音声言語情報処理 2−13
（1994. 7. 8）

自然発話の意味理解と対話システム

山本幹雄　肥田野勝　伊藤敏彦　甲斐充彦　中川聖一

豊橋技術科学大学
〒441愛知県豊橋市天伯町雲雀ヶ丘1−1

自然な発話を許す音声対話システムでは、ユーザの発話を表現する文法が書き言葉と比べてかなり緩くなり、しかも間投詞、言い直しなどの現象も多く生じるため、音声認識率がどうしても低くなる。このような自然な発話を音声認識部だけで対処することは現在のところ無理があるため、ある程度誤認識された文でも正しく意味解析ができる文字レベルの構文・意味解析部が必要である。本報告では、助詞落ち、倒置などの現象を含む自然な発話を理解できるだけでなく、音声認識部で誤認識された文（例えば、助詞の誤認識）にもある程度対応できる構文・意味理解システムと、それを応用した対話システムについて報告する。助詞落ち、助詞誤り、倒置にはいくつかのヒューリスティックスで対応する。また、タスクや場面設定のような文脈的情報も誤りを含む文を理解する場合は重要であるため、これを利用するためのフィルタリングの手法とトップダウン的なキーワードをもとにした意味抽出を用いている。

Spontaneous Speech Understanding and Dialog System

Mikio YAMAMOTO, Masaru HIDANO, Toshihiko ITOH,
Atsuhiko KAI and Seiichi NAKAGAWA

Toyohashi University of Technology
1-1, Tempaku, Toyohashi, Aichi, 441, JAPAN

This paper describes the spoken dialog system for spontaneous speech. It is difficult to recognize and understand spontaneous speech, because spontaneous speech has many phenomena of ambiguity such as ommitions, inversions, repairs and so on. Since there is a trade-off between looseness of linguistic constraints and recognition precision, the recognition rate of speech recognizer is limited. Therefore, the interpretation part must cope with not only spontaneous sentences but illegal sentences with recognition errors. We developed the robust interpretation method and applied it to the dialog system. The interpretation method use some heuristics for ommitions of post-position and inversions and top-down context knowlege.

1. はじめに

ユーザに自然な発話を許す音声対話システムは、これまで音声認識でテスト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。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。さらに、間投詞や言い直しなどの問題によって、認識率はさらに下がる。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[Dowding94]、どこかで妥協するしかない。そこで、音声認識結果を受け取る意味解釈部や応答生成部は、文法的にかなり自由なユーザの発話を処理できるだけでなく、文の一部が誤認識された音声認識システムの認識結果にもある程度対処することが期待される。例えば、かなり厳密に作成された意味文法でも、これからランダムに文を生成すれば、構文的・意味的に不自然になることの方が多い。また、言語モデルとしてbigramやtrigramを用いる音声認識では、やはり不適格な認識結果を出力する場合が多くなると考えられる。

本報告では、音声認識システムの一部に誤認識された文字列を含む認識結果にもある程度対応した、自然な発話を意味理解するシステムとそれを応用した対話システムについて報告する。

第2章では、試作した音声対話システムの概要を述べる。第3章では、本対話システムの構文・意味解析部について述べる。

2. 対話システム

2.1 概要

今回作成した対話システムの概要を図1に示す。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、対話システムに送られる。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、文脈処理を行った後、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。知識を統一してすべての処理（音声認識や言語理解）を同時に行う方法も提案されているが[Mast94]、我々は音声認識と言語理解に必要な知識の使用目的がかなり異なっており、また、実験システムとしては、各部分が独立に動いていた方が各部分の検証には使いやすいと考えている。各部分の概略を以下にまとめる。基本的にはこれまで開発してきたシステムを踏襲している[Yamamoto93]。

2.2 音声認識部

今回使用した音声認識システムは、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対処する。未知語処理では、これらの部分を任意の音節系列による認識尤度スコアを用いる。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未知語が生じると仮定している。自然な音声に対する認識精度は特定話者に適応化したHMMの音節モデルを使って、意味理解率が75%であるが、簡単な助詞の誤りは正解とした場合の理解率であり、これを誤認識とすると、より精度は落ちる。文法をもう少し強めの制約で書けば、これらの誤認識の多くはなくなると思われる。しかし、完全に正しい助詞しか受理しない文法を書くのはきわめて困難であり、構文・意味理解部で助詞の誤りにも対応しなければならないことに変わりはない。また、細かい文法を書くと文法数が多くなってしまい、プログラムの実行に必要なメモリーも膨大になるのも問題である。音声認識部に関する詳しい内容は文献[甲斐他94]を参照されたい。

