of successfully recovered sentences with and without context knowledge/vocabulary.

	文脈知識なし		文脈知識あり	
	受理可	受理不可	受理可	受理不可
語彙非提示	185 文 (61.7%)	15 文 (33.3%)	217 文 (72.3%)	27 文 (60.0%)
語彙提示	189 文 (63.0%)	10 文 (22.2%)	203 文 (67.7%)	25 文 (55.6%)

(括弧内は訂正成功率)

文脈知識の有無による訂正成功率は，どの被験者でも文脈の知識を使用できる方がよくなっており，文脈知識のない場合より平均 10%上がっている．特に未知語を含んだ文，受理不可の文に対して文脈の知識の効果がよく表れている．これらから誤認識結果の訂正には，文脈の知識の使用が有効であることが分かる．

原文と意味が異なる文に認識された入力文 80 文を 6 人が訂正した 480 文で，正しく復元訂正された文は文脈知識なしで 40.8%，文脈知識ありで 55.4%であった．また，認識結果が原文と意味的に同等な文（35 文を 6 人が訂正した 210 文）であったにもかかわらず，修正によって意味の異なった文に変更された場合は，文脈知識なしで 2.9%，文脈知識ありで 1.9%あった．

表 3 に原文と意味的に異なる原因となった誤認識部分をどれだけ修正できたかを示す．各欄のスラッシュの右側の数値は原文と意味的に異なった原因の誤認識の数であり，左側の数値は修正に成功した誤認識の数である．助詞に関しては助詞の置換，欠落はおおむね訂正できていることが分かる．助詞の混入についてはあまり訂正に成功していないが，これは文法が文節単位に作られているため，助詞が単独で混入される場合より他の誤認識，たとえば名詞の混入と同時に表れる場合が多いからである．また助詞に関しては文脈の知識の有無による変化はほとんど見られない．このことから助詞の訂正に関しては，文脈の知識に依存しない助詞誤認識復元ヒューリステックスで十分であると考えられる．ただし，終助詞は，文脈知識の有無による差が大きく表れている．特に欠落に関しては文脈知識ありの方がかなりよく修正できている．このことから

表3 誤認識箇所の修正成功内訳
Table 3 The number of successful recoveries from recognition errors.

	文脈知識なし		文脈知識あり	
原因	受理可	受理不可	受理可	受理不可
助詞の置換	72/84	3/6	72/84	3/6
欠落	80/108	12/36	77/108	24/36
混入	31/72	1/12	37/72	2/72
終助詞の置換	32/60	0/6	48/60	1/60
欠落	1/6	4/12	6/6	8/12
名詞の置換	12/54	4/30	16/54	9/30
欠落	11/42	4/24	17/42	7/24
混入	25/90	0/12	38/90	2/12
述部の置換	5/66	3/18	10/66	9/18
欠落	0/0	12/24	0/0	20/24
混入	0/12	0/6	1/12	0/6
疑問詞の置換	6/18	0/0	8/18	0/0
欠落	1/36	2/12	4/36	5/12
混入	2/6	3/12	4/6	7/12

*各欄は，復元訂正できた誤認識/意味が異なった文の誤認識

対話システムのタスクによってリジェクトできる文，復元訂正できる誤認識の文を想定できる．後章でこのストラテジー（フィルタリング）についても述べる．

名詞，述部，疑問詞等の訂正は非常に難しい．被験者は名詞，述部等はできるだけそのまま使用し文脈的，意味的におかしい場合にそれらを訂正しようとしていた．逆に重要な名詞，述部，疑問詞（つまりキーワード）がきちんと認識されていると，全体的に誤認識された文（復元が困難な文）からでも意味的に正しい文に復元訂正できるものも多くあった．このキーワードからの文の復元についてのストラテジーについても後章で述べる．

表 4 に音韻情報，つまり未知語，間投詞等に認識された音節列を利用することにより原文と意味が異なる文を原文と意味同等な文に訂正できた文数を示す．音韻情報を利用した訂正では，受理不可文 15 文を 6 人が訂正した 90 文のうち，文脈知識なしで 20%，文脈知識ありで 43%が復元に成功している．未知語や生成規則外の文はその部分が音節列として音韻情報が出現しやすいとしても，かなりよく訂正できているといえる．特に文脈知識ありの場合，音韻情報を利用して訂正を行った文の 81%が原文と意味同等な文に訂正されており，そのほとんどが音韻情報として，文の理解に重要な名詞，疑問詞，述部の単語の音節列そのままか，似た音節列が表れているものであった．ほかに，音韻情報を用いての訂正に成功したのは，文脈知識のある場合でシステムの応答から得られた情報，特に地名，宿泊施設名等の名詞と，誤認識された単語や未知語として表れた音節列が似ているためにうまく訂正できたものがあった．

表4 音韻情報を利用しての訂正
Table 4 Successful estimation of unrecognized part using phonetic information.

