

文脈を考慮した音声認識結果絞り込み手法

山岡 孝行 飯田 仁

ATR 自動翻訳電話研究所

梗概

音声言語処理システム実現において、言語解析の効率化のために音声認識結果の曖昧さを解消する問題がある。この問題に対して、一文を越えた文脈からの情報を適用して解決する手法について述べる。まず、階層型プラン認識モデルにより、対象対話の対話構造の解析を行なう。続いて、そのモデルを活用した次発話の予測手法により、次発話に関する情報を抽出する。そして得られた情報を文脈からの次発話に対する制約と捉え、その制約に従い音声認識候補の選択を行なう。質問-応答に関するローカルなプランと、与えられた情報に対応する言語表現に関する知識を設定した実験により、曖昧さ解消における文脈的判断の効果を示す。

Using Contextual Information to Select Correct Speech-Recognition Candidates

Takayuki YAMAOKA and Hitoshi IIDA

ATR Interpreting Telephony Research Laboratories

Abstract

In order to make a language processing module in a spoken-language integrated processing system more efficient, a method of contextual information application to select the correct candidate from a speech recognition results is proposed. Initially, the target spoken dialogue structure is analyzed by a multi-layered plan recognition model for dialogue understanding. Secondly, the next utterance is possibly predicted using the constructed dialogue structure, and information about the next utterance can be obtained. Finally, using this information as a constraint for a possible expression of the next utterance, correct candidates which satisfy the constraint can be selected.

1 はじめに

自動翻訳電話などの音声言語処理システム実現のためには、音声認識技術と言語処理技術の統合が必要である。双方の技術はそれぞれの分野で成果を挙げている。しかしながら、従来の言語処理システムでは、多くの場合誤りのない入力を仮定している。一方、それに対して現状の音声認識技術では、完全な音声認識を行なうことは不可能であり、言語処理への誤りのない入力は期待できない。入力が曖昧であることは、言語処理の負担を増加させ効率を落すことになる。

このような問題に対して、一文内での言語情報を使って文内の整合性をみる手法により、音声認識結果の候補削減を行なう方法が提示されている [5]。しかしこのような文節間の係受け関係だけの言語情報では、文章としての的確な候補がいくつか出現し、曖昧さを残すことが多い。また、係受け処理のみで効果を上げようとすれば、膨大な量の係受けデータを必要とする。

例えば以下のような例は、非文ではないので候補として曖昧さを残す。

(例 1)

「名前は、鈴木真弓です。」 (1-a)

「名前は、鈴木真弓ですか？」 (1-b)

(例 2)

「わかりました。」 (2-a)

「ありました。」 (2-b)

特に、話し手の発話の意図を表す表現の曖昧さ (例 1) や、会話に特有な簡略化された慣用的な表現の曖昧さ (例 2) は、一文の整合性のみでは解消できない。しかし、滑らかな対話翻訳を目指す時、これらの表現の解釈は重要な役割を占める。このような問題を解決する一手段として、一文を越えたいわゆる文脈の情報の利用が考えられる。そのような音声認識への文脈情報利用の試みとして、MINDS[6] や SPURT-I[7] などがあるが、いずれもトピックに関連した文の命題内容を構成する要素 (単語) に対してのものであり、上の例のような話し手の意図を表す表現の選択には適用できない。

処理の対象となる対話を限定すれば、対話を実現する際の常識を利用し、対象発話の前対話の理解状態を参考にして後続する発話のタイプを予測し、それに対応する表現としての適切さを評価することにより、表現の曖昧さを絞り込むことができる。例えば (例 1) は、名前を尋ねる質問に回答している発話であれば、(1-a) の表現が正しい。また (例 2)

は、情報交換を行なう対話において、あるトピックについての情報交換が成立した状況であれば、情報の受け手が情報交換の内容について理解した時の確認の合図に良く用いられる表現として、(2-a) が適切である。

本報告では、協調的目標指向型対話を対象とした、音声認識結果の候補選択を行なうシステムについて述べる。このシステムでは、対話理解機構として階層型プラン認識モデル [1][2] による対話構造構築を行ない、そのモデルを活用した次発話の予測手法 [3][4] により得られる次発話に関する情報を、文脈情報として音声認識候補の選択に利用する。特に本報告では、話し手の意図を表わす表現についての曖昧さを解消するための、知識について詳しく考察を行なう。

