

事前説明によるシステムへの入力発話の変化と 誤認識結果の人間による復元

伊藤敏彦 大谷耕嗣 肥田野勝 山本幹雄 中川聖一

豊橋技術科学大学 情報工学系
〒441 豊橋市天伯町字雲雀ヶ丘1-1

あらまし 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、音声認識率はどうしても低くなる。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どこかで妥協する必要があり、このため話者の入力文と受理可能な文にはギャップが生じる。また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り、誤認識された入力文にも対処することが意味理解部に要求される。本報告では以上の問題点に関する検討のための基礎データを集めることを目標に行なった3つの実験について述べる。一つ目は音声対話システムを使用するユーザへの事前の説明を変えることによって対話システムへのユーザの入力がどのように変化するかを調べた。二つ目はユーザの入力文数に対して異なり単語数がどのように変化するかを調べ、あるタスクでどの程度の単語数が必要かを検討した。三つ目の実験は音声認識部によって生じた誤認識を人間はどれくらい原文と意味的に同じ文に訂正できるかである。

和文キーワード 自然な発話、音声対話システム、事前説明、誤認識、復元

Effects of a prior explanation on the speaker's utterance and recovery strategies of humans from misrecognition

Toshihiko ITO, Koji OTANI, Masaru HIDANO,
Mikio YAMAMOTO and Seichi NAKAGAWA

Toyohashi University of Technology
Department of Information and Computer Sciences
1-1, Hibirigaoka, Tenpaku-cho, Toyohashi-shi, Aichi, 441, Japan

Abstract It is difficult to recognize and understand spontaneous speech, because spontaneous speech has many phenomena of ambiguity such as omissions, inversions, repairs and so on. Since there is a trade-off between the looseness of linguistic constraints and recognition precision, the recognizer cannot perfectly recognize the completely free speech of the user on the current art of speech recognition. Therefore some problems arise. First problem is that there are gaps between sentences a dialog system can accept and sentences the user wants to say. Second problem is that the semantic analyzer has to understand sentences with misrecognition that human never utters. In this paper, we describe three experiments concerning the problems of spontaneous speech dialog systems and their results. First experiment is about effects of a prior explanation of the system's limit on the speaker's utterance. Second experiment is about the relationship between the number of different words and the number of inputs. Third experiment is about recovery strategies of human to understand correct meanings of misrecognized sentences.

英文 key words Spontaneous speech, speech dialog system, prior explanation, misrecognition, recovery

1 はじめに

最近では、あたかも機械と対話しているように思わせるなりすまし方式 (Wizard of Oz 方式) による自然発話のデータの分析が盛んに行なわれている [Fraser91][上條他 94][中里他 94]。一般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、未知語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。しかしこのため、話者が発話したい文と受理可能な文には大きなギャップが生じる。また認識率の低下によって、音声認識結果には単語の置換、挿入、欠落などの誤認識が増大する。

そこで、話者の発話したい文と対話システムが受理可能な文のギャップをできるだけ減らすことと、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。

本稿では以上の問題点に関する検討のための基礎データを集めることを目標に行なった3つの実験について述べる。一つはユーザの特性を調べる実験で、音声対話システムを使用する被験者への事前の説明を変えることによって対話システムへの入力発話に変化が生じるかどうか、もう一つはユーザの入力文数に対して、異なり単語数がどのように変化するかを調べる実験である。最後は対話システムの意味解釈部の能力の限界を調べるために行なったもので、発話を実際に音声認識システムで認識させ、その結果生じた誤認識を人間はどれくらい原文と意味的に同じ文に復元できるかを調べる実験である。

本報告では、3つの実験の内容と結果、および音声対話システムに应用する場合の検討を報告する。

2 事前説明による話者の発話の変化

2.1 事前説明を変える実験方法

この実験では、被験者への入力文の発話に関する事前の説明 (入力文、発話の条件) を変化させることにより被験者の発話する入力文がどのように変化するかを調べた。これは、対話システムが受理できる文をユーザに発話させるためには、どれくらいの説明や制限をユーザにする必要があるのかを判断するためである。

各被験者には、研究室旅行の計画をするために富士山周辺の観光 (観光地・宿泊施設) について聞くことができる富士山観光案内対話システムを使用すると想定してもらった。入力文に対する制約等の説明後、観光地と宿泊施設に関する質問文等の、対話システムへの入力文を20文前後思いのまま発声してもらい、その音声をDATで記録した。さらにその音声から書き起こしたデータを分析に用いた。

