

対話構造木コーパスを用いた発話意図予測

加藤真吾[†] 松原茂樹^{††} 山口由紀子^{††} 河口信夫^{††}

[†] 名古屋大学情報科学研究科 〒 464-8601 名古屋市千種区不老町
^{††} 名古屋大学情報連携基盤センター 〒 464-8601 名古屋市千種区不老町
E-mail: †gotyan@el.itc.nagoya-u.ac.jp

あらまし 音声対話システムがユーザとの協調的な対話を進める上で、次の発話の意図を予測することは有用である。ユーザの発話意図を予測できればそれを用いた音声認識精度の向上が期待でき、システムの発話意図の予測はそのまま次発話の決定に利用できる。本稿では、発話が入力されるたびに対話の構造を漸進的に作り上げ、対話構造が付与された対話データを参照することにより、次発話の意図を予測する手法について述べる。CIAIR 車内音声対話コーパスを用いて次発話意図予測の実験を行ったところ、対話構造が与えられた状況下での本手法の利用可能性を確認した。キーワード 対話モデル、対話理解、構造解析、対話制御、音声対話システム

Speech Intention Prediction based on Structurally Annotated Spoken Dialogue Corpus

Shingo KATO[†], Shigeki MATSUBARA^{††}, Yukiko YAMAGUCHI^{††}, and Nobuo KAWAGUCHI^{††}

[†] Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan
^{††} Information Technology Center, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan
E-mail: †gotyan@el.itc.nagoya-u.ac.jp

Abstract It is important for a cooperative spoken dialogue system to predict the intention of the next utterance. If a system can predict the intentions of user's utterances, it would contribute to an improvement of accuracy of speech recognition and could be used for the decision of system's utterances. In this paper, we propose a method of speech intention prediction based on structurally annotated spoken dialogue corpus. The method constructs the dialogue structures incrementally, and predicts the intention of the next utterance statistically. As a result of an experiment, 93% was provided as the prediction accuracy.

Key words dialogue model, dialogue understanding, structural analysis, dialogue control, spoken dialogue system

1. はじめに

近年、音声処理技術の発展にともない、ユーザの自然な発話に適切に回答し、協調的に対話を遂行するシステムの実現を目指した研究が盛んに行われている(例えば [1] [2] [3])。音声対話システムが協調的に対話を進める上で、ユーザの発話意図を理解し、対話の目的とその達成状況を随時把握することが重要である [4]。そのための一つの方法として、システムが対話の進行にともなって漸進的に対話の構造を作成し、途中までに作成された対話構造に基づいて、対話の状況を把握する方法が考えられる。タスク指向型の対話では、対話全体のタスクの達成のために必要となるサブタスクの達成の連続として、対話が遂行さ

れていると考えられるが、対話の途中の段階では、最終的なタスクはもちろん、直面しているサブタスクや、別のサブタスクとの関係などを随時把握することにより、より適切な対応が行える可能性がある。対話において、ある発話が行われた後に、システムが次の発話の意図を予測することは、話者交代を円滑に行ったり、データベース検索を事前に行うなどにより応答の即時性を高めるなど、自然な対話の遂行に有用である。さらに、次発話の意図がユーザのものであったときに、その意図に関連する単語が出現するなどの予測ができ、音声認識の精度を高めることができる可能性がある。また、次発話の意図がシステムのものであれば、それをもとにした対話制御はより適切かつ柔軟性の高いものになると予想できる。

0074 - 05:00:476-05:02:787 M:D:I:C:
 この辺で & コノヘンデ
 たいが & タイガ
 食べれるところは & タベレルトコロワ
 ないかなあ<SB> & ナイカナア<SB>
 0075 - 05:05:895-05:06:156 M:O:I:C:
 はい<SB> & ハイ<SB>
 0076 - 05:08:206-05:10:671 M:O:I:C:
 (F エー) & (F エー)
 お料理は & オリョーリワ
 どんな & ドンナ
 お料理が & オリョーリガ
 よろしかったですか<SB> & ヨロシカッタデスカ<SB>