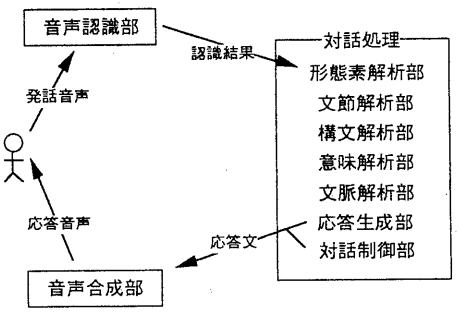

図1　対話システムの構成

2.3 構文・意味・文脈解析部

形態素解析結果をもとに、文節解析と構文の解析を行う。構文の解析は係り受けに基づいた文節間の依存構造を解析する。解析がうまく行かなかったときに、助詞落ちや倒置などを認めた処理で部分解析の結果を使うために、チャートデータ構造[Gazdar89]を利用したものとなっている。さらに、タスクや対話設定による知識を使った解析も行う。これまで、様々な文法的に不適格な文を理解する手法が提案されているが[松本94]、我々の方法はこれらのいくつかを組み合わせたものとなっている。構文・意味解析部は3章で詳しく述べる。

文脈解析部は直前までの対話から得られた文脈情報を意味解析結果に加えることを行う[Yamamoto93]。

2.4 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。図2に構成を示す。応答生成部は問題解決器、知識データベース、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。図3に知識データベースの一部を示す。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。関係ＨＡＳは所有、関係ＮＡＭＥは部分集合の関係、関係ＰＬＡＹはある場所とそこでできることの関係などを意味する。この例では、河口湖にはレイグランドホテル、河口湖ホテルがあり、河口湖ホテルではテニスができることを示している。

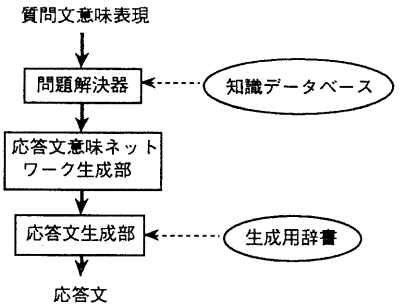

図2　応答文生成部の構成

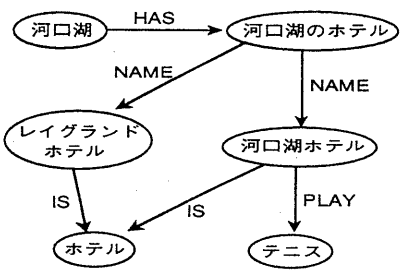

図3　知識データベースの一部

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なう。問題解決器は入力された意味ネットワークの動詞の種類、文の種類、動詞と結ばれているアークの種類、そしてWH型の質問の場合は疑問詞のあるアークの種類を調べる。質問文がYN型の質問の場合は動詞と結ばれている各アークに対して、WH型の質問の場合は疑問詞のかかっていない各アークに対して部分検索を行なう。部分検索は例えば、”テニスのできるホテルはどこにありますか”という質問文における”テニスのできるホテル”の検索である。部分検索は知識データベースと検索用意味ネットワークとのマッチングによって行なわれる。この検索用意味ネットワークは意味ネットワークの形によって決定される。この例の場合、知りたいのはテニスのできるホテルの名前であるから、ノード”ホテル”からＩＳアークを遡ったノード（ホテル名）と、ノード”テニス”から行動概念であるＰＬＡＹアークを遡ったノード（テニスのできる場所）の両方に属するノードを検索結果とするような検索用意味ネットワークが決定される。この場合の検索結果は”河口湖ホテル”となる。次にこれらの部分検索結果を用い、部分検索と同じ過程で検索用意味ネットワークを決定し、最終検索が行なわれる。そこでマッチした検索結果が最終結果として出力される。

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なっている。

応答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。どのような入力パターンに対してどのように応答するかをif-thenルールのような形で書かれた知識に基づいて動作する[Yamamoto93]。