まず次章で、対象とする協調的目標指向型対話の発話の特徴について述べる。3 章で、システム全体の構成と各モジュールの機能の概要と、音声認識候補の選択処理への文脈情報の適用の枠組みについて述べる。4 章で、話し手の意図を表わす表現を扱うための知識について整理を行なう。5 章では、サンプル対話 7 対話について行なった実験結果について報告する。6 章で、現状では扱っていない対象・現象について、今後の課題を示す。

2 対象対話の分析 (発話の表現の分類)

本報告では、ある目的を達成するために協調的な情報交換が営まれる話し言葉対話 (協調的目標指向型対話) を対象としている。協調的目標指向型対話における発話の表現は、下記のように分類することができる。

- 一般的な陳述を行なう表現
 - 発話の意図を表す表現 (文末)
 - 発話の命題内容を構成する表現 (自立語)
- 対話で固定化された表現 (断片的一発話)

() の中は、その表現が最も一般に現れる文章中の位置を示している。

一般に情報交換における陳述は、伝達したい情報の内容を表すの命題部分と、発話の目的に従った話し手の意図を表す部分とに、切り分けて考えることができる。命題内容は、述語とその格要素から構成される。それら構成要素のエンティティは、自立語として表現される。一方、発話の意図を表す表現は、文末の付属語や補助的な語で表現されることが多い。例えば (例 1) では、「名前は鈴木真弓である」と

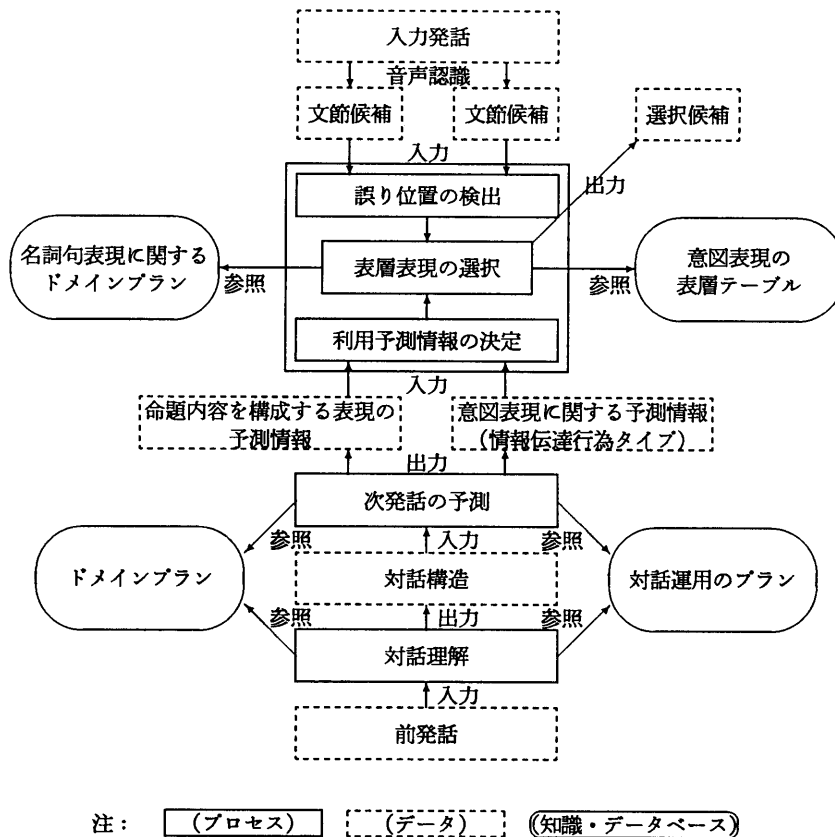


図 2: 文脈情報による絞り込みシステムの構成

3.2.1 対話理解

対話理解部は、以下の3つの対話運用に関するプランと1つの話題領域固有のプランを利用した階層型プラン認識モデル [1][2] により対象対話の対話構造を構築する。

- 対話運用に関するプラン

インタラクションプラン： 協調的な応答実行のための知識を記述したプラン。

コミュニケーションプラン： ある話題についての協調的な対話を実現するための知識を記述したプラン。

ダイアログプラン： 対話の対局的な展開の知識を記述したプラン。

- 話題領域固有のプラン

ドメインプラン： 話題の行為達成のための行動を記述したプラン。

このモデルでは、2つのスタックを使って理解状態と未充足プランにより、対話構造を管理する。すなわち、入力され認識された発話の解釈は理解状態のスタックにより理解事項として管理される。そして、その発話が関与する目的の達成に必要でありかつ未出現の内容を表す発話に関する情報が、未充足プランのスタックにより管理される。（図3）

