

## 話しことば対話によるテキスト内容の伝達

杉山 聡 堂坂 浩二 川端 豪

NTT 基礎研究所

〒 243-0198 神奈川県厚木市森の里若宮 3-1

{sugiyama,dohsaka}@atom.brl.ntt.co.jp kaw@idea.brl.ntt.co.jp

本稿では、コミュニケーションにおける話しことばと書きことばの特性を比較する実験結果に基づき、話しことばは、書きことばと比較して、少ない情報量で必要な情報を伝達できる特性をもつことを示す。話しことばによるコミュニケーションと書きことばによるコミュニケーションを比較するために、書きことばテキストの内容を音声で伝達するというタスクを設定し、このタスクを話しことばで遂行するシステム(話しことば対話システム)と、書きことばで遂行するシステム(読み上げシステム)を作成した。この2つのシステムを使って書きことばと話しことばの特性を比較した実験の結果、話しことばには少ない発話モーラ数で必要な情報を伝達できるという特性があることが分かった。

キーワード: 話しことば, 自然言語生成, 話しことば生成, 音声対話

## Communicating the Content of Written Text in Spoken Dialogue

Akira Sugiyama Kohji Dohsaka Takeshi Kawabata

NTT Basic Research Laboratories

3-1 Morinosato-wakamiya, Atugi-shi, Kanagawa, 243-0198 Japan

This paper shows that the necessary information can be communicated using a smaller amount of information in spoken dialogue than in written language communication based on the result of the experiment in which the property of spoken dialogue was compared with that of written language communication. In that experiment, we counted the number of the moras that were uttered to communicate the content of a given written text in either of spoken dialogue or written language communication. The result of the experiment shows that the information that hearers require can be communicated using a smaller number of moras in spoken dialogue than in written language communication.

keyword: spoken language, natural language generation, spoken language generation, spoken dialogue

## 1 はじめに

話しことばによるコミュニケーションは人間同士の基本的な情報伝達手段であり、我々は、話しことばを用いて快適に情報を交換することができる。この話しことば対話が人間同士のコミュニケーションで大きな役割を果たしているのと同様に、計算機と人間とのコミュニケーションにおいても大きな役割を果たすことが期待される。これまでに、話しことばの特徴を分析した報告<sup>4), 8)</sup>や、話しことばの特徴をまねた話しことば生成を計算機で行うことを目指した研究<sup>1)</sup>が行われている。しかし、話しことばの特徴をもったことばでコミュニケーションを行うことが、計算機と人間との情報伝達にどのように貢献するのかわかっていない。話しことばによるコミュニケーションの特性を調べることは、今後の話しことば対話の研究に対する指針となると考える。

話しことばによるコミュニケーションの一般的な特性を捉えることは、その対象が広すぎるため困難である。本研究では、書きことばによるコミュニケーションと比較することによって、話しことばによるコミュニケーションの特性を得ることを目指す。書きことばによるコミュニケーションとの比較のため、書きことばで記述されたテキストの内容を計算機がユーザに音声で伝達するというタスクを設定した。このタスクで、テキストを話しことばで伝達する実験と、書きことばで伝達する実験を行い、それぞれの実験で課題の達成に要した発話モーラ数と課題達成率を比較対象として比較することで、話しことばによるコミュニケーションの特性を調べる。本研究ではこの実験のために、テキストの内容を話しことばに変換して伝達するシステム(話しことば対話システム)と、テキストを書きことばのまま伝達するシステム(読み上げシステム)を作成した。これらのシステムを用いて実験を行い、話しことばによるコミュニケーションには、書きことばによるコミュニケーションに比べて少ない発話モーラ数で必要な情報を伝達できるという特性があることが分かった。

以下、書きことばで記述されたテキストから話しことばを生成する際にとりいれた話しことばの特徴を2節で説明し、3節でその特徴を持った話しことばを書きことばから生成する実験システムの説明をする。4節で話しことばによるコミュニケーションと書

きことばによるコミュニケーションの特性を比較する実験の内容を説明し、5節でその実験の結果の報告と分析をする。

## 2 話しことばの特徴

人間同士の音声対話では、話しことばという、書きことばとは異なった特徴を持つことばが用いられる<sup>7)</sup>。本研究では、話しことばの特徴の中で書きことばとは特に異なる特徴として以下の3つに注目した。