この実験では、以下のように事前の説明を変化させ、被験者の発話の変化を調査した。

被験者共通の事前説明の内容を以下に示す。

- 間投詞の使用の許可。
- 基本的に標準語を使用。
- 対話システムから得られた情報に関する質問も許可。(実際には対話システムから応答は無いので、被験者に対話システムからの応答があったと想定してもらった。)

被験者によって変化させた事前説明の内容を以下に示す。

条件1 共通の事前説明以外の説明は行なわない。

条件2 共通の事前説明+対話システムで使える例文を提示。

条件3-A 共通の事前説明+例文の提示+復文の使用の禁止。

条件3-B 共通の事前説明+例文の提示+システムが受理する単語の公開。

条件4 共通の事前説明+例文の提示+復文の禁止+単語の公開。

(例文の提示の場合、例文は15文提示し、例文と同じ文の使用は禁止する。対話システムが受理する単語の提示の場合、対話システムで使える名詞、動詞の原形、形容詞、疑問詞 (計149単語) を提示し、動詞と名詞に関しては提示された単語から使用してもらった。)

以上のような実験を、本学の学生 (音声認識の研究を行っていない他の研究室に所属) に各条件5人ずつ計20人に協力してもらい、一人あたり20文前後の入力文を発話してもらった。

また比較のため、同一の被験者に、以下のようなパターンで事前説明を変化させて入力文を発話してもらう実験も行なった。

パターン1 条件1 → 条件2 → 条件3-A → 条件4

パターン2 条件1 → 条件2 → 条件3-B → 条件4

ただしこの場合、各条件で5文ずつ計20文の入力文を発話してもらい、被験者は各パターン3人ずつである。

2.2 実験結果

対話システムが受理できた文 (受理率) の変化を調べた実験の結果を表1及び表2に示す。(使用した対話シ

表 1: 事前説明の違いによる受理文の変化

	被験者数	発話文数	受理率
条件 1	11	156	7.1%
条件 2	11	136	5.9%
条件 3 - A	9	129	5.4%
条件 3 - B	8	130	32.3%
条件 4	11	127	52.0%

表 2: 使用単語の詳細

	単語数			異なり単語数	
	総数	平均	未登録	登録	未登録
条件 1	1587	10.2	392(24.7%)	112	213
条件 2	1468	10.8	414(28.2%)	108	214
条件 3 - A	1189	9.2	307(25.8%)	89	177
条件 3 - B	944	7.3	65(6.9%)	149	48
条件 4	908	7.1	35(3.9%)	139	17

表 3: 受理不可能となった原因の分類

	生成規則	未登録単語 A	未登録単語 B
条件 1	15(10.3%)	123(84.8%)	7(4.8%)
条件 2	8(6.3%)	114(89.0%)	6(4.7%)
条件 3 - A	7(5.7%)	103(84.4%)	12(9.8%)
条件 3 - B	46(52.2%)	35(40.0%)	7(8.0%)
条件 4	36(59.0%)	21(34.4%)	4(6.6%)

テムは、単語数 241 単語、倒置、助詞落ちも受理し文節の順序も自由でパープレキシティ約 74 の文法を使用。)。

表 1 に各条件で実験を行なった時に、発話してもらった文数と対話システムが受理できた受理率を示す。発話文数は実際に発話してもらった文の数であり、受理率は発話された文の何パーセントが実際に音声認識システムの文法で受理されたかを示している。

表 2 には各条件で被験者が発話した単語の総数と異なり単語数について示す。表の平均は一文当たりの平均使用単語数を示す。

対話システムが受理できる単語を公開した場合の効果は、結果としてはっきりと表れている。条件 3 - B と条件 4 において未登録単語として出現した単語は、語彙として提示していない形式名詞、形容詞、副詞、接続詞などがほとんどで、残りは被験者の見間違いによるものであった。また条件 3 - B と条件 4 の違いは複文を禁止するかどうかについての違いのみである。また一文当たりの平均単語数もほぼ同じであるし、条件 3 - B で実際に発話してもらった入力文に複文は 130 文中 8 文しか使われていない。しかしながら受理率には 20% 近くの差が出ている。この原因は主に未知語の数が条件 3 - B より条件 4 の方が 30 個も少ないためと考えられる。(全部の登録単語をチェックしなかったため) 未知語の出現頻度は被験者の個人差によるものが大きいので実際の受理率の差はもっと少ないとも考えられる。複文を禁止することにより、一文当たりの平均使用単語数は約 1 単語程度減少していることから、発話文に影響を与えていることはわかるが受理率には現れなかった。