	文脈知識なし		文脈知識あり	
訂正後	受理可	受理不可	受理可	受理不可
原文と意味同等な文	25 文 (4.1%)	18 文 (20.0%)	29 文 (4.8%)	39 文 (43.3%)
原文と意味が異なる文	9 文 (1.5%)	13 文 (14.4%)	14 文 (2.3%)	9 文 (10.0%)
合計	34 文	31 文	43 文	48 文

表5 異なる意味の文として認識された文の訂正
Table 5 The number of successfully recognized sentences which are recognized as sentences with different meaning.

	文脈知識なし		文脈知識あり	
訂正後	受理可	受理不可	受理可	受理不可
原文と意味同等な文	34 文 (21.8%)	5 文 (13.9%)	52 文 (33.3%)	14 文 (38.9%)
原文と意味が異なる文	122 文 (78.2%)	31 文 (86.1%)	104 文 (66.7%)	22 文 (61.1%)
合計	156 文	36 文	156 文	36 文

表 5 に音声認識部での認識結果が日本語的には正しいが実際に発話された文の意味とは異なっている文（つまり，実際に発話された文が「スキーがしたいのですが」で，認識結果が「スキーがしたいのですか」となった場合等）を，実際に発話された意味の文に正しく訂正できた文の数を示す．文法的（日本語的）に正しいが原文とは意味が異なる文になった場合，正しく復元訂正できたのは文脈知識なしで平均 20%，文脈ありで平均 34%程度であった．また，正しく復元訂正できた文のほとんどが終助詞の置換と欠落の誤認識によるもので，終助詞以外の誤認識に関しては復元訂正は大変難しいことが分かった．文法的に正しい文になりやすいのは名詞の置換，疑問詞の欠落，述部の置換等であるが，文脈知識のない場合では被験者はこれらの誤認識文の修正はほとんど行わず，正しい文と見なしていた．

以上の結果はシステムについてまったく知らない被験者のものである．音声対話システムを熟知しているエキスパートが同様な実験を行った場合，よく誤認識される単語のパターンを用いていたり，単語の音響的類似性を考慮して訂正を行うことを確かめている．

単語数 241，パープレキシティ 74 の文法で，音声認識部だけでの平均文認識率は 57.4%（音声認識用文法受理率は 87.0%）であったが，人間による訂正によって意味理解率は 80%に向上した．またエキスパートによる訂正の場合は意味理解率は 87%まで上昇した[7]．

2.3 まとめ

音声認識システムで誤認識された認識結果を人間に訂正してもらう誤認識訂正実験より，以下のストラテジーを用いていることが分かった．

(1) 名詞，動詞，疑問詞等の自立語の訂正は難しい．
(2) 助詞の誤認識，終助詞の誤認識の修正は人間にとってそれほど難しいことではない．
(3) 人間は名詞，動詞をできるだけそのまま使用し，文脈的，意味的にそれではおかしい場合にそれらを訂正しようとしている．
(4) 名詞，動詞，疑問詞の誤認識を意味理解部で修正することは，文脈を利用することによりある程度可能になる．
(5) 未知語，言い直し，間投詞として誤認識された重要な名詞，動詞，疑問詞をその部分の音声認識結果である不完全な音韻系列から推測し訂正している．
(6) 訂正困難と思われる文は重要と考えられる名詞，動詞，疑問詞から文全体を想像する．

以上のストラテジーはシステムについてまったく知らない被験者による実験からの結果である．システムを熟知しているエキスパートはさらに以下のようなストラテジーも使っている．