- 発話単位が短い
- 任意のタイミングで問い返すことができる
- 情報を段階的に詳細化する

### 2.1 人間同士の対話に見られる話しことばの特徴

上記の3つの特徴を、人間同士の対話<sup>2)</sup>をもとに説明する。

発話単位が短い 話しことばでは、書きことばで明示的に示される単位である文よりも短い単位で発話される。人間同士の対話<sup>1)</sup>では、次のように短い単位で発話が区切られ、対話相手の発話<sup>2)</sup>が挿入される様子が観察される。

十二月の はい 二十六日から はい 二十九日まであつ二十八日までですか はい その三日間はやっておりません あーそうですか

任意のタイミングで問い返すことができる 聞き手は、話し手の発話に対して任意のタイミングで問い返すことができる。話し手は、問い返しが行われた時点で伝達を中断し、問い返しに対する応答を行う。

夜の うん ファンティリュージュ うん これは新しいから あつ ぜひお薦め 何時からですか これは七時半からです

<sup>1</sup>東京ディズニーランドに関しての初心者とエキスパートとの電話による対話。初心者は知りたい情報についてエキスパートに質問し、エキスパートは自身の知識や資料に基づいて説明する。

<sup>2</sup>下線の引かれていない部分は情報を提供する側の発話を表し、下線の引かれた部分は対話相手の発話を表す。

情報を段階的に詳細化する 話し手はすべての情報を一度に伝達するのではなく、少しずつ情報を詳細化しながら伝達する。次の例では、対象(トゥーンタウン)を話題にあげてから、その属性を1つずつ加えている。

トゥーンタウンっていいまして はい あの  
トゥモローランドの はい 奥にできました  
はい で あの ミッキーやミニーが住んでい  
る町並みを再現したような形で はい

また、聞き手からの「より詳しく知りたい」という要求を受けることで詳細化をすることもある。

## 2.2 話しことばでのテキストの伝達

これらの話しことばの特徴を、書きことばで記述されたテキストを音声で伝達するというタスクで、テキストを話しことばで伝達すると、どのような伝達が可能になるのか説明する。

発話単位が短い 書きことばは句点によって明示的に文という単位に分けられる。例えば、図1の書きことばで記述されたテキストは4つの文で構成される。書きことばで伝達する場合は、この文を単位として発話し、文単位で確認をとりながら伝達を進める。一方、話しことばで伝達する場合は、文の途中で適度にポーズを置き、聞き手のあいづちを認識することで聞き手の理解を確認しながら伝達を進める。

話しことばによるコミュニケーションでは、書きことばの場合よりも頻繁に聞き手の理解を確認しながら伝達を進める特徴がある。

任意のタイミングで問い返すことができる 書きことばで伝達する場合、問い返しに対して文単位で応答する。話しことばでの伝達では、文の伝達の途中でも問い返しを受け付け、即座に応答を返す。応答の単位も、図2の対話例<sup>3</sup>のように、文よりも小さい単位で、応答として適切な部分を推定し、その部分だけを伝達する。

情報を段階的に詳細化する 書きことばでの伝達の場合はテキストに書かれてある通りに読み上げる。

<sup>3</sup><P>は話し手がポーズを置いたことを表す。

話し手: NTTでは<P>  
聞き手: はい  
話し手: 光源として<P>  
聞き手: はい  
話し手: 精密制御色素レーザーを用いた実験を行い<P>  
話し手: 約20秒のホログラフィ動画の記録再生に  
聞き手: もう一度  
話し手: 約20秒のホログラフィ動画の

図2: 対話例

話しことばでの伝達の場合は、詳細な情報を記述している部分は省略しながら伝達することが可能である。例えば、図1の1文目の伝達として、図3のように最初の伝達では、「結晶材料」の内容を詳細に説明している部分を省略して伝達し、問い返されてから省略した詳細部分を伝達するということが行われる。

話し手: NTTでは<P>  
聞き手: はい  
話し手: 新しい結晶材料を発見しました<P>  
聞き手: 結晶材料  
話し手: 結晶材料というのは  
聞き手: はい  
話し手: 超高速ホログラフィ動画記録を可能にする  
というものです

図3: 対話例

## 3 実験システム

話しことばによるコミュニケーションの特性を調べるための実験道具として、2節で説明した特徴を持った話しことばで、書きことばで記述されたテキストの内容を伝達する話しことば対話システムと、書きことばのまま伝達する読み上げシステムを作成した。本節ではそれぞれのシステムを簡単に説明する。