2.5 音声合成部

応答生成部が生成する応答文の文字列をワークステーション上で動いている音声合成サーバに送り、音声を出力している。音声合成は富士通研究所のシステム[小林他94]を使用させていただいており、辞書には富士山観光案内特有の固有名詞の発音を追加している。

3. 自然な発話の音声認識結果に対する構文・意味解析部

3.1 自然な発話の音声認識結果の検討

自然な発話は文法的な自由度が大きく、さらに音響的な曖昧さも加わるため、ユーザの発話した文を正確に認識することはきわめて困難である。たとえ正しく認識された発話であっても、省略、断片文、助詞落ち、倒置などがあるため、これまでの文法的な知識に大きく依存する自然言語処理手法では対処できない。また、正確な音声認識が困難であるため、意味理解部に渡されるユーザの発話文は、誤認識による音声認識特有の非文法性が混入したものもある。

我々は、そのような音声認識結果を機械処理するための検討を行うために、人間がどの程度理解できるかの調査を行った。音声認識結果のうち、完全に認識できなかった文（文字列レベルで発話と完全に一致しなかったもの）を、人間がもとの正しい文に修正できるかを調べた。文脈として、これらの発話は富士山観光・宿泊案内システムへのユーザの発話であることを被験者には伝えてある。１０文づつ５人の被験者で調査した。調査に用いた、ユーザの発話した文（間投詞や言い直しは省いてある）と誤認識結果の例を以下に示す。(a)～(f)は正しく理解された。(g)は元の意味と異なった意味で理解された。(h)は意味不明と判断されたものである。

「富士急ハイランドにはどんなアトラクションがありますか」 =>(a)「富士急ハイランドやどんなアトラクションがありますか」 [助詞誤り]

「河口湖ホテルの料金はいくらかかりますか」
=>(b)「河口湖ホテルの料金()かかりますか」
[内容語の脱落] [助詞落ち]
(c)「河口湖ホテルの料金はいくらかありますか」
[述部の誤認識]

「西湖で宿泊したいんですが」
=> (d)「西湖で宿泊したいんですか」 [文末表現の誤認識]

「民宿に食事とかは付いているんでしょうか」
=> (e)「民宿に食事とかは付ける」
[述部の付属語列の誤認識]

「富士博物館の入館料はいくらなんですか」
=> (f)「富士博物館を入館料いくらの付く」
[重要な内容語以外は誤認識]

「どんなホテルがありますか」
=> (g)「どのホテル()ありますか」
[疑問に関する単語の誤認識] [助詞落ち]

「どんな宿泊施設がありますか、河口湖の周辺には」
=> (h)「ボート、ボートの近くってかかりますかねその九千円には」 [文全体に渡る誤認識]

(a)は助詞「には」が誤認識しているが、全員正しく修正できた。「富士急ハイランド」と「どんなアトラクション」は並列の関係に成りにくいという知識と、「ありますか」との関係で修正できたと思われる。(b)は、「いくら」が脱落しており、しかも誤認識結果は文法的には正しい。しかし、ホテルで料金がかかるのは当たり前であり、yes-or-no型の質問とするとおかしいために、「いくら」を補っている。(c)は最後の動詞がおかしいが、文の前半から容易に「ありますか」を「かかりますか」に修正できる。「料金」、「いくら」、「ありますか」で、理解可能な文を作ることが困難であり、「かかりますか」が推測されるためであろう。(d)は最後の文末表現が誤認識され「が」が「か」に変わったことにより、質問文に変形してしまっている。しかし、宿泊案内システムに「宿泊したい」かどうかを尋ねるのはおかしいため、質問文でなかったと推測している。これは、タスクと発話者がユーザであるという知識を知っているために可能となる修正である。(e)は逆に質問文がただの叙述文になった例である。しかし、

これも、観光案内システムがタスクの知識をユーザに教えるのはありえるが、その逆はほとんどありえないため、質問文への修正ができたと思われる。(f)「富士博物館」、「入館料」、「いくら」以外のほぼすべてが誤認識されている。しかし、これでも人間は正しい意味を推定できる。「いくら」から質問のタイプ、「富士博物館」と「入館料」から質問の対象を推定できると思われる。(g)は助詞「が」が脱落したため、その助詞の推定で人によって個人差が出た。「が」と「に」の曖昧さがある。これを正しく推定するためには、より局所的な文脈（直前の対話など）が必要であると思われる。(h)は重要な内容語が誤認識されているため、もとの文は全く推定できない例である。

