

## 自発的音声対話における話者交替の制御に関わる 発話末の統語的・韻律的特徴

堀内 靖雄 小磯 花絵 土屋 俊 市川 薫  
千葉大学

**概要** 千葉大学地図課題対話コーパス中の各発話の最後の部分の統語的・韻律的特徴を次発話者の予測の観点から分析し、以下の結果を得た。(1) 発話末にくる、いくつかの品詞は後続発話の前の無音区間とは無関係に話者交替と高い相関がある、(2) ある特有の表現は、発話末を予測させるシグナルを持つ、(3) ある種の韻律的特徴をもつ助詞は話者交替に相関がある、(4) 山型ピッチパターンは話者交替を伝える、(5) 文脈情報に関与する、ある種の現象は今後の課題である、(6) ラッチング発話を分析することにより、聞き手は聞きながら、話し手の発話進行を予測する。

**キーワード** 千葉大学地図課題対話コーパス・課題指向的対話・自発的音声対話・話者交替・韻律的特徴

## Some Utterance-final Syntactic and Prosodic Characteristics Relevant to the Control of the Speaker Shift Phenomena in Spontaneous Spoken Dialogues

Yasuo Horiuchi, Hanae Koiso, Syun Tutiya and Akira Ichikawa  
Chiba University

**Abstract** Syntactic and prosodic features of the final part of preutterances in Chiba Map Task Dialogue Corpus have been analyzed that are useful in the prediction of the speakerhood of the next pre-utterance. A preutterance is defined as a stretch of speech bounded by pauses. Major findings are (1) that some words from the restricted set of part of speech are highly relevant to the shifting of the speakerhood regardless of the silence preceding the second utterance, (2) that some particular expressions give signals for the approach of the end of the utterance, (3) that certain prosodic features of the utterance of postparticles are relevant to the shifting of speakerhood, (4) that the occurrence of the "hat" type of pitch pattern, namely "rise-then-fall" convey information on the shifting of speakerhood, (5) that certain phenomena concerning contextual information are left to be unanalyzed in future and (6) that the analysis of the "latching" utterances shows that the hearer are anticipating the initiator's course of utterance while listening.

**keywords:** Chiba Map Task Dialogue Corpus, task-oriented dialogue, spontaneous spoken dialogue, speaker shift, prosodic features

## 1 はじめに

人間のように自然で協調的な対話を行なえる対話システムの構築のためには以下の二つの機能が必要となる。

1. システムがユーザの発話停止を予測し、さらに、ユーザの次の発話行動を予測する
2. システムが自分の発話を維持し、あるいは、ユーザの発話を誘うために必要なシグナルを制御する

本稿では、発話者の交替／非交替の観点から千葉大学地図課題コーパスを分析し、人間同士の対話における発話制御の方法について考察する。その際、話者の交替／非交替に関与する要因について、(1) 語彙レベル、(2) 韻律レベル、(3) 文脈レベルを想定し、それぞれのレベルごとに分析を行なう。また、人間同士の対話ではラッチング発話（相手の発話終了の直前から発話を開始する現象）が頻繁に生じるが、このラッチング発話について、対話コーパスを用いて分析する。最後に、以上の分析で得られた結果から、対話システム構築のための発話生成理解者モジュールに必要な機能をまとめる。

## 2 発話の時間的關係と話者の交替／非交替

本稿では 0 秒を越える無音区間に挟まれた一続きの音声連続<sup>1</sup>を発話断片 (utterance fragment) と定義し、分析の対象とした<sup>2</sup>。二人の話者による対話において、二つの発話 ( $u^1, u^2$ ) があるとき、開始時刻をそれぞれ  $t_1^1, t_1^2$ 、終了時刻を  $t_2^1, t_2^2$  とする。 $t_1^1 < t_1^2$  かつ、 $t_1^1$  と  $t_2^2$  との間にこれら二つの発話以外の発話が存在しないとき、 $u^1$  を「先行発話」、 $u^2$  を「後続発話」と呼ぶ<sup>3</sup>。また、先行発話の発話者を  $s^1$ 、後続発話の発話者を  $s^2$  とすると、 $s^1 \neq s^2$  の場合を話者交替、 $s^1 = s^2$  の場合を話者非交替と呼ぶ。先行発話と後続発話の関係としては、話者の交替／非交替の関係以外にも、それらの間の時間的關係がある。 $\Delta = t_2^2 - t_1^1$  を定義すると、先行発話と後続発話との時間的關係には次のような可能性が存在し、それらの後続発話を以下のように分類・命名する<sup>4</sup>。