(7) よく誤認識される単語のパターンを用いて訂正を行う．また単語の音響的類似性を考慮して訂正する．

以上の結果のうち，(2)から助詞の誤認識・欠落修正アルゴリズム，(4)から背景的な知識を用いたフィルタリングストラテジー，(6)からキーワードによる文の理解ストラテジーを考案し，システムに組み入れた．これらについては4章で述べる．

3. 音声対話システム

本章では，復元訂正実験より得られた結果をもとに，音声認識システムで一部誤認識された認識結果にもある程度対応した，自然な発話の意味理解システムを応用した対話システムについて述べる．

3.1 音声対話システムの概要

ユーザの発話は音声認識部で認識され，文字列に変換された後，対話システムへ送られる．対話システムは，認識結果を形態素解析，文節解析，構文解析，意味解析，省略された動詞の補完と代名詞の補完の処理である文脈解析を行った（以上までの解析を行う部分を意味理解部と呼ぶ）後，応答生成部が応答を文字列として生成する．生成された応答文は音声合成部で音声合成され，ユーザに音声で応答する．それぞれの部分はおおよそシーケンシャルに処理がなされる．なお，本対話システムが対象としている単語の品詞分類はJumanの品詞分類に，意味分類はIPALの意味素性にしたがっている．以下に音声認識部，応答文生成部・対話制御部の概要を述べる．

3.2 音声認識システム

対話システムの音声認識部は，HMMを音節のモデルとして用い，文脈自由文法の構文解析法とフレーム同期型連続音声認識の統合アルゴリズムを基礎としたものである．さらに不要語や言い直しの部分を未知語処理に基づいて処理する．未知語処理では，これらの部分を任意の音節系列により表現し，その認識尤度スコアにペナルティを設ける．文脈自由文法は自然な対話音声を認識するために，助詞落ちや倒置を含む文を受理するように作成した．未知語処理は文節の境界で間投詞や言い直しが生じると仮定している．認識部の語彙数は241，受理できる文集合のテストセットでパープレキシティは74である．音声認識部に関する詳しい内容は文献8)を参照されたい．

3.3 応答生成部・対話制御部

応答生成部は，ユーザの発話の意味表現を受け取り応答文を生成する．応答生成部は問題解決器，知識データベース，応答文生成用意味ネットワーク生成部，応答文生成部から構成されている．知識データベースは意味概念とその間の関係による意味ネットワークで表現されている．問題解決器は入力された文の要求するデータの検索を行う．応答文生成用意味ネットワーク生成部では，入力文意味ネットワークからの応答文生成に必要な情報と問題解決器のデータ検索結果を使用し応答文生成部へ入力するための応答文意味ネットワークを生成する．応答文生成部では入力された生成用意味ネットワークの形から応答文用テンプレートを選択し，生成用意味ネットワークの各ノードをテンプレートに埋め込んでいく方法で応答文の生成を行っている．応答文生成部はユーザの質問に対して応答を生成するが，対話制御部は主に叙述文が入力されたときにどのような応答を返すかを決定する．

4. 自然な発話の音声認識結果に対する構文・意味解析

本章では，自然な発話や誤認識文に対しても正しく構文解析・意味解析を行うために訂正復元実験の結果を参考にして開発された構文・意味解析のヒューリスティックスについて述べる．この構文・意味解析部では，前章で述べた対話システム部の「構文解析」,「意味解析」を行う．

```
(ある (FORM WH-Q)(TARGET (AT-LOC))
      (NEGATION NIL)
      (AT-LOC (富士山))
      (OBJ (観光地 (Q-OBJ (WH-RENTAI)))))
```

図3 意味表現（内部表現）例
Fig. 3 An example of a semantic representation.

ここでの意味解析部における意味理解の定義とは，発話内容を対話システムの問題解決器が解釈できる内部表現（意味表現）に変換することである．つまり，発話内容を正確に内部表現（もちろんタスクやその規模に依存する）に変換する（表現の曖昧性は除く）ことにあるので，事実関係の判断等は問題解決器の部分で処理される．たとえば，「富士山にはどんな観光地がありますか.」という質問文の場合，その内部表現は図 3 のようになる．

