

言語・非言語モダリティを統合した
会話エージェントのプレゼンテーション行動自動生成
Automatic Generation of Presentation Behavior
of Embodied Conversational Agents
by Integrating Verbal and Nonverbal Modalities

京都大学情報学研究科

Graduate School of Informatics, Kyoto University

西田 豊明

Toyoaki Nishida

東京大学大学院情報理工学系研究科

School of Information Science and Technology, University of Tokyo

黒橋 禎夫 久保田 秀和 遠山 義洋 大泉 敏貴 柴田 知秀 李 清
Sadao Kurohashi Hidekazu Kubota Yoshihiro Toyama Toshitaka Oizumi Tomohide Shibata Li Qing

The goal of this research is to develop a suit of techniques for automatically producing an interactive presentation by integrating verbal and nonverbal modalities for embodied conversational agents. The result is five-fold. Firstly, we have investigated utterances of recorded oral presentation and extracted the behavior of prosodic parameters under varying discourse structures. Secondly, we have designed the basic architecture for controlling the prosody based on the discourse analysis of natural language tests. Thirdly, we have developed a written-language-to-spoken-language text conversion method. Fourthly, we have developed a method of generating summary slides that supplement presentation. Finally, we have developed a method of automatically highlighting multimodal synchronization.

Key Words: Embodied Conversational Agents, Presentation Behavior, Integration of Verbal and Nonverbal Modalities, Discourse Analysis, Written-Language-To-Spoken-Language Text Conversion

1. 研究の目的

本研究では、自然言語テキストからの談話情報抽出に基づく韻律生成手法を中心に、言語・非言語モダリティを統合した会話エージェントのプレゼンテーション行動自動生成方式の実現に取り組んだ。

本研究は、参加者のコミュニケーション行為の連続的な流れである会話を、会話量子と呼ぶ会話の結節点を表す「会話の粒」の離散系列として近似する会話量子化のアプローチに基づいている。現在の実装では、会話量子を知識カードと呼ばれる自然言語テキストと

参照イメージの複合体によって表現する。本研究では、主として一人の会話参加者が所与の話題についてプレゼンテーションを行う状況を対象にした会話量子を表現する知識カードを入力として、会話エージェントの発話の韻律情報と非言語行動の自動生成手法を研究開発を行う。また、既存文書からの知識カード自動生成を目指して、書き言葉を話し言葉に変換する手法と、プレゼンテーション補助資料を生成する手法の研究開発も行う。

本研究で得られた主な成果は次の通りである。

第一に、実データを分析して、対話テキストの内容を合成音声によつて的確に表出する能力を持つ音声対話生成方式の基本設計を行った。読み上げ音声を聞くだけでは理解しにくい箇所を抽出して、言語処理・音声処理によつて聞きやすく変換するアルゴリズムを実現することを目標に、話題や焦点などの対話内容の理解の手がかりとなる言語表現と韻律の関係の分析を行った。

第二に、自然言語テキストの談話構造解析に基づく韻律制御システムのアーキテクチャの設計を行った。この方式では、与えられた自然言語テキストを解析して、エージェントのジェスチャーや顔表情などの非言語的なモダリティを決定し、汎用的なエージェントアニメーション自動生成システム CAST (The Conversational Agent System for neTwork applications) における会話エージェントのプレゼンテーション行動を生成する。

第三に、話し言葉・書き言葉変換の基本的な手法を組み込んだ自然言語処理システムを試作した。このシステムでは、話し言葉を、違和感の生じる表現、音声に適さない表現、複雑な構造をもつ表現を含まない、「音声として聞いたときにわかりやすい表現」と定義し、機能語的な表現を中心に、普通体から丁寧体への文法的な変換を行う。

第四に、自然言語テキストからプレゼンテーション補助資料を生成する手法の開発を行った。この手法では、数文から数十文の意味的まとまりを持つ自然言語テキストを入力とし、文章構造の解析、主題や重要な説明表現の抽出、見出し語・重要説明表現の配置を順次行うことによつて、要約スライドを作成する。

