

TR-IT-0199

対話処理部と音声認識部を統合する
ための予備実験とその分析

A Preliminary Study for Integrating
a Dialogue Processing Module with a Speech Recognizer

竹澤 寿幸 巖寺 俊哲 荒川 直哉 加藤 直人

Toshiyuki TAKEZAWA Toshiaki IWADERA Naoya ARAKAWA Naoto KATOH

1996. 10

内容梗概

対話処理部で用意できる発話タイプ情報と話題の情報を用いて、対話音声認識結果が実際にどのように扱えるかに関する予備実験を行なった。今後の研究を進める上で貴重な資料となるので、その分析結果の内容を報告する。

ATR 音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

© 株式会社 エイ・ティ・アール音声翻訳通信研究所

© 1996 by ATR Interpreting Telecommunications Research Laboratories

目次

1	まえがき	1
1.1	目的と背景	1
1.2	目視により分析する対象の対話音声認識実験結果ファイル	1
1.3	発話タイプの付与された対話の書き起こしファイル	3
2	文末述語に関して尤もらしい認識候補を選ぶ手法	5
2.1	準備	5
2.1.1	表層表現を用いた表層発話行為の認定・ラベル付与	5
2.1.2	文中要素が発話行為の決定に寄与する発話の取り扱い	6
2.2	文末述語に関して尤もらしい認識候補を選ぶ手法の概要	7
2.3	見い出せたヒューリスティックス事例	8
2.3.1	表層表現パターン	9
2.3.2	分類方法	9
2.3.3	C類に対するヒューリスティックス	10
2.3.4	A類に対するヒューリスティックス	11
2.3.4.1	主体尊敬表現の形式	11
2.3.4.2	ドメインに依存したヒューリスティックス	11
2.3.5	A類/B類決定のヒューリスティックス	12
2.3.6	特殊な場合(1): [そう で す] クラス	17
2.3.7	特殊な場合(2): [* お 願 い し ま す] クラス	19
2.4	効果の見積り	19
2.4.1	適用実験	19
2.4.2	正解を特定できない事例	21
2.4.3	誤る事例	23
3	話題情報を用いて内容語を制御する指針	27
3.1	ねらいと基本的な考え方	27
3.2	枠組	28
3.2.1	対話の構造化	28
3.2.2	話題の認識	31
3.3	話題情報を用いて内容語を制御する指針の概要	31
3.3.1	用語	31
3.3.2	アプローチ	32
3.3.3	TOPIC 情報を用いた語彙制御の性質	32

3.3.4	制御の持続期間	33
3.4	<n-spel> の調査と制御の指針	34
3.4.1	基礎調査	34
3.4.2	<n-spel> の制御の指針	36
3.5	<n-num-ex> の調査と制御の指針	36
3.5.1	基礎調査	36
3.5.2	<n-num-ex> の制御の指針	38
3.6	要約と今後の課題	39
4	音声認識に対する対話処理部の貢献度の見積り	41
	謝辞	41
	参考文献	43
A	発話タイプ (ACT ラベル)	45
B	構造化された対話例 (TAS12009)	49

第 1 章

まえがき

1.1 目的と背景

対話処理部で用意できる発話タイプ情報と話題の情報を用いて、対話音声認識結果が実際にどのように扱えるかに関する予備実験を行なった。今後の研究を進める上で貴重な資料となるので、その分析結果の内容を報告する。

具体的には、対話処理部と音声認識部の統合に関して、現在次の二点を検討している。

(1) 文末述語に関して尤もらしい認識候補を選ぶ手法

(2) 話題情報を用いて内容語を制御する指針

今回はその予備実験として、部分木に基づく構文解析規則とポーズ情報を用いた対話音声認識手法 [竹澤他 96] による実験結果¹を使って、対話処理部の発話タイプと話題の情報 [巖寺他 94] がどの程度役に立つのかに関する評価を、人手で目視により確認および分析した。本報告書ではその概要を述べる。

1.2 目視により分析する対象の対話音声認識実験結果ファイル

目視により分析する対象の対話音声認識実験結果ファイルを表 1.1 に示す。音声認識の対象となるのは 5 対話 (TAC23034, TAS12009, TAS22001, TAS33002, TCC22074) であるが、ビーム幅 (B500, B1000, B2000, B5000) の違い、および文法の枠組 (ag, al) の違いにより、計 40 ファイルが得られる。この 40 ファイルが本報告書の対象となる。なお、上記の各ファイルについて言及する時は、「TAC23034.B1000.ag」のように略して示す。

¹この実験では部分木に基づく文法のみを使っている。前終端記号バイグラムは併用していない。

表 1.1: 対話音声認識実験結果ファイル

TAC23034.FMAKA.B1000.cfg-ag-2-1.sss
TAC23034.FMAKA.B1000.cfg-al-2-1.sss
TAC23034.FMAKA.B2000.cfg-ag-2-1.sss
TAC23034.FMAKA.B2000.cfg-al-2-1.sss
TAC23034.FMAKA.B500.cfg-ag-2-1.sss
TAC23034.FMAKA.B500.cfg-al-2-1.sss
TAC23034.FMAKA.B5000.cfg-ag-2-1.sss
TAC23034.FMAKA.B5000.cfg-al-2-1.sss
TAS12009.FYUYO.B1000.cfg-ag-2-1.sss
TAS12009.FYUYO.B1000.cfg-al-2-1.sss
TAS12009.FYUYO.B2000.cfg-ag-2-1.sss
TAS12009.FYUYO.B2000.cfg-al-2-1.sss
TAS12009.FYUYO.B500.cfg-ag-2-1.sss
TAS12009.FYUYO.B500.cfg-al-2-1.sss
TAS12009.FYUYO.B5000.cfg-ag-2-1.sss
TAS12009.FYUYO.B5000.cfg-al-2-1.sss
TAS22001.FHITA.B1000.cfg-ag-2-1.sss
TAS22001.FHITA.B1000.cfg-al-2-1.sss
TAS22001.FHITA.B2000.cfg-ag-2-1.sss
TAS22001.FHITA.B2000.cfg-al-2-1.sss
TAS22001.FHITA.B500.cfg-ag-2-1.sss
TAS22001.FHITA.B500.cfg-al-2-1.sss
TAS22001.FHITA.B5000.cfg-ag-2-1.sss
TAS22001.FHITA.B5000.cfg-al-2-1.sss
TAS33002.FYOAS.B1000.cfg-ag-2-1.sss
TAS33002.FYOAS.B1000.cfg-al-2-1.sss
TAS33002.FYOAS.B2000.cfg-ag-2-1.sss
TAS33002.FYOAS.B2000.cfg-al-2-1.sss
TAS33002.FYOAS.B500.cfg-ag-2-1.sss
TAS33002.FYOAS.B500.cfg-al-2-1.sss
TAS33002.FYOAS.B5000.cfg-ag-2-1.sss
TAS33002.FYOAS.B5000.cfg-al-2-1.sss
TCC22074.FHITA.B1000.cfg-ag-2-1.sss
TCC22074.FHITA.B1000.cfg-al-2-1.sss
TCC22074.FHITA.B2000.cfg-ag-2-1.sss
TCC22074.FHITA.B2000.cfg-al-2-1.sss
TCC22074.FHITA.B500.cfg-ag-2-1.sss
TCC22074.FHITA.B500.cfg-al-2-1.sss
TCC22074.FHITA.B5000.cfg-ag-2-1.sss
TCC22074.FHITA.B5000.cfg-al-2-1.sss

1.3 発話タイプの付与された対話の書き起こしファイル

対話処理部で利用している発話タイプ (ACT ラベル) を付与した対話の書き起こしファイルも参照した。その対話データは、対話音声認識実験を行なった5対話を含む ATR 音声言語データベース中の 200 対話である。その 200 対話の話題を次に示す。

- a. ホテルの予約・変更・キャンセル
- b. ホテルの紹介
- c. ホテルでのサービス・トラブル
- d. 会議室利用の問い合わせ・申し込み
- e. 飛行機のフライトの予約・変更・解約
- f. バス・列車の切符の問い合わせ
- g. 交通手段の問い合わせ
- h. レンタカーの問い合わせ・申し込み
- i. バスツアー・旅行パックの問い合わせ・申し込み
- j. 演劇・コンサートのチケットの予約・購入
- k. レストランの予約

第 2 章

文末述語に関して尤もらしい認識候補を選ぶ手法

2.1 準備

対話処理部では、「表層表現を用いた表層発話行為の認定・ラベル付与」[巖寺他 94]を行なう。ラベルは、あらかじめ定義した表層表現パターン(主に文末形式)と発話行為との対応テーブルを用いて付与される(詳細については[巖寺他 94]を参照)。そこで、まず、文献[巖寺他 94]の手法で採用している発話タイプ(ACT ラベル)と、その手法について簡単に述べる。

2.1.1 表層表現を用いた表層発話行為の認定・ラベル付与

「表層発話行為の認定」は、対話のインタラクション構造を認識するために、ACT 単位(発話の最小単位)に発話行為の種類を示す ACT ラベルを付与することにより、ACT 単位の遂行する発話行為を認定する手法である(ACT ラベルの種類・定義については付録 A を参照されたい)。ACT ラベルは、あらかじめ定義した表層表現パターンと発話行為との対応テーブルを用いて付与される。このテーブルを ACT 辞書と呼ぶ。

表 2.1: ACT 辞書 (一部)

表層表現パターン	ACT ラベル
はい	(<acknowledge> <alert> <greet> <yes> <accept>)
* か	(<yn-question> <confirmation-question> <inform>)
* ございま す か	(<yn-question> <confirmation-question>)
* ございま す が	(<inform> <offer> <suggest>)
* で す	(<inform>)
* すいません	(<alert> <apology> <thank>)
* た い ん で す	(<desire>)
* どのような * で す か	(<wh-question>)

表 2.1において、表層表現パターン中の「*」は、表現中のその位置に任意の表現を許すことを示している。また、複数の ACT ラベルが記述されているものは、行為の多義があることを示す。なお、ACT 辞書は、形態素情報(形態素の区切り: 「|」で示す)は保持しているが、品詞情報は保持していない。

例えば、「八日|と|九日|の|分|で|ございま|す|か」という発話の認識候補にこの手法を適用すると、表2.2のようになる。

表2.2の例において、表層表現パターンが〔デフォルト〕となっているものは、適用される表層表現パターンがACT辞書に存在しないということを示している。したがって、特定のACTラベルも付与されず、<default>というラベルが付与されることになる。

表 2.2: 音声認識候補への適用例「八日|と|九日|の|分|で|ございま|す|か」

候補 順位	音声認識候補	適用される 表層表現パターン	付与される ACTラベル
1	八日 と 九日 の 分 で ございま す か	〔* ございま す か〕	(<yn-question> <confirmation-question>)
2	八日 と 九日 の 分 で ございま す か が	〔デフォルト〕	(<default>)
3	八日 と お 九日 の 分 へ ございま す と あ っ	〔デフォルト〕	(<default>)
4	八日 と 九日 の 分 で ございま す が	〔* ございま す が〕	(<inform> <offer> <suggest>)
5	八日 と お 九日 の 分 で ございま す と あ っ	〔デフォルト〕	(<default>)
6	八日 と お 九日 の お 分 へ ございま す と あ っ	〔デフォルト〕	(<default>)
7	八日 と お 九日 の お 分 で ございま す と あ っ	〔デフォルト〕	(<default>)
8	八日 と 九日 の 分 で ございま す か か	〔* か〕	(<yn-question> <confirmation-question> <inform>)
9	八日 と 九日 の 分 で ございま す と あ っ	〔デフォルト〕	(<default>)
10	八日 と 九日 の 分 で ございま す が あ っ	〔デフォルト〕	(<default>)
11	八日 と 九日 の 分 で ございま す が あ	〔デフォルト〕	(<default>)
12	八日 と 九日 の 分 で ございま す と あ あ っ	〔デフォルト〕	(<default>)
13	八日 と 九日 の 分 で ございま す と あ あ	〔デフォルト〕	(<default>)

本稿では、発話(認識候補)に適用される表層表現パターンを〔 〕で、ACTラベルを< >で囲んで示す。また、適用されるACT辞書中の表層表現パターンを単に「表層表現パターン」呼ぶ。

2.1.2 文中要素が発話行為の決定に寄与する発話の取り扱い

文献[巖寺他94]の手法を音声認識結果に適用するにあたり、一つ注意すべき点がある。それは表層表現パターンに含まれる文中要素の取り扱いである。

ACT辞書に記載されている表層表現パターンの大部分は文末要素だが、時に文中要素を含む表層表現が存在する。<wh-question>における「*(不定語)* (述語)か」がその典型例である。ここで「不定語」とは、「何」「だれ」「いつ」など不定の対象を指し示す語を言う。品詞としては、副詞、普通名詞、代名詞、連体詞など多岐に渡る。

表 2.3: 文中要素を含む表層表現パターンとACTラベル(一部)

* いかが * ま しよ う か	(<wh-question>)
* いくら * で しよ う か	(<wh-question>)
* いつ * で しよ う か	(<wh-question>)
* どう い う * で す か	(<wh-question>)

このような文中要素は発話行為の認定に不可欠な要素であるが、運用上さまざまな問題が発生するので、今回の調査では除外した。つまり、今回は第一段階として文末表現に注目した。

2.2 文末述語に関して尤もらしい認識候補を選ぶ手法の概要

文末述語の情報を用いて音声認識候補の中から尤もらしい認識候補を選ぶ手法を検討する。まず、前節で述べたようにして、すべての音声認識候補に対して発話タイプを付与する。表 2.4 に例を示す。

表 2.4: 音声認識結果例 (1)