また表から条件 1、つまり条件をなにも指定しない場合の方が、例文だけを提示した場合や例文の提示と複文の禁止を指定した場合より良い結果が得られていることがわかる。この原因として例文を提示する際に例文と全く同じ文は作らないようにという条件を指定したために、例文と似た文を使用しにくくなったからではないかと考えられ、実験の再検討が必要である。しかしながら例文の提示には、被験者の入力文を考える時間の軽減の効果がはっきり見られた。

表 3 に、対話システムが被験者の発話した文を受理できなかった場合の原因を示す。表の「生成規則」は、発話された文が生成規則の不完全が原因で受理できなかった文の数である。「未登録単語 A」は単語の未登録が原因で受理できなかった文の数であり、「未登録単語 B」には、今回の実験で使った文法内の単語に似た単語のカテゴリーが存在し、そのカテゴリーにその未登録単語を登録すれば受理できる文になる単語数である。この表から、当然であるが語彙を見せない場合は、受理できない原因は未登録単語にあり、見せた場合は文法の原因が増加する。見せた場合に文法の原因の数が絶対的にも増加しているのは単語に制限があるとその単語をいろいろなバリエーションで使うためであると考えられる。また、語彙を提示しない場合の未登録単語の 8 割近くは名詞と動詞であった。

条件 3 - A と条件 1、条件 2 との比較から複文を禁止することによって、原因が未登録単語 B である割合が上昇している。このことから複文の禁止をすることによる効果が多少みられる。

同一の被験者に条件を段々厳しくさせながら入力文を発話してもらったパターン 1、2 の実験では、条件を厳しくする毎に一文当たりの使用単語数が減る傾向があることが分かった。対話システムの受理率に関しては、前述の結果とほぼ同様に対話システムで使用している単語を提示する条件以外はこれといった特別な効果はみられなかった。

また、今回の実験では助詞落ちや倒置については、事前説明で特に指定はしなかったが、全 678 文中助詞落ちが 27 文、倒置が 4 文あった。

以上の結果より、対話システムへの入力条件を段々厳しくすることによって 1 文当たりの平均使用単語数は減少していくが、音声認識システムの受理できた文の割合自体は音声認識システムで受理できる単語の提示以外では、あまり効果はなかった。これは今回の使用している辞書が 241 単語という非常に小さいものであるため、単語数の多い辞書を使用すれば例文の提示や複文の禁止などの条件だけでも対話システムが受理できる文の割合は上昇していくと考えられる。しかしながら、その割合はそれほど期待できるものではないかもしれない。

3 単語カバー率

この実験では、富士山観光案内対話システムでユーザが自由に発話する文を受理するためには、どれくらいの数の単語が必要か調べた。タスクが扱う対象は2章で述べた実験より広い。方法としては、同記事前説明で対話システムへ入力する入力文を被験者に考えてもらい、入力数に対する対話システムの未登録単語の種類の増加傾向について調査した。具体的には富士山周辺の観光に関することを聞くことのできる富士山観光案内システムを使用すると想定してもらい、対話システムに入力する入力文を考えてもらい、紙に記述してもらった。その際、事前の説明は以下のような内容である。

- 口語体で記述。
- 標準語を使用。
- 関投詞の使用を禁止。
- 助詞落ちや倒置については通常使用する範囲内。

以上のような条件で44人(約20文/人)に入力文を考えてもらった。100文を1セットとして、セットを増やす毎に登録及び未登録単語の異なり数がどのように変化するかについて調査した結果を表4に示す。

表4: 登録単語を増やした時の異なり単語数の変化

セット	登録単語		未登録単語		カバー率(%)
	異なり	総数	異なり	総数	
1	-	-	266	821	-
2	108	661	135	173	44.4/79.3
3	113	646	77	97	59.5/86.9
4	127	826	72	86	63.8/89.6
5	143	738	74	89	66.1/89.2
6	143	782	72	85	66.5/90.0
7	148	733	64	72	69.8/91.0