4.1 解析手順

構文解析・意味解析は文節解析を行った結果の係り受けに基づく文節間依存関係を解析する．解析の途中結果はチャートデータベースに格納され，1度行った部分解析結果を保存するようになっている．このとき，助詞落ち，助詞誤り，倒置に対応するためにいくつかのヒューリスティックスを用いている．係り受け解析が成功した場合は，再帰的に文節の意味表現を組み合わせて文の意味を作る．意味表現はドメインやタスクといった背景的な観点からおかしなところがないかどうかチェックされる．これをフィルタリングと呼ぶ．詳しくは 4.3 節で述べる．さらに，ボトムアップに意味表現を得ることができない場合は，トップダウン的にキーワードによる意味の抽出を行う．パターンに記述された制約に適合する単語を探すことにより，全体の意味表現を得る．詳しくは 4.4 節で述べる．全体の処理は次のような手順で行われる[9]．

(1) 以下の処理を順次行っていき，解析が成功した時点で（2）へ行く．すべての処理で失敗した場合は（3）へ行く．
  (a) 助詞落ち，倒置を禁止して解析
  (b) 助詞落ちを許可して解析
  (c) 助詞落ち，倒置を許可して解析
  (d) 助詞の誤りを認めて，倒置を許可して解析（助詞が誤っていると仮定した助詞は省略されたと見なし，助詞落ちを解析するヒューリスティックスを用いる.）
(2) 文脈的な知識によって，正しい内容かどうかをチェックする（フィルタリング）．
  (a) 正しくない場合
    修正用のヒューリスティックスがある場合は，それを適用し，解析結果とする（解析終了）．修正用のヒューリスティックスがない場合は，(3) へ行く．
  (b) 正しい場合
    得られた意味表現を解析結果とする（解析終了）．
(3) 部分解析結果を用いてキーワード解析を行い，その結果を解析結果とする．

4.2 助詞落ち，助詞誤り，倒置の解析

我々は，比較的小規模なタスクでは助詞落ちと倒置の 90%を解析可能とする以下のようなヒューリスティックスを提案している[6]．

助詞落ち用

- 助詞が省略された名詞文節は最も近くの述部に係る．
- 述部に係る場合は，必須格を候補として考える．
- 文頭の助詞落ち名詞文節には「は」を補った文節も文節切り出しの結果の 1 つとして追加する．
- 述部を飛び越さない次の名詞文節に「の」が省略されているものとして係ることができる．だだし，これは述部に係ることができない場合に限る．
- 助詞の省略された名詞が，次の名詞文節と概念階層上で 1 つ上の親概念を持つ場合，並列の「と」の省略として係ることができる．

倒置用

- 文の先頭を含み，終止形の述部で終わる最も長い部分解析木から順番に倒置でない部分を候補とする．
- 任意の部分解析木は直前（文の左隣）の部分解析木に係ることができる．

解析システムでは，助詞落ち，倒置がないものとして解析を行い，それに失敗した場合，これらのヒューリスティックスが使われる．さらにヒューリスティックスを使っても解析に失敗した場合，助詞が誤っているものとして解析を試みる．誤っていると仮定した助詞は省略されたと見なし，上記のヒューリスティックスを用いることによってこの処理を行っている．誤りと仮定した助詞が少ない解析結果を候補として優先する．

4.3 タスクの情報を利用したフィルタリング

フィルタは，認められない意味表現を記述したパターンとして登録されている．修正して正しく変形することが可能なものは，修正手続きをパターンとともに登録してある．このパターンに一致した場合，意味表現はリジェクトされ，修正手続きがある場合は，その手続きが適用される．パターンの例を図 4 に示す．

```
filter1: (pattern: ((ある かかる) (form assert))
          modify-fun: (change 'form yn-q))
```

図4 意味表現フィルタの記述例
Fig. 4 An example of a semantic representation filter.

```
(prototype: (?aru (form wh-q) (target (obj))
                  (obj ?org)
                  (at-loc ?loc))
   binding: (?aru (imi (ある)))
                  (?org (sem-features org))
                  (?loc (sem-features loc)))
```

図5 キーワードパターンの例
Fig. 5 An example of a keyword pattern.

pattern の後にはマッチング用のパターン，modify-fun の後には修正用の関数が記述される．たとえば，図 4 の filter1 の例では，「『ある』または『かかる』が主動詞の発話は質問であるはずである」という知識を記述している．このような知識を使用して終助詞の誤認識などの復元訂正を行っている．