3.2.2 次発話の予測

協調的目標指向型対話を仮定すれば、上述モデルの未充足プランのスタックを参照することにより、次発話を予測することが可能である [3][4]。なぜなら、未充足プランのスタックに残っている内容は、対話の目的のために必要であるので、後に発話されることが期待される。よって、未充足プランのスタックに入っている発話に関する情報は、次発話に関する予測情報として利用することができる。

各プランのもつ機能から、発話の意図に関する表

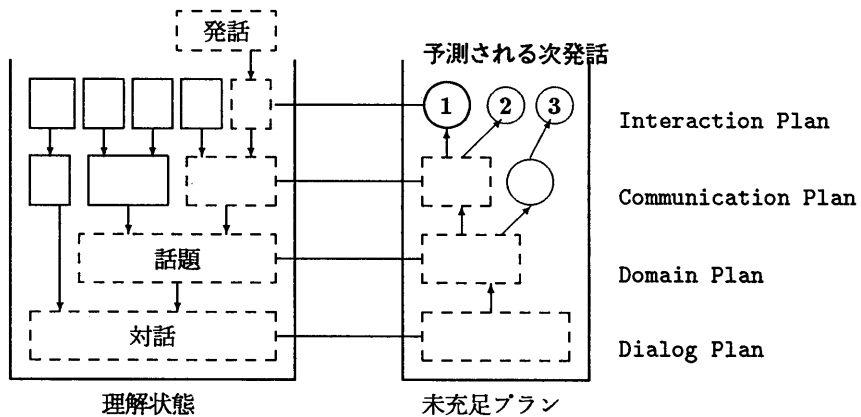


図 3: 階層型プラン認識モデルの対話の管理

現および対話で固定化された表現はインタラクションプランから情報伝達行為のタイプとして、命題内容を構成する表現はドメインプランに記述された話題として、抽象化された内容として予測される。予測された情報は、プランの連鎖により下位のプランに伝搬される。従って各発話の予測情報は、情報伝達行為のタイプと命題内容に関する話題との組で表される。

3.3 次発話予測情報による絞り込み処理

予測情報の優先順位： 次発話の予測情報は、対話理解モデルの未充足プランの管理に従った優先順位を持つ。例えば図3の例では、予測情報の優先順位は、1-2-3の順になる。この優先順位により、対話の流れに沿った話題のより下位のプランから充足されるであろう予測を行なっていることになる。

誤り位置の検出： 文脈情報を適用した選択を行なうために、まず音声認識結果の誤り位置を検出する。絞り込み処理部へは、音声認識の出力の文節候補のラティスが入力となる。誤り位置は、各文節候補の自立語に誤りがあるか付属語に誤りがあるかで区別する。さらに、最終文節にあるか否かも区別する。すなわち、誤り位置は以下の3つの位置で区別される。

1. 文中の自立語の誤り。
2. 文中の付属語の誤り。
3. 最終文節（文末の表現）の誤り。

このような構文的区別で表現との関係付けを行なうことは、MINDSのなどの構文パタンの利用や、

SPURT-IのASPのような専用パーサに比べ、意味的な精密さは落ちる。しかし、構文解析でなく、音声認識結果の曖昧さ解消の処理としては、簡便な処理が望ましい。ここでの処理は、文節・単語ラティスによる入力と、発話の表現の分類と位置の対応から、簡便な処理による表現の意味的区別が可能である。

利用情報の決定： 誤り位置により、利用する予測情報を決定する。すなわち、誤り位置が、最終文節であれば情報伝達行為のタイプに関する予測情報、自立語であれば命題内容に関する情報となる。なお、文中の付属語に関しては、現在扱っていない。

誤り位置が複数ある時は、最終文節の誤りからの処理を優先する。これは、プランの階層性に従っている。なお、固定化表現を含む予測情報がある時は、それを一文単位で優先的に採用する。

表層表現の選択： 最後に、予測情報に対応付けられる表層表現のものを候補の中から選択する。予測情報と表層表現の対応は、表層テーブルで管理する。

4 意図表現に関する知識

ここでは、特に話しての意図を表わす表現の曖昧さを解消することに焦点を当て、それを解決するための知識として、情報伝達行為と表層表現の対応とインタラクションプランの整理を行なう。