表 1: 発話間の時間的關係による後続発話の分類

時間的關係	後続発話の分類
$\Delta > 100$	intervalled utterance
$100 \geq \Delta > 0$	immediate utterance
$0 \geq \Delta > -100$	latching utterance
$-100 \geq \Delta$	interrupting utterance

以上の分類に従って、千葉大学地図課題対話コーバ

<sup>1</sup>本稿では促音などによる無音区間は音声連続として扱った。  
<sup>2</sup>ただし、以下では簡単のため、発話断片のことを単に発話 (utterance) と呼ぶ。

<sup>3</sup> $t_1^1 = t_1^2$  のときには同時発話開始と呼ぶが、本稿では分析の対象外とする。

<sup>4</sup>ただし、本稿では interrupting utterance は分析の対象外としている。

ス中の一対話 (jiniu) における冒頭 5 分間のデータを後続発話の話者交替／非交替とその時間的關係について調べた (表 2)。

表 2: 発話間の時間的關係と話者の交替／非交替

時間的關係	後続発話	
	話者非交替	話者交替
intervalled utterance	99	95
immediate utterance	34	40
latching utterance	—	30
interrupting utterance	—	50

この観察から、 $\Delta > 100, 100 \geq \Delta > 0$  のいずれの場合にも、話者が交替する場合としない場合がほぼ同じ頻度で生じている。このことは、いわゆる無音区間の存在／非存在や長短は、話者の交替／非交替とは中立であることを示唆する。そこで、 $\Delta$  の分析区間を細かくし、同様の分析を行なったが、同じ傾向がみられた (図 1)。

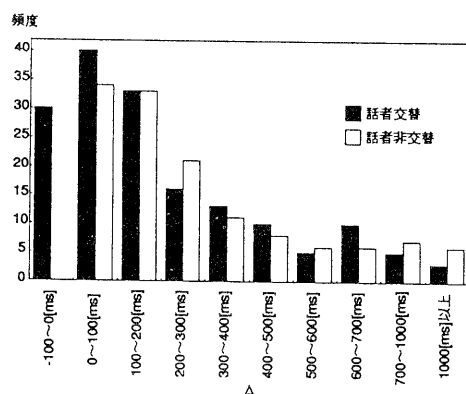


図 1: 話者の交替／非交替と先行発話に対する時間的關係

対話システムにおいて、先行話者の発話の終了時 ( $t_2^1$ ) 付近で話者が交替するか否かを予測・判断することは、対話の円滑な進行には欠かすことができない。しかし、以上の分析結果は、話者の交替／非交替の予測には無音区間の存在があまり寄与しない可能性を示唆するものである。そのため、話者の交替の有無を予測するという観点からは、無音区間以外の情報への着目が必要であると考えられる。そこで本稿ではまず、先行話者の発話の終了 ( $t_2^1$ ) と後続発話の交替／非交替の予測に関与する要因として、(1) 語彙レベル、(2) 韻律レベル、(3) 文脈レベルを想定し、それらの観点から対話コーパスの分析を行なう。

表 3: Δ による話者交替/非交替の品詞別頻度 (話者交替 T; 話者非交替 H)

Δ	名詞		動詞		副詞		接続詞		連体詞		感動詞		間投詞	
	T	H	T	H	T	H	T	H	T	H	T	H	T	H
-100~0 ミリ秒	2	-	6	-	0	-	0	-	0	-	5	-	0	-
0 ~ 100 ミリ秒	4	4	2	0	0	3	0	3	0	0	10	1	1	6
100 ~ 200 ミリ秒	5	6	3	1	1	6	0	3	0	1	14	4	0	2
200 ~ 300 ミリ秒	0	0	1	1	0	2	0	3	0	1	8	3	0	2
300 ~ 400 ミリ秒	1	2	0	0	0	3	0	1	0	0	6	2	0	1
400 ~ 500 ミリ秒	2	2	0	0	0	3	0	0	0	0	7	1	0	0
500 ~ 600 ミリ秒	0	1	0	0	0	0	0	0	0	1	3	1	0	1
600 ~ 700 ミリ秒	0	1	0	0	0	2	0	0	0	0	8	2	0	0
700 ~ 1000 ミリ秒	0	2	0	0	0	4	1	0	0	0	3	0	0	1
1000 ミリ秒以上	0	0	0	0	0	0	0	0	0	0	1	4	0	0
合計	14	18	12	2	1	23	1	10	0	3	65	18	1	13