### 3.1 話しことば対話システム

話しことば対話システムは、書きことばで記述されたテキストの内容を話しことばの特徴を持ったこ

NTTでは、超高速ホログラフィ動画記録を可能にする新しい結晶材料を発見しました。この新材料を用いると、極低温において、書き込みレーザー光の周波数を変えて記録媒体に情報を重ね書きする「周波数多重記録」が可能となり、静止画像なら約1000万枚(テレビ映像換算100時間分)が記録できます。さらに光反応速度が速いため、周波数を連続的に変化させることにより実際に動いている物体の記録も実現でき原理的には10億分の1秒の高速記録も期待されます。NTTでは、光源として精密制御色素レーザーを用いた実験を行い、約20秒のホログラフィ動画の記録/再生に成功しました。

図1: 書きことばで記述されたテキストの例(テキスト1)

とばに変換して音声で伝達し、ユーザの発話も音声で受け取る音声対話システムである。

話しことば対話システムの構成を図4に示す。システムは「書きことば構造抽出部」「書きことば構造から話しことばへの変換部」「言語表現理解部」「発話実行部」の4つのモジュールで構成される。

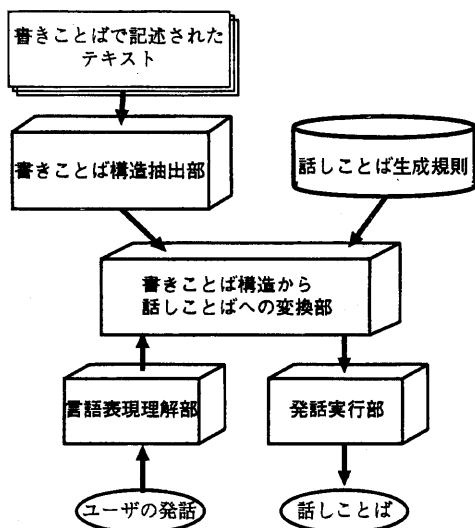


図4: 話しことば対話システムの構成

書きことば構造抽出部では、話しことばへの変換に必要な情報(書きことば構造)をテキストから抽出し、この書きことば構造を付加したテキストを出力する。この処理は対話が実行される前に前処理として行う。言語表現理解部ではユーザの発話を音声認識し、認識した発話文字列から発話意図を生成する。音声認識システムには、NTTヒューマンインタフェース研究所音声情報研究部で開発した音声認

識エンジンREX<sup>6)</sup>を使用した。書きことば構造から話しことばへの変換部では、発話履歴とユーザの発話意図とテキストから話しことばを文字列で生成する。発話実行部は、この文字列に対応した音声データを用いて音声を出力する。今回の実験では、あらかじめキャプチャーした音声を生データとして利用した。

2.2節で説明した、書きことばで記述されたテキストから話しことばを生成する方法を説明する。

発話単位が短い情報を伝達する単位として文節を利用した。1つの文節を伝達するごとに、ポーズを置くか置かないかの判断と次に伝達する文節を決定する。このことによって、話しことばのように短い単位でユーザの理解を確認しながら、次に伝達することを決めることができる。

テキストの形態素解析にはJUMAN<sup>5)</sup>を用い、構文解析にはKNP<sup>3)</sup>を利用した。

任意のタイミングで問い返すことができる この特徴を実現するためには、以下の2つのことができないなければならない。

- ユーザの発話に即座に対応すること
- 問い返しに対して適切な応答を生成すること

前者に関しては、音声認識器とは別に音声区間を監視する装置を用意し、音声区間の始まりを検知した時点で発話実行部が音声データを出力していた場合はこの出力を中断する。ユーザの発話が終了して音声認識器が結果を出力するまで発話生成を中断し、認識結果を用いて次に伝達内容を計算し直す。

後者については、ユーザの問い返しに対する適切な応答部分を推定するために、テキストの係り受け構造を利用した。例えば、図1のテキストの4文目を

構文解析すると図5の係り受け構造が得られる。図2の対話例での聞き手の問い返しに対して、伝達していた文節(記録/再生に)に係る文節(ホログラフィ動画の)を辿り、木構造の葉の文節(約20秒の)から応答する。