4.1 情報伝達行為と発話対

情報伝達行為は、電話による問合せなどにおいて、話し手が何らかの情報を聞き手に与える発話を行な

う時に観察される行為である。情報伝達行為は、その行為の話し手の発話欲求の信念と聞き手の信念に与える影響の違いに着目すれば、大きく次のようにに区別することができる。まず、聞き手に与える影響が、与えられた命題内容に関する聞き手の発話を促す要求 (Demand) の情報伝達行為と、それに促されて発話される、すなわち要求の発話に対する応答 (Response) の情報伝達行為とに分けられる。また、情報交換の成立の確認の意味で発せられる発話の情報伝達行為を、確認 (Confirm) の情報伝達行為とする。その他、上記3つに含まれない発話の情報伝達行為を陳述 (Inform) の情報伝達行為とする。

ここで、協調的に行なわれる情報交換を仮定すれば、Demand の情報伝達行為に対する聞き手の『応答の義務』が生じる。すなわち情報交換成立のためには、聞き手が Demand により受けた信念に対して何らかの形で情報を与えること、すなわち Response の発話を行なわなければならない。このような、ある話題 (命題内容) についての情報交換成立のための、Demand と Response の情報伝達行為を持つ発話の対を『発話対』と呼ぶ。発話対には、要求した話者が情報交換成立に対する確認の意味で、Confirm の発話を行なうことがある。

発話対は、対話全体の構造の基本的構成要素となると考えられる。ここでは、『発話対の構築』を基本的立場として、極力表層表現から判断できる手がかりに着目して、情報伝達行為の分類・設定ならびに発話対の設定を行なう。発話対はインタラクションプランとして記述される。

4.2 情報伝達行為の設定

情報伝達行為を以下のように分類する。

情報伝達行為は、いわゆる発話の意図の部分とトピックとなる話題の属性で表現する。ここで話題の属性を導入する理由は、同様の意図による発話でも、その話題となる事柄の属性により、特に応答表現に違いが出ると考えるからである。

発話の意図の部分を表わすタイプには、以下のものを設定する。これらは、おもに文のモダリティから判断できる。() 内は代表的な文形式を表している。

INFORM: 事柄の事実について述べる。(平叙文)
 ASK: 未知の事柄について質問する。
 (疑問詞を伴う疑問文)
 CONFIRM: 事柄の真偽について質問する。(疑問文)
 REQUEST: 相手に行動を依頼する。(命令文)
 OFFER: 自分の行為を拘束する。
 (動作動詞を述語とした平叙文)
 GREETING: あいさつ。(おもに固定化された表現)

以下のものは、INFORM の下位に属すると考えられるが、対話での機能上特定の情報伝達行為を設定する。

AFFIRMATIVE: 話題の内容を肯定する。
 NEGATIVE: 話題の内容を否定する。
 ACCEPT: 行為を受け入れる。
 REJECT: 行為を拒む。
 CONFIRMATION: 情報交換成立の確認。

話題の属性には以下のものを設定する。

ACTION: ある行為・行動についての話題
 VALUE: ある事物の属性についての話題
 STATEMENT: ある事物の状態についての話題

上のような分類から、以下のような情報伝達行為のタイプを設定することができる。また、これらの情報伝達行為に対応する代表的な表層表現を示す。

(イタリック文字は変項を表し、それぞれ ACT は ACTION に関する内容、OBJ は OBJECT に関する内容、STA は STATEMENT に関する内容、WH は疑問詞を表す。)

Demand Class:

ASK-ACTION: 「*ACT* は *WH* ですか。」
 「*ACT* について教えてください。」
 CONFIRM-ACTION: 「*ACT* (するの) ですか。」
 REQUEST-ACTION: 「*ACT* して下さい。」
 「*ACT* していただけますか。」
 OFFER-ACTION: 「*ACT* します。」
 ASK-VALUE: 「*OBJ* は *WH* ですか。」
 「*OBJ* をお願いします。」
 CONFIRM-VALUE: 「*OBJ* は *VAL* ですか。」
 「*OBJ* は *VAL* ですね。」
 ASK-STATEMENT: 「*STA* は *WH* ですか。」
 CONFIRM-STATEMENT: 「*STA* ですか。」
 GREETING-OPEN: 「もしもし。」「はい。」
 GREETING-CLOSE: 「失礼します。」
 「さようなら。」