第五に、比喩的な意味の可視化と、重要語のスーパーインポーズによつて、言語モダリティと非言語モダリティの同期を強調する手法を開発した。与えられた自然言語テキストは、まず日本語解析エンジンを用いて解析され、非言語的なモダリティによつて強調すべき部分が同定され、その部分をどのような非言語モダリティによつて強調するのかが決定される。例えば、重要な概念について述べるときには、強調のジェスチャーを用い、眉を上げて目を大きく見開くというプランが用いられる。次に、自然言語テキストが音声合成システムに入力され、音声ナレーションが作成される。同時に、音声とエージェントアニメーションを同期さ

せるためのタイムスケジュールが計算される。

2. 対話テキストの構成と韻律情報

本研究では、テキストの内容に応じて、動的に制御した合成音声を生成することによつて、対話の構造、文脈を明示化し、理解しやすい発話を生成することをめざす。このため、テキストを見るだけではまた読み上げ音声のみを聞くだけでは理解しにくい箇所を抽出し、言語処理、音声処理を施す[1]。

2.1 対話テキストの構成と韻律情報

複数の話者が双方向にやりとりした発話をテキストに書き起こしたものを対話テキストと呼ぶ。対話テキストには、独話も含めて複数発話からなる談話構造があり、そのうち2人以上の話者による談話構造を対話構造と呼ぶ。また、対話内における文同士の関係を形成する構造として修辞構造がある。以下ではこれらをまとめてテキスト構造と呼ぶ。

2.2 対話テキストに出現する言語情報

人間同士の対話には、いわゆる短文を「て」などの接続助詞でつないだ複文が多数存在する。対話テキストは書き起こした文書であり実際の発言内容そのままではないにせよ、1文に複数の内容を含めることは理解の弊害になると考えられる。こうした長文を短文に変換することはとくに修辞構造の推定、制御に役立つものであると考える。そのほか対話の内容を的確に伝える際に弊害となりうる表現について分析、検討を行い、対話内容の表現方法について検討する。

2.2.1 対話構造

談話構造という観点からみると、対話とは、複数人による発話の意図・目的の達成によつて展開される。一方、質問と返答などの一定の行為の交換のパターン（以下対話対と呼ぶ）によつて対話にある種の構造が与えられることもある。前者の大局的な対話構造は、通常の意味での談話構造と一致する。一方、後者の双方向の局所的な対話構造である対話対は談話、意図とは異なる構造を持つ。本研究では対話対を分析、表出の対象とする。

対話対の中でもっとも基本的な相互行為を達成するものを隣接ペアと呼ぶ。隣接ペアは以下のような2つの発話X、Yの連続である[2]。

- 1) XとYは隣接した位置にある。
 - 2) XとYは異なる話者が算出する。
 - 3) 第1部分Xは第2部分Yに先行する。
 - 4) XはYを特化する(Xは決まった形のYを要求する)。
- X = 質問, Y = 返答はその一例である。

局所的な対話構造をつくるものとして、対話対のほか、一方的な働掛、応答の機能をもつものも存在する。前者の例として提案や主張、後者の例として感想などを挙げるができる。

2.2.2 修辞構造

修辞構造とは、文間関係を用いたテキスト中に出現する文、段落の木構造を意味する。修辞構造は、1対話内で閉じた構造であり、1対話中に1つ以上存在する。

2つの対話テキスト¹⁾について分析を行い、接続詞、副詞などの手がかり語をもとにした分析結果では対話テキスト1では全文の42.7%、対話テキスト2では33.5%の文に対する修辞関係を確認できた。意味内容にまで踏み込むと、対話テキスト1では81.5%、対話テキスト2では69.5%の修辞関係を確認できた。修辞関係のない文は、対話の先頭文や対話構造を構成している文、作者の意図を読みとれなかった文である。