Δ	格助詞		接助詞		終助詞		その他助詞		助動詞		その他		合計	
	T	H	T	H	T	H	T	H	T	H	T	H	T	H
-100~0 ミリ秒	6	-	6	-	4	-	0	-	4	-	0	-	30	-
0 ~ 100 ミリ秒	4	6	6	1	7	2	0	1	3	0	0	4	40	34
100 ~ 200 ミリ秒	5	5	1	2	2	0	2	2	0	0	1	33	33	
200 ~ 300 ミリ秒	5	5	0	1	1	1	1	0	0	0	1	16	21	
300 ~ 400 ミリ秒	0	1	0	0	4	1	0	0	2	0	0	13	11	
400 ~ 500 ミリ秒	1	1	0	0	0	0	0	1	0	0	0	10	8	
500 ~ 600 ミリ秒	0	1	0	1	2	0	0	0	1	1	0	5	6	
600 ~ 700 ミリ秒	0	0	1	0	0	1	0	0	0	0	0	10	6	
700 ~ 1000 ミリ秒	0	0	0	0	0	0	0	1	1	0	0	5	7	
1000 ミリ秒以上	0	1	2	0	0	1	0	1	0	0	0	3	7	
合計	21	20	16	5	20	8	1	6	13	1	6	165	133	

### 3 語彙レベルの分析

先に言及した5分の対話における各発話の発話末の品詞と Δ に関する話者交替/非交替の頻度を表 3 に示す。この表から、話者交替/非交替と品詞との間に、相関がみられるものが存在することがわかる。たとえば、動詞や助動詞の場合、終止形や仮定形、命令形が発話末に多く出現するが、これらの品詞の後では話者の交替が生じやすい傾向にある。また、「はい」などのあいづち表現が大半を占める感動詞や「ね」「よね」「か」など終助詞でも同様の傾向が見られた。一方、副詞や接続詞、「えっとー」、「あの一」などのいわゆる言いよどみ表現を含む間投詞のあとでは、話者交替はほとんど生じなかった。このように、発話末の品詞には、話者交替/非交替との相関の高いものが存在しており、これらの表現が話者交替/非交替に関わる重要な signal の一つとなっていると考えることができる。

次に先行発話と後続発話との時間的関係の観点から分析を行なうと、副詞や間投詞などは Δ の値が 0 ~ 1 秒の広い範囲に分布していることがわかる。このことは、これらの品詞で終了する発話の大半は話者非交替となり、相手話者の発話が起りにくいということを示唆している。つまり、これらの品詞のあとの無音区間の存在は、話者交替の signal とはならないという

ことを意味する。

一方、たとえば「ください」に対する後続発話は、大半が latching utterance、つまり先行発話の終了を待たずに発話を開始するような発話となっている (表 4)。これは、「ください」のようにある程度時間長を伴って

表 4: 話者非交替における「ください」と「か」の後続発話の分類

発話タイプ	「ください」	「か」
latching utterance	6	1
immediate utterance	2	3
intervaed utterance	3	8
発話数	11	12
平均 Δ	45[ms]	216[ms]

発話される語彙は、発話の終了を待たずに signal となりうるためであると考えられる。仮に以上の考察が正しいとすると、話者交替の signal となる表現の中でも、より短い時間長で発話されるものは、Δ の値が正の値をとることが予想される。そこで、同様に文末表現であり、かつ話者交替の生じやすい終助詞「か」に関しても後続発話のタイプを調べたところ (表 4)、「ください」にくらべて「か」のほうは latching utterance はそ

れ程多くなく、 $\Delta$  の値が正の値をとることが多かった。

以上の観察から、次の2点を指摘することができるのではないかと考えられる。

- 聞き手にとっては語彙が話者交替の signal となっており、先行発話が停止したという情報、つまり発話後の無音区間は、むしろ話者が交替するまでに要した時間、すなわち話者交替の結果である。
- 話者交替に関与する signal は、先行発話の終了時刻 ( $t_{1/2}$ ) よりも前に存在しうる。

本節では、話者の交替/非交替に関わる要因として、発話末の語彙に着目して分析を行ったところ、いくつかの品詞は話者の交替/非交替に相関を示すことが確認された。しかし、名詞や格助詞、接続助詞のように、品詞と話者の交替/非交替との相関が低いものも存在する。とくに、今回の分析の対象とした対話データには、発話末には助詞がくるのがかなり多いにもかかわらず（全発話の 32%）、終助詞以外の助詞に関しては、話者の交替/非交替との相関はみられなかった。そのため、話者交替の制御という観点からは、このような品詞に関しては、別の観点から分析する必要があると考えられる。