Response Class:

INFORM-VALUE: 「*OBJ* は *VAL* です。」
 INFORM-ACTION: 「*ACT* して下さい。」
 「*ACT* です。」
 INFORM-STATEMENT: 「*STA* です (が) 。」
 AFFIRMATIVE: 「はい。」「そうです。」
 NEGATIVE: 「いいえ。」「ちがいます。」
 「～ません。」

ACCEPT-ACTION: 「わかりました。」
 REJECT-ACTION: 「ACTできません。」
 ACCEPT-OFFER: 「(よろしく) お願いします。」
 REJECT-OFFER: 「(いいえ) 結構です。」

Confirm Class:

CONFIRMATION: 「わかりました。」
 「そうですか。」

以上のような、対応する表層表現(「ですか」「で
 しょうか」「ごさいましようか」などのバリエーショ
 ンを含めて)を、表層テーブルに記述しておき、音
 声認識候補の選択に利用する。

対話で固定化された表現の設定 目標指向型対話に
 おいて慣用的に発話される固定化された表現として、
 以下のようなものを設定する。これらは、併せて示
 すように、それ自体固有の情報伝達行為を持つ。

もしも GREETING-OPEN
 はい GREETING-OPEN, AFFIRMATIVE
 そうです AFFIRMATIVE
 いいえ NEGATIVE
 そうですか CONFIRMATION
 わかりました ACCEPT-ACTION, CONFIRMATION
 失礼します GREETING-CLOSE
 さようなら GREETING-CLOSE

これらの表現は、併せて示した情報伝達行為によ
 る発話をなす時、対話中に断片的な一発話として現
 れることが多い。従ってこれらの情報伝達行為が予
 測されている時は、一文に対して優先的に対応を見
 るようにする。

4.3 インタラクシヨンプランの設定

4.1 節で述べたように協調的なやり取りを仮定す
 れば、4.2 節で設定した情報伝達行為は、表 1 のよ
 うな発話対を構成する。

これらをインタラクシヨンプランのスキーマの
 decomposition として、Demand Class, Response
 Class, Confirm Class の順で記述し、対話理解並
 びに次発話の予測に用いる。

5 実験結果

3 章で述べた枠組みの実験プロトタイプシステム
 を作成し、4 章で説明した知識とダイアログプラン
 の一部を利用して、7 モデル対話について固定化表
 現および応答・確認発話の文末表現の絞り込みを対
 象とする、実験・評価を行なった。

実験プロトタイプは、TI-ExplorerII 上に Com-
 mon Lisp によりインプリメントされている。実験

表 1: 発話対の構成

Demand Class	Response Class
ASK-ACTION	INFORM-ACTION
CONFIRM-ACTION	AFFIRMATIVE, NEGATIVE INFORM-ACTION
REQUEST-ACTION	ACCEPT-ACTION REJECT-ACTION
OFFER-ACTION	ACCEPT-OFFER REJECT-OFFER
ASK-VALUE	INFORM-VALUE
CONFIRM-VALUE	AFFIRMATIVE, NEGATIVE INFORM-VALUE
ASK-STATEMENT	INFORM-STATEMENT
CONFIRM-STATEMENT	AFFIRMATIVE, NEGATIVE INFORM-STATEMENT
GREETING-CLOSE	GREETING-CLOSE

の入力は、係受け処理を行なった後の音声認識候補
 の文節ラティスである。入力の内訳を以下に示す。

総発話数	138 発話
要求発話数	56 発話
応答発話数	58 発話
確認発話数	12 発話
その他	12 発話

入力の内、音声認識・係受け処理後、曖昧な候補
 を残したものと、その誤り位置の内訳は以下の通り
 である。

複数候補を残した発話数	113 発話
固定化された表現の誤り	32 発話
文末表現の誤り	26 発話
(うち応答・確認発話	10 発話)

この内、固定化表現はすべて応答・確認発話ある
 いは対話の開始・終了に関する発話である。

そして応答・確認発話の誤りについての、実験に
 よる絞り込みの結果は以下の通りである。結果は、
 優先度 1 位で選択された発話の数を示している。(実
 験の入力のほとんどは、係受け処理により 2 つの候
 補まで絞り込まれていた。)