表4の登録単語は例えばセット2を調査する時はセット1までに出現した単語を辞書に登録し、その辞書を用いた場合に登録されていた単語の総数と異なりである。カバー率の左の値は、異なり単語をどれだけカバーしているかを示し、右の値は頻度まで考慮した出現単語をどれだけカバーしているかを示している。これを見れば明らかだがカバー率は徐々に上がってはいるが600文ではまだ未登録単語の飽和を確認するには至らなかった。図1と図2にタスクやシステムによる単語カバー率の変化を示す。図の横軸は辞書登録に用いた文の数である。カバー率が下がっている場合もあるが、これはカバー率の評価に用いた文に依存するためである。図中のFuji1は今回の実験で使用した文(8.1単語/文)の結果である。Fuji2はWizard of Oz方式で収集した富士周辺の宿泊施設案内システムへの入力文(2.9単語/文)の結果であり、AirlineはWizard of Oz方式で収集した航空座席予約システムへの入力文(2.7単語/文)の結果である。またFuji2とAirlineはシステム主導型システムへの入力文である[山本(誠)他94]。システム主導型のシステム

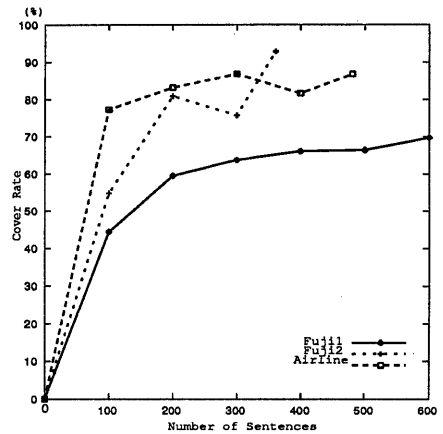


図1: 異なり単語カバー率の変化

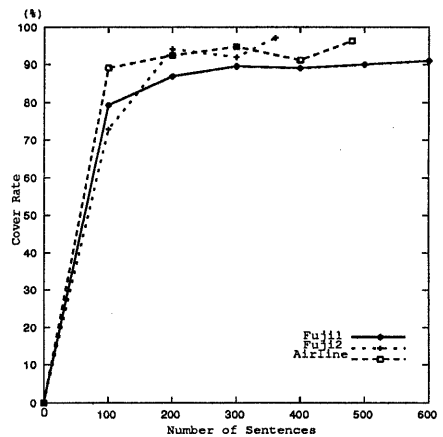


図2: 単語カバー率の変化(出現頻度考慮)

への入力文は、今回の実験で収集した文と比べ一文当たりの単語数が少ないことから新たに登録される単語数も少ないのに関わらず、単語カバー率、異なり単語カバー率ともに高い。理由としてタスクが小さいことと、システム主導型であるために使用される単語がかなり制限されるためと考えられる。またタスクによる変化はあまり見られなかった。今回の実験で収集した文はユーザー主導型システムへの入力文と考えられたため、ユーザー主導型のシステムではかなりの単語が必要なが分かる。

4 誤認識結果の人間による復元

この章では、いろいろな誤認識を含んだ音声認識結果を被験者へ提示し、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を調べる実験について報告する。

4.1 音声認識システム

今回認識結果を得るために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処理する。未知語処理では、これらの部分を任意の音韻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。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[甲斐他94]を参照されたい。

4.2 復元実験

実験には、富士山周辺の観光地、宿泊施設に関する115文を使用した。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(10文)、未知語を含む文(5文)、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(5文)から構成され、全く同じ文は含んでいない。未知語としては名詞、動詞、助詞、終助詞などを使用している。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なった。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3度(不特定者モデル・メルケプストラム・回帰係数、話者適応モデル・メルケプストラム・継続時間制御、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)、115文の認識を行っている。平均認識率は、それぞれ40.5%、52.6%、57.4%であった。そのようにして得られた1文あたり6つの認識結果(合計690文)から実験で使用する115文の認識結果を選択した。選択方法はよく生じる間違い、未知語処理が行なわれたものを優先的に、しかし全体的にはバラエティーに富んだものになるように選択した。115文中100文はなんらかの誤認識を含んでいる。使用した認識結果の内訳と、認識結果が原文と意味的に全く同じでない80文の原因の内訳を表5に示す(著者らの主観で分類)。表の受理可能文とは、倒置文、間投詞、助詞落ち、言い直しを含むが文法で受理可能な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、生成規則外文、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3(図4も参照)に示す。認識システムから得られた情報は全て被験者に提示し、被験者に分かり易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。また今回、語彙の提示の有無と文脈の知識の有無による訂正率の変化も見ると同時に、使用する115文の音声認識結果を全くのランダムに並