図 1 書き起こしテキストの例

そこで本稿では、対話の構造が付与されたコーパスを用いて、次の発話の意図を予測する手法を提案する。発話が入力されるたびに、対話構造を解析して構造を作り上げ、作られた構造を持つ対話をコーパスから取り出し、それらの対話から統計的に得られた意図を予測結果とする。

本稿の構成は以下の通りである。2 節では、使用した対話構造木コーパスについて説明する。3 節では、漸進的な対話構造解析を、4 節では、対話構造に基づく発話意図予測手法を述べる。最後に、5 節では、評価実験について報告する。

2. 対話構造木コーパス

本節では、発話意図の予測において使用した、対話構造木コーパス [9] について述べる。対話構造木コーパスは、名古屋大学 CIAIR において作成した意図タグつき車内音声対話コーパス [5] [6] に対して、部分対話のタスクや部分対話間の関係を明らかにすることを目的として構造が付与されている。

2.1 意図タグ付き車内音声対話コーパス

名古屋大学統合音響情報研究拠点 (CIAIR) では、実環境における頑健な音声対話システムの実現を目的として 1999 年度から 2003 年度までの間に、実走行車内における音声及び対話を収録した [5] [6]。このデータベースは 800 名を越える被験者に対してその音声収録されており、道路案内や店情報検索などをタスクとする対話が、対人、対 W O Z、対システム、の 3 種類を相手にして行っている。また、収録した音声データは人手によって書き起こしており、多様な言語情報を付与している。図 1 に書き起こしデータの例を示す。

さらに、これらの対話のうち、ドライバとオペレータの対人の飲食店内タスク対話には、意図タグを付与している [7] [8]。意図タグとは、話者の発話の意図を表すタグであり、談話行為、動作、対象、詳細情報の 4 階層で構成し、原則として 1 発話に対して 1 つ付与している。意図タグが付与された対話例を表 1 に示す。ここで、話者の “D” はドライバによる発話を、“O” はオペレータによる発話を意味する。

2.2 対話構造

本研究における対話構造とは、意図が付与された発話に対して、発話と発話の対応関係、発話が含まれるサブタスク、及び、最終的なタスクとの関係を表すものである。これらの構造を明

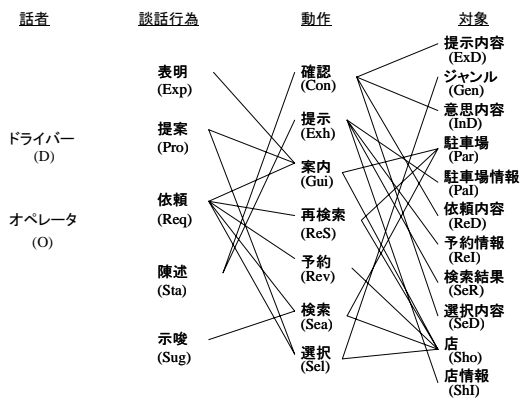


図 2 対話構造木コーパスに使用されている意図タグ体系

表 2 POD の種類と内容

POD	内容
GENRE	店のジャンルを決定する
GUIDE	店や駐車場に案内する
P_INFO	駐車場情報、駐車場の有無、近隣の駐車場等に関する情報のやり取りを行う
P_SRCH	駐車場の検索を行う
S_INFO	店情報値段、メニュー、場所、定休日等に関する情報のやり取りを行う
SLCT	店や駐車場を選択する
SRCH	店を検索する
SRCH_RQST	検索依頼をする
RSRV	予約を行う
RSRV_DTL	予約依頼の詳細情報のやり取りを行う
RSRV_RQST	予約依頼をする