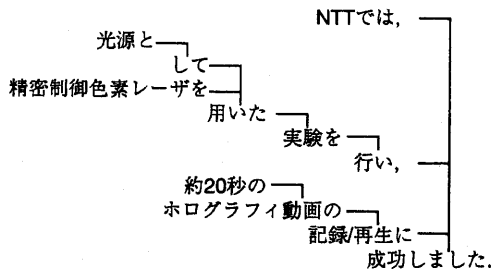


図5: 係り受けの木構造

情報を段階的に詳細化する ある文節に対する詳細情報となっている箇所を特定するために、テキストの連体修飾節の構造を利用した。例えば、図1のテキストの1文目では「超高速ホログラフィ動画記録を可能にする」の部分が「結晶材料」の連体修飾節である。連体修飾節の部分を飛ばして伝達することによって、簡略した情報伝達を行う。問い返された場合は省略した部分を伝達する。

### 3.2 話しことば対話システムの実行例

話しことば対話システムを用いてテキストを伝達した実行例を図6に示す。図中の上側がシステムの発話で下側がユーザの発話であり、四角はそれぞれの音声の発話区間を表す。矢印はユーザの音声の始まりを検知したことによって、システムの発話に中断がかかったことを示している。このように、文節を単位として発話し、ユーザの問い返しに対して即座に応答することができている。

### 3.3 読み上げシステム

書きことばで記述されたテキストを書きことばで伝達する読み上げシステムは以下のように動作する。

- テキストを一文ずつ音声で読み上げる。

- 「文の始めから」という発話を認識した場合、同じ文を再度読み上げる。
- 「前の文」という発話を認識した場合、1つ前の文を読み上げる。
- 「一番最初から」という発話を認識した場合、テキストの先頭の文を読み上げる。
- 読み上げの最中に発話を認識しても中断はしない。

## 4 実験

話しことばによるコミュニケーションの特性を、書きことばによるコミュニケーションとの比較で得ることを目的とした実験を行った。本節ではこの実験の内容を説明する。

実験のタスクは、書きことばで記述されたテキストの内容を音声で伝達することであり、テキストの伝達には3節で説明した2つのシステムを用いた。被験者には、テキストから獲得すべき情報を質問形式で与えた。どちらかのシステムを使ってテキストに記述された情報を獲得し、与えられた質問に答えることを被験者の課題とした。

### 4.1 実験条件

- テキスト: NTT R&D 報道発表
- 記事数: 4 記事
- 記事の大きさ: 344 文字～561 文字, 3 文～6 文
- 被験者: 30 名 (大学生)
- 課題: システムと対話してテキストの内容に関する質問用紙を埋める

図1のテキストは実験で使った記事の1つである。一度実験をしたらそのテキストに対する知識が残るので、同一の被験者が同一のテキストに対して2つのシステムで実験をすることはしない。したがって、被験者は話しことば対話システムとの対話で2つのテキストの課題を行い、読み上げシステムを使って2つのテキストの課題を行った。

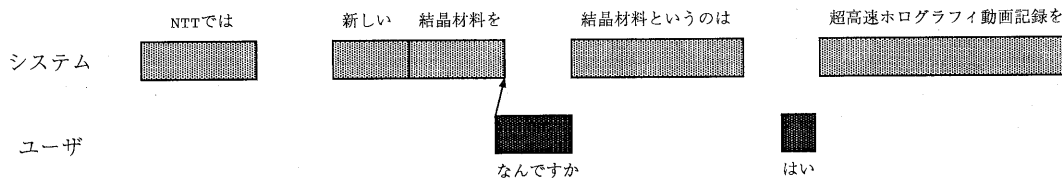


図 6: 話しことば対話システムの実行例

システムを使う前に課題として与える質問用紙を渡し、記事から獲得しなければならないことを把握してもらい、システムを使う際の被験者のモチベーションとした。

#### 4.2 実験環境

実験は図 7 の環境で行った。被験者はヘッドホンからシステムの発話を聞き、マイクに向かって発話をする。また、対話をしながら質問用紙を埋めることができる。

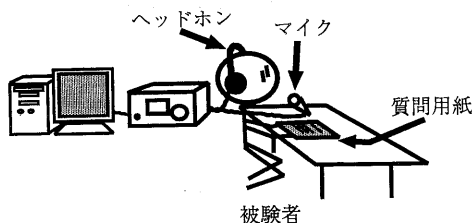


図 7: 実験環境

システムの発話モーラ数は、ユーザに伝達される情報量の近似になると想定している。話しことばで伝達する場合、小さな発話単位であいづちによる確認をとりながら伝達し、情報を詳細化しながら伝達するため、書きことばで伝達する場合よりも少ない情報量で伝達できることが期待される。

課題達成率は、テキストの内容に関する質問事項への解答の正答率で測った。また、それぞれのアンケートは 5 段階評価で答えてもらった。アンケート 1 は、冗長性が少ないと感じた場合が 5 で冗長性が多いと感じた場合は 1 とし、アンケート 2 は、容易と感じた場合は 5 で困難と感じた場合は 1 である。