2.2.3 対話内容の表現方法

対話テキストを対話音声とする際に、加工すべき表現方法について分析する。

対話の内容を簡潔に表現し、その発話意図を明確にすることは内容の理解に重要である。短文分割によって、1文内に複数の主張が盛り込まれないようにするとともに、修辞関係をより明確にする [3]。

長文の定義は曖昧であるが、その指標として文内に出現する単語数、用言数、有効語数、接続詞数、句点数などが挙げられている 図1に有効語数をもとに行った分析結果を示す。正確な閾値は存在しないが、例え

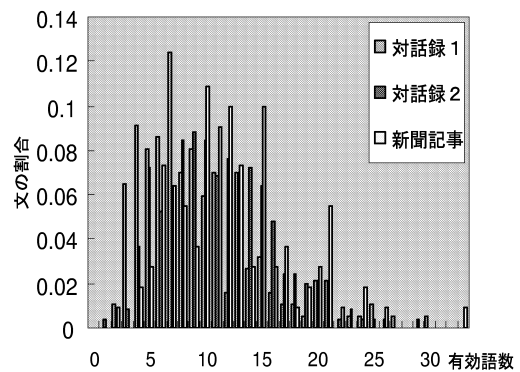


図1: 対話テキスト中の文の有効語数

*新聞記事: 今井賢一「21世紀型企業」, 1996年1月1日~8日付日本経済新聞に掲載。

ば有効語数が15以上の文には、対話テキスト1の15.0%、対話テキスト2の24.1%、新聞記事の31.8%の文がそれぞれ当てはまる。

2.3 自然対話音声の分析

実在の対話音声データを対象に、テキスト構造に代表される言語情報と韻律情報との関係、また話題や焦点などの対話内容の理解に手がかりとなる語句と韻律情報との関係について分析した。分析に用いたデータは以下の3文書である。

- ・ メメックス研究会第2フェーズ・第2期報告書 2000年9月 第1章(約30,000文字)
- ・ メメックス研究会第2フェーズ・第2期報告書 2000年9月 第8章(約18,000文字)
- ・ スタンフォード大学日本センターシンポジウム『IT企業集積と地域経済』2000年9月(約18,000文字)

対話音声については、研究会およびシンポジウムの講演者の付けたピンマイクからテープに収録したものをを使用した。標本化周波数16kHz、量子化精度16bitでA/D変換し、ESPS/waves+によって韻律情報およびポーズ長の分析を行った。

今回使用した対話音声と対話テキストは、言語的な発言内容において一致していない。話し言葉には松本ら[4]の示す通り、フィルターや非流暢性、非文法性などが存在する。本研究で対象とする対話テキストには、書き起こしの段階でこれらの現象が取り除かれている。音声の分析の際にもこれらの現象の部分を省き、対話テキストの内容に沿って解析を行った。

¹⁾ 対話テキスト1: 雑誌「実業の日本」1998年5月号の特集『「小さい会社」が日本を救う! 新しい日本産業の岩盤を探す』での対談記事。

対話テキスト2: メメックス研究会初年度報告書(1996年)12章9節『次年度以降の研究の進め方について』での質疑応答/議論の記録。

表2：ポーズ長と立ち上がり F0

対象	ポーズ長(s)	立ち上がり F0 (Hz)
段落間	1.03	213.4
順接	0.28	216.9
説明	0.89	149.7
例示	0.91	164.4
並列	1.45	205.0

表3：話題，焦点語句の F0，発話速度

対象	最高 F0 (Hz)	最低 F0 (Hz)	発話速度 (秒/モーラ)
話題語1	230.7	117.8	0.093
話題語2	206.0	115.4	0.099
指示詞	195.9	144.0	0.058
接続詞	245.1	118.0	0.156
有意な動詞	212.0	129.6	0.080
無意な動詞	152.0	101.1	0.137