4.4 キーワード抽出による意味解釈

上記の意味解釈の方法がすべて失敗した場合，トップダウン的な意味解釈を行う．各文節の意味表現を抽出し，その意味表現とマッチする変数（キーワード）を持つパターンを使って意味表現を作る．**図 5** にこの手法の知識（キーワードパターン）の例を示す．prototype の後は結果としての意味表現のもととなるパターン，binding の後は変数の束縛条件である．変数の前には「?」が付けてある．図 5 では ?aru という変数は「(ある)」という意味を持つ文節の意味表現に束縛される．その値が，パターンの変数の値として展開される．図 5 のキーワードパターンは「富士山にどんなホテルがありますか」などの文に対応する意味表現を抽出するもので，「富士山」などの場所の名詞，「ホテル」などの施設の名詞，「ありますか」などの「ある」という意味を持つ動詞が，音声認識結果に含まれていれば，強制的に上記のような意味表現に変換する．このようにして人間が重要な単語から文全体の意味を想定するストラテジーを対話システム上に実現している．もちろん，すべての意味解析をキーワード抽出のみで行うことも考えられるが，小さなタスクにおいては意味理解用キーワードをすべて用意することも可能であるが，大きなタスクにおいてすべてのキーワードを用意することは現実的ではない．

表6 各モードでの意味理解率
Table 6 Understanding rate under each of modes.

モード	意味理解率 [%]	
	音声認識部が正解認識	音声認識部が誤認識
通常モード	20.2	9.2
助詞落ちモード	7.6	2.5
助詞誤りモード	0.0	2.5
フィルタモード	0.0	2.5
キーワードモード	0.8	9.2
合計	28.6	25.9

5. 評価実験

本章では，実際に雑音のある実験室で対話システムを使用し収集したデータをもとに行った音声対話システムの意味理解部の評価実験について述べる．

5.1 評価実験 1

意味理解部の最初の評価として，富士山観光案内対話システムについてある程度知っている研究室内の 5 人の学生に，実際に対話システムを使用してもらった．被験者には，1 泊 2 日の富士山周辺への研究室旅行を計画すると想定してもらう．それから富士山周辺の観光（観光地，宿泊施設）について聞くことができる音声対話システム（富士山観光案内システム）を使用し，いくつかの項目（1，2 日目の目的地とそこでのプラン，宿泊施設の場所，種類，料金，名前の計 8 項目）を決めてもらうタスクを課した．5 人の被験者のタスク達成率は 100%であった．実験結果を**表 6** に示す．また，5 人の被験者の全発話 119 文中 89 文が音声認識用文法で受理可能であった．

対話システムの意味理解部の評価方法として，被験者が実際に発話した入力文の書き起こしをシステム開発者が意味表現（意味ネットワーク）に変換したもの（意味表現 1）と，入力文の認識結果を対話システムの意味理解部に入力し，結果として出力した意味表現（意味表現 2）が一致した場合，正解とし評価を行った．

表 6 の「通常モード」はノーマルな意味解析モード，つまり特別な理解処理なしに正しい意味表現を出力できた割合を示している．表より，全体の 9.2%が音声認識部で誤認識となっても正しく意味理解できることが分かる．たとえば入力文が「富士山に泊まりたいですけど」に対し出力が「富士山に泊まりたいんですが」の場合などである．「助詞落ちモード」は助詞落ちモードによる処理によって正しい意味表現を出力することができた割合を示している．「助詞誤りモード」，「フィルタモード（フィルタパターン数 7）」，「キーワードモード（キーワードパターン数 14）」もそれぞ

れのモードの処理によって正しく意味理解された割合を示している．したがって，正しく意味理解できた全発話の54.5%（28.6%+25.9%）のうち，ノーマルモードによる29.4%（20.2%+9.2%）のほかに，リカバリーモードによって25.1%が正しく意味理解されたことを示している．これは音声認識部の文認識率は30.3%であるが（文認識率は30.3%であったが，このうち正しく意味理解できたのは94%（全体の28.6%）であった.），リカバリーモードによって意味理解率は54.5%まで上昇したことを意味している．

5.2 評価実験2

評価実験1の終了後，その実験結果などを用いてさらにフィルタモード，キーワードモード等の知識の追加（フィルタパターン数2，キーワードパターン数5追加）を行い，評価実験2を行った．被験者は音声対話システムに関してまったく知識のない大学院生7人である．被験者に課したタスクは評価実験1とまったく同じである．被験者には富士山観光案内システムについて簡単な説明を行い，さらに実際の富士山観光案内システムの対話例を提示し参考にしてもらった．