正解: 千 二 百 円 ちょうだい いた し ます	[* ちょうだい いた し ます]	(<action-request> <information-request> <inform>)
1 千 二 百 円 ちょうだい で し ます	[* ます]	(<inform> <offer> <promise>)
2 あ 千 二 百 円 ちょうだい で し ます	[* ます]	(<inform> <offer> <promise>)
3 千 二 百 円 ちょうだい いた し ます	[* ちょうだい いた し ます]	(<action-request> <information-request> <inform>)
4 あ 千 二 百 円 ちょうだい いた し ます	[* ちょうだい いた し ます]	(<action-request> <information-request> <inform>)
5 千 二 百 円 ちょうだい で し ます	[* ます]	(<inform> <offer> <promise>)
6 税 に 百 円 ちょうだい で し ます	[* ます]	(<inform> <offer> <promise>)
7 千 二 百 円 ちょうだい が し ます	[* ます]	(<inform> <offer> <promise>)
8 あ 千 二 百 円 ちょうだい が し ます	[* ます]	(<inform> <offer> <promise>)
9 千 二 百 円 ジョン が いた し ます	[* いた し ます]	(<inform> <promise> <accept> <permission-request>)
10 あ 千 二 百 円 ジョン が いた し ます	[* いた し ます]	(<inform> <promise> <accept> <permission-request>)

(認識候補に適用される表層表現パターンを [] で、ACT ラベルを < > で囲んで示す。表層表現パターン中の「*」は、表現中のその位置に任意の表現を許すことを示している。)

多くの認識結果を分析した結果、次の三つの段階からなる方略によって、正解候補を絞り込んでいくことができそうな感触を得た。これらは、実際に人手で目視により作業した手続きに対応している。

- a. 方略 1: 音声認識候補に適用される表層表現パタンのうち、{終助詞/接続助詞} を除いた部分が最長のものを選ぶ。
- b. 方略 2: 表層表現パターン別に分類し、文脈情報・発話内容から、パターン別のヒューリスティックスを用いて誤認識を識別する。
- c. 方略 3: 残ったものの中から最短の表層表現パターンを選ぶ。

いまの対話音声認識手法では、短い要素である助詞などが過剰につながったり、複数の品詞カテゴリの可能性のある要素が別のカテゴリとして認識されたりといった誤認識が多いため、まず、方略 1 で、{終助詞/接続助詞} 以外の要素で長いものを選ぶ。これは定義されている表層パターンに「*」という記述を許しているため、そのパタンのうちなるべく長いものを優先的に選んで調査するほうが有益であったという背景もある。対話音声認識器の性能や、表層パターンの定義の仕方によっては不要ないし見直してよい段階であるが、とにかく量が多いので、今回はこうして候補を絞り込んでから目視により分析する方針とした。なお、例えば表 2.4 の結果に対してはこの手続きで正解を選ぶことができる。

方略2では、主体尊敬表現の形式・名詞句の「なわ張り」[神尾90]・話者情報(話者の役割; サービス提供者かサービス享受者か)などの概念をもとに、ヒューリスティックスにより言語表現としてあり得ない誤認識を識別する。

例としてまず話し言葉で多用される「です/ございます」系について概要を説明する。

表 2.5: [*|で|す] クラスの分類例

A類 <inform>系	[* で す] [* で す が] [* で す から] [* で す よ]
B類 <question>系	[* で す か] [* で す ね] [* で す な]
C類 題目提示的	[* で す が]

表 2.5のように発話ラベルを分類し、その文(発話)中の直前の名詞句との関係を検査する。話題情報を参照し、名詞句内の名詞の出所を調べることで、より尤もらしいものを選べる場合(例えば、ホテル役の人が「【客の名前】様ですが」とは言わない等)と、タスクドメインの制約からわかるもの(例えば、客が「満室です」とは言わない等)がある。

また別の例として「そうです」系については次のような傾向が見られた。概要を述べる。

〔そう|で|す〕クラスを含む発話の直前にある話し相手による発話が、名詞文以外の<yn-/confirmation-question>である場合、〔そう|で|す|が〕はつながりにくい。

例えば、「ほか|の|お|部屋|は|どう|で|す|か|。」の後には、〔そう|で|す|が〕はつながりにくいので、「そうですね」が優先的に選ばれる。「これはあなたの本ですか」の後には「そうです」「そうですが」がつながってもよいので、「名詞文以外の」と限定している。

最後に、方略2までで正解を特定できない場合には、最短の表層表現パターンを選ぶことで多くの場合正解に至ることが多い(方略3)。残りやすい例としては、〔*|お|願|い|し|ま|す|ね〕、〔*|お|願|い|し|ま|す|よ〕が挙げられるが、この中では〔*|お|願|い|し|ま|す〕というパターンが尤もらしいことが多い。

2.3 見い出せたヒューリスティックス事例

方略1と方略3は「経験的方略」に相当するため、対話音声認識器の性能や定義した発話タイプに依存する面がある。しかし、逆に言えば、「経験的方略」は対話音声認識器や発話タイプの変更があれば再実験しさえすれば良い。むしろ定量的な評価をしやすと考えられる。

一方、方略2は「論理的方略」に相当するため、ある程度の一般性があると考えられる。またそこで見い出せたヒューリスティックスは、そのヒューリスティックスをそのまま何らかの方法でインプリメントするかどうかは別にしても、今後の研究の方向を検討する上で役立つ可能性がある。逆になかなか簡単に定量的な評価がしにくい側面がある。

したがって、見い出せたヒューリスティック事例は、必ずしも体系的に整理できてはいないが、基礎資料として活用価値があると考えられる。そこで、できる限り多くこの場に掲載することとする。もしその詳細に興味のない読者は本節を読み飛ばして良い。

2.3.1 表層表現パターン

[述語] 部分に共通性を持つ(例えば、[* | ま | す] [* | ま | す | か] [* | ま | す | が] …) 表層表現パタンの集合を「クラス」と呼ぶことにする。例えば、[述語] が「ます」で共通している表層表現パタンの集合を「[* | ま | す] クラス」と名付けると、方略 1 を終えた段階で多くの可能性が残っている代表的な表層表現パターンとして、[* | お | 願 | い | し | ま | す] クラス、[* | ご | ざ | い | ま | す] クラス、[* | で | す] クラス、[* | ま | す] クラス、[* | ん | で | す] クラス、[そう | で | す] クラスなどがある。

2.3.2 分類方法

まず、前処理として、残る候補を表層表現パターン別に A 類・B 類・C 類に分類する。ここで A 類とは、[述語] に「が」「から」など(「と」を除く) 接続助詞が付いた表層表現パターン、[終助詞/接続助詞] が付いていない表層表現パターン、さらに終助詞「よ」の付いた表層表現パターンをいう。また、B 類とは、[述語] に「よ」を除く終助詞、または接続助詞「と」が付いた表層表現パターンをいう。大まかに言うと、A 類は <inform> 系、B 類は <question> 系の表層表現パターンである。ただし、A 類/B 類の区別は、あくまで表層表現パターンによる分類であって、ACT ラベルによる分類ではない(接続助詞「と」が付いた表層表現パターンは B 類に属することに注意)。A 類と B 類の区別は、話し相手に関する情報(話し相手の行動/名前など)を表すことができるかどうかの区別である。話し相手に関する情報を述べるには、「…か」「…ね」など質問の形を取るか、「…と」などでマークし仮定のこととして述べるかしなければならない。つまり、B 類の表層表現パターンでなければならない。一方、話し相手に関する情報以外のものを伝達する時には、質問/仮定の形式で伝達することはなく、必ず A 類の表層表現パターンで表す。A 類と B 類の区別はここにある。A 類と B 類は二者択一的であり、必ずどちらか一方だけが選ばれることになる。なお、この区別は[神尾 90] の「情報のなわ張り」の概念を基にしている。大まかに言うと、A 類は「話し手のなわ張り」に属する情報を表現する時、B 類は「聞き手のなわ張り」に属する情報を表現する時に用いられる形式である。

さらに、[* | で | す] クラスの [* | で | す | が]、[* | の | で | す] クラスの [* | の | で | す | が]、[* | ご | ざ | い | ま | す | が] クラスの [* | ご | ざ | い | ま | す | が] は、上のものとは独立に、第三のカテゴリ C 類にも分類することにする。これは、「ですが」等には題目提示的な用法があるためである。

題目提示的な「ですが」の例

ユーザ:	えー ホテル の 電話番号 で す が
ユーザ:	え
ユーザ:	二 零 三 の
ユーザ:	四 四 三 の
ユーザ:	一 七 零 零 で す

〔* | で | す〕クラスを例にとると、表2.5のように分類される。表2.5で、〔* | で | す | が〕がA類とC類の両方に分類されていることに注意されたい。

A類とC類、またはB類とC類も、概念上は二者択一的なのだが、そのうちのどちらかを選ぶには、かなり細かい道具立てが必要になってくるので、C類に関しては、確実に可能性を取り除けるもののみを減らすことにし(この時統語構造を参照する)、それでは可能性を取り除けないものについては、次の方略(方略3)で扱うことにする。

ここでの方略の目的は、C類の統語構造が適格かどうか調べる(不適格ならば可能性を取り除く)と、表層表現パターンがA類とB類のどちらであるかを決定することの二つである。具体的には、やはりヒューリスティックスをいくつか設けて、正解を絞り込んでいくという方法を取る。

この方略の目的から、ヒューリスティックスを用いなくても、分類するだけでこの段階の方略が終了する場合もある。次の例を考えてみよう。

表 2.6: 音声認識結果例(2) (一部)

正解:	鈴木 和子 と い い ま す	〔* と い い ま す〕	<<inform>
1	鈴木 和子 と い い ま す	〔* と い い ま す〕	<<inform>
4	鈴木 和子 と い い ま す が	〔* と い い ま す が〕	<<inform>

〔* | と | い | い | ま | す〕クラスにはC類に分類される表層表現パターンはない。また、上の二つの表層表現パターンは、定義上A類に分類され、B類には分類されない。したがって、この時点で残っているすべての表層表現パターンがA類に分類されることが確定する。このように、C類に関するチェックが終わり、A類あるいはB類のうちいずれか一方に特定できた場合、その時点でこの段階の方略(方略2)は終了し、次の方略(方略3)に移る。そうではない場合は、以下の手順を踏む。

2.3.3 C類に対するヒューリスティックス

まず、C類に対するヒューリスティックスから検討する。C類、つまり題目提示的な用法では、〔* | で | す | が〕 〔* | の | で | す | が〕 〔* | ご | ざ | い | ま | す | が〕の命題部分は必ず名詞句である。そうでなければ、題目提示的な用法にはならない。したがって、その発話の命題部分(言い替えれば、適用される表層表現パターンの前方部分)を調べることで、C類として適格かどうかを判断できる。ここで次のようなヒューリスティックスを立てる。

C類に対するヒューリスティックス

命題部分全体が名詞句ではない場合、C類ではない。

- a. (名詞句以外)+ 「で | す | が」
- b. (名詞句以外)+ 「な」 + 「の | で | す | が」
- c. (名詞句以外)+ 「で」 + 「ご | ざ | い | ま | す | が」

この時点でC類の可能性を取り除けない場合、方略3で処理されることになる。

2.3.4 A 類に対するヒューリスティックス

前述の通り、この時点まで残っている表層表現パターンが、すべて A 類 (あるいは B 類) に分類される場合には、その時点でこの段階の方略は終了するのだが、A 類に属する表層表現パターンと B 類に属する表層表現パターンが認識候補中に混在している時は、何らかのヒューリスティックスを用いて、そのどちらかの可能性を取り除けることが望ましい。

ここで、A 類に適用されるヒューリスティックスを考えてみよう。これは、主体尊敬表現を利用したヒューリスティックスである。

2.3.4.1 主体尊敬表現の形式

a. 「お+(動詞連用形)+になる」

- お|部屋|代|は|三|部屋|とも|そちら|で|お|支払|い|にな|り|ま|す|か|。[*|ま|す|か] (<yn-question>)
- それとも|往復|を|お|買|い|にな|り|ま|す|か|。[*|ま|す|か] (<yn-question>)

b. 「ご+(サ変名詞)+になる」

- で|、|三|名|様|で|ご|利用|にな|り|ま|す|と|、[*|ま|す|と] (<inform>)
- それで|いつ|の|ご|予約|にな|り|ま|す|か|。[*|いつ|*|ま|す|か] (<wh-question>)

c. 「お+(動詞連用形)+だ/です」

- 現在|、|ニューヨークシティホテル|に|お|泊ま|り|で|す|ね|。[*|で|す|ね] (<confirmation-question>)
- なに|か|御|希望|は|お|あ|り|で|す|か|。[*|で|す|か] (<yn-question>)

主体尊敬表現で尊敬の対象となるのは、その発話が表す動作の主体であり、典型的には主語である。ここで我々の考察しているドメインから次のヒューリスティックスを立てる。

2.3.4.2 ドメインに依存したヒューリスティックス

対話の登場人物はシステムとユーザの二人だけである。

上のヒューリスティックスを立てると、主体尊敬表現における尊敬の対象が自動的に決まる。つまり、二人しかいないのならば、話し手自身を尊敬の対象とすることはないのでから、自動的に、もう一人の人物、つまり話し相手が尊敬の対象になるということになる。このように決まれば、主体尊敬表現を含む発話が表す動作は話し相手に関する情報だということにもなるので、そのような発話は A 類で表すことができない。

ドメインに依存したヒューリスティックス (1)

[*|で|す] または [*|ございま|す] クラスが候補として残っていて、かつその直前の語が動詞連用形の時、A 類ではない。

ドメインに依存したヒューリスティックス (2)

[*|ま|す] クラスが候補として残っていて、かつそれらが適用される発話 (認識候補) が「{|動詞連用形/サ変名詞| + になり(ます)」という構造の時、A 類ではない (※注 1)。

(※注1) 「お+(動詞連用形)+になる」という形式で、文末がB類ではない例がATR対話データベース(200対話)中1例だけ存在した。

ユーザ: それと、|トラベラーズチェック|は|使え|ます|で|しよ|う|か|。

[*|ま|す|で|しよ|う|か] (<yn-question>)

システム: はい|、|[はい] (<yes>)