表 3 対話構造木コーパスの情報

対話数	789
発話数	8150
構造木の種類数	659
付与した POD ののべ数	7329
構造木の高さ (最大)	15
構造規則の種類	297

らかにすることにより、ある時点での対話の状況を把握する。

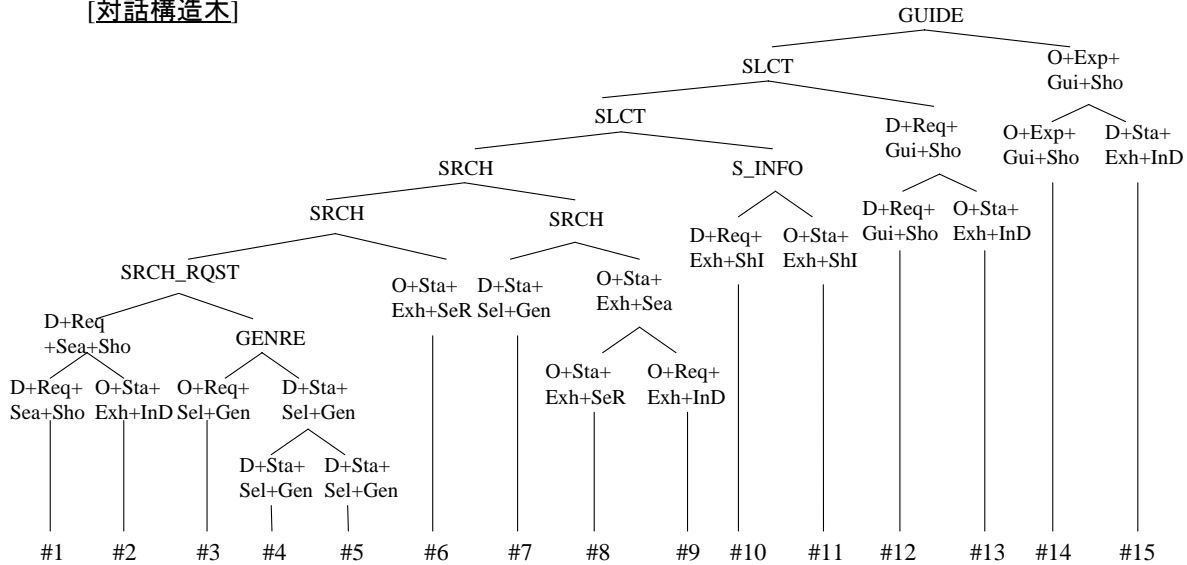
まず、意図タグを拡張する。本来、意図タグは 4 つのレイヤーで構成されているが、第 4 レイヤーは、意図に関する詳細情報を表し、対話構造の記述には必ずしも必要ではない。そこで、本研究では、第 3 レイヤーまでを使用することとし、また、対話では話者情報も対話構造の決定に重要であるので、話者情報に意図タグの 3 つのレイヤーを合わせたものを意図タグの拡張として使用した。図 2 に、本研究で用いる意図タグ体系を示す。意図タグは全部で 41 種類存在している。例えば、表 1 の発話番号 1 「この辺でたいが食べれるところはないかなあ」という発話に対して、「店を検索してほしい」という意図であると考えられるので、意図タグ “D+Req+Sea+Sho” が与えられる。

さらに、部分対話とその対話全体との関係を明らかにするために、木構造の各ノードに、POD (Part-Of-Dialogue) と呼ばれるカテゴリをラベルとして与える。意図タグもまた POD で