#### 5.2 結果

それぞれの評価尺度ごとの平均と標準偏差を表 1 に示す。また、システムの発話モーラ数を横軸にとり、課題達成率を縦軸にとってプロットしたグラフを図 8 に示す。図中では、話しことば対話システムで伝達した結果を○で示し、読み上げによる伝達システムで伝達した結果を×で示した。

#### 5.3 考察

表 1 から、いずれのテキストでも、話しことば対話システムで伝達した方が、読み上げシステムで伝達した場合よりも課題達成に要した発話モーラ数の平均が少なく、標準偏差も小さいことが分かる。この傾向は、図 8 から、○印の方は分散の幅が小さく、グラフの左側でまとまりがあることに対し、×印の方はグラフ全体に広がっていることから読み取れる。

課題達成率に関しても表 1 から、話しことば対話システムで伝達した場合でも、読み上げシステムと同じ程度かそれ以上の課題達成率が得られたことが読

### 5 実験結果

#### 5.1 評価尺度

実験の評価として次の 4 つの尺度を用いた。

- システムの発話モーラ数
- 課題達成率
- アンケート 1 (システムの発話の冗長性は?)
- アンケート 2 (知りたい部分の情報を得るのは容易でしたか?)

	発話モーラ数		達成率(%)		アンケート1		アンケート2	
テキスト1	377							
話しことば	1024	366	91.4	8.7	4.1	0.7	3.9	0.9
書きことば	1389	439	90.0	13.5	4.1	0.8	3.9	1.0
テキスト2	513							
話しことば	2084	413	92.4	7.5	4.0	0.7	3.5	1.2
書きことば	2531	756	92.1	8.3	3.7	1.0	2.8	1.1
テキスト3	344							
話しことば	1368	348	95.7	5.7	3.9	0.8	2.7	1.2
書きことば	1907	693	77.1	18.8	3.0	1.4	3.1	1.2
テキスト4	561							
話しことば	1219	309	95.2	5.6	4.1	0.9	3.9	1.2
書きことば	1695	573	91.4	14.6	3.3	0.9	3.6	0.9

(平均) (標準偏差)