システム: お|使え|にな|り|ま|す|。[*|ま|す] (<inform>)

ただし、規範的には「お|使|い|になれ|ま|す」というべきだと思われる。

なお、上のヒューリスティックスはもちろんいつも正しいわけではないが、第三の人物が登場したとしても、その人物が「ユーザの友達・妻・主人」等、ユーザ側の人物の場合には、ここで述べた方法は依然として有効であることが多い。というのは、システムにとっては、ユーザでもユーザの知人でも、尊敬の対象となることには変わらない、というのが一点、さらに、ユーザの知人のことに関してはシステムよりもユーザの方がよく知っているというケースが多いというのがもう一点である。後者についていうと、そのような場合には、システムは必ずB類の文末構造を使って主体尊敬表現を用いなければならない(※注2)。

(※注2) 上のヒューリスティックスが成り立たないのは、ユーザ側の人物のことを、ユーザよりもシステムの方がよく知っている場合である。例えば、以下の例を参照。

ユーザ1: 私の友人の田中が今日そちらのホテルに泊まる予定なんですが、

ユーザ2: もう来てますか。

システム3: はい、

システム4: お越しになっています。

この場合、「田中がもう来ているかどうか」という情報は、ユーザよりもシステムがよく知っている情報である。したがって、システム4の発話では、主体尊敬表現の文末がB類ではなくA類で表される。

2.3.5 A類/B類決定のヒューリスティックス

次に、システムとユーザの間でやりとりされる情報に着目して、A類かB類かを決定する。ここで「情報」の担い手は、基本的に名詞句である(※注3)。したがって、名詞文を作りうる[*|で|す] [*|ございま|す]クラスがヒューリスティックスの対象となる。[*|で|す] [*|ございま|す]クラスが伝える情報、つまり(例1)のX(名詞)の性質により、候補を絞り込む。

(例1) a. X |です/ですが/ですよ/ですね/ですか/ですと|

b. X で |ございます/ございますが/ございますよ/ございますね/ございますか/ございますと|

(※注3) 前述の通り、ここでは[神尾90]の「情報のなわ張り」の概念を活用している。しかし、神尾のいう「情報」とは基本的に「命題」であるのに対し、ここでは「情報」の担い手を名詞句としていることに注意。これには以下のような理由がある。[神尾90]では、情報(命題)が話し手/聞き手に近くなる条件(例えば、「話し手自身が直接体験によって得た情報」[神尾90] (ibid: 33))を考察しているが、情報(命題)がそれらの条件に当てはまるかどうかを判定するには、かなり複雑な枠組が必要となると思われる。一方、名詞句の場合、それが表す情報が話し手/聞き手のどちらの「なわ張り」に属するかが(命題に比べ)はっきりしているケースが多く、命題の場合よりは「なわ張り」の判定が容易である(もっとも、これにはドメイン/タスク/対話者の役割があらかじめ限定されたものであるからという事情もあろう)。このような理由から、ここでは名詞句が表す情報に限定する。

対話を「情報のやりとり」という視点からみてみよう。対話において情報のやりとりをするからには、当然 *information gap* (対話者の一方はある情報を知っているが、他方はその情報を知らないという状態) が存在すると想定される。そして対話の進行上そのような情報を必要なだけ対話の場に提供することは不可欠である。この時、対話者のどちらがその情報を持っているか、また、情報にはどのような種類があるか、などは基本的にドメインにより決まる。例えば、「ホテルの予約」というドメインならば、名前(ユーザが提供)、滞在期間(同)、部屋の金額(システムが提供)等が必要な情報である。別のドメインでは、また別の種類の情報がやりとりされることになる。

どのような情報がどれだけあるかをドメイン別に調査するのは本稿の目的ではないが、我々の考察している方略に関係することが一つある。それは、情報を〔* | で | す〕〔* | ございま | す〕クラスで述べる場合、話し手が自ら提供する情報か、話し相手から提供された情報かで、文末形式を使い分けなければならないという事実である。すなわち、話し手が〔* | で | す〕〔* | ございま | す〕クラスを使って情報を提供する時はB類を用いずA類を用いる。一方、提供された情報について〔* | で | す〕〔* | ございま | す〕クラスで表現する時には、A類は使わずB類を用いる。以下の例を参照されたい。

- (例2) a. 私はジョン・フィリップス | です/ですが/ですよ | 。
 b. * 私はジョン・フィリップス | ですか/ですね/ですと | 。

- (例3) A: 和食をお願いします。
 B: a. * 和食 | です/ですよ |
 b. 和食 | ですね/ですか/ですと (えー、千二百円になります) |
 c. 和食ですが (えー、千二百円で、えー、少々お時間をいただくことになりましたが)

(例2)では、やりとりされる情報が名前「ジョン・フィリップス」であり、自ら提供する情報である。その情報を〔* | で | す〕クラスで提供する際には、A類の文末形式を用いなければならない。また、(例3)ではやりとりされる情報が「和食」で、これは話し相手から提供された情報である。この場合にはA類ではなくB類の文末形式を用いて表現しなくてはならない。(例3Bc)の「ですが」が適格なのは、A類ではなくC類の「ですが」と解釈されるためである。なお、(例2)(例3)の「です」を「ございます」に変えても、適格性に違いはない。

以上の観察より、次のヒューリスティックを設ける。

A 類 / B 類決定のヒューリスティックス (1)

- a. [* | で | す] [* | ございま | す] の直前の名詞句が、話し手が自ら提供する情報の場合には B 類の可能性を取り除く。
- b. [* | で | す] [* | ございま | す] の直前の名詞句が、話し相手から提供された情報の場合には A 類の可能性を取り除く。

このヒューリスティックスを設けるためには、その前提として「話し手が自ら提供する情報」「話し相手から提供された情報」の判定が行なえる枠組が必要である。どのような情報がやりとりされ、どれだけそのような情報が存在するかをドメイン / タスク別に調査し、それを (音声対話システムの構築に必要となるであろう) 意味処理・談話処理部門に組み込めれば、正しい文末述語の同定に役立つと思われる (※注 4)。

(※注 4) しかし、一対話内で、同じ名詞句が、自ら提供する情報であったり話し相手から提供された情報であったりする場合があるので、問題は単純ではない。

- A1: 朝食は、和食と洋食がございます。
- B2: 和食と洋食ですか
- A3: はい、和食と洋食です
- B4: それでは、和食をお願いします
- B5: 和食でございますね

A の視点から見ると、(A3) (A5) の「和食」の情報のステータスが違う。(A3) の「和食」は自ら提供する情報であり、(A5) の「和食」は話し相手から提供された情報である。

A 類 / B 類決定のヒューリスティックス (1) を次の例で確かめてみよう (先行文脈とともに示す)。

表 2.7: 音声認識結果例 (3)

システム:	お 客 様 、	[* 様]	<<vocative>
システム:	お 名前 を お 願 い いた し ま す 。	[* お 願 い いた し ま す]	<<information-request>
ユーザ:	はい、	[はい]	<<accept>
ユーザ:	ジョン ・ フィリップス と 申 し ま す 。	[* と 申 し ま す]	<<inform>
システム:	ファーストネーム が ジョン 様 で 、 ラストネーム が フィリップス 様 で ございま す ね 。	[* ございま す ね]	<<confirmation-question>
1	ございま す ね	[* ございま す ね]	<<confirmation-question <inform>
25	ございま す が	[* ございま す が]	<<inform> <offer> <suggest>
28	ございま す	[* ございま す]	<<inform>

上の例で、[* | ございま | す] クラスの直前にある名詞句は「フィリップス | 様」である。この対話のドメインは「ホテルの予約」であるので、その名詞句 (ユーザの名前) は、対話上必要な情報である。ユーザの名前は、システムにとっては提供された情報であるから、A 類 / B 類決定のヒューリスティックス (1) b が機能し、A 類の表現である [25] [* | ございま | す | が]、[28] [* | ございま | す] が正しく取り除かれる。

ただし、以下の例は話し相手から提供された情報だけではうまく処理できない。

表 2.8: 音声認識結果例 (4)

ユーザ:	あと二日間滞在を 延長したいんですけど。	[* た い ん で す け れ ど も]	<<desire>>
システム:	はい。	[はい]	<<acknowledge>>
システム:	エイミー・ハリス様ですね。	[* で す ね]	<<confirmation-question>>
システム:	七日までのご予約ですので、	[* で す の で]	<<inform>>
システム:	八日と九日の分でございますか。	[* ご ざ い ま す か]	<<confirmation-question>>
1	八日と九日の分でございますか	[* ご ざ い ま す か]	<<yn-question>> <<confirmation-question>>
4	八日と九日の分でございますが	[* ご ざ い ま す が]	<<inform>> <<offer>> <<suggest>>

この場合、[* | ご | ざ | い | ま | す | か] が伝達(確認)する情報は「八日 | と | 九日 | の | 分」であるが、これを「話し相手から提供された情報」と認識させるには、実際に話し相手(ユーザ)から提供された「二 | 日 | 間」という情報と、その他に年月日の情報(計算)が必要になる。

上の方略とは似ているが、それよりもやや単純な方略がある。上の方略では、情報は対話内で得られる(対話/ドメインによって変わる)ものであった。しかし、名詞によっては、その名詞の語彙特性から、どのような対話でも常にシステムからユーザに提供される情報であるものがある。例えば、「ホテルの予約」にしばしば現れる「満室」という名詞である。

- (例 4) システム: a. あいにく満室で | ございます / ございますが |
b. * あいにく満室で | ございますか / ございますね / ございますと | (※注 5)
- (例 5) ユーザ: a. * 満室 | です / ですよ |
b. 満室 | ですか / ですね / ですと |
c. ? 満室ですが (※注 6)

(※注5) ここでの「ございますね」はACTラベル<confirmation-question>が付与されるものとして解釈されたい。

本稿では、[* | で | す | ね] [* | ございま | す | ね] をB類に分類していることから分かるように、定義上A類に属する(つまり、話し相手に関する情報を伝達しない) [* | で | す | ね] [* | ございま | す | ね] を考慮に入れていない。このような用法をも方略に組み込もうとすると、方略がかなり複雑なものになると予想される。

ちなみに、このような [* | で | す | ね] [* | ございま | す | ね] には以下の二つの用法がある。

- (i) 話し相手に情報を伝える。ACTラベル<inform>が付与される。[* | で | す] [* | ございま | す] とほぼ同じ用法である。しかし、心的な計算が要求される文脈では自然に用いられるが、計算の必要がなく、思い出すのに労力がかからない情報に対しては使いにくい、という特徴を持つ [田窪他96]。

a.	システム:	カード の 期限 は いつ で す か 。 [* いつ で す か] (<wh-question>)
	ユーザ:	[え - と] [え - と] (<temporizer>)
	ユーザ:	来年 の 四月 まで で す ね 。 [* で す ね] (<inform>)
b.	ユーザ:	次 の 列車 は 、 何時 で す か 。 [* 何時 で す か] (<wh-question>)
	システム:	次 の 列車 まで は [デフォルト] (<inform>)
	システム:	. . . [あ] [あ] (<expressive>)
	システム:	四 分 しか な い で す ね 。 [* で す ね] (<inform>)
c.	甲:	あなたの奥さんの名前は?
	乙:	直美です(?ね)。 [田窪他96] (ibid: 37)

この用法を方略に組み込むには、「心的な計算が必要な情報か否か」を判定する枠組が必要になる(なお、(ia) (ib)において、[* | で | す | が] を含む発話の直前が「えーと」(間投詞)「. . . あ」(同)であることは示唆的である)。

(例4b)において「あいにく満室でございますね」が適格であると判断されるのは、満室かどうかを調べるのに心的な計算が要求されたと解釈されるためであろう。

- (ii) 間投用法。間投詞的に用いられる。終助詞「ね」の間投用法の丁寧体である。ACTラベル<inform>が付与される。

a.	システム:	[え -] [え -] (<temporizer>)
	システム:	通常 は で す ね 、 [* で す ね] (<inform>)
	システム:	メニュー の 変更 や 交換 は し な い ん で す が 。 [* ん で す が] (<inform>)
b.	システム:	[え -] 通常 は メニュー の 変更 や 交換 は し な い ん で す が 。 [* ん で す が] (<inform>)
c.	システム:	[え -] 通常 は ね 、 メニュー の 変更 や 交換 は し な い ん で す が 。 [* ん で す が] (<inform>)

(iia) が間投用法の「ですね」の例である。このような間投用法と、文末形式の「ですね」は、以下の点で性質が異なる。まず、間投用法の場合、(iib)のように「ですね」がなくても統語的に適格である。さらに(iic)のように間投用法の「ですね」を「ね」に変えても、丁寧さは失われるものの依然適格である。

(※注6) 「満室」という名詞と題目提示的用法の「ですが」は共起しにくいようだ。「満室」という題目に対して陳述(コメント)する事柄が想定できないからであろうか。

このような名詞は、あらかじめ用法を限定しておくことができる。そこで、以下のヒューリスティックスを設ける。

A 類 / B 類決定のヒューリスティック (2)

「 α [* | で | す] クラス」 「 α | が / で | [* | ごぞいま | す] クラス」という構造で、 α がある種の名詞の場合、

- a. 発話者がユーザの時は A 類の可能性を取り除く
- b. 発話者がシステムの時は B 類の可能性を取り除く

ただし、ある種の名詞 α : 「満室」「空き」等

このヒューリスティックによって、次のような例で誤認識が取り除ける。

表 2.9: 音声認識結果例 (5)

システム:	和室 か 、 ツインルーム の シングルユース	[* ごぞいま す が]	(<inform>)
	で し たら 空き が ごぞいま す が 。		
1	空き が ごぞいま す が	[* ごぞいま す が]	(<inform> <offer> <suggest>)
13	空き が ごぞいま す か	[* ごぞいま す か]	(<yn-question> <confirmation-question>)