このようにして収集した音声データの内訳を**表7**に示す．7人が発話した全発話数232文のうち，140文が音声認識用文法で受理でき，その文法を用いて認識したとき，文認識率は38.8%であった（評価実験1より音声認識用文法受理率に対して文認識率が良い(38.2%対64.2%）のは認識条件が異なるためである．評価実験1では話者適応モデル・メルケプストラム・継続時間制御で音声区間の切り出し誤りを含むが，評価実験2では話者適応モデル・メルケプストラム・回帰係数・継続時間制御で音声区間の切り出し誤りはない)．

項目（タスク）達成率は98.2%である．全体的に間投詞や言い淀みなどが少ないのは，被験者が機械対人間の対話と意識し過ぎているためと思われる．

対話システムの意味理解部の評価方法は評価実験1と同じ方法で行った．またこの評価実験は条件を変化させ2度行った．1度目は7人の発話データをそのまますべて用いて評価を行った．それから7人の発話データの半分（116文）をまったくランダムに抜き出し，その認識結果と書き起こしデータを意味理解部の開発者に渡した．それをもとにさらに意味理解部の知識データの追加（フィルタパターン数27，キーワードパターン数2追加）を行ってもらい，残りのデータ116文（232文–116文）で2度目の評価を行った．**表8**にその結果を示す．

表の「テキスト入力」とは被験者の発話文を書き起こしたもの（つまり音声認識率100%の場合で，間投詞，言い直し部分等を除いたもの）を直接意味理解部に入力した場合の結果である．テキスト入力でも正しく意味理解できなかった理由として，約半分が意味表現生成用規則がないためであり，残り半分は意味理解部の解析用語彙への単語の未登録のためであった（登録単語のうち自立語は284である）．この表を見ると第1回の評価では全体の発話文の50%，第2回目の評価で全体の発話文の63%が正しく意味理解できていることが分かる．この結果は復元訂正実験の結果より悪い．この理由として訂正実験で用いた入力発話の音声認識用文法受理率は87.0%（115文中100文）であるのに対して，この評価実験の音声認識用文法受理率は60.3%と約27%も低いことがあげられる．そこでこの評価実験2の認識結果を対話システムに関する知識のない被験者2人に復元訂正してもらったところ，訂正成功率は64.0%程度であった．このことから意味理解部は人間に近い意味理解率であることが分かる．

音声認識用文法で受理できなかった文（92文）のうち未知語が原因で文が受理できない文が65文（未知語26文＋未知語＆文法規則外39文）もあり，未知語に対する処理の必要性がかなり重要であることが分かった．このほか，文法規則が不備なため受理できない文が20文あった．そこで，音声認識用文法の改良を行い，タスク範囲外の文2文を除いた90文を認識用文法で受理できるまで拡大した．これによって単語数は275，パープレキシティは102となった．この改良した文法を用いてDATに録音してある発話データをもう1度，音声認識システムに認識させた．この

表7 評価用データ
Table 7 Evaluation data.

被験者数	7人
項目達成率	98.2%
発話数	232文
文法受理率	140文(60.3%)
文認識率	90文(38.8%)
間投詞	15文
言い淀み	8文
言い直し	3文
言い間違い	5文

表8 評価結果
Table 8 Evaluation results.

評価	意味理解文数	評価文数
第1回	117文(50.4%)	232文
	51文(44.0%)	116文
第2回	146文(62.9%)	232文
	65文(56.0%)	116文
テキスト入力	182文(78.4%)	232文
	96文(82.7%)	116文

表 9 文法改良後の評価結果
Table 9 Evaluation results after grammer improvement.