以上、まとめると、次のような知識を使って文の修正をしていると言える。（1）文形的な一般的な知識、（2）観光地案内システムというタスク（の発話であるという前提）の知識、（3）ユーザの（問い合わせ）発話であるという前提の知識。これらの知識を使って以下のような処理が人間には可能と思われる。

（1）内容語が正確に認識されている場合は、多少の助詞などの機能語の修正はおおよそ正しくなされている。

（2）文の主動詞とその動詞に係る複数の名詞が正しく認識されている場合は、かなり大きく機能語が誤認識されていても正しい意味をとらえることができる。

（3）意味内容において中心的な役割を果たす単語（内容語）の誤認識があると、誤った意味に理解されるか、意味不明となる。

以下では、上記の知識を用い、人間と同様な処理を行う構文・意味理解システムについて検討する。助詞落ちや助詞誤り(a,b)は3.3節、文末表現や述部の誤認識(c,d,e,g)は3.4節、ボトムアップに意味表現を得ることが出来なかった場合の対応(f)は3.5節でそれぞれ述べる。

3.2 構文・意味理解部の概要

音声認識結果を京都大学で開発されたJUMAN[松本他93]で形態素解析する。接続ルールは自然な発話および誤認識結果にも対応するように書き、富士山観光案内に必要な固有名詞を辞書に追加したものを使っている。形態素解析した結果を文節にまとめ、構文解析する。構文解析は、係り受けに基づく文節間の依存関係を解析するシステムである。解析の途中結果はチャートデータベースに格納され、一度行った部分解析結果を保存するようになっている。構文の解析が成功すれば、その構文をもとに意味表現を生成する。構文・意味理解部の構成を図4に示し、以下にそれぞれの部分の概略を述べる。

文節解析：文節解析の前処理を行った形態素解析の結果を用いて、文節解析される。前処理は、連続する数詞をまとめて1つの数詞とすることなどを行う。文節解析は、「（接頭語）＋自立語＋次の自立語の直前までの付属語の列」を1つの文節とする非常に単純なものある。付属語の列はJUMANの解析で接続可能とされたものに限る。音声認識結果は音節の少ない助詞や接尾辞などに多くの誤認識が含まれるため、あまり厳密な解析はせず、単純な規則でまとめる。

構文解析：文節のそれぞれについて、自立語と付属語の意味から構文解析に必要な情報を、素性と素性値のペアとして集めた後に、チャートデータベース[Gazdar89]に文節に対応する部分解析木として登録し、係り受けをボトムアップに行う。係る文節と、係られる文節の単語情報によって係り受けのチェックを行う。動詞と名詞の解析には、動詞の格フレームと名詞の階層的な意味素性を使っている。意味素性はIPAL辞書の単純な意味素性を使っている。こ

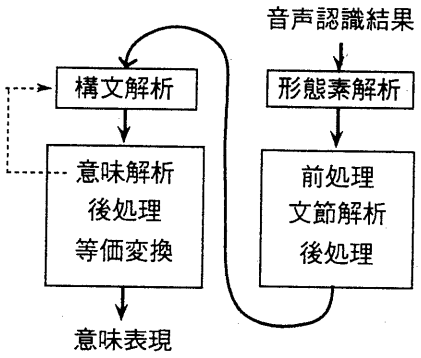

図4　構文・意味解析部の構成

の単純な意味素性でも助詞が省略された際の推定にも、タスクが小さい場合は充分であることを既に確認している[山本他92]。また、助詞落ち、助詞誤り、倒置に対応するために、3.3節で述べるヒューリスティックスを用いている。

意味解析：係り受け解析が成功した場合は、構成性原理を使って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。実際の実現は構文解析木に沿って、各部分木の意味解釈規則を呼び出して行っている。この方法であれば、全体の解析が失敗した場合、部分解析木から同じように部分的な意味表現を得ることができる。意味解析が終了した時点で、意味解析の後処理を行う。後処理は、等価表現を探して標準形に変換することや、応答生成部に便利な情報を付加することを行う。