「空き」は必ずシステムからユーザに提供される情報なので、システムが発話者の時は、その発話の文末は必ず A 類になる。したがって、B 類に属する [13] 「空き | が | ごぞいま | す | か」 [* | ごぞいま | す | か] を取り除ける。

常にシステムからユーザに提供される名詞は、現時点では「満室」「空き」の二つだけだが、このような特徴を持つ名詞(または逆に、常にユーザからシステムに提供される名詞)をさらに発見することができれば、この方略の有用性は高まると思われる。

2.3.6 特殊な場合 (1): 「そう | で | す」クラス

次に、「そう | で | す」クラスに適用されるヒューリスティックを考察する。「そう | で | す」クラスには以下のものがある。

表 2.10: 「そう | で | す」クラス

「そう で す ねえ」	(<temporizer> <yes> <acknowledge>)
「そう で す か」	(<acknowledge> <confirmation-question>)
「そう で す が」	(<yes>)
「そう で す な」	(<confirmation-question> <yes> <temporizer> <acknowledge>)
「そう で す ね」	(<confirmation-question> <yes> <temporizer> <acknowledge>)
「そう で す」	(<yes>)

(括弧で囲まれた表層表現パターンは、現時点での ACT 辞書には登録されていないが、登録してあったとしてもおかしくないものを表す。以下、それらも登録されているものとみなす)

このうち、ACT ラベル <yes> を担う 「そう | で | す」 「そう | で | す | が」 の性質に着目し、それらに関するヒューリスティックを設ける。

ACT ラベル <yes> を担う 「そう | で | す」 「そう | で | す | が」 には、名詞文の <yn-/confirmation-question> に対する返答としてのみ用いられ、形容名詞文・形容詞文・

動詞文の <yn-/confirmation-question> の返答としては用いられないという用法上の制限がある。

- (例6) a. これはあなたの本ですか [名詞文] — (はい、) そうです
 b. 彼は元気ですか [形容名詞文] — * (はい、) そうです
 c. そのりんごはおいしいですか [形容詞文] — * (はい、) そうです
 d. 机の上りんごはありますか [動詞文] — * (はい、) そうです

ただし、「のです」「わけです」は、構文的には名詞文なので(準体助詞「の」「わけ」により命題が体言化される)、「そうです」で返答することができる。

- (例7) a. アメリカに行ったのですか [ノデスカ] — (はい、) そうです
 b. アメリカに行ったわけですね [ワケデスネ] — (はい、) そうです

同様に、「でよろしいですね」「でよろしいですか」「でよろしいでしょうか」「でいらっしゃいますね」等も、意味的には名詞文(「…なのか、…でないのか」を問う疑問文)として解釈されるので、「そうです」で返答することができる。

- (例8) a. お客様は二階か三階をご希望で、お名前が田中様でよろしいですね。 [デヨロシイデスネ]
 — (はい、) そうです
 b. ご連絡先はケーアンドケー運輸の鈴木様、でよろしいですか。 [デヨロシイデスカ]
 — (はい、) そうです
 c. これでよろしいでしょうか。 [デヨロシイデショウカ]
 — (はい、) そうです
 d. 田中健史様でいらっしゃいますね。 [デイラッシャイマスネ]
 — (はい、) そうです

〔そう | で | す〕〔そう | で | す | が〕のこのような性質を使って、それらの解釈の可能性を削減する。まず、〔そう | で | す〕クラスのすぐ前にある話し相手による発話が <yn-/confirmation-question> か否かを調べる。話し相手による <yn-/confirmation-question> ではない場合には、〔そう | で | す〕〔そう | で | す | が〕を取り除く。<yn-/confirmation-question> があっても、それが名詞文ではない場合には〔そう | で | す〕〔そう | で | す | が〕を取り除く。〔そう | で | す〕〔そう | で | す | が〕が候補として残るのは、直前の話し相手の発話が名詞文の <yn-/confirmation-question> である場合に限る。規則の形で書くと、以下のようなになる。

〔そう | で | す〕クラスのヒューリスティックス

〔そう | で | す〕クラスを含む発話の直前にある話し相手による発話が、名詞文の <yn-/confirmation-question> でない場合、〔そう | で | す〕〔そう | で | す | が〕を取り除く。

このヒューリスティックスを設けることにより、以下の例で正解を特定できる。

表 2.11: 音声認識結果例 (6) (デフォルトとなる発話は除く)

通訳者:	ほか の お 部屋 は どう です か 。	[* どう です か]	<wh-question>
担当者:	[そう です ね]	[そう です ね]	<temporizer>
1	そう です ね	[そう です ね]	<confirmation-question> <yes> <temporizer> <acknowledge>
13	そう です	[そう です]	<yes>
15	そう です が	[* です が]	<inform>

この例では、[そう | です] クラスの直前にある話し相手による発話が <wh-question> なので、[そう | です] [そう | です | が] は取り除かれる。

2.3.7 特殊な場合 (2): [* | お | 願 | い | し | ま | す] クラス

表 2.12: [* | お | 願 | い | し | ま | す] クラス

[* お 願 い し ま す]	<action-request> <information-request> <permission-request> <desire> <inform> <farewell>
[* お 願 い し ま す ね]	<action-request> <permission-request> <desire> <inform> <confirmation-question> <farewell>
[* お 願 い し ま す よ]	<action-request> <permission-request> <desire> <inform> <farewell>

表 2.12 から、[そう | です] クラスと同様に、必ずしも終助詞が発話行為を決定しているわけではないことが分かる。ただし、[そう | です] クラスと違い、付与される ACT ラベルはどれもほぼ同じである (上記三つの [* | お | 願 | い | し | ま | す] クラスの ACT ラベルに関する相違点としては、<confirmation-question> を担う可能性のあるのは [* | お | 願 | い | し | ま | す | ね] のみ、また <information-request> を担えるのは [* | お | 願 | い | し | ま | す] のみである、という二点である)。これら三つの表層表現パタンの違いは、フォーマル/インフォーマルといった文体上のものであり、実質的な違いはないとしてもよいであろう。したがって、意味や発話行為の観点から誤認識を識別するのではなく、音声認識候補の傾向から誤認識を識別するのが妥当だと思われる ([* | お | 願 | い | し | ま | す] クラスに限っては方略 2 をスキップすることにしたのは、このためである)。方略 3 に従うと、上記三つのうち [* | お | 願 | い | し | ま | す] が認識されていれば、必ずそれが選ばれることになる。

2.4 効果の見積り

2.4.1 適用実験

効果を見積もるために、実際の対話音声認識結果を人手で目視により分類した。結果を表 2.13 に示す。

表 2.13: 効果の見積り

	UTTR	PRE	S1	S2	S3	UNIDENTIFIABLE	WRONGLY-IDENTIFIED
TAC23034.B1000.ag	17	8	7	1	—	1	—
TAS12009.B1000.ag	11	5	5	—	1	—	—
TAS22001.B1000.ag	13	4	4	1	3	1	—
TAS33002.B1000.ag	21	7	12	1	—	1	—
TCC22074.B1000.ag	15	10	1	—	4	—	—
	77	34 44.2%	29 37.7%	3 3.9%	8 10.4%	3 3.9%	—
TAC23034.B1000.al	15	8	7	—	—	—	—
TAS12009.B1000.al	10	5	5	—	—	—	—
TAS22001.B1000.al	15	2	7	1	4	1	—
TAS33002.B1000.al	16	7	7	—	2	—	—
TCC22074.B1000.al	14	7	3	—	4	—	—
	70	29 41.4%	29 41.4%	1 1.4%	10 14.3%	1 1.4%	—
TAC23034.B2000.ag	17	8	7	1	—	1	—
TAS12009.B2000.ag	12	4	6	—	2	—	—
TAS22001.B2000.ag	14	3	4	1	—	1	—
TAS33002.B2000.ag	23	4	17	1	—	1	—
TCC22074.B2000.ag	15	9	2	—	4	—	—
	81	28 34.6%	36 44.4%	3 3.7%	11 13.6%	3 3.7%	—
TAC23034.B2000.al	16	7	8	—	—	1	—
TAS12009.B2000.al	10	5	5	—	—	—	—
TAS22001.B2000.al	15	2	5	1	5	2	—
TAS33002.B2000.al	17	6	8	—	2	1	—
TCC22074.B2000.al	15	8	3	—	4	—	—
	73	28 38.4%	29 39.7%	1 1.4%	11 15.1%	4 5.5%	—
TAC23034.B500.ag	16	9	6	—	—	1	—
TAS12009.B500.ag	10	6	3	—	1	—	—
TAS22001.B500.ag	13	6	3	1	2	1	—
TAS33002.B500.ag	18	6	10	1	—	1	—
TCC22074.B500.ag	14	10	1	—	3	—	—
	71	37 52.1%	23 32.4%	2 2.8%	6 8.5%	3 4.2%	—
TAC23034.B500.al	14	9	5	—	—	—	—
TAS12009.B500.al	10	6	4	—	—	—	—
TAS22001.B500.al	12	2	5	1	3	1	—
TAS33002.B500.al	13	7	4	—	2	—	—
TCC22074.B500.al	11	7	3	—	1	—	—
	60	31 51.7%	21 35.0%	1 1.7%	6 10.0%	1 1.7%	—
TAC23034.B5000.ag	18	5	9	1	2	1	—
TAS12009.B5000.ag	13	4	7	—	2	—	—
TAS22001.B5000.ag	14	2	3	—	6	—	3
TAS33002.B5000.ag	28	6	17	3	—	2	—
TCC22074.B5000.ag	16	6	6	—	4	—	—
	89	23 25.8%	42 47.2%	4 4.5%	14 15.7%	3 3.4%	3 3.4%
TAC23034.B5000.al	17	6	9	1	—	1	—
TAS12009.B5000.al	9	4	5	—	—	—	—
TAS22001.B5000.al	15	2	4	1	5	3	—
TAS33002.B5000.al	24	6	16	—	2	—	—
TCC22074.B5000.al	15	6	4	—	4	1	—
	80	24 30.0%	38 47.5%	2 2.5%	11 13.8%	5 6.3%	—
total sum	601	234 38.9%	247 41.1%	17 2.8%	77 12.8%	23 3.8%	3 0.5%

音声認識の対象となるのは5対話 (TAC23034, TAS12009, TAS22001, TAS33002, TCC22074)

であるが、ビーム幅 (B500, B1000, B2000, B5000) の違い、および文法の枠組 (ag, al) の違いにより、計 40 ファイルを参照した。正解より下位の候補に対しても分析対象とした。

601 事例を分析したところ、表層表現パターンが適用されない候補 ([デフォルト] となる) が全体の 38.9% (234 件) あった。方略 1 で扱えるものが 41.1% (247 件)、方略 2 で扱えるものが 2.8% (17 件)、方略 3 で扱えるものが 12.8% (77 件)、その他残りが 3.8% (23 件)、誤るものが 0.5% (3 件) であった。

UTTR:	認識候補に正解の文末を含む発話の数
PRE:	前処理 (デフォルトとなる認識候補を除外) を行なった時点で正解を特定できる発話の数
S1:	方略 1 を適用した時点で正解を特定できる発話の数
S2:	方略 2 を適用した時点で正解を特定できる発話の数
S3:	方略 3 を適用した時点で正解を特定できる発話の数
UNIDENTIFIABLE:	方略をすべて適用しても、二つ以上候補が残る (正解を特定できない) 発話の数
WRONGLY-IDENTIFIED:	誤認識を正解としてしまう発話の数

正解を特定できない発話 (認識候補) が 23 例、誤認識を正解としてしまう発話が 3 例あった。

2.4.2 正解を特定できない事例

正解を特定できない事例は 23 例あるが、それらは以下の四種類のパターンに分類できる。

1. (動詞)+ [* | ま | す] クラス
2. 間投詞/感動詞
3. B 類/C 類の中から一つに特定できないもの
4. 複雑な名詞句

(動詞)+ [* | ま | す] クラス

表 2.14: 音声認識結果例 (7) (一部)

(TAS22001.B1000.ag-23)			
正解:	トラベラーズチェック は 使え ま す か		
1	トラベラーズチェック は 使え ま す か	[* ま す か]	(<yn-question> <confirmation-question> <offer> <promise> <invite>)
3	トラベラーズチェック は 使え ま す が	[* ま す が]	(<inform> <offer> <suggest> <permission-request> <promise>)
4	トラベラーズチェック が 使え ま す か	[* ま す か]	(<yn-question> <confirmation-question> <offer> <promise> <invite>)
5	トラベラーズチェック は 使え ま す か か	[* か]	(<yn-question> <confirmation-question> <inform>)
6	取 ら な い 明日 客 は 使え ま す か	[* ま す か]	(<yn-question> <confirmation-question> <offer> <promise> <invite>)
8	トラベラーズチェック は 作 り ま す か	[* ま す か]	(<yn-question> <confirmation-question> <offer> <promise> <invite>)
12	トラベラーズチェック は 使え ま す か が か	[* か]	(<yn-question> <confirmation-question> <inform>)

動詞文に関する方略は特に設けていないので、上のような、いわば無標の動詞文に対しては候補を一つに特定できずに方略が終了してしまう。

間投詞/感動詞

表 2.15: 音声認識結果例 (8) (一部)

(TAS22001.B2000.al-11)			
正解: え			
1	ええ	[ええ]	(<acknowledge> <temporizer> <expressive> <wh-question> <yes> <accept>)
2	え	[え]	(<expressive> <temporizer> <wh-question>)
9	で	[で]	(<acknowledge> <temporizer> <expressive>)

候補がすべて間投詞/感動詞からなる場合は、その中で最短のものを選ぶとしたが、上の場合、最短の間投詞/感動詞が [2] と [9] の二つあり、正解を特定できない。

B 類/ C 類の中から一つに特定できないもの

表 2.16: 音声認識結果例 (9) (一部)