#### 4 韻律レベルの分析

本節では、ほぼ同じ割合で、話者の交替と非交替が生じる格助詞を対象に、韻律的な観点から分析を行う。

まず話者交替時における先行発話末の1モーラの韻律的特徴を観察したところ、(1) ピッチパターンが山型<sup>5</sup>で長音化する [6]、(2) パワーが大きい（最後のモーラでパワーが落ちない）、という2つの特徴を持つ発話が数多く観察された。

次に、話者非交替の場合を観察したところ、(1) 最後の1モーラのピッチパターンが平坦である、(2) パワーが小さくなる、(3) 長音化しない、あるいは長音化してもパワーは小さくなり、ピッチも下降する、という3つの傾向が多く観察された。

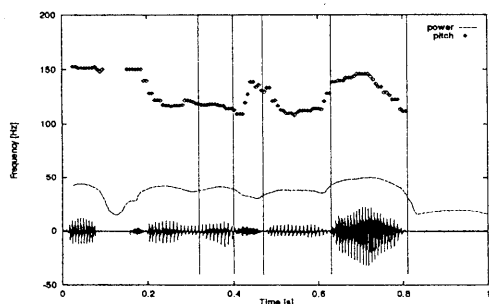


図 2: 山型のピッチパターンの例（「ますのでー」）

<sup>5</sup>山型のピッチパターンの例を図 2 に示す。

このように、発話末に存在する格助詞の韻律的特徴を調べたところ、話者交替を引き起こす発話と引き起こさない発話との間に顕著な差がみられた。以上の観察から、先行発話の発話末における韻律的特徴が話者交替/非交替に関与していることが示唆された。

とくに、図 2 にあるような、最後 1 モーラのピッチパターンが山型になり、かつ長音化されているものに関して調べたところ（表 5）<sup>6</sup>、ほとんどが話者交替であった（91%）。また、話者交替の後続発話はほとんどがあいづちであり（話者交替のうちの 95%）、latching utterance または immediate utterance がその 85% を占めた。

表 5: 山型ピッチパターンの後続発話

発話タイプ	話者交替発話		話者非交替発話
	あいづち	以外	
latching utterance	10	0	—
immediate utterance	7	0	1
intervaled utterance	3	1	1
発話数	20	1	2

さらに山型ピッチパターンの極大点の時刻からあいづちの開始時刻までの時間を測定したところ、ほぼ 200 ~ 300 ミリ秒であった。一方、あいづち以外が発話される場合には、それよりも大きな値を示した。このことは、山型ピッチパターンに関して、特徴が最も顕著なのはその極大点であり、その後ろに続く長音の長さや  $\Delta$  で特徴付けられるものではないことが示唆される。さらに、山型ピッチパターンの発話に対する聞き手の発話があいづち以外、あるいは発話非交替の場合には  $\Delta$  が大きくなる傾向がある。

このことから、山型のピッチパターンについて、以下のような発話制御が推測される。

山型ピッチパターンの発話は、何らかの応答を期待する発話である。すなわち、このパターンの発話の後続発話として「はい」などのあいづちが発話されれば、先行発話の意図は聞き手に伝わったものとして、聞き手のあいづちに続けて、先行発話の話者が発話を続けることができる。もし「いいえ」を意味するような応答発話、あるいは応答がまったくない（あるいは一定以上待っても応答がない）場合には先行発話の意図が伝わっていないと考えられ、その発話を補足するような発話を先行話者は行なう。

以上をまとめると、山型のピッチパターンは話者交替/非交替に関して、重要な情報を持っていることが分かる。すなわち、山型のピッチパターンは (1) 相手

<sup>6</sup>対話コーパス 3 分のデータを対象に分析を行った。

の反応を何う発話であり、(2) その特徴は山型ピッチパターンの極大点にあり、(3) その発話に対する応答発話によって、話者の今後の発話の展開が変化する、ということが示唆された。今後の韻律ラベルの分析としては、他の品詞の分析を行なうと同時に、最後の1モーラに限定せずに分析を行なっていく必要があると考えられる。