意味表現は文脈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。フィルタリングは間違っている意味表現をマッチング用パターンで記述し、それにマッチした場合リジェクトする。ただし、修正用の関数が付随している場合は、その関数を適用し、正しい意味表現に修正される。詳しくは3.4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、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3.5節で述べる。

全体の処理は次のような手順で行われる。

（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。
(1-1)助詞落ち、倒置を禁止して解析
(1-2)助詞落ちを許可して解析
(1-3)助詞落ち、倒置を許可して解析
(1-4)助詞の誤りを認めて、倒置を許可して解析

（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)。
(2-1)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。
(2-2)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。

3.3 助詞落ち、助詞誤り、倒置の解析

我々は、小さ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[山本他92]。

助詞落ち用

(1)助詞が省略された名詞文節は最も近くの述部に係る。
(2)述部に係る場合は、必須格を候補として考える。
(3)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結果の１つとして追加する。
(4)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。ただし、これは述部へ係ることができない場合に限る。
(5)助詞の省略された名詞が、次の名詞文節と概念階層上で同じ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

(1)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の候補とする。
(2)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。

これらのヒューリスティックスを用いて以下のような文が解析可能となる。

(例1)助詞落ちの例
河口湖ホテルの料金（は、が）いくらかかりますか。
値段（を）教えて下さい。
(例2)倒置の例
どんな観光地があるんですか、河口湖って。
(例3)助詞落ちと倒置の融合した例
付いてますか、食事。

今回の解析システムは、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。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、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優先する。

3.4 タスクの情報を利用したフィルタリング

3.2節の分析で考察したように、人間は認識結果の文に対して、一般的な構文的知識や文脈的知識を用いて、文が正しいかどうかを判断し、修正できる。これを計算機で行うために、簡単なフィルタリングを実現した。フィルターは、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共に登録してある。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。パターンの記述は変数を含む意味表現として記述され、マッチングはユニフィケーションに似たマッチングアルゴリズム（意味表現はslot-valueのペアであるため、ペアの順序は問わないマッチング）を利用している。パターンの例を図5に示す。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。パターンは意味表現と同じように記述されるが、ノードの値としてリストが許され、その中の値ならどれでもマッチング可能である。例えば、図5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。filter2では、「「ある」という動詞は値段を尋ねるwq型の質問にはならない」という知識を表してる。

```
filter1: (pattern: ((ある かかる) (form assert))
          modify-fun: (change 'form yn-q))
filter2: (pattern: (ある (form wh-q)
                         (cost (wh))))
```