(TAS33002.B2000.ag-48)	
正解 (先行文脈と共に示す)	
ユーザ: それから朝食もお願いいししたいの です が 、	[* お 願 い し た い の で す が]
(<desire>)	
ユーザ: できます す か 。	[* で き ま す か]
(<yn-question>)	
システム: はい 、	[はい]
(<yes>)	
システム: ご 用 意 で き ま す け れ ど も 、	[* ま す け れ ど も]
(<inform>)	
システム: 宿泊 料 と は 別 に 千 二 百 円 ち ょう だ い い た し ま す 。	[* ち ょう だ い い た し ま す]
(<inform>)	
システム: 和食 、 洋食 共 に 千 二 百 円 と な っ て お り ま す 。	[* と な っ て お り ま す]
(<inform>)	
ユーザ: 結構 で す 。	[* 結 構 で す]
(<acknowledge>)	
ユーザ: それでは 和食 で お 願 い し ま す 。	[* お 願 い し ま す]
(<action-request>)	
システム: 和食 で ご ざ い ま す ね 。	[* ご ざ い ま す ね]
(<confirmation-question>)	
1 和食 で ご ざ い ま す ね	[* ご ざ い ま す ね]
(<confirmation-question> <inform>)	
4 和食 で ご ざ い ま す ね ね	[* ね]
(<inform> <confirmation-question>)	
14 和食 で あ ご ざ い ま す ね か	[* か]
(<yn-question> <confirmation-question> <inform>)	
15 和食 で ご ざ い ま す が	[* ご ざ い ま す が]
(<inform> <offer> <suggest>)	
16 和食 で ご ざ い ま す	[* ご ざ い ま す]
(<inform>)	
25 和食 で ご ざ い ま す か	[* ご ざ い ま す か]
(<yn-question> <confirmation-question>)	

まず、方略1により [4] [14] は取り除かれる。次に方略2に進むと、この文脈で「和食」というのはシステムにとっては話し相手(ユーザ)から提供された情報なので、A類の文末表現である [15] は取り除かれる。C類に属する [*|ご|ざ|い|ま|す|が|] は、統語的に適格なので取り除けず、[16] は残る。したがって、この時点で残るのは、B類 [*|ご|ざ|い|ま|す|ね|] [*|ご|ざ|い|ま|す|か|] と C類 [*|ご|ざ|い|ま|す|が|] である。しかし、方略3に進んでも、残る表層表現パタンの文字数はすべて同じなので、これ以上候補を絞り込めない。

複雑な名詞句

表 2.17: 音声認識結果例 (10) (一部)

(TAS33002.B2000.al-42)		
正解:	お 聞 き な い た い 事 件 が ご ざ い ま す か	
1	お 聞 き な い た い 事 件 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
2	お 聞 き な い て 方 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
3	お 聞 き な い た い 事 件 な ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
4	お 聞 き な い た い 方 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
5	お 聞 き の い た い 事 件 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
6	お 聞 き の い て 方 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
7	お 聞 き な い て 五 ド ル ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
8	お 聞 き な い て 方 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
9	お 聞 き な い て 方 な ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
10	お 聞 き の い た い 事 件 な ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
11	お 聞 き な い て 中 が ご ざ い ま す か ((<yn-question> <confirmation-question>))	(* ご ざ い ま す か)
31	お 聞 き な い た い 方 が ご ざ い ま す が ((<inform> <offer> <suggest>))	(* ご ざ い ま す が)

表層表現パターンが〔* | ご | ざ | い | ま | す〕クラスの場合は、やりとりされる名詞句に着目して誤認識を取り除くというヒューリスティックスを立てた。しかし、表 2.17 の場合、名詞句が連体修飾節を含む複雑な構造をしていて、かつその主要部名詞が形式名詞「こと」である。このような構造の名詞句では、それが情報として「自ら提供した」ものなのか「話し相手から提供された」ものなのかを判定するのは難しいだろう。さらに、ここではその名詞句が正しく認識されていない。これでは、名詞句の情報を使って正解を特定するのは非常に困難である。

以上、四種類見てきたが、最後まで残る複数の認識候補を観察すると、正解候補は誤認識候補よりも常に認識順で上位にあることが分かる。そこで、次のヒューリスティックスを設けてもよいかもしれない。

ヒューリスティックス (暫定)

方略 1-3 がすべて終了しても正解発話を特定できない時には、その時点で残っている候補のうち認識順がより上位にあるものを優先する。

2.4.3 誤る事例

誤る事例は三つある。その事例を掲載するが、この三つは B5000.ag ファイルにしか見られない、極めてまれな事例である。

表 2.18: 誤る事例 (1) (一部)

(TAS22001.B5000.ag-3)	
正解:	お 願 い し た い ん で す け れ ど も
1	お 願 い し た い ん で す け れ ど も [* お 願 い し た い ん で す け れ ど も] (<action-request> <information-request> <desire> <permission-request> <inform>)
2	お 願 い し た い ね け れ ど も [* け れ ど も] (<inform>)
3	お 願 い し て る で す け れ ど も [* で す け れ ど も] (<inform>)
5	お 願 い し た い ん で す け ど [* お 願 い し た い ん で す け ど] (<action-request> <information-request> <desire> <permission-request> <inform>)

表 2.18の場合、方略2が終った時点で [*|お|願|い|し|た|い|ん|で|す|け|れ|ど|も] [*|お|願|い|し|た|い|ん|で|す|け|ど] の二つが残るが、方略3を適用すると、誤って [*|お|願|い|し|た|い|ん|で|す|け|ど] を正解としてしまう。しかし、この二つには実質的な意味の違いがない。

表 2.19: 誤る事例 (2) (一部)

(TAS22001.B5000.ag-14)	
正解:	え え ホ テ ル の 電 話 番 号 で す が
1	え え ホ テ ル の 電 話 番 号 で す が [* で す が] (<inform>)
3	え え ホ テ ル の 税 は 番 号 で す が [* で す が] (<inform>)
4	え え ホ テ ル の 電 話 は ど う で す が [* で す が] (<inform>)
6	え え ホ テ ル の 電 話 番 号 で す か [* で す か] (<yn-question> <confirmation-question> <permission-request>)
13	え え あ ホ テ ル の 電 話 番 号 で す か [* で す か] (<yn-question> <confirmation-question> <permission-request>)
14	え え ホ テ ル の 税 は 番 号 で す か [* で す か] (<yn-question> <confirmation-question> <permission-request>)
15	え え ホ テ ル の 電 話 は ど う で す か [* ど う で す か] (<suggest> <permission-request> <offer> <wh-question> <yn-question>)

文中要素を含む表層表現パターンについては本稿では考察しないとしたが、表 2.19の [15] に適用される表層表現パターン [*|ど|う|で|す|か] は、不定語「どう」が含まれるものの文中要素として機能していないので(※注7)、通常の表層表現パターンとして扱われる。そうすると、ここでの最長の表層表現パターンは [*|ど|う|で|す|か] になり、方略1により、それを誤って正解と認識してしまう。

(※注7) この表層表現パターンが ACT 辞書に登録されているのは、「(お茶は) どうですか」「(行ってみては) どうですか」など、ものを勧めたり、提案したりする用法があるためである(このような場合、ACT ラベルは、<wh-question>ではなく <suggest> が付与される)。

表 2.20: 誤る事例 (3) (一部)

(TAS22001.B5000.ag-14)	
正解:	トラベラーズチェック は 使え ま す か
1	トラベラーズチェック は 使え ま す か [* ま す か] (<yn-question> <confirmation-question> <offer> <promise> <invite>)
3	トラベラーズチェック は 使え ま す が [* ま す が] (<inform> <offer> <suggest> <permission-request> <promise>)
5	トラベラーズチェック は 使え ま す か か [* か] (<yn-question> <confirmation-question> <inform>)
9	取 ら れ る 明 日 客 は 使え ま す か [* ま す か] (<yn-question> <confirmation-question> <offer> <promise> <invite>)
24	トラベラーズチェック も 使え ま す か [* ま す か] (<yn-question> <confirmation-question> <offer> <promise> <invite>)
37	トラベラーズチェック は 使え ま す か ね [* ま す か ね] (<yn-question>)
41	トラベラーズチェック は 使え ま し た [* ま し た] (<inform>)

我々の定義によると、表 2.20 の [41] に適用される [* | ま | し | た] の [述語] 部分の文字数は 3 である。他の表層表現パタンの [述語] 部分の文字数は 2 なので、方略 1 により、最長の表層表現パターン [* | ま | し | た] を誤って正解と認識してしまう。

第 3 章

話題情報を用いて内容語を制御する指針

3.1 ねらいと基本的な考え方

音声認識候補の誤認識に含まれる語の中には、対話における出現位置が決まっているものがある。そのような語とそれが出現しうる状況(主に対話の任意の時点における「話題」との関連性を調査し、その結果をもとにした制約ないし選好を設けることで、誤認識を減らすことが考えられる(「話題」の認識方法については [巖寺他 94] を参照)。

ところで、表 3.1 に示すように、アルファベットや数字はしばしば誤認識の原因となり、そのため正解発話が得られないという事態が生じることがある。そこで、対話における話題(topic)情報と、音声認識結果ファイルに頻出する誤認識のうち <n-spel> (アルファベット) および <n-num-ex> (番号) との関連性を調べる。さらに、このような関連性を用いて、音声認識結果ファイルに頻出する誤認識発話の出現を防ぎ、音声認識の精度を向上させる方法を検討する。

語彙制御を検討するにあたっては、ATR 対話データベース (200 対話) を参照した。

表 3.1: アルファベットや番号がたくさん含まれる音声認識結果の例

正解:	お 支払い は どのように し たらよろし い で し よ う か
1	お 支払い は どのように し た い もし エイチ オー か
2	お 支払い は どのように し た い もし エス オー か お
3	お 支払い は どのように し た い い う もし エス オー か
4	お 支払い は どのように し た い もし エス オー か あ
5	お 支払い は どのように し た い お もし エス オー か
6	お 支払い は どのように し た い もし エス オー か お お
7	ご 支払い は どのように し た い もし エス オー か
8	お 支払い は どのように し た い もし エス オー か お あ
9	お 支払い は どのように し た い もし エイチ オー か お
10	お 支払い は どのように し た い よう シー エス オー か お
11	お 支払い は どのように し た い もし エス オー か 五
12	お 支払い は どのように し た い もし エス オー か あ っ
13	お 支払い は どのように し た い もし エス オー か お あ っ

3.2 枠組

語彙制御のためにここで用いる方法において重要なのは、

- (1) 質問—応答のような隣接ペア (adjacency pair) を認識できること
- (2) 各発話時点での話題 (topic) を認識できること

である。この二つの条件を満たす枠組として、「対話のインタラクション構造と話題の認識」[巖寺他 94]がある。そこで、まず、その枠組について簡潔に述べる。

3.2.1 対話の構造化

以下のステップに従って対話を構造化する(※注8)。

- (3) a. ACT単位が遂行する発話行為の認定
- b. ACT単位のACTクラスへの分類
- c. MOVEの形成
- d. EXCHANGEの形成

(※注8) [巖寺他 94]にはこれらの手順の前に「前処理」という手順があるが、本稿では省略する。

それぞれのステップについて見ていこう。

(3a)では、各ACT単位(ほぼ一文に相当する発話の最小単位)の遂行する発話行為を示すACTラベルを付与する。ACTラベルは、あらかじめ定義した表層表現パターンと行為の対応テーブル(ACT辞書)を用いて付与される。現在のところ、31種類のACTラベルがある。例えば、

- (4) そちらの|ホテル|の|(予約|で)|予約|を|し|た|い|の|で|す|が|。
- (5) お|客|様|の|お|名前|の|スペル|を|いただけ|ま|す|で|しよ|う|か|。
- (6) いつ|ご|滞在|の|ご|予定|で|しよ|う|か|。

(4)には<desire>、(5)には<information-request>、(6)には<wh-question>というACTラベルが付与される(ACTラベルについては、付録Aを参照)。

次に(3b)で、各ACT単位をACTクラスに分類する。あらかじめ定義したACTラベルとACTクラスの対応テーブル(ACTクラス辞書)を用いて分類される。ACTクラスは、ACTを伝達上の機能で分類したものである。これには、<<INITIATE>>、<<RESPONSE>>、<<NEUTRAL>>の3つのクラスがある。それぞれ次のように定義される。

- (7) a. <<INITIATE>>: 質問、依頼といった聞き手に働き掛ける発話、あるいは、返答の義務を負わせる発話
- b. <<RESPONSE>>: 返答の発話、相手の発話によって引き起こされ、自発的には生起しない発話
- c. <<NEUTRAL>>: 発話表現単独では、上記のどちらとも分類できない発話、あるいはどちらとも分類できる発話、すなわち、どちらかに決定するためには、先行する発話の系列を参照する必要がある発話

ACT クラスへの分類が終了すると、(3c) のステップに移る。ここでは MOVE が形成される。MOVE とは、同一の話者が産出したインタラクション機能 (後述する [開始] [応答] [補足]) が同一とみなされる一個以上の連続する ACT から構成される。例えば、

- (8) システム 1: はい
 システム 2: わかりました
 システム 3: では、至急送らせて頂きます

(8) のような場合、三つの ACT 単位からなるが、これらは全体で単一のインタラクション機能 ([応答] または [補足]) を持つ。したがって、ひとまとまりにして取り扱う必要がある。これら三つの ACT 単位は全体で一つの MOVE として扱われる。

MOVE の形成は次の手順で行なわれる。まず ACT 単位を ACT クラスに従って次のように MOVE クラス [INITIATE] または [RESPONSE] に分類する。

- (9) a. <<INITIATE>> → [INITIATE]
 b. <<RESPONSE>> → [RESPONSE]

ACT クラス <<NEUTRAL>> は、ACT クラス <<NEUTRAL>> 解釈ルールに従って、[INITIATE] または [RESPONSE] のいずれかに解釈される。

ACT クラス <<NEUTRAL>> 解釈ルール

ACT の ACT クラスが <<NEUTRAL>> のとき、MOVE 形成時に以下の条件に従って [INITIATE] または [RESPONSE] のいずれかに解釈される。