	意味理解文数	評価文数
認識文法	146 文 (62.9%)	232 文
改良前	65 文 (56.0%)	116 文
認識文法	162 文 (69.8%)	232 文
改良後	83 文 (71.6%)	116 文
テキスト	182 文 (78.4%)	232 文
入力	96 文 (82.7%)	116 文

SYS1: 富士山観光案内システムです．ご用件をお願いします．
USR2: えーと，富士山の周りには何がありますか．
(富士山の [ま] 旅館何ありますか．)
SYS2: データベースにないので，お答えできません．
USR3: 富士山の辺りに何がありますか．
(富士山の辺りに何ありますか．)
SYS3: 富士山には河口湖や山中湖，西湖，本栖湖，精進湖があります．
USR4: 山中湖にスポーツできる所はありますか．
(山中湖にスポーツできる所はありますか．)
SYS4: データベースにないので，お答えできません．
USR5: えー，山中湖に宿泊施設はありますか．
(山中湖に宿泊施設はありますか．)
SYS5: 山中湖にはニュー山中湖や山中湖ホテルがあります．
USR6: えー，ニュー山中湖，のー，宿泊料金はいくらですか．
(ニュー山中湖運動宿泊料金はいくらですか．)
SYS6: ニュー山中湖は 8000 円かかります．
USR7: えー，山中湖ホテルの宿泊料金はいくらですか．
(山中湖ホテルの宿泊料金はいくらですか．)
SYS7: 12000 円かかります．
USR8: 精進湖に，では何ができすか．
(精進湖では何 [え] できますか．)
SYS8: 精進湖ではサイクリングやスケートができます．

図 6 対話例
Fig. 6 An example of dialogue.

ときの文認識率は 52.1%であった．この認識結果を意味理解部に入力し評価を行った結果を**表 9** に示す．

音声認識用文法を改良し，文法受理率を改良前の 60.3%から 99.1% (230/232 文) まで向上させた場合，意味理解率は 7%程度上昇した．受理可能な文に対する文認識率はパープレキティの増大のために減少しているが (64.2%から 52.6%へ)，全体としての意味理解率は向上していることからも未知語の減少や未知語に対する処理の必要性が分かる．音声認識用文法改良後の意味理解率 70%は，文認識率が 52.1%である割にテキスト入力の意味理解率に近い結果であるから，意味理解部のロバスト性はかなり向上したといえる．

5.3 システム動作例

実際にシステムを使った評価実験を行ったときの対話例を**図 6** に示す．"SYS" はシステムの発話・応答であり，"USR" はユーザの発話である．ユーザの発話の下の括弧内の文は，対話システムの意味理解部に実際に入力された音声認識結果である．現在の対話システムの意味理解部へ入力される音声認識結果は，未知語として認識された部分を削除したものである．実際に認識された文中の [ ] でかこまれた部分は認識部が間投詞として認識した部分である．間投詞 (USR2, 5, 6, 7) や言い直し (USR8)，言い淀み (USR6) の発話への対処，助詞の誤認識による脱落 (USR3) への対処がうまく動作していることがわかる．(質問応答用の知識データベースはまだ整備が不十分である．)

6. システムの拡張

今回は訂正復元実験によって得られた人間の修正ストラテジーの一部である，助詞の修正，フィルタリングによる修正，キーワードによる修正を実現し評価した．しかしながらそのほかに以下のようなストラテジーも意味理解率の向上に有効であると考えている．

- 音韻情報の利用．
  音声認識部が未知語として音節系列を出力する場合があるが，入力単語の認識部での結果が不確かなために入力を棄却しその部分を未知語として出力した場合と，本当に音声認識部にとって未知語であった場合の 2 通りがある．我々のシステムの音声認識部の辞書と文法は，その認識結果に対して理解する言語理解部の辞書と文法とは異なるものを使っている．このため音声認識部では未知語である単語でも，言語理解部では未知語でないものもありうる (言語理解部の方が単語辞書を大きくしておく場合)．このため，音声理解部で未知語となったものは言語理解部で復元できる可能性がある．そのため未知語，間投詞として認識された音韻や誤認識された単語の音韻情報を利用して元の単語に復元する．これによって重要な名詞，動詞，疑問詞等を復元できる可能性がある．
- 誤りやすい誤認識パターンの利用．
  よく間違えられる誤認識のパターンをシステムに与える．これによって上記の場合と同様に重要な名詞，動詞，疑問詞などの復元の可能性がある．
- 動的文脈の利用．
  現在のシステムは静的な文脈のみを利用しているが，訂正実験では人間は以前になされた発話も非常に重要視していることが確認されている．これ

らの動的文脈の使用も考えている.

- N-best の利用[11].
  今は認識結果は最適と考えられるもののみを対象にしているが，いくつかの認識候補結果である N-best を利用することもロバストなシステムには有用であると思われる.