べたものと、音声認識結果を対話形式(20対話)で並べた2種類を用意した。

表5: 実験に使用した115文の認識結果の内訳

原文と認識結果の比較	受理可能文	受理不可能文
意味同等な文に認識	35文	0文
同じ意味でない文に認識	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
動詞の置換	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、りよ…、民宿はありますか。
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3: 認識結果の例

復元実験の被験者の人数は6人で、実験を行なう前に実験の目的、訂正の方法、認識結果の見方、入力文の内容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が認識できる241単語の一覧表も渡した。最初に文脈知識の無い場合の実験として、例文115文を1文ずつ全くランダムに提示し、復元訂正してもらった。次に文脈知識のある場合として、対話形式に並べ替えた認識結果を同じく1文ずつ提示し、復元訂正してもらった。この時、1文を訂正する度に訂正した文の原文とそれに対するシステムの応答を提示し次の文の訂正を始めてもらった。実験終了後、被験者の訂正文から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。

4.3 実験結果

図4に、被験者が訂正に成功した文をいくつか示す。例1の原文は定型文である。認識結果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、言い直しを含んでいる文である。認識結果は動詞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「山中湖ホテルを近くしたいんですが」と考えた。次にその文では意味がおかしいので動詞を「近く」から「宿泊」に変更し訂正に成功している。次に例文3は未知語「ペンションマリエ」を含んだ文である。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」に変更している。

語彙の提示、文脈の知識の有無による訂正成功率の変化の結果を表6に示す。表の受理可は、倒置文、間投詞、助詞落ち、言い直しを含むが文法で受理可能な600文(100文*6人)、受理不可は未知語を含んだ文、生成規則外文90文(15文*6人)を評価した値である。

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、動詞、疑問詞を変更、挿入する場合である。しかしながら、今回の実験の結果では名詞、動詞、疑問詞の変更、挿入が正しく修正された文は全体に比べてかなり少ない。つまり、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、動詞、疑問詞の修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の文自体、又は文脈から変更、挿入されるべき単語が連想できるものが多く、語彙の提示の有無による復元率の差が表れなかったと考えられる。他の理由として241個の単語一覧表を修正を行なう度に参照することは非常に困難であるため活用されなかったとも考えられる。

表6: 文脈知識と語彙提示の有無による訂正成功文数 (括弧内は訂正成功率)

	文脈知識なし		文脈知識有り	
	受理可	受理不可	受理可	受理不可
語彙	185文	15文	217文	27文
非提示	(30.8%)	(16.7%)	(36.2%)	(30.0%)
語彙	189文	10文	203文	25文
提示	(31.5%)	(11.1%)	(33.8%)	(27.8%)

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が有る方がかなり良くなっており、平均で文脈知識の無い場合より10%上がっている。特に未知語を含んだ文、生成規則外の文に対して文脈の知識の効果

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4: 認識結果の訂正例

が良く表れている。

次に復元成功文の内訳を表7に示す。音声認識によって原文とは意味的に異なる文に誤認識された480文の、文脈知識なしで41%、文脈知識有りで55%が原文と意味同等な文に復元訂正されている。

次に音韻情報、つまり未知語、間投詞などに認識された音節列を利用することにより原文と相違な文を原文と意味同等な文に訂正できた文数と、文法的には正しいが原文と意味が異なる文の訂正に成功した文数を表8に示す。音韻情報を利用しての訂正に未知語を含む文や生成規則外文である受理不可文の文脈知識なしで20%、文脈知識ありで43%が成功している。これは未知語や生成規則外の文はその部分が音節系列として音韻情報が出現しやすいためであるとはいえ良い結果と言える。特に文脈知識ありの場合、音韻情報を利用して訂正を行なった文の81%が原文と意味同等な文に訂正されている。音韻情報を利用して訂正に成功した文を分析してみると、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、動詞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。このことから名詞、疑問詞、動詞が文の理解に重要な要素であることがわかる。他に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、宿泊施設などの名詞と、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場合、うまく訂正できたものもいくつかあった。