- a. ACT が次のいずれかを満たす場合は [INITIATE] である
- (i) ACT が対話開始時の第一 ACT である
 - (ii) ACT が話者交替直後の第一 ACT でかつ開始されている EXCHANGE が存在しない
 - (iii) 直前 ACT の話者と同一話者であり、直前 ACT の ACT クラスが [INITIATE] である
- b. 上記以外の場合は [RESPONSE] である

MOVE クラスが決定したら、MOVE 形成ルールに従って MOVE を形成する。

MOVE 形成ルール

次の条件に従って MOVE は形成される。

- a. 次のいずれかを満たす ACT は、新たな MOVE を開始する
- (i) 対話開始時の第一 ACT
 - (ii) 話者交替直後の第一 ACT
 - (iii) 直前 ACT と同一のターン内の発話であり、MOVE クラスが直前の ACT と異なる場合
- b. 新たな MOVE を開始する直前の ACT で MOVE は終了する
- c. 直前 ACT と同一話者でかつ同一 MOVE クラスである場合、直前 ACT の属する MOVE が継続する

最後に (3d) のステップで EXCHANGE が形成される。ここで、話しかけることをインタラクションの [開始]、応答することを [応答]、補足・評価を [補足] と呼ぶことにすると、対話に次のようなパタンの繰り返しを見ることができる。

- a. [開始] + [応答]
- b. [開始] + [応答] + [補足]

このようなパターンがインタラクション上の単位を形成している。このインタラクションの単位を EXCHANGE という。

EXCHANGE 形成の手順は以下の通りである。まず MOVE のインタラクション機能をインタラクション機能決定ルールに従って決定する。

インタラクション機能決定ルール

次の条件に従ってインタラクション機能が決定される。

- a. MOVE クラスが [INITIATE] の時は、[開始] である
- b. 対応する [応答] MOVE が無い [開始] MOVE が存在する時、[RESPONSE] は [応答] である (このときこの [応答] MOVE の対応する [開始] MOVE は、もっとも新しく [開始] された MOVE である)
- c. 先行するすべての [開始] MOVE の各々に対応する [応答] MOVE が各一個ずつ存在する時、[RESPONSE] は、[補足] である (このときこの [補足] MOVE は、[開始] MOVE と [応答] MOVE が対応付けられた (つまり終了した) EXCHANGE の中で最新のものの一部を形成する)

インタラクション機能決定後、EXCHANGE 形成ルールに従って EXCHANGE を形成する。

EXCHANGE 形成ルール

次の条件に従って EXCHANGE は形成される。

- a. インタラクション機能 [開始] を持つ MOVE が新たに開始された時、新たな EXCHANGE が開始される
- b. インタラクション機能 [応答] を持つ MOVE が終了した時、EXCHANGE は終了する
- c. MOVE が継続している場合は、EXCHANGE は継続する
- d. 先行するすべての EXCHANGE が終了しており、インタラクション機能 [補足] を持つ MOVE が新たに形成された場合、終了した EXCHANGE で最新のものを再度継続する

以上の手順を踏むと、対話が構造化される (以上の手順の適用例として、付録 B を参照)。

3.2.2 話題の認識

上の手順で形成された対話のインタラクション構造を用いて各発話時点毎に話題を認識する。話題には、持続時間の違いにより以下の二種類が設けられる。

- a. LT-TOPIC: 1つの EXCHANGE を越えて持続し得る話題 (Long-Term Topic)。たとえば、トピックマーカ「は」によって導入される話題である。
- b. ST-TOPIC: 1つの EXCHANGE 中でのみ持続する話題 (Short-Term Topic)。たとえば、トピックマーカ「が」、「に」、「を」によって導入される話題である。

ただし、インタラクション機能 [応答] または [補足] の MOVE で導入された話題は、すべて ST-TOPIC として扱われる。また、インタラクション機能 [開始] をもつ MOVE を開始する ACT において話題が全く導入されない場合は、直前に終了した EXCHANGE の LT-TOPIC のみが継続され、ST-TOPIC は継続されない。

3.3 話題情報を用いて内容語を制御する指針の概要

3.3.1 用語

本稿で用いる用語には、次のものがある。

- (1) ST-TOPIC
- (2) 当該 EXCHANGE で得られた LT-TOPIC (※注9)
- (3) ACT ラベル

(※注9) ここで、(2)の LT-TOPIC の扱いに注意されたい。(2)はつまり、先行する EXCHANGE から継続される LT-TOPIC は用いないということである。これには次の理由がある。実際の対話を観察すると、新しい LT-TOPIC が導入されずに以前の LT-TOPIC がいくつもの EXCHANGE にわたって継続され、その結果、EXCHANGE によっては不適切な TOPIC になることがある。この時、継続された LT-TOPIC を使って語彙制御を行なうと、不必要な語彙であるのにアクセス可能な状態になっている、または必要な語彙であるのにアクセス不能な状態になっているという事態が生じるおそれがある。この事態を避けるため、本稿では LT-TOPIC を語彙制御のための概念としては用いず、当該 EXCHANGE で得られたものに限ってそれを使用する。

(1)(2)から、TOPIC は一つの EXCHANGE 内でのみ有効な概念であるということが言える。したがって、本稿では ST-TOPIC と LT-TOPIC を区別する理由がなくなる。以下ではそれら二つをまとめて TOPIC と呼ぶ。

さらに、TOPIC 情報だけでは対処しきれない場合もあるので、次のような概念も用いる。

- (4) システムからユーザへ伝えられる情報
- (5) 第一 EXCHANGE(対話の冒頭)

(4)(5)の概念を具体的にどのように用いるかは、制御を考察する段階で述べる。

3.3.2 アプローチ

アプローチとしては、大まかに以下の二種類が考えられる。

- (6) a. 特定の語が TOPIC として出現した場合、辞書の一部を一時的にアクセス可能にする (普段はアクセス不可)
- b. 特定の語が TOPIC として出現した場合、辞書の一部を一時的にアクセス不可能にする (普段はアクセス可能)

(6a) は、語彙の性質や対話の推移から設定する方法で、(6b) は、誤認識の傾向から設定する方法であるといえよう。

この二つの方法のうち、本稿では (6b) の方法は考察しない。これは、以下の理由による。(6a) の方法を取る場合、まず語彙に着目し、それが出現しうる状況を (TOPIC 情報を用いて) 規定するという手順を踏む。一方 (6b) の方法を取る場合は、まず誤認識候補に着目し、それに含まれる語彙の出現を (TOPIC を用いて) 禁止するという手順を踏むことになる。このような手順の違いから、対象語彙の設定に関して、以下のような重大な相違点が生じる。つまり、(6a) の場合には対象とする語彙を容易にかつ積極的な形で規定できるのに対し、(6b) の場合は、誤認識候補に含まれる語という非常に消極的な規定しかできず、制御対象となる語彙を文法カテゴリ・意味カテゴリから規定することが非常に困難である。(6b) の制御方法の対象となる語彙は、したがって、お互いに何の共通性も持たないものとなり、一般性を欠くことになる。このような事情から、本稿では (6b) の方法を取らないことにする。

以上から、本稿では (6a) の方法を採用する。したがって、制御対象となる語彙をまず設定して、その語彙が出現すべき箇所を規定するという手順を踏む。本稿での語彙制御の対象は <n-spel> と <n-num-ex> であるから、ここで次の基本的な制御が設けられる。

<n-spel> および <n-num-ex> は、特に規定されない限りアクセス不可能な状態になっている

以下では、それらが出現し得る (しなければならない) のはどのような状況かを調査し、その調査結果をもとにこの制御に対するメタ的な制御を設定することになる。

3.3.3 TOPIC 情報を用いた語彙制御の性質

TOPIC の認識と発話時との関係についてここで見ておく。一般に、TOPIC が得られるのはインタラクション機能 [開始] 終了直後であり、その TOPIC を音声認識時に活用できるのはインタラクション機能 [応答] の発話のみである ([応答] が終了すると、EXCHANGE は (一旦) 終了するので、[補足] の発話時には当該 EXCHANGE における TOPIC を活用できない)。以下の例を参照されたい。

```
-----EXCHANGE-08-----
[開始] -----
通訳者 34: クレジットカード | 番号 | の | 方 |、 | 確認 | さ | せ | て | いた | だ | い | て | よ | ろ | し | い | で | す | か |。
          (<information-request><<INITIATE>>[INITIATE])
[応答] -----
申込者 35: マスターカード | で |、 | 五 | 二 | 七 | 九 |、 | 三 | 九 | 二 | 零 |、 | 二 | 四 | 六 | 九 |、 | 零 | 零 | 九 | 八 | で | す |。
          (<inform><<NEUTRAL>>[RESPONSE])
[補足] -----
通訳者 36: はい |、 (<acknowledge><<RESPONSE>>[RESPONSE])
```

通訳者 37: こちら | の | ファイル | の | 番号 | と | 同じ | です |。 (<inform><<NEUTRAL>>[RESPONSE]
 通訳者 38: [| え |] (<temporizer><<NEUTRAL>>[RESPONSE]
 通訳者 39: おそれ | います | が |、 (<apology><<NEUTRAL>>[RESPONSE]
 -----EXCHANGE-09-----
 [開始] -----
 通訳者 40: 有効 | 期限 | も | お | 伺 | い | でき | ます | か |。 (<information-request><<INITIATE>>[INITIATE]
 [応答] -----
 申込者 41: はい |。 (<accept><<RESPONSE>>[RESPONSE]
 申込者 42: 九 | 十 | 六 | 年 | 四 | 月 | まで | に | な | っ | て | い | ます |。 (<inform><<NEUTRAL>>[RESPONSE]

対話の途中におけるインタラクション構造の形成と TOPIC 認識との時間的關係について見てみよう。通訳者 34 の発話終了時に「クレジット | 番号」という TOPIC が得られるので、申込者 35 の発話の前にはその情報を利用できる。その後、通訳者 36 の発話により、申込者 35 の発話の一つで MOVE を形成し、かつそれが [応答] と解釈されるので、EXCHANGE が終了する。EXCHANGE が終了すると、TOPIC が消去されるので、通訳者 36 以降の発話では新しい TOPIC が導入されることになる。しかし、[補足] において新しい TOPIC が導入されることはほとんどない(実際、この場合でも TOPIC は導入されない)ので、それ以降では TOPIC は認識されない状態になる。この状態は、通訳者 40 の発話終了時まで続く。通訳者 40 の発話が認識されると、新しい TOPIC (「有効 | 期限」) が導入される(同時に ACT ラベル = ACT クラスから新しい EXCHANGE の始まりが認識される)。ここで、通訳者 40 の発話そのものは、TOPIC がいないまま認識されることに注意されたい。

インタラクション機能 [開始] に属する発話の音声認識時には TOPIC が認識されていないことがある(※注 10)。

(※注 10) 実際には、[開始] の発話直前に TOPIC を認識できることもある。例えば(※注 11)に挙げる (i) などである。しかし、この場合に得られるのは、その [開始] が属する EXCHANGE における TOPIC ではなく、その直前の EXCHANGE における TOPIC であり、適切な TOPIC とは言えないことに注意されたい。

我々がここで用いる方法は、特定の TOPIC の出現により語彙の出現を許すというものである。このような方法を取る限り、対象となる語彙は必ず先行発話で得られた TOPIC を活用できる発話に含まれるものでなければならない。したがって、対象となる語彙は以下の性質をもったものに限られる。

TOPIC を使った語彙制御において、制御の対象となる語彙は、インタラクション機能 [開始] には出現し得ないものでなければならない

これから、[開始] に現れる語彙の制御には、TOPIC 以外の情報を使った方法が必要になる。

3.3.4 制御の持続期間

次に、制御の持続期間について考察する。TOPIC を使った制御では、TOPIC の持続時間に合わせて次のように定める。

TOPIC を使った制御の持続期間は、制御が適用されてから、当該 EXCHANGE が終了するまでである (※注 11)。

(※注 11) 厳密に言えば、「当該 EXCHANGE が終了するまで」ではなく「当該 EXCHANGE の終了が認識されるまで」としなければならない。というのは、以下のような現象があるためである。

(i)

---EXCHANGE-16-----

システム 57: お|客|様|の|連絡先|の|ご|住所|と|電話番号|を|お|願|い|でき|ます|か|。

(<information-request><<INITIATE>>[INITIATE]

ユーザ 58 : [| あ - |] (<temporizer><<NEUTRAL>>[RESPONSE]

ユーザ 59 : 今|ホテル|に|滞在|して|る|ん|です|けれど|も|、 (<inform><<NEUTRAL>>[RESPONSE]

---EXCHANGE-17-----

ユーザ 60 : そち|ら|の|連絡先|で|よろ|し|い|です|か|。 (<permission-request><<INITIATE>>[INITIATE]

システム 61: はい|、 (<accept><<RESPONSE>>[RESPONSE]

ユーザ 59 の発話とユーザ 60 発話はそれぞれ別の EXCHANGE に属しているが、このこと (言い替えると、ユーザ 59 の発話で EXCHANGE-16 が終了すること) が認識されるのはどの時点であるかを考えてみよう。

ユーザ 60 の発話開始直後 (「そちらの…」) では、ユーザ 60 は直前発話 (ユーザ 59) と同一話者なので、MOVE 形成ルールの条件に当てはまらない。したがって、新しい MOVE が形成されない。

ユーザ 60 の発話が完了した時は、ユーザ 59 の MOVE クラスとユーザ 60 の MOVE クラスが異なるので、新しい MOVE が形成される。さらにユーザ 60 の MOVE クラスが [INITIATE] なので、インタラクション機能が [開始] となる。ここで初めて新しい EXCHANGE が形成されることになる。

したがって、ユーザ 59 の発話で EXCHANGE が終了すると認識されるのは、ユーザ 60 の発話が終了した時点である。ここで、ユーザ 60 の発話も EXCHANGE-16 で適用される語彙制御の対象となることに注意されたい。最終的な結果を見ると、ユーザ 59 の発話で EXCHANGE-16 が終了するのだが、その終了の認識と発話の進行とは必ずしもシンクロナイズ (synchronize) しない。