表 1: 実験結果

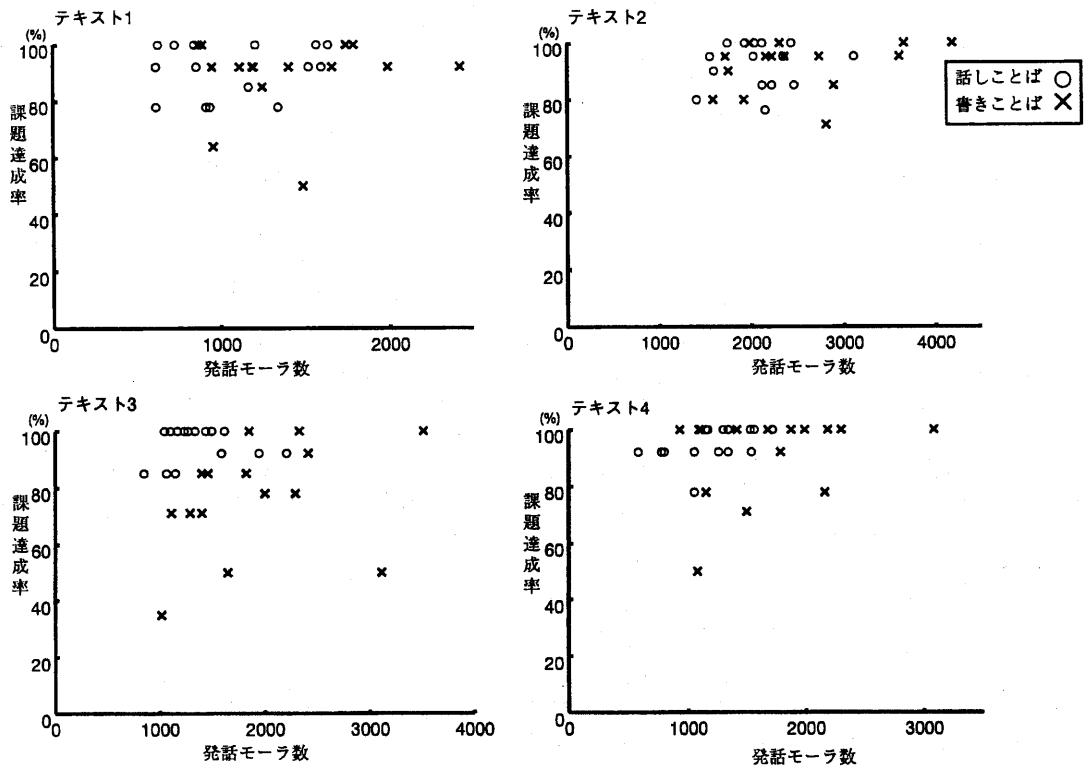


図 8: システムの発話語数と課題達成率のプロット図

み取れる。このことは、図8では、○印がグラフの上の方にまとまっていることでも示されている。

また、アンケートの結果に関しては、テキスト3とテキスト4で、話しことば対話システムの方が冗長性が少ないという傾向が見られたが、それ以外では違いは見られなかった。

以上の分析から、以下の2つがいえる。

- 話しことば対話システムで伝達した場合でも、読み上げシステムと同じ程度の課題達成率が得られる。
- 話しことば対話システムで伝達すると、読み上げシステムで伝達した場合に比べて少ない発話モーラ数で課題が達成される傾向がある。

このことから、

- 話しことばによるコミュニケーションでは、書きことばによるコミュニケーションに比べて少ない発話モーラ数で必要な情報を伝達できる。

ということが分かった。この話しことばによるコミュニケーションの特性は、「情報のブラウジングなど、情報を聞き流しながら対話を行い、知りたいところだけ深く聞くような場面で話しことばによる伝達が有効である」という知見が得られる。

## 6 おわりに

本稿では、話しことばによるコミュニケーションの特性を、書きことばによるコミュニケーションと比較することによって獲得することを目的とし、話しことばによるコミュニケーションと書きことばによるコミュニケーションの特性を比較する実験を行った。実験では、書きことばで記述されたテキストの内容を音声で伝達することをタスクとし、このタスクを話しことば、または書きことばで遂行することを行った。このための実験道具として、書きことばで記述されたテキストを話しことばで伝達する話しことば対話システムと、書きことばで伝達する読み上げシステムを作成した。この2つのシステムを使って書きことばと話しことばの特性を比較した結果、話しことばによるコミュニケーションには、書きことばによるコミュニケーションに比べて少ない発話モーラ数で必要な情報を伝達できるという特性があることが分かった。

## 謝辞

日頃ご指導いただく石井健一郎情報科学研究部長、および、討論して頂いた対話理解研究グループの皆様、北陸先端科学技術大学院大学 島津 明 教授、対話システムの作成にご協力頂いたNTTアドバンステクノロジー(株)久保田 哲也氏、実験にご協力頂いた日本アイアール(株)柴田 葉子氏に感謝します。

## 参考文献

- 1) 堂坂 浩二, 島津 明. タスク指向型対話における漸次的発話生成モデル. 情報処理学会学会誌, Vol. 37, No. 12, pp. 2190-2200, 1996.
- 2) 川森 雅仁, 島津 明, 堂坂 浩二, 中野 幹生. 対話処理のためのコーパス作成. 電子情報通信学会技術研究報告 NLC97-5, pp. 31-38, 1997.
- 3) Sadao Kurohashi and Makoto Nagao. Kn parser: Japanese dependency/case structure analyzer. In *Proceedings of the International Workshop on Sharable Natural Language Resources (SNLR 1994)*, pp. 48-55, 1994.
- 4) Willem J. M. Levelt. *Speaking: From Intentions to Articulation*. The MIT Press, 1989.
- 5) 松本裕治, 黒橋禎夫, 山地治, 妙木裕, 長尾真. 日本語形態素解析システム JUMAN 使用説明書, version 3.11, 11 1996.
- 6) 野田 喜昭, 山口 義和, 山田 智一, 今村 明弘, 高橋 敏, 松井 知子, 相川 清明. 「音声認識エンジン REX の開発」. 1998 年電子情報通信学会総合大会, 情報・システム 1, D-14-9, March 1998.
- 7) 島津 明, 小暮 潔, 堂坂 浩二, 中野 幹生. 対話処理のための分散協調モデル. 言語処理学会第 1 回年次大会発表論文集, pp. 81-84, 1995.
- 8) 島津明, 川森雅仁, 小暮潔. 対話の分析 一問投詞的応答に注目して. 情報処理学会 自然言語処理研究会, Vol. 95, No. 9., 1993.