7. むすび

本論文では，人間は誤認識された音声認識結果をどのくらい原文と意味的に正しい文に訂正できるかという実験を行い，その考察を基に開発した頑健な音声対話システムを述べた．単語数241，パープレキシティ74の文法で，音声認識部だけでの平均文認識率は57.4%であったが，認識結果からのエキスパートの復元訂正によって文理解率は87%に向上した．この人間の訂正ストラテジーを用いた自然発話の意味理解システムの開発を行い，その評価実験を行った．語彙数275，パープレキシティ102の文法を用いて，最終的にシステムの利用に関して助詞落ちや言い直し，間投詞を含む初心者の発話した自然言語に対して，文音声認識率52%，意味理解率72%を得た.

今後は6章で述べたような意味理解部の拡張や，ユーザがタスクを達成するためにかかる負担を減らすための協調的な応答生成システム[12]，検索結果を表示したり音声以外の入力手段を持つマルチモーダルな対話システム[13]の開発を行う予定である.

参考文献

1) Dowding, J. et al.: Gemini: A Natural Language System for Spoken-Language Understanding, *Proc. the 31st Annual Meeting of the ACM*, pp.54–61 (1994).
2) Kuhn, R. and Mori, R.De: The Application of Semantic Classification Trees to Natural Language Understanding, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17, No.5, pp.449–460 (May 1995).
3) Jackson, E., Bear, J., Moore, R. and Podlozny, A.: A Template Matcher for Robust NL Interpretation, *Proc. Speech and Natural Language Workshop*, Morgan Kaufmann, pp.190–194 (Feb. 1991).
4) Ward, W. and Young, S.: Flexible Use of Semantic Constraints in Speech Recgnition, *Proc. ICASSP 93*, Vol.II, pp.49–50 (Apr. 1993).
5) 伊藤，大谷，肥田野，山本，中川：事前説明によるシステムへの入力発話の変化と認識結果の人間による復元，情報処理学会，音声言語情報処理研究会，94-SLP-4-7 (1994. 12).
6) 山本，小林，中川：音声対話文における助詞落ち・倒置の分析と解析手法，情報処理学会論文誌，Vol.33, No.11, pp.1322–1330 (1992).
7) 肥田野，伊藤，山本，中川：音声対話システムにおける自然発話の頑健な一理解法，第50回情報処理学会全国大会論文集，7R-7 (1995. 3)
8) 甲斐，間宮，中川：自然発話の認識・理解のための解析・照合手法の比較，情報処理学会，音声言語情報処理研究会，94-SLP-2-12 (1994. 7).
9) 山本，肥田野，伊藤，甲斐，中川：自然発話の意味理解と対話システム，情報処理学会，音声言語情報処理研究会，94-SLP-2-13 (1994. 7).
10) 荒木，河原，西田，堂下：キーワード抽出に基づく意味解析による音声対話システム，電子情報通信学会，音声技法，SP91-94 (1991).
11) 肥田野，中川：音声対話システムにおけるn-best文認識結果の一利用法，第52回情報処理学会全国大会論文集，4D-2 (1996. 3).
12) 伊藤，中川：音声対話システムにおける協調応答，情報処理学会，音声言語情報処理研究会，96-SLP-10-19 (1996. 2).
13) 傳田，中川：日本語音声による観光案内システムのマルチモーダルインターフェイス化，第52回情報処理学会全国大会論文集，4D-3 (1996. 3).

(平成7年9月11日受付)
(平成8年2月 7 日採録)

山本 幹雄（正会員）

昭和61年豊橋技術科学大学大学院情報工学専攻修了．同年（株）沖テクノシステムズラボラトリ入社．昭和63年豊橋技術科学大学情報工学系教務職員．平成4年助手．平成7年筑波大学電子・情報工学系講師．工学博士．音声・言語処理の研究に従事．電子情報通信学会，人工知能学会，日本音響学会，ACL，AAAI各会員.

伊藤 敏彦（正会員）

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

肥田野　勝（正会員）

平年6年豊橋技術科学大学情報工学課程卒業．平成8年同大学大学院情報工学専攻修了．現在伊藤忠テクノサイエンス株式会社勤務．在学中は自然言語処理の研究に従事．

中川　聖一（正会員）

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