上のような現象が存在するため「EXCHANGE が終了するまで」という言い方は厳密さを欠くのだが、以下では便宜上そのような言い方をする。

TOPIC 以外の情報を用いた制御の持続期間については、その都度指定する。

3.4 <n-spel> の調査と制御の指針

3.4.1 基礎調査

<n-spel> は「名前やイニシャルの綴りを説明する時に使うアルファベット」 [竹沢他 95a] に与えられるカテゴリである。<n-spel> に属する語には、以下のものがある。

表 3.2: <n-spel>

エイ	ビー	シー	ディ	イー	エフ	ジー
エイチ	アイ	ジェー	ケー	エル	エム	エヌ
オー	ピー	キュー	アール	エス	ティ	ユー
ヴィ	ダブリュー	エックス	ワイ	ゼット		

これら <n-spel> が実際にどのように使われているかを見ておこう。

表 3.3: <n-spel> の使われ方 (括弧内は <n-spel> を含む EXCHANGE の数)

名前の綴り	(22)
イニシャル	(2)
料理コース	(1)
ルート (道順)	(3)
席 (コンサート)	(3)

それぞれの例を以下に挙げる。

名前の綴り

システム: お|客|様|の|お|名前|の|スペル|を|いただけ|ます|で|し|ょう|か|。 (<information-request>)
 ユーザ : はい|。 (<accept>)
 ユーザ : ティー|エー|エヌ|エー|ケイ|エー|で|す|。 (<inform>)

イニシャル

ユーザ : その|上|に|エー|エイチ|という|イニシャル|が|書|か|れ|て|ます|。 (<inform>)

料理コース

システム: はい|。 (<greet>)
 システム: フロント|で|ござい|ます|。 (<inform>)
 ユーザ : 四|百|二|十|一|号|室|の|エイミー|・|ハリス|で|す|。 (<inform>)
 ユーザ : きょう|の|七時|ぐらい|に|ディナーコース|の|ビー|を|二|人|前|お|部屋|の|ほう|に|運|ん|で|いた|だ|き|た|い|ん|で|す|が|。 (<action-request>)

ルート (道順)

ユーザ : その|バス|は|どこ|から|乗|れ|ば|いい|ん|で|す|か|。 (<wh-question>)
 システム: この|ホテル|の|正面|から|で|て|い|ます|。 (<inform>)
 システム: エム|六|の|系統|で|す|。 (<inform>)
 ユーザ : ホテル|から|出|て|い|る|ん|で|す|か|。 (<confirmation-question>)
 ユーザ : エム|六|で|す|ね|。 (<confirmation-question>)
 ユーザ : 分か|り|ま|し|た|。 (<acknowledge>)

席 (コンサート)

システム: 四日|の|ビー席|ないし|シー席|で|し|たら|ご|用意|でき|ます|が|。 (<inform>)
 ユーザ : そう|で|す|か|。 (<acknowledge>)
 ユーザ : 料金|は|エー|、|ビー|、|シー|それぞれ|お|いくら|な|ん|で|す|か|。 (<wh-question>)

このうち、料理コース、ルート (道順)、席 (コンサート) については、<n-spel> を用いない別の言い回しがある。以下を参照。

- (例 1) システム: ディナーコースビー|を|二|人|前|、|七時|に|お|部屋|の|ほう|に|お|持|ち|いた|し|ま|す|。 (<confirmation-question>)
 (例 2) システム: ルートエム|六|の|バス|に|乗|っ|て|いた|だ|い|て|、|バッテリー公園|で|降り|て|くだ|さ|い|。 (<instruct>)
 (例 3) システム: エー席|、|ビー席|、|シー席|と|ござい|ます|が|、 (<inform>)

3.4.2 <n-spel> の制御の指針

要約すると、<n-spel> が出現するのはおおむね次のような状況に限られる。

- (1) サービス提供者から綴りを求められた時
- (2) サービス提供者から名前を求められた時

この状況を定式化し、次のような制御を設けることが考えられる。

- (3) <n-spel> は、特に規定されない限りアクセス不可能な状態になっている。
- (4) サービス提供者による発話 U が (i) (ii) の条件を満たす時、その次の発話から <n-spel> をアクセス可能にする。
 - (i) 発話 U に ACT ラベル <information-request> が付与される。
 - (ii) 発話 U で「スペル」「つづり」「スペリング」「名前」のいずれかが TOPIC として認識される。

このような制御を設けることで、誤認識の出現を防ぎ、音声認識の精度を向上させることができるかと期待できる。なお、この制御を実装する場合、決定論的に 0/1 で扱う方法と、確率論的に出現しやすさを変えるものがある。

3.5 <n-num-ex> の調査と制御の指針

次に、<n-num-ex> に対する制御を考える。

<n-num-ex> に対する制御は、先ほどの <n-spel> に比べて複雑なものになる。というのは、<n-spel> よりも用法が多様で、出現位置も一定していないためである。

3.5.1 基礎調査

<n-num-ex> は「電話番号やクレジット番号」[竹沢他 95a] (p. 24) などの番号に与えられるもので、「一万四千元」「一本」などの数字とは区別される。<n-num-ex> には以下のものがある。

表 3.4: <n-num-ex>

零	[ree]
〇	[maru]
一	[ichi]
二	[ni]
三	[sang]
四	[jong]
五	[go]
六	[roku]
七	[nana]
八	[hachi]
九	[kjuu]

先ほどと同じように、初めに <n-num-ex> がどのように使われているかを見ておこう。

表 3.5: <n-num-ex> の使われ方 (ブラケット内は出現位置; 括弧内は <n-num-ex> を含む EXCHANGE の数)

番号 (電話番号/クレジット番号/ファックス番号)	[開始] [応答]	(86)
荷物の預かり票	[応答]	(5)
清算書ナンバー	[応答]	(3)
部屋番号	[応答]	(26)
小数	(すべてシステムの発話)	---
飛行機の便名	[開始] [応答]	(8)

それぞれの例を以下に挙げる。

番号 (電話番号/クレジット番号/ファックス番号)

システム: カード|の|ナンバー|を|お|願|い|でき|ます|で|し|よ|う|か|。 (<information-request>)
 ユーザ: はい|。 (<accept>)
 ユーザ: [|えー|] (<temporizer>)
 ユーザ: 四|九|八|零|、|零|四|五|九|、|九|一|九|九|、|五|三|一|三|で|す|。 (<inform>)
 システム: はい|、|かしこまりました|。 (<acknowledge>)

荷物の預かり票

システム: では|、|お|客|様|が|お|受|取|り|に|な|れ|た|預|か|り|票|の|ナ|ン|バ|ー|を|教|え|て|ら|え|ま|す|か|。 (<information-request>)
 ユーザ: はい|、 (<accept>)
 ユーザ: [|えーと|] (<temporizer>)
 ユーザ: 十|五|の|一|七|八|で|す|。 (<inform>)
 システム: はい|、|かしこまりました|。 (<acknowledge>)

清算書ナンバー

システム: お|客|様|の|清|算|書|ナ|ン|バ|ー|を|お|願|い|でき|ま|す|か|。 (<information-request>)
 ユーザ: [|あ|] (<temporizer>)
 ユーザ: 失礼|。 (<apology>)
 ユーザ: 一|一|六|二|八|九|で|す|。 (<inform>)
 システム: はい|。 (<acknowledge>)

部屋番号

システム: ありがとうございます、(<greet>)
 システム: ニューヨークシティホテルでございませう。(<inform>)
 ユーザ: すいません、(<alert>)
 ユーザ: ワシントンホテル五〇七号室に滞在しております田中と申しませう。(<inform>)

小数

ユーザ: 席を予約したいんですけど。(<desire>)
 ユーザ: そちらで食事をすると、どれくらい掛かりますかね。(<wh-question>)
 システム: [そうですね] (<temporizer>)
 システム: お料理が、一品につき、二、三十ドルから三、四十ドルになります。(<inform>)
 システム: お飲物は別料金になります。(<inform>)
 システム: それからテーブルチャージが五ドル。(<inform>)
 システム: それから税金が八、二五パーセントになります。(<inform>)
 ユーザ: 分かりました。(<acknowledge>)

飛行機便名

システム: そちらのお名前と現在お持ちの予約便名をお願ひいたします。(<information-request>)
 ユーザ: はい、(<accept>)
 ユーザ: [え] (<temporizer>)
 ユーザ: 鈴木和子と言います。(<inform>)
 ユーザ: [え] (<temporizer>)
 ユーザ: 現在の予約が八月十日のニューヨーク発ロサンゼルス行き二〇一便です。(<inform>)
 ユーザ: 十三時発です。(<inform>)
 システム: 分かりました。(<acknowledge>)

表 3.5 に示すように、「番号」「飛行機の便名」は [開始] でも使われている。[開始] に含まれる語は TOPIC では処理できない。したがって、それらに対する制御には TOPIC 以外の概念も利用されることになる。

以下、それぞれの用法別に <n-num-ex> の制御方法について検討する (※注 13)。

(※注 13) 今回はユーザの発話に調査対象を限定したため、システムの発話中にしか存在しない「小数」は考察しない。

3.5.2 <n-num-ex> の制御の指針

要約すると、<n-num-ex> が出現するのはおおむね次のような状況に限られる。

- 番号 (電話番号/クレジット番号/ファックス番号)
- 荷物の預り票
- 清算書ナンバー
- 部屋番号
- 飛行機便名

「部屋番号」は対話の [開始] でも使われることがあるが、それ以外はおおむね [応答] の位置にのみ現れる。そこで次のような制御を設けることが考えられる。

- (1) 対話の第一 EXCHANGE が終了するまでは、<n-num-ex> をアクセス可能とする。
- (2) サービス提供者による発話 U が (i) (ii) の条件を満たす時、その次の発話から <n-num-ex> をアクセスしやすくする。
 - (i) 発話 U に ACT ラベル <information-request><yn-question><wh-question> のいずれかが付与される。
 - (ii) 発話 U で (iii)-(vii) に挙げる語が TOPIC として認識される。
 - (iii) ナンバー／電話番号／カードナンバー／住所／連絡先／部屋 | 番号／カード | 番号／クレジットカード | 番号／支払い | 方法／ファックス | 番号／番号／連絡
 - (iv) 預かり票／ナンバー／荷物
 - (v) 清算書／ナンバー
 - (vi) 連絡先／住所／部屋 | 番号／部屋／番号／ホテル／ファックス番号／電話番号
 - (vii) 便名／航空便／フライト／予約 | 便名
- (3) サービス提供者による発話 U が (i) の条件を満たす時、その次の発話から当該 EXCHANGE が終了するまで <n-num-ex> をアクセスしやすくする。
 - (i) 発話 U に ACT ラベル <confirmation-question> が付与される。
- (4) サービス提供者が新たに番号／飛行機便名を対話に導入したならば、ユーザによる確認 (*confirmation*) が終了するまで、<n-num-ex> をアクセスしやすくする。
- (5) それ以外の状態では、<n-num-ex> はアクセスしにくくする。

このような制御を設けることで、誤認識の出現を防ぎ、音声認識の精度を向上させることができるかと期待できる。なお、この制御を実装する場合、決定論的に 0/1 で扱う方法と、確率論的に出現しやすさを変えるものがある。

3.6 要約と今後の課題

今回の調査では、主に TOPIC 情報 (と ACT ラベル) を用いて語彙を制御したが、この方法には次のような課題がある。

- (1) 文法カテゴリと意味カテゴリが対応しているものしか扱えない (例: 「エイ、ビー、シー、…」アルファベットの使用の場は限られる; 文法カテゴリは <n-spel>) (もちろん、個々の語に対して制御を設けることはできるが、そうすると制御が一般性のないものになる)
- (2) 制御対象となる語彙は、インタラクション機能 [開始] には出現し得ないものでなければならない ([開始] においては、発話認識時に TOPIC 情報を活用できないため)

このような限界を持っているため、他の意味カテゴリにこの種の制御を適用させることは難しい。したがって、語彙制御をさらに進めるためには、他の方法も考える必要がある。

現時点では、次のような方法が考えられる。例えば、語彙にはシステム/ユーザどちらかしか使えない語がある。「当」(rentai/prefix-hutu-hutu) という語について、200 対話を調べた結果、「当」を含む発話はすべてシステムによるものであることが分かった (それも、「当 | ホテル」という表現でのみ現れている)。これは、「当」という語の持つフォーマル

さに起因するものだと思われる。一方、200 対話中「妻」(pow-n-hutu)という語は、ユーザの発話にしか存在しない。これは、この語の持つ「視点」や対人関係によるものであろう。このように、音声認識候補に頻出する語で、このうちどちらかに分類できるものが多くあれば、制御の方法として採用できるのではないだろうか。

さらに、次のような方法も考えられる。今まで考察してきたのは EXCHANGE 内での語彙制御であった。しかし、それよりも大きい単位で語彙の制御はできないだろうか。その単位とは、タスク、すなわち一つの対話である。タスク別に対話を見ると、「あるタスクではまず使われないであろう語彙」または「特定のタスクでしか使われない語彙」が存在することが分かる。例えば、「熱」「医者」「診る」などは、「ホテルの予約」には使われない語彙だと思われる。タスクがシステムに伝わった時点で、そのような語彙を制御できれば、かなり効果的な制御になるとと思われる。

第 4 章

音声認識に対する対話処理部の貢献度の見積り

対話処理部で用意できる発話タイプ情報と話題の情報を用いて、対話音声認識結果が実際にどのように扱えるかに関する予備実験を行なった。対話処理部と音声認識部の統合に関して、特に次の二点を検討した。