- * モーラ：日本語の仮名 1 文字．拗音は 2 文字．
- * 話題語 1：出現頻度の高い名詞 (tf 値の高い名詞)．
- * 話題語 2：特定性のある名詞 (idf 値の高い名詞)．

表4：文中の位置情報

対象	最高 F0 (Hz)	最低 F0 (Hz)	発話速度 (秒/モーラ)
文頭	271.6	123.0	0.109
読点后	210.5	98.1	0.088
文中	163.5	86.8	0.105

テキスト構造の境界について分析を行った結果を表2に示す．テキスト構造として，段落間(談話構造)，修辞関係を対象とした．段落間や並列など話題の変換点ではポーズ長が長く，立ち上がり F0値が高いことが分かる．話題の変わらない順接ではポーズ長が短い．修辞関係の種類や対話構造を対象としてさらに分析を進めていく必要があるが，ポーズ長に関しては話題の転換と関係がありそうである．

話題語句やテキスト構造を表す指示詞，接続詞についての分析を行った結果を表3に示す．ここで，各語句についての F0，発話速度の分析結果である．特徴的な韻律情報を示した語句を対象とおり，無意な動詞は比較のための資料である．F0については，いずれもピッチ範囲が広く，平均 F0が高いという特徴を持っている．無為な動詞と比べると歴然としている．発話速度に関しては指示詞が速く 接続詞が遅いという結果を得た．

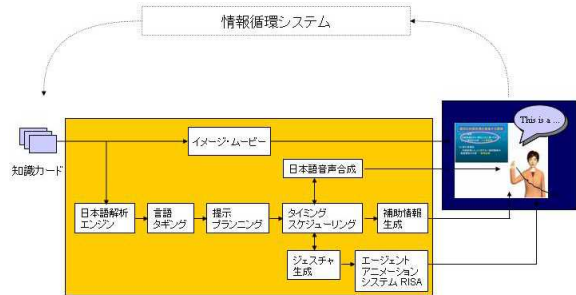


図2:知識カードを入力とするマルチモーダルプレゼンテーション自動生成システム CAST のアーキテクチャ

個人の特徴や文の正確もあるのでさらに分析が必要かと思うが，一定の傾向は出現していると言える 話題，焦点語句について文中の位置情報による差異についても分析を行った．その結果を表4に示す．

発話速度に大差はないものの，文頭や読点后ではピッチ範囲が広く平均 F0が高いのに対して，文中ではピッチ範囲が狭く平均 F0が低いという結果を得た．

3. 言語・非言語モダリティを統合した会話エージェントのプレゼンテーションの自動生成のアーキテクチャ

与えられた自然言語テキストから，言語・非言語モダリティを統合した会話エージェントのプレゼンテーション行動を自動生成するためには，言語的なモダリティによる行動と，非言語的なモダリティによる行動を連携させる必要がある．

そのための土台として，本研究では，汎用的なエージェントアニメーション自動生成システム CAST (The Conversational Agent System for neTwork applications)[5]をベースとして用いる．CAST では，与えられた知識カード中の文章からエージェントのジェスチャーや顔表情などの非言語的なモダリティを決定する(図2)．与えられた文章は，まず日本語解析エンジン[6]を用いて解析され，非言語的なモダリティによって強調すべき部分が同定され，その部分をどのような非言語モダリティによって強調するのかが決定される．例えば，重要な概念について述べるときには，強調のジェスチャーを用い，眉を上げて目を大きく見開く，といった非言語情報を用いた強調が行われる．次に文章が音声合成システムに入力され，音声ナレーションが作成される．同時に，音声とエージェントアニメーションとを同期させるためのタイムスケジュールも計算される．

4. 書き言葉話し言葉変換

知識カードの作成はテキストベースのエディタを用いて行うため、書き言葉になっている。会話エージェントで利用するためには、書き言葉を話し言葉に変換することが望ましい。ここでは、書き言葉話し言葉変換と呼ぶ処理を行い、音声発話生成システムに入力するテキストそのものを書き言葉から話し言葉に変換する手法[7]について述べる。