表 7: 訂正後の文の内訳
(a) 修正後、原文と意味同等な文 (修正成功)

分類	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
意味同等 文の修正	115 文 (30.7%)	0 文 (0%)	98 文 (23.3%)	0 文 (0%)
意味同等文 の不修正	88 文 (23.5%)	0 文 (0%)	108 文 (25.7%)	0 文 (0%)
意味が異なる 文の修正	171 文 (45.7%)	25 文 (100%)	214 文 (51.0%)	52 文 (100%)
合計	374 文	25 文	420 文	52 文

(b) 修正後、原文と意味が異なる文 (修正失敗)

分類	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
意味同等 文の修正	6 文 (2.2%)	0 文 (0%)	4 文 (2.2%)	0 文 (0%)
意味が異なる 文の修正	148 文 (65.5%)	41 文 (63.1%)	109 文 (60.6%)	22 文 (57.9%)
意味が異なる 文の不修正	45 文 (20.0%)	11 文 (16.9%)	47 文 (26.1%)	10 文 (26.3%)
復元を断念	27 文 (12.0%)	13 文 (20.0%)	20 文 (11.1%)	6 文 (15.8%)
合計	226 文	65 文	180 文	38 文

次に認識結果が文法的には正しいが原文とは意味が異なる文になった場合、どれだけ訂正できたかを調べた。正しく訂正できたのは、文脈知識なしで20%、文法ありで34%程度であった。しかしながら、訂正できた文のほとんど全てが終助詞の置換と欠落によるもので、終助詞以外の誤認識に関しては訂正は大変難しいことが分かる。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、動詞の置換などであるが、文脈知識のない場合、被験者はこれらの文の修正をほとんど行っていないかった。

表9に原文と意味的に異なった原因の誤認識部分をどれだけ修正できたかを示す。各欄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。まず助詞に関してだが助詞の置換、欠落はおおむね訂正できていることがわかる。助詞の混入についてはあまり訂正に成功していないが、これは助詞が単独で混入された場合より他の誤認識、例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。次に終助詞に関してだが、文脈知識の有無による差が大きく表れている。特に終助詞の欠落に関しては文脈知識ありの方がかなりうまく修正できている。また終助詞の訂正率に関して、被験者個々の結果を見てみると、文脈知識の有無によって大きく差が出ているものと全く差の出ないものがある。そこで被験者の訂正してもらった結果を分析してみると、文脈知識の有無によって差が大きい被験者は、文脈知識なしの場合に失敗した文のほとんどが、例えば「サイクリングをしたいんですが」の終助詞の部分の「が」

表 8: 音韻情報の利用と文法的に正しい文の訂正
(a) 音韻情報を利用し訂正

分類	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
訂正後、原文と 意味同等な文	25 文 (73.5%)	18 文 (58.1%)	29 文 (67.4%)	39 文 (81.3%)
訂正後、原文と 意味が異なる文	9 文 (26.5%)	13 文 (41.9%)	14 文 (32.6%)	9 文 (18.8%)
合計	34 文	31 文	43 文	48 文

(b) 文法的で意味の異なる文の訂正

分類	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
訂正後、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%)
訂正後、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 9: 誤認識箇所での修正成功内訳

原因	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
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
動詞の置換	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

が「か」に置換され「サイクリングしたいんですか」と疑問文に変わっていたものであった。今回の実験では、認識結果は富士山観光案内システムへの入力文であるからこのような文をシステムへ入力することは不自然である。そのことに気づいた被験者と気づかなかった被験者で復元成功文数に差がでたと考えられる。最後に名詞、動詞、疑問詞に関しては、文脈知識の有無による差は表れているが、全般的に修正を必要とする誤認識に対して修正に成功した数がかなり少ない。

5 音声対話システムへの応用

この章では、前章までの実験のまとめと対話システムを構築する場合の検討を述べる。実験の結果を以下にまとめる。

- 音声認識システムの辞書のリストの提示により受理率は大幅に上昇。

- 発話に対する制限の増加に伴い、1文当たりの単語数は減少。
- 600文では、未登録単語の飽和には至らなかった。
- 文脈知識の利用により、文の訂正成功率は上昇。
- 助詞、終助詞の訂正は比較的容易。
- 自立語の誤認識の修正は困難。
- 未知語処理による音韻情報も訂正の助けになる。