### 5 文脈レベル

以上のように、語彙的あるいは韻律的な特徴から、話者交替/非交替の分析を行ってきたが、これらの観点からは説明のつかない現象も存在する。例えば、以下の対話例にあるように、先行の発話の一部にあたかも同調するようにして発話したり、あるいは先行発話に内容的に続けるようにして発話する例などが地図課題対話コーパスには少なからず観察される<sup>7</sup>。しかし、これらの現象が、どのようなメカニズムで生じているかは、語彙や韻律的なレベルのみからでは説明がつかない。これらは、語彙や韻律などの局所的な側面のみならず、文脈レベルからも考察する必要があるという点で、今後の重要な課題とされる。

- (1) G: その右側のところから下に<256>
- (2) G: ていう感じで<-96>
- (4) G:           \*のー
- (3) F: みぎがわー\*のところ<144>
- (5) G: からし\*たに<352>
- (6) F:           \*した
- (7) F: はい<352>

### 後続発話が先行発話を補完(補完発話)

話者 G の発話 (5) は話者 F の発話 (3) に続けて発話されている。これら二人の話者による発話を合わせると (1) の発話を繰り返したものになっている。これらは、Δ が正の値をとる場合もあるが、先行発話の一部から重複しつつ補完を行うような形式も観察される。先行発話に割り込み、同じあるいは類似した発話をほぼ同時に行なう発話(同調的発話)

(4) の発話は相手の発話と同じ発話をほぼ同時に行なっている例である。この対話例のように短い発話のみならず、かなり長い時間、同時に話す状況も比較的多数観察された。これらの現象では、聞き手側が発話のタイミングと表現内容をかなり正確に予測する必要があるが、とくに他の部分に比べて発話速度が遅くな

り、長音化や促音部分などで起こりやすい傾向にあることが、定性的に観察された。

### 先行発話の一部を繰り返す発話(反復発話)

(3) の発話は (2) の発話末の「で」とほぼ同時に発話されている。これは、(1) の発話の一部、すなわち「右側」を繰り返すことにより、確認を行なっていると考えられる。このように、先行発話の一部を繰り返す発話は、多くが相手あるいは自己の確認として機能する。

### 6 latching utterance

二人の人間による対話は語彙的、韻律的、文脈的レベルにより、話者制御されていることを示したが、これらは、話者交替か非交替かという次話者の予測と同時に、先行発話の終了の予測も重要な課題となる。表 3 にみられるように、地図課題対話では、latching utterance やあるいは immediate utterance のように、先行発話の終了を確認してから発話しているのではなく、終了を予測して発話を開始しているものが非常に多く存在する(話者交替の 42%)。この先行発話の終了の予測には、その終了を示唆する何らかの signal (語彙・韻律・文脈レベル) が重要とされる。そこで、千葉大学地図課題コーパス、約 5 時間のデータ<sup>8</sup>に出現した 479 個の latching utterance を対象に、その先行発話の発話末付近の特徴を調べた。まず第一に、先行発話末の語彙的情報について調べてみたところ、表 6 に示す結果が得られた。

表 6: 全 latching utterance の先行発話の分類

文末表現	「下さい」	20
	「よ」「ね」等の終助詞	130
	その他	70
あいづち		62
その他		197

この表と上述の分析から、latching utterance の多く(220 発話 46%) は語彙的なレベルである文末表現から予測されることが示唆される。また、先に述べたように、たとえば先行発話が山型のピッチパターンの発話の場合、そのあとには「あいづち」が挿入されやすい。これは、「あいづち」が発話された場合、その次の発話は山型のピッチパターンを持つ発話に続く発話であるため、すでに発話準備が行なわれており、即座に発話可能であるため、latching utterance や immediate utterance が多いと考えられる。「その他」にあたる 197 発話に対しては、韻律的、文脈的なものなどが予測に関与していると推測される。

また、全 latching utterance の 57% にあたる 274 の latching utterance があいづちであった。そこで、

<sup>8</sup>お互いに親近性のない男性ペアによる 30 対話

<sup>7</sup>対話テキスト中で、行末の「<>」内の数字は発話間の無音区間をミリ秒単位で表わしており、負の値はラッチングを意味している。また「\*」は割り込みを表わしており、後続話者が割り込んだ場所に記述されている。

これら「あいづち」がラッチングした先行発話について分類を行なうと表7のようになった。

表7: ラッチングしたあいづちの先行発話

文末表現	「下さい」	9
	「よ」「ね」等の終助詞	86
	その他	37
あいづち		8
その他		134

統語的な情報により132発話(48%)は分析されるが、それ以外についても、多くの発話はその先行発話の韻律的特徴により分析可能である。韻律的特徴には、当然のことながら、上述の山型のピッチパターンを持つものが圧倒的に多い[6]。