4.1 書き言葉と話し言葉

話し言葉の基本は話すという行為、つまり、聞き手を相手とする話し手によってなされる行為であり、音声による表現の行動である。これに対して、書き言葉は、読み手を相手とする書き手によってなされる文字列による表現の行動である。

聞き手は話し手によって現前する相手である。話し手を直接目の当たりにすることで聞き手は文章そのものでなく、音声による表現や身振り手振りを情報として受け取ることができる。一方、読み手が書き手の前にいることはない。読み手は文字列だけから情報を読み取らなければならない。

音声による話し言葉の表現は、瞬間的に聞き手の両各を刺激しながらも時間とともに流れ去ってしまうため、空間的に定着することはない。そのため聞き手は発話を刻々に捉えなければならず、話しても聞き手の制約のことを考えて表現をしなければならない。文字列による書き言葉の表現では、読み手の制約も緩く、書き手も自分のペースで表現を展開しやすい。

話し言葉には書き言葉には見られない特徴がいくつかある。ここではテキストレベルでの書き言葉と話し言葉の違いについて議論する。

音声で伝えられる話し言葉では聴いてわかりにくい語の使用は自然に避けられる。また、「私立」と「市立」などのような同音異義語では、読み替えの工夫がなされたりしている。口調と側面でも話し言葉特有のものがある。例えば、「...理事会を開き、この問題に関し、正式に遺憾の意を...」という文のなかに出現する連用形止めの表現は、日常的な話し言葉では用いられることは少なく、「開いて」、「関して」と接続助詞のついた形がより一般的である。さらに、聞き手を対象とする話し言葉では、相手に対する敬意や親しみを表す

表現など何らかの待遇表現が用いられる。不特定多数を相手にしている書き言葉では、論説文や新聞記事などをはじめほとんどが、「だ・である体」で書かれた脱待遇の表現である。また、品詞や単語などについてみたときに、話し言葉では感動詞・応答詞・終助詞・間投助詞・指示語・融合形などが多いことが指摘されている[8]。

以上の考察に基づき、話し言葉を、「音声として聞いたときにわかりやすい表現」と定義し、テキスト中の用語や構造を変換することに焦点をあてた書き言葉話し言葉変換の自然言語処理システムの構築を目標とした取り組みを行った。

4.2 普通体の違和感を解消するための変換方式

普通体を音声にしたときに少じる違和感を解消するための変換方式について検討した。

普通体を音声にしたときに違和感を感じてしまうのは、普通体が相手の存在を前提としない不特定多数に対する脱待遇の表現だからである。そこで、公の場での発話に一般的に用いられる丁寧体を、普通体の代わりに用いることで待遇表現を補うこととした。丁寧体とは、一般に「です・ます体」と呼ばれる表現のことである。

普通体から丁寧体への変換は、内容に直接関係しない機能語的な変換であると言える。そこで、文法的に変換を行うアルゴリズムを用いることとした。つまり、次のようなアルゴリズムによって変換を行う。

1. 入力として与えられるデータに対して、JUMANとKNPを用いて構文解析し、各形態素の形態素情報と構文情報を得る。
2. 形態素情報と構文情報を参照し、変換規則と合致する形態素列を規則で定められた形態素列に置き換える。
3. 変換処理を実行し、出力する。

現在、手入力で作成した50個ほどの変換規則を用いている。変換規則に用いた情報としては、形態素名、品詞の種類、活用の種類、活用形、構文情報などがある。変換規則の適用例を図3に示す。

変換を地層カード134枚に適用した結果、すべてのカードにおいて意図した結果が得られた。実際の変換例を図4に示す。ここからわかるように、語尾の変換がほとんどであり、変換量もごくわずかで、これをもつ