表 1 意図タグつき対話の例

発話番号	話者	発話内容	意図タグ			
			第一 (談話行為)	第二 (動作)	第三 (対象)	第四 (詳細情報)
#1	D	この辺でたいが食べれるところはないかなあ	依頼	検索	店	
#2	O	はい	陳述	提示	意思内容	肯定
#3	O	お料理はどんなお料理がよろしかったですか	依頼	選択	ジャンル	
#4	D	生系がいいかな	陳述	選択	ジャンル	
#5	D	なべが食べたいです	陳述	選択	ジャンル	
#6	O	はいこの近くですとちゃんこなべとおでん会席あとしゃぶしゃぶな どがございますが	陳述	提示	検索結果	ジャンル
#7	D	おでん会席がいい	陳述	選択	ジャンル	
#8	O	はいそうしますとマルというお店になりますが	陳述	提示	検索結果	店名
#9	O	よろしかったでしょうか	依頼	提示	意思内容	
#10	D	予約は必要ありますか	依頼	提示	店情報	予約制
#11	O	あ予約のほうは予約されなくてもお店には入ることができますが	陳述	提示	店情報	予約制
#12	D	あじゃあそこまで案内お願いします	依頼	案内	店	
#13	O	かしこまりました	陳述	提示	意思内容	了解
#14	O	それではマルまでご案内いたします	表明	案内	店	
#15	D	はい	陳述	提示	意思内容	肯定

[対話構造木]



[対話構造規則]

GUIDE	->SLCT	O+Exp+Gui+Sho	GENRE	->O+Req+Sel+Gen	D+Sta+Sel+Gen
SLCT	->SLCT	D+Req+Gui+Sho	S_INFO	->D+Req+Exh+ShI	O+Sta+Exh+ShI
SLCT	->SRCH	S_INFO	D+Sta+Sel+Gen	->D+Sta+Sel+Gen	D+Sta+Sel+Gen
SRCH	->SRCH	SRCH	D+Req+Gui+Sho	->D+Req+Gui+Sho	O+Sta+Exh+InD
SRCH	->SRCH_RQST	O+Sta+Exh+SeR	D+Req+Sea+Sho	->D+Req+Sea+Sho	O+Sta+Exh+InD
SRCH_RQST	->D+Req+Sea+Sho	GENRE	O+Exp+Gui+Sho	->O+Exp+Gui+Sho	D+Sta+Exh+InD
D+Sta+Exh+SeR	->O+Sta+Exh+SeR	O+Re+Exh+InD			

図 3 対話構造木の例

あるとする。コーパスの作成にあたり、41 種類の意図タグのほかに 11 種類の POD を定めた。表 2 に意図タグ以外の POD の種類と内容を示す。なお、コーパスでは、簡単のため、対話構造は二分木であるとする。

2.3 対話構造付与

対話を発話意図の系列とみなし、以下の操作を繰り返すこと

により、対話構造を作成する。

発話の融合 意図タグ系列において、意味的に対応関係にある隣接する 2 つの発話を発話対とする。その対に最も合致する POD をラベルとして付与する。

部分対話の融合 意味的に対応関係にある隣接する部分対話に対して、その POD を対話対とし、その対話対に最も合致する POD をラベルとして付与する。

なお、対象とする対話コーパスのドメインは飲食店案内であり、飲食店を案内することがゴールである。そこで、全体の対話構造に対する POD は案内を意味する“GUIDE”であった。

上記の方針で付与された構造は図 3 に示すように二分木で表される。また、この対話構造木から得られる構造規則を図 3 に示す。作成された対話構造木コーパスは 789 対話 8150 発話の規模である。コーパスの基礎統計を表 3 に示す。

3. 対話構造の解析

3.1 解析手法

本研究では、対話構造の解析過程をチャートで表現する。チャートとは、対話を構成する発話系列 $u_1 \dots u_n$ に対して、発話間に位置する $n+1$ 個の節点の集合 $\{0, 1, \dots, n\}$ とノード間に張られた弧の集合からなるグラフである。弧には、項 (term) とよばれる対話の構造がラベル付けされる。項は以下の 1~3 で定義される。

- (1) 発話 u の意図を l とするとき、 $[u]_l$ は項である。
- (2) 意図 l に対して、 $[?]_l$ は項である。
- (3) X が POD または意図、 $\sigma_k (1 \leq k \leq n)$ が項のとき、 $[\sigma_1, \dots, \sigma_n]_X$ は項である。