図5　意味表現フィルターの記述例

3.3 キーワード抽出による意味解釈

上記までで述べた、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。これは、キーワードスポッティングによる音声理解の方法[荒木他91][Takebayashi93]や、自然言語処理におけるパターンマッチングによる意味理解のアプローチ[Carbonell87]に近い。図6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6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6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。なお、束縛条件にはandやorの条件も使える。

```
(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6　キーワードパターンの例

4. 対話例

試作した対話システムのドメインは富士山観光案内タスクのサブタスクで、観光地案内と宿泊案内だけを対象としたものである。音声認識部の語彙数は210単語、構文・意味理解部の語彙は約1000単語であるが、発話内容に関係があり、意味解釈に使われる単語は約200である。以下に入力文の例と認識結果を示す。括弧内の文は音声認識された文であり、中括弧内は間投詞、言い直しなどの未知語と認識された結果である。カギ括弧の中は、注釈で、(誤)認識結果の特徴、あるいは利用された処理と知識を表す。

(a)えーと、どこにあるんですか、河口湖ホテル。
({えと}どこにあるんですか、河口湖ホテル)
[助詞落ち][倒置][音声認識部による間投詞の認識]

(b)どんなホテルが河口湖のあたりにありますか
(どんなホテルが河口湖のあたりがあります)
[助詞誤り]

(c)西湖で宿泊したいんですが
(西湖で宿泊したいんですか)
[文末表現の誤認識]

(a)は正しく認識された自由な発話の例である。間投詞は音声認識部で正しく認識され、これを省いて意味解釈を行う。助詞落ちと倒置を処理するためのヒューリスティックスによって正しく理解される。(b)は助詞の誤りを含むが、「が」格が2つあるので、構文解析に失敗し、助詞誤りの仮定がおかれ、正しく解析される（「ホテル」は「が」と「に」の可能性があるが、「あたり」には「に」しか可能性がないので、正しく解析される）。(c)は「したい」という欲求を表す表現が質問になることはないというフィルタリングの知識でリジェクトされた後に、キーワード解析によって、正しく意味解釈される。

以下にすべての文に対して意味解釈に成功した対話例を示す。

USER> えーと、どんな観光地あるんですかね、河口湖って。
(({えと}、どんな観光地ってあるんですか、河口湖で)
[助詞誤り、倒置]
SYSTEM> 河口湖には富士博物館や産屋ヶ崎や河口湖遊園があります。
USER> 富士博物館の入場、入館料はいくらなんですか。
(富士博物館は金はいくらんですか{ま})
[キーワードによる意味抽出]
SYSTEM> 富士博物館の入館料は５００円かかります。
USER> それと、ど、どんなホテルがありますか。
(それじゃ、{えと}どのホテルがありますか)
[言い直しの音声認識システムによる除去]
SYSTEM> 河口湖グランドホテル、河口湖ホテルなどがあります。
USER> 食事は付いているんでしょうか。
(食事とかは付ける)
[フィルタリングによって質問文に変形]

5. おわりに

自然な発話を許す音声対話システムについて報告した。構文・意味解析部では、自然な発話を理解するだけでなく、音声認識部の誤認識結果をもある程度解釈できるようにした。このため、助詞や倒置に関する解析用ヒューリスティックスを、助詞誤りにも適用できるように、さらに、フィルタリングの操作により、タスクの文脈的な制約で受け入れられない文をリジェクト、および修正できる機構を持っている。ボトムアップの解析がすべて失敗した場合は、トップダウン的なキーワードによる意味抽出を行っている。今後はこのシステムの評価と、このシステムを使って自然な発話を理解するために必要な知識の検証を行っていく。

謝辞

形態素解析に京都大学長尾研で開発されたJUMAN,音声合成には富士通研究所の音声合成システムを使用させていただきました。ここに感謝いたします。

参考文献

[荒木他91]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」,信学技報, SP91-94, pp.25-32, (1991).

[甲斐他94]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、音声言語処理研究会報告,(1994.7).

[小林他94]小林、片江、松本、木村、加世田、大山：「高品質日本語テキスト音声合成システムの開発」、情報処理学会第４９会全国大会（平成６年秋）発表予定, (1994).

[松本他93]松本、黒橋、宇津呂、妙木、長尾：日本語形態素解析システムJUMAN使用説明書Version1.0, (1993).

[松本他94]松本、今一：「頑健な自然言語処理の研究動向と課題」、情報処理学会、音声言語情報処理研究会資料、SLP94-2, pp.8-14, (1994).

[山本他92]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).

[Dowding94]J.Dowding et al: "Gemini: a natural language system for spoken-language understanding", Proceedings of the 31st Annual Meeting of the ACM, pp.54-61, (1994).

[Gazdar89]G.Gsxfst snf C.Mellish, "Natural Language Processing in LISP", Addison-Wesley, (1989).