応答・確認発話の絞り込み結果

固定化表現の選択	32 発話中 28 発話 (87.5%)
文末表現の選択	10 発話中 9 発話 (90%)

この結果から、インタラクシヨンプランという基本的かつローカルな対話の知識の記述により、応答・確認発話の意図を表す表現について音声認識の曖昧性が解消できることを確認した。

6 実験の考察と今後の課題

本節では、実験でうまく処理できなかった対象についての考察と、その解決に向けての今後の課題を議論する。

情報交換における応答には、ここで記述したような協調的な応答の他に、質問という応答の要求の行為自体に対する応答をする場合がある。例えば、あいつちの「はい」などはこのレベルの応答であると考えられる。実験対象発話の中でこのような例は3発話あった。すべて、ASK-VALUEに対するものであり、上述の発話対の設定では、予測できないものである。

このような対話現象を扱うためには、応答のメカニズムのより細かいレベル分けと、知識としての切り分けが必要とされる。しかしながら、きめ細かな知識の設定は、予測情報の増大につながり、候補選択の制約を緩めることになる。計算機モデルとして効率的な処理を実現するためには、メタレベルの知識の導入や、対話参加者の相互信念や理解状態を援用するための形式化を検討しなければならない。

また逆に、インタラクシヨンプランは情報交換における規範的な知識として考えることができるが、実際の対話では、発話対の埋め込みや応答の省略、言い替えなど多様な現象が表れる。その多様性を、対話における間接的な推論の実行と捉え、その推論に関する知識の記述のためにインタラクシヨンプランより上位のプラン枠組の詳細化を検討する必要がある。つまり、より上位のプランによりインタラクシヨンプランの実行制御を行なうことで、不完全なやり取りについても認識・理解を可能にする。これにより、応答・確認発話のみならず、要求の発話の処理や、情報伝達行為と表層が一致しない表現の処理が可能となる。

一方、命題内容に関する表現の具体的な絞り込みを可能にするため、名詞句の参照[8]、あるいは連想機構の研究[7]などの成果を取り入れる手法の検討を行なっていく。

7 おわりに

階層性プラン認識モデルによる対話理解と次発話の予測手法を活用して、文脈情報を適用した音声言

語処理システムについて述べた。特に、文章としては正しいが対話の流れとして不適切な文末表現や会話で固定化された表現の絞り込むための知識として、情報伝達行為と発話対について考察し、定義を行なった。そして、プロトタイプによる実験・評価を行ない有効性を確認した。

今後は、多様なやり取りを扱えるためのより詳細なプランの記述を考え、その処理性能の向上をはかるとともに、命題内容についての表現を扱える枠組みを研究していく必要がある。

謝辞 本研究の機会を与えて下さった ATR 自動翻訳電話研究所 樽松明社長に感謝の意を表します。また、日夜有効な議論、助言を与えて下さる同研究所言語処理研究室の皆様ならびに三菱電機(株)の有田英一氏に感謝します。

参考文献

- [1] 飯田, 有田: 4 階層プラン認識モデルを使った対話の理解, 情報処理学会論文誌, Vol.31, No.6 (1990)
- [2] Iida, H., Yamaoka, T. and Arita, H.: Three Typed Pragmatics for Dialogue Structure Analysis, COLING'90(1990). to appear.
- [3] 山岡, 有田, 飯田: 階層型プラン認識モデルによる対話理解と次発話の予測手法, 情報処理学会「談話理解とその応用」シンポジウム (1989)
- [4] Yamaoka, T. and Iida, H.: A Method to Predict the Next Utterance Using a Four-layered Plan Recognition Model, ECAL'90(1990). to appear.
- [5] 柿ヶ原, 森元: SL-TRANS における文節候補の削減 - 係受け関係を用いた文節候補の選択 -, 情報処理学会第 39 回全国大会 (1989)
- [6] Hauptmann, A. G., Young, S. R. and Ward, W. H.: Using Dialog-Level Knowledge Sources to Improve Speech Recognition, Proceedings of AAAI'88(1988)
- [7] 堀, 辻野, 溝口, 角所: 音声理解システム SPURT-I - 動的クラスタリング方式と文節発声による性能評価 -, 電子情報通信学会論文誌 D-II, Vol.J72-D-II No.8(1989)
- [8] 野垣内, 飯田: キーボード会話における名詞句の同一性の理解, 情報処理学会自然言語処理研究会報告 72-1(1989)