任意の意図 l に対して $[?]_l$ を未決定項 (undefined term) という。項 σ における最も左に位置する未決定項を最左未決定項といい、 $lut(\sigma)$ で表す。ラベル付けされた項が未決定項を含む弧を活性弧といい、含まない項を不活性弧という。項 σ が $[u]_l$ 、または、 $[?]_l$ のとき、 $r(\sigma)$ で意図 l を表し、 σ が $[\sigma_1, \dots, \sigma_n]_X$ のとき、 $r(\sigma)$ で POD または意図 X を表すものとする。

対話構造解析は以下の手順で実行する。 $\{0\}$ を節点集合とするチャートを初期状態とし、 k 番目の発話 u_k が入力されるたびに、以下の処理を実行する。

(1) 節点 k を追加する。 u_k の意図 l_k を推定し、節点 $k-1$ と k の間に項 $[u_k]_{l_k}$ でラベル付けされた不活性弧を張る。

(2) 任意の節点 $j (j < k)$ に対して、 $X = r(\sigma_k)$ なる項 σ_k でラベル付けされた j と k の間のすべての不活性弧に対して、以下の処理 (a),(b) を可能な限り実行する。

(a) 対話構造規則 $A \rightarrow X Y \dots Z$ があれば、節点 j と k の間に項 $[\sigma_k, [?]_Y, \dots, [?]_Z]_A$ でラベル付けされた活性弧を張る。

(b) 節点 $i (i < j)$ と j の間の項 γ の最左未決定項が $[?]_X$ であれば、 γ の最左未決定項を σ_k で置き換えた項でラベル付けされた弧を節点 i と $k+1$ の間に張る。

最終的に、節点 0 から n に張られている不活性弧が存在しなければ、それは対話構造解析に失敗したことを意味する。

3.2 対話構造の解析例

3.1 節で述べた解析手法を用いて、表 1 の発話番号 2「はい」が入力されたときの処理の説明を図 4 に示す。の解析を行う。

発話 u_2 が入力されると、節点 1 と 2 の間に項 $[u_2]_{O+Sta+Exh+InD}$ でラベル付けされた不活性弧を張る。一方、 $lut(\sigma_1^2) = O + Sta + Exh + InD$ なので、 $lut(\sigma_1^2)$ を $[u_2]_{O+Sta+Exh+InD}$ で置き換えて、 σ_2^2 とし、 $r(\sigma_2^2) = D +$

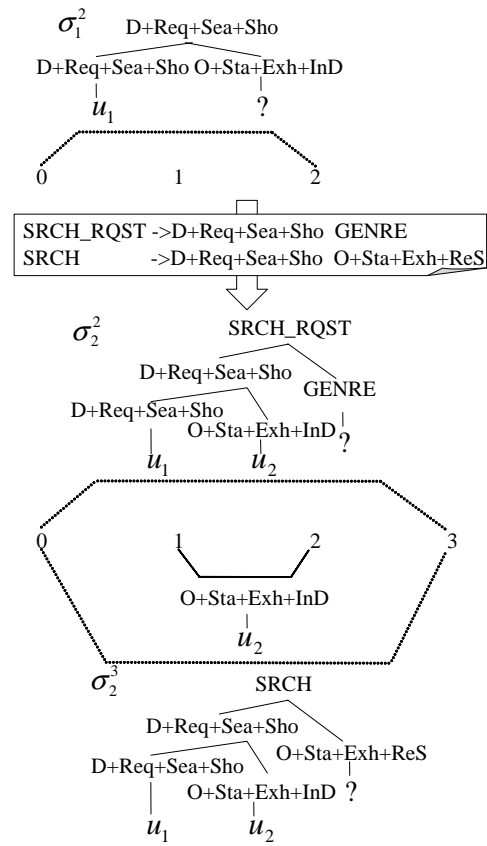


図 4 対話構造解析の例

$Req + Sea + Sho$ に対して、対話構造規則 “ $SRCH_RQST \rightarrow D + Req + Sea + Sho \text{ GENRE}$ ”、及び、“ $SRCH \rightarrow D + Req + Sea + Sho \ O + Sta + Exh + ReS$ ” をそれぞれ適用し、 σ_2^2, σ_3^2 とする。