[Mast94]M.Mast et al: "A speech understanding and dialog system with a homogeneous linguistic knowledge base", IEEE Tran. on Pattern Analysis and MachineInteligence, Vol.16, No.2,(1994).

[Yamamoto93]M.Yamamoto et al: "A spoken dialog system with verification and clarification queries", IEICE Trans., Vol.E76-D, No.1, pp.84-94, (1993).

「なぜ音声認識は使われないか・どうすれば使われるか？」討論報告

嵯峨山 茂樹
NTT ヒューマンインタフェース研究所

5 月 20 日の第 1 回音声言語情報処理研究会において、筆者は「なぜ音声認識は使われないか・どうすれば使われるか？」という題目で事前 E-mail 討論内容を基に報告・議論した。この時、参加者の方々を対象に、以下に示す各項目の趣旨を筆者が説明し、直ちに挙手により賛意を表明して頂き、その数 (票数) を数えた。当日その場で集計されたもの (に筆者自身の票も追加した票数) を、各項目ごとに示す。

これらの項目は、事前 E-mail 討論のために筆者が用意した 22 項目 ((a)-(v)) に加えて事前討論内容から抽出した 6 項目 ((A)-(F)) を新たに加えたものである。挙手投票終了時には会場には 83 名の方々がおられたが、調査の途中で来場された方もおられたので、投票参加者数は 80 名強と思われる。

当日会場での挙手投票による調査

◆質問「あなたは、なぜ、これまで音声認識技術が期待するほど使われて来なかったと思いますか？ 以下の選択肢の中に当てはまると思うものがあれば、複数回答でお答え下さい。」

(a) 認識率が低い。音声認識は技術が未熟である。まだ使いものにならない。……36
(b) 自由発話音声認識ができるまでは本格的に使えない。いまの連続音声認識は使いものにならない。…23
(c) 話者の間で認識性能の差が大き過ぎる。認識率が低い話者がいるのでは使えない。……13
(d) 扱える語彙サイズ (語数) がまだ小さ過ぎる。…11
(e) 語彙制約があるのでは使えない。任意の音声・文字変換できなければ不自由。……4
(f) ワードスポッティング技術が肝要。……18
(g) リジェクト力が弱いことが最大の問題。……40
(h) 雑音や発声変形や回線変動などに対するロバストネスが不足。……46
(i) ヒューマンインタフェースが未熟だから使いにくい。……41
(j) うまい対話制御が重要。これがうまくいっていないから使われない。……22
(k) 人は機械に向かって話すのには抵抗がある。音声認識技術は本質的に嫌われる。……7
(l) 音声認識誤りがどのように起こるのか、どう発声すれば避けられるのか分からない。この不透明感がユーザにとって最も辛い。……41
(m) キーボードなどの他手段に比べて入力効率が決して良くない。……17
(n) 音声認識機能とアプリケーションのインタフェースが確立していない。……24
(o) コストの問題。現状では、音声認識技術は高価過ぎる。……11
(p) まだ速度が十分でない。……6
(q) マイクロフォンが口の近くになければ動作しないのでは応用が限られる。……10
(r) 音声認識の応用について知恵が足らない。現在の技術レベルでも工夫すれば使えるはず。……37
(s) 言語処理が問題。音響処理はかなりのレベルに達しているが言語モデルあるいは言語処理 (構文解析、意味理解。状況、知識、社会常識) が遅れている。……18
(t) 音声認識への要求条件が厳し過ぎる。もっと育てるつもりで使えるところから使うべきだ。……0
(u) 音声認識を使う人間を訓練する必要がある。キーボードだって一日で使えるようにはならない。……6
(v) 音声認識の使い方の社会的コンセンサスがまだない。……11
(A) 音声研究者が実用化など考えていないからだ。この技術は行ける、という確信がなければ実用化はできない。……22
(B) 提供されるサービスが十分に知的 (intelligent) である必要がある。「対話」が知的であるかではなく「サービス」が知的であることが肝要。……21
(C) 次の 3 つの条件がまだうまく噛み合っていない。
  I. 音声認識理解の性能と機能がある程度のレベル (応用による) に達していること
  II. システムに組み込む人と言う意味でのユーザが容易に組み込める為のサポートツールが用意されていること
  III. 音声認識理解を利用することを前提とした応用技術が開発されていること (例えば、音声認識を前提としたＣＡＩ技術など)
……34
(D) 長年の研究成果があまり開かれたものになっていない。市場を立ちあげるためには、基礎技術以外にも製品やサービスとして完成させるアイデアや技術や根性や資本が不可欠。……7
(E) 音声認識というものは既に人間が非常な夢を描き、理想が先行してしまった技術であるために、なかなか便利であると認めて貰えない。発想の転換が必要。……4
(F) メンタルモデル (システムがユーザにどのように見えているか) とシステムイメージのギャップを埋める手段が確立していない。……36

以上のうち 30 票を超えるほど賛同が多かった項目は (a), (g), (h), (i), (l), (r), (C), (F) であり，これらから、参加者の大勢の問題認識は、

「音声認識に関してまだ基本性能が十分でなく、リジェクト力、雑音や回線変動に対するロバストネスを含めた基本性能がまだ不足である。応用の可能性については楽観的だが、応用については知恵不足。また、応用する上でヒューマンインタフェースとメンタルモデルが現実的問題である。応用開発ツールを用意することも重要。」

とまとめることができるだろう。

当日この調査に御協力下さった参加者の皆様に深く感謝致します。