#### 7 対話制御モデルの構築にむけて

本稿では後続発話に対する先行発話の語彙的・韻律的な特徴についての分析を行なった。その結果、発話間の時間的・話者の関係についてのいくつかの知見が得られた。人間のように自然で協調的な対話システムを実現するための対話制御モデルには以下のような機能が必要となると考えられる。

##### ・次の発話行動の予測

ユーザの発話中の特徴により、ユーザの発話停止後の発話の計画を推測する。もし、ユーザが続けて発話をした(話者非交替)のであれば、システムは発話を行わず、逆にユーザが相手すなわちシステムに反応を求めているのであれば、それに促した発話を行なう。

##### ・ユーザの発話停止の予測

ユーザの発話中の特徴により、発話停止を予測する。これにより、発話が停止してから、それに応答する動作を起こすのではなく、若干早いタイミングで次の発話に関する動作を開始することが可能となる。例えば、もしユーザが山型ピッチパターンで発話を行なった場合、システムは状況に応じて、即座に反応することが必要である。すなわち、(1) あいづちをうつ、(2) 質問や反対の意見を言う、(3) 反応をせずに保留する、という三つの動作のどれかを選択する必要があるということである。

これらの機能を有するモデルを構築することにより、対話システムの発話制御をより協調的なものにするができると考えられる。

#### 8 まとめ

本報告では、発話者の交替/非交替の観点から、人間同士の対話における発話制御について、コーパスを用いて分析した。その際、話者の交替/非交替に関する要因として、語彙レベル・韻律レベル・文脈レベルの立場から分析を行なった。結果として、以下のような分析結果が得られた。

- いくつかの品詞は話者交替/非交替に相関を示し、また、発話末における韻律的特徴も話者交替/非交替に相関を示すものがある。
- 話者非交替と相関の高い品詞は  $\Delta$  が大きくても話者交替が起こりにくい。
- latching utterance の分析により、聞き手が話し手の発話を予測しながら聞いているということが示唆され、また「ください」のような文末表現には発話終了を待たずに予測可能なものが存在する。
- 山型のピッチパターンは話者交替/非交替に関して、特に重要な情報を持っており、その発話は通常、相手の応答を要請するものが多い。また、山型ピッチパターンの特徴は極大点にあり、その時点からの応答の分類(あいづちあり/それ以外の応答あり/応答なし)によって、話者のそれ以降の発話展開が変化することが推測された。
- 文脈レベルで話者交替/非交替に関与する情報がある。

以上の結果から、対話システム構築のための対話制御モデルには以下の機能が必要となることを述べた。

- 次の発話行動の予測
- ユーザの発話停止の予測

今後は本稿で述べたさまざまな分析をさらに多くの対話例で行ない、より一般的な特徴の抽出を行なうと同時に、発話制御のモデル化を行なう予定である。

#### 参考文献

- [1] Levinson 1983: *Pragmatics*. Cambridge.
- [2] J. B. Pierrehumbert and M. E. Beckman 1988: *Japanese Tone Structure*. The MIT Press.
- [3] Sacks H., E. A. Schegloff and G. Jefferson 1974: A simplest systematic for the organization of turn-taking in conversation. *Language* 50(4), 696-735.
- [4] Sacks, H., E. A. Schegloff and G. Jefferson 1978: A simplest systematic for the organization of turn-taking in conversation. *Studies in the organization of conversational interaction*, ed. by J. Schenkein, 7-55. New York: Holt.
- [5] 小磯花絵 他 1994: 地図課題コーパス(中間報告). 人工知能学会研究会資料 SIG-SLUD-9402. 25-30.
- [6] 小磯花絵, 堀内 靖雄, 土屋 俊, 市川 薫 1995: 下位発話単位の音声的特徴と「あいづち」との関連について. 人工知能学会研究会資料 SIG-SLUD-9502.
- [7] 杉藤 美代子 1989: 談話におけるポーズとイントネーション. 杉藤 美代子編. 『日本語の音声・音韻(上)』. 講座日本語と日本語教育(第2巻). 明治書院.
- [8] 中尾 信弥, 塚田 元 1993: 協調的対話における発話パターンの特徴分析. 人工知能学会研究会資料, SIG-SLUD-9302-1, 1-8.
- [9] メイナード 1993: 『会話分析』. くろしお出版.