4. 発話意図の予測

発話系列 $u_1 \dots u_n$ における発話 u_k が入力されたときの次発話 u_{k+1} の意図予測は、その時点までに作成されている対話構造 T_k を用いて実行する。ここで、項 σ_j にラベルつけられた弧の始点を $s(\sigma_j)$ 、終点を $e(\sigma_j)$ で表すとき、対話構造 T_k とは、以下を満たす項の系列 $\sigma_1 \dots \sigma_m$ である。

- (1) $s(\sigma_1) = 0, e(\sigma_m) = k$
- (2) 任意の連続する項 $\sigma_{j-1}, \sigma_j (1 < j \leq m)$ に対して (a),(b) を満たす

- (a) $s(\sigma_j) = e(\sigma_{j-1})$
- (b) σ_{j-1} の最左未決定項を $lut(\sigma_{j-1})$ とするとき、 $r(\sigma_j)$ から $r(lut(\sigma_{j-1}))$ に到達可能。

なお、一般に、このように定義された対話構造は、解析途中のチャートに複数含まれているが、本稿では、ある対話構造が選択されたときに、それを用いて次発話の意図を予測する方法について述べる。すなわち、対話構造を T_k とするとき、次発話 u_{k+1} の意図 \hat{l}_{k+1} は、

$$\hat{l}_{k+1} = \arg \max_{l_{k+1}} P(l_{k+1} | T_k)$$

表 4 発話意図予測の結果

最左未決定項のみ	最左未決定項, 親節点	最左未決定項, 親節点の子節点	最左未決定項, 親節点とその子節点
0.89	0.89	0.93	0.93

として計算する.

対話構造 $T_k = \sigma_1 \cdots \sigma_m$ のもとで次発話の意図が l_{k+1} である確率 $P(l_{k+1}|T_k)$ の計算について述べる. 対話構造のある節点の POD α に対して, その親節点の POD を $p(\alpha)$ と記し, $p(\alpha)$ の子節点のうち最も左に位置する節点の POD を $b(\alpha)$ と記す. すなわち, 最左未決定項 $lut(\sigma_j)$ が $Y = r(lut(\sigma_j))$ であり, かつ, それが σ_j において対話構造規則 $A \rightarrow X \cdots Y \cdots Z$ によって出現するとき, $b(r(lut(\sigma_j))) = X$, $p(r(lut(\sigma_j))) = A$ となる. 対話構造における以上の節点を用いて, 確率 $P(l_{k+1}|T_k)$ を以下のように計算する.

$$\begin{aligned}
P(l_{k+1}|T_k) & \\
&\approx P(l_{k+1}|\sigma_m) \\
&\approx P(l_{k+1}|r(lut(\sigma_m)), b(r(lut(\sigma_m))), p(r(lut(\sigma_m)))) \\
&= \frac{f(l_{k+1}, r(lut(\sigma_m)), b(r(lut(\sigma_m))), p(r(lut(\sigma_m))))}{f(r(lut(\sigma_m)), b(r(lut(\sigma_m))), p(r(lut(\sigma_m))))}
\end{aligned}$$

ただし, f は頻度関数である.

5. 発話意図予測実験

本節では, 発話意図予測手法の有用性を確認するために行った発話意図予測実験について述べる.

対話構造木コーパスの全 789 対話から 689 対話を学習セット, 100 対話を評価セットとし, この評価セットの 100 対話の全 997 発話の意図を予測する. 本実験では, 入力となる T_k は, 評価セットで与えられているものを使用する. この条件のもとでの発話意図予測精度について調べる.