対話システムが使用できる単語を提示しても、対話システムが受理できた文の割合は52%であり、また受理できなかった原因の約60%は生成規則の無いものであった。このことから、富士山観光案内というある程度小さいタスクにおいても人間の発話する文のバリエーションの多さ、その文法の作成の難しさが分かる。おそらく、ユーザに自由な発話を許す場合、数千以上の単語が必要であると考えられる。そこで、対話システムでユーザの発話をほとんど受理するようにするには、システム主導型の対話システムを構築するか、ユーザに対してかなり厳しい入力制限、特にシステムで使用できる単語の提示(主に名詞と動詞)を行ないその単語の範囲内で発話してもらうか、数千語の単語の文法を持つことである。使用できる単語の提示という考え方は魅力的であるが[山本(幹)他94a]、単語を提示し単語の範囲内で発話してもらうことは、ユーザに負担が大きい上聞きたい事が聞けないような問題が起こる場合が多い。単語数の多い文法の使用は、自由な発話をユーザに許すことができるが、文法の作成の難しさや、その文法で正しく音声認識できるかという問題がある。一つの解決法は比較的制約のある文法と制約のゆるい文法(例えば、bigram)の併用法であると考えている。

また、誤認識訂正実験より人間は名詞、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た。つまり名詞、動詞、疑問詞などの自立語の訂正は難しいと考えられるから、音声対話システムの音声認識部は、これらを正確に認識することが重要である。名詞や動詞を正確に認識するために、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]を併用することも考えられる。逆に助詞の誤認識、終助詞の誤認識の修正はそれほど難しいことではないと思われる。実際、我々は小さなタスクでは、助詞落ち、助詞誤りを解析可能とする手法を提案している[山本(幹)他94b]。また名詞、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いくらかは可能になると思われる。特に未知語や間投詞として誤認識された重要な名詞、動詞、疑問詞の音韻情報を利用した誤認識の修正はかなり効果があるのではないかと考える。

今後はこれらの結果や被験者からの訂正戦略に関するインタビューから対話システムに应用できる音声認識の誤認識訂正アルゴリズムの開発、実現を行なっていきたいと考えている。我々は、この実験で得られた結果が

対話システムの訂正の限界ではないと考えている。実験の被験者は音声対話システムに関しての特別な知識などは持っていない。仮に誤認識しやすい単語のパターンなどの知識を持っていれば復元率はより上昇したと思われる。これらの知識の音声対話システムの応用も考えている。

6 むすび

3つの実験、事前説明によってシステムへのユーザの発話がどのように変化するかについての実験、ユーザの入力数に対して異なり単語がどの程度増加していくかについての実験、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なった。単語数241、パープレキシティ74の文法で、音声認識部だけの平均文認識率は不特定者モデル・メルケプストラム・回帰係数、話者適応モデル・メルケプストラム・継続時間制御、話者適応モデル・メルケプストラム・回帰係数・継続時間制御でそれぞれ40.5%、52.6%、57.4%であったが、認識結果からの復元訂正成功まで加えると文理解率はそれぞれ74%、78%、80%に向上する。今後はこの人間の訂正戦略をシステムに組み込む予定である。

参考文献

- [Fraser91] N.M.Fraser and G.N.Gilbert:Simulating speech system,Computer Speech and Language,vol.5,pp.81-99(1991)
- [上條他94] 上條、秋葉、伊藤、田中「音声対話データの分析と発話理解の応用」、情報処理学会、音声言語情報処理研究会報告、94-SLP-3-6(1994.10)
- [中里他94] 中里、亀山、白井「音声対話システム使用時のユーザ発話の分析」、人工知能学会、全国大会論文集,pp.93-96,(1994)
- [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)
- [山本(誠)他94] 山本(誠)、山本(幹)、中川「音声による対話システムの評価法における一考察」、情報処理学会、全国大会論文集, pp.61-62(1994)
- [山本(幹)他94a] 山本(幹)、高木、中川「メニューに基づく音声対話システムとその評価」、電子情報通信学会、音声技報、SP93-130(1994)
- [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、音声言語情報処理研究会報告、94-SLP-2-12(1994.7)
- [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」、信学技法、SP91-94,pp.25-32,(1991).
- [山本(幹)他94b] 山本(幹)、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報告、94-SLP-2-13(1994.7)