- (1) 文末述語に関して尤もらしい認識候補を選ぶ手法
- (2) 話題情報を用いて内容語を制御する指針

効果を見積もるために、実際の対話音声認識結果を人手で目視により分類した。音声認識の対象となるのは5対話 (TAC23034, TAS12009, TAS22001, TAS33002, TCC22074) であるが、ビーム幅 (B500, B1000, B2000, B5000) の違い、および文法の枠組 (ag, al) の違いにより、計40 ファイルを参照した。正解より下位の候補に対しても分析対象とした。

また、対話データとして、(上記5対話を含む) ATR 音声言語データベース (200 対話) を参照した。

601 事例を分析したところ、表層表現パターンが適用されない候補 ([デフォルト] となる) が全体の 38.9% (234 件) あった。このうちの多くが発話末に「番号」がつながるものであるから、内容語を制御する指針は役に立つと期待できる (もちろん他にも解決手段はあろうが)。

方略 1 で扱えるものが 41.1% (247 件)、方略 2 で扱えるものが 2.8% (17 件)、方略 3 で扱えるものが 12.8% (77 件)、その他残りが 3.8% (23 件)、誤るものが 0.5% (3 件) であった。

謝辞

本研究を進めるにあたり、協力をいただいた西川賢哉、衛藤純司 両氏に感謝します。また、助言と協力をいただいた ATR 音声翻訳通信研究所 第四研究室 森元 逞 室長、石崎雅人 主任研究員に感謝します。

参考文献

- [巖寺他 94] 巖寺俊哲・石崎雅人・森元逞 (1994) 「対話のインタラクション構造と話題の認識」『自然言語処理 104-16』 119-126.
- [神尾 90] 神尾昭雄 (1990) 『情報のなわ張り理論：言語の機能的分析』大修館書店.
- [金水 91] 金水敏 (1991) 「伝達の発話行為と日本語の文末形式」『神戸大学文学部紀要 18』 23-41.
- [田窪他 96] 田窪行則・金水敏 (1996) 「対話と共有知識：談話管理理論の立場から」『月刊言語 25-1』大修館書店: 30-39.
- [竹沢他 95a] 竹沢寿幸・田代敏久・衛藤純司 (1995a) 「部分木を単位とする音声認識用日本語文法」『ATR テクニカルレポート』 TR-IT-0110 (1995-04).
- [竹沢他 95b] 竹沢寿幸・田代敏久・加藤直人・衛藤純司 (1995b) 「音声対話システムに特化したキーワードと意味カテゴリ設計のための基礎調査」『ATR テクニカルレポート』 TR-IT-0140 (1995-10).
- [竹沢他 96] 竹沢寿幸・森元逞 (1996) 「部分木に基づく構文規則と前終端記号バイグラムを併用する対話音声認識手法」『信学論 D-II』 Vol. J79-D-II, No. 12 (1996-12).
- [益岡他 92] 益岡隆志・田窪行則 (1992) 『基礎日本語文法：改訂版』くろしお出版.

付録 A

発話タイプ (ACT ラベル)

現在のところ、以下の 31 種類の発話タイプ (ACT ラベル) を使用している。

ACT-LABEL : <accept>

DEFINITION : 要求、依頼、助言などを受理する／受け入れる／承知する

ACT-LABEL : <acknowledge>

DEFINITION : 聞き手が話し手に談話を継続させるためにあいづちをうつ

ACT-LABEL : <action-request>

DEFINITION : 話し手が聞き手に行為を要求する

ACT-LABEL : <alert>

DEFINITION : 聞き手の注意を引く

ACT-LABEL : <apology>

DEFINITION : 謝罪の言葉

ACT-LABEL : <believe>

DEFINITION : 思っている／信じていることを伝える

ACT-LABEL : <confirmation-question>

DEFINITION : 話し手が聞き手に確認をする

ACT-LABEL : <desire>

DEFINITION : 希望、欲求を表明する

ACT-LABEL : <do-you-understand-question>

DEFINITION : 話し手が聞き手に話し手の発話を理解したかどうか確認する

ACT-LABEL : <expressive>

DEFINITION : 感嘆詞をする

ACT-LABEL : <farewell>

DEFINITION : 挨拶の言葉

ACT-LABEL : <good-wishes>

DEFINITION : 聞き手の幸福・安泰などを願う言葉

ACT-LABEL : <greet>

DEFINITION : 挨拶の言葉

ACT-LABEL : <inform>

DEFINITION : 話し手は聞き手に情報を与える

ACT-LABEL : <information-request>

DEFINITION : 情報を要求する

ACT-LABEL : <instruct>

DEFINITION : 話し手が聞き手の利益になる方法で目的に到達するよう指図する

ACT-LABEL : <invite>

DEFINITION : 話し手が聞き手に何かを一緒にするよう誘う

ACT-LABEL : <no>

DEFINITION : YN-question の否定応答

ACT-LABEL : <offer>

DEFINITION : 話し手が聞き手の利益になることを提案する

ACT-LABEL : <offer-follow-up>

DEFINITION : 話し手が聞き手の利益になることを再度提案する

ACT-LABEL : <permission-request>

DEFINITION : 話し手が聞き手に許可や了解を求める

ACT-LABEL : <promise>

DEFINITION : 話し手が聞き手の利益になることを約束する

ACT-LABEL : <reject>

DEFINITION : 要求、依頼、助言などを拒絶する／拒否する

ACT-LABEL : <suggest>

DEFINITION : 話し手が間接的な方法で聞き手に意見を述べる

ACT-LABEL : <temporizer>

DEFINITION : 話し手が躊躇を示す

ACT-LABEL : <thank>

DEFINITION : 聞き手に感謝の言葉を述べる

ACT-LABEL : <thanks-response>

DEFINITION : 感謝の言葉に応える

ACT-LABEL : <vocative>

DEFINITION : 名前などを呼びかける

ACT-LABEL : <wh-question>

DEFINITION : 話し手が聞き手にいつ、どこ、だれが、いかに、だれにを尋ねる

ACT-LABEL : <yes>

DEFINITION : YN-question の肯定応答

ACT-LABEL : <yn-question>

DEFINITION : 話し手が聞き手に「はい」か「いいえ」で応えられる質問をする

付録 B

構造化された対話例 (TAS12009)

第 II 部 9.2 節で述べた対話の構造化の手法を ATR 対話データベース TAS12009 に適用すると、以下のような構造が得られる。

EXCHANGE_01

[開始] -----
ACT_01 通訳者: はい | , , (<greet>), <<NEUTRAL>>, [INITIATE]
ACT_02 通訳者: ホリデイインクラウンプラザ | で | ござい | ます | 。 , (<inform>), <<NEUTRAL>>, [INITIATE]
[応答] -----
ACT_03 申込者: きょう | 予約 | し | て | お | り | ま | し | た | 鈴木 | 直子 | と | 申 | し | ます | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
ACT_04 申込者: 都合 | で | 泊 | ま | れ | な | く | な | っ | て | し | ま | っ | た | の | で | , , (<inform>), <<NEUTRAL>>, [RESPONSE]
ACT_05 申込者: [| あ | の |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_02

[開始] -----
ACT_06 申込者: キャンセル | し | た | い | ん | で | す | が | 。 , (<desire>), <<INITIATE>>, [INITIATE]
[応答] -----
ACT_07 通訳者: [| あ | あ |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]
ACT_08 通訳者: そう | で | す | か | 。 , (<acknowledge>), <<RESPONSE>>, [RESPONSE]
ACT_09 通訳者: [| え |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_03

[開始] -----
ACT_10 通訳者: 当日 | の | キャンセル | で | す | と | , | 五 | 十 | パーセント | , | キャンセル | 料 | が | 掛 | か | る | こと | は | ご | 存 | じ | で | す | か | 。 , (<yn-question>), <<INITIATE>>, [INITIATE]
[応答] -----
ACT_11 申込者: はい | , , (<yes>), <<RESPONSE>>, [RESPONSE]
ACT_12 申込者: 分 | か | っ | て | しま | っ | ます | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
ACT_13 申込者: [| と |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_04

[開始] -----
ACT_14 申込者: キャンセル | 料 | は | お | い | くら | に | な | り | ます | で | し | ょ | う | か | 。 , (<wh-question>), <<INITIATE>>, [INITIATE]
[応答] -----
ACT_15 通訳者: はい | , , (<acknowledge>), <<RESPONSE>>, [RESPONSE]
ACT_16 通訳者: 鈴木 | 様 | , , (<vocative>), <<NEUTRAL>>, [RESPONSE]
ACT_17 通訳者: 確認 | いた | し | ます | の | で | , (<inform>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_05

[開始] -----
ACT_18 通訳者: 少々 | お | 待 | ち | くだ | さ | い | 。 , (<action-request>), <<INITIATE>>, [INITIATE]
ACT_19 通訳者: [| はい | , , (<alert>), <<INITIATE>>, [INITIATE]
ACT_20 通訳者: えー |] , (<temporizer>), <<NEUTRAL>>, [INITIATE]
ACT_21 通訳者: ご | 予約 | は , (<inform>), <<NEUTRAL>>, [INITIATE]
ACT_22 通訳者: [| えー |] , (<temporizer>), <<NEUTRAL>>, [INITIATE]
ACT_23 通訳者: 百 | 八 | 十 | 六 | ドル | の | シングル | ルーム | で | す | 。 , (<inform>), <<NEUTRAL>>, [INITIATE]
ACT_24 通訳者: [| え |] , (<temporizer>), <<NEUTRAL>>, [INITIATE]
ACT_25 通訳者: です | の | で | , | キャンセル | 料 | は | 九 | 十 | 三 | ドル | に | な | り | ます | 。 , (<inform>), <<NEUTRAL>>, [INITIATE]
[応答] -----
ACT_26 申込者: はい | , | 分 | か | り | ます | し | た | 。 , (<acknowledge>), <<RESPONSE>>, [RESPONSE]

EXCHANGE_06

[開始] -----
 ACT_27 申込者: お | 支払い | は | どのように | し | たら | よろし | い | で | し | よ | う | か | 。 , (<wh-question>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_28 通訳者: [| え |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]
 ACT_29 通訳者: ご | 予約 | の | 際 | に | クレジットカード | 番号 | を | 何 | っ | て | お | り | ま | す | の | で | , (<inform>), <<NEUTRAL>>, [RESPONSE]
 ACT_30 通訳者: [| え |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_07

[開始] -----
 ACT_31 通訳者: そ | ち | ら | の | 口 | 座 | の | 方 | に | 請 | 求 | さ | せ | て | い | た | だ | き | ま | す | 。 , (<permission-request>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_32 申込者: 分 | か | り | ま | し | た | 。 , (<accept>), <<RESPONSE>>, [RESPONSE]
 [補足] -----
 ACT_33 通訳者: お | そ | れ | い | り | ま | す | が | , (<apology>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_08

[開始] -----
 ACT_34 通訳者: クレジットカード | 番号 | の | 方 | に | 確 | 認 | さ | せ | て | い | た | だ | い | て | よ | ろ | し | い | で | す | か | 。 , (<information-request>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_35 申込者: マスターカード | で | , | 五 | 二 | 七 | 九 | , | 三 | 九 | 二 | 零 | , | 二 | 四 | 六 | 九 | , | 零 | 零 | 九 | 八 | で | す | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
 [補足] -----
 ACT_36 通訳者: はい | , (<acknowledge>), <<RESPONSE>>, [RESPONSE]
 ACT_37 通訳者: こ | ち | ら | の | ファイル | の | 番号 | と | 同 | じ | で | す | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
 ACT_38 通訳者: [| え |] , (<temporizer>), <<NEUTRAL>>, [RESPONSE]
 ACT_39 通訳者: お | そ | れ | い | り | ま | す | が | , (<apology>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_09

[開始] -----
 ACT_40 通訳者: 有 | 効 | 期 | 限 | も | お | 何 | い | で | き | ま | す | か | 。 , (<information-request>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_41 申込者: はい | 。 , (<accept>), <<RESPONSE>>, [RESPONSE]
 ACT_42 申込者: 九 | 十 | 六 | 年 | 四 | 月 | ま | で | に | な | っ | て | い | ま | す | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
 [補足] -----
 ACT_43 通訳者: か | し | こ | ま | り | ま | した | 。 , (<acknowledge>), <<RESPONSE>>, [RESPONSE]

EXCHANGE_10

[開始] -----
 ACT_44 通訳者: で | は | キャンセル | 料 | は | こ | ち | ら | の | 口 | 座 | の | 方 | に | 請 | 求 | さ | せ | て | い | た | だ | き | ま | す | 。 , (<permission-request>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_45 申込者: よ | ろ | し | く | お | 願 | い | た | します | 。 , (<farewell>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_11

[開始] -----
 ACT_46 申込者: 念 | の | た | め | に | , | そ | ち | ら | 様 | の | お | 名 | 前 | を | お | 聞 | かせ | い | た | だ | け | ま | す | か | 。 , (<information-request>), <<INITIATE>>, [INITIATE]
 [応答] -----
 ACT_47 通訳者: はい | , (<accept>), <<RESPONSE>>, [RESPONSE]
 ACT_48 通訳者: [| え |] , (<acknowledge>), <<RESPONSE>>, [RESPONSE]
 ACT_49 通訳者: わ | た | く | し | は | ジョン | ・ | フィリップス | と | 申 | し | ま | す | 。 , (<inform>), <<NEUTRAL>>, [RESPONSE]
 [補足] -----
 ACT_50 申込者: [| と |] , (<expressive>), <<NEUTRAL>>, [RESPONSE]

EXCHANGE_12

[開始] -----
 ACT_51 申込者: ジョン | ・ | フィリップス | さ | ん | で | す | ね | 。 , (<confirmation-question>), <<INITIATE>>, [INITIATE]
 ACT_52 申込者: で | は | , | よ | ろ | し | く | お | 願 | い | た | します | 。 , (<farewell>), <<NEUTRAL>>, [INITIATE]

EXCHANGE_13

[開始] -----
 ACT_53 通訳者: あ | り | が | と | う | ご | ざ | い | ま | した | 。 , (<thank>), <<INITIATE>>, [INITIATE]