評価方法は, 評価セットで与えられている各発話の意図タグ (正解意図タグ) と出力された意図タグとの一致を調べる.

$$\text{正解率} = \frac{\text{正解発話数}}{\text{実験対象の発話数}}$$

また, 提案手法との比較のために, 以下の 3 つの近似式を含めた 4 つの場合で正解率を調べた.

$$\begin{aligned}
P(l_{k+1}|T_k) &\approx \frac{P(l_{k+1}|r(lut(\sigma_m)))}{f(l_{k+1}, r(lut(\sigma_m)))} \\
&= \frac{P(l_{k+1}|r(lut(\sigma_m)), p(r(lut(\sigma_m))))}{f(l_{k+1}, r(lut(\sigma_m)), p(r(lut(\sigma_m))))} \\
&= \frac{P(l_{k+1}|r(lut(\sigma_m)), b(r(lut(\sigma_m))))}{f(l_{k+1}, r(lut(\sigma_m)), b(r(lut(\sigma_m))))} \\
&= \frac{P(l_{k+1}|r(lut(\sigma_m)), b(r(lut(\sigma_m))), p(r(lut(\sigma_m))))}{f(l_{k+1}, r(lut(\sigma_m)), b(r(lut(\sigma_m))), p(r(lut(\sigma_m))))}
\end{aligned}$$

結果を表 4 に示す. これより, 最左未決定項の周辺情報を使用することにより高い精度で意図予測できることを確認した. その他の近似方法においても, 子節点の情報が有用であることがわかった.

6. おわりに

本稿では, 対話構造木コーパスを用いた発話意図予測手法について述べた. 対話構造木コーパスから得られた構造規則を用いて対話構造を漸進的に作成し, この対話構造に基づいて発話意図の予測を行う. 対話構造木コーパスを参照するとき生じるデータスパースネス問題に対処するため, データの参照は近似的に実行する. 予測実験の結果, 対話構造が与えられた状況下での本手法の利用可能性を確認した.

今後の課題としては, この手法を対話システムに組み込み, 柔軟な対話制御について検討することがあげられる.

謝辞

本研究の一部は, 日本学術振興会科研費基盤研究 (B)(2)(No. 15300045) によります.

文 献

- [1] A. Rudnicky, C. Bennett, A. Black, A. Chotomongcol, K. Lenzo, A. Oh and R. Singh: Task and domain specific modeling in the Carnegie Mellon communicator system. Proc. International Conference on Spoken Language Processing, 2000.
- [2] L. Lamel, S. Rosset, J. Gauvain, S. Bennacef, M. Garnier-Rizet and B. Prouts: The LIMSI ARISE system, Speech Communication, Vol. 31, pp. 339-353, 2000.
- [3] S. Seneff, Response planning and generation in the MERCURY flight reservation system, Computer Speech and Language, vol. 16, pp. 283-312, 2002.
- [4] D. J. Litman and J. F. Allen: Discourse Processing and Commonsense Plans. Phillip R. Cohen, Jerry Morgan, Martha E. Pollack, editors. Intentions in Communication. pp.365-388, MIT Press, Cambridge, MA, 1990.
- [5] N. Kawaguchi, K. Takeda, and F. Itakura: Multimedia corpus of in-car speech communication. J. VLSI Signal Processing, vol.36, no.2, pp.153-159, 2004.
- [6] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura: CIAIR In-Car Speech Corpus -Influence of Driving States-, IEICE Trans. on Information and System, vol. E88-D, No.2, pp.578-582, 2005.
- [7] Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki: Speech Intention Understanding based on Decision Tree Learning, Proceedings of 8th International Conference on Spoken Language Processing, Cheju, Korea, 2004.
- [8] 入江, 松原, 河口, 山口, 稲垣: 音声対話コーパスにおける発話意図タグの設計と評価, 電子情報通信学会論文誌, vol.J88-DII No.10, pp.2169-2173, 2005.
- [9] S. Kato, S. Matsubara, Y. Yamaguchi, and N. Kawaguchi: Construction of Structurally Annotated Spoken Dialogue Corpus, Proceedings of 5th Workshop on Asian Language Resources(ALR-05), pp.40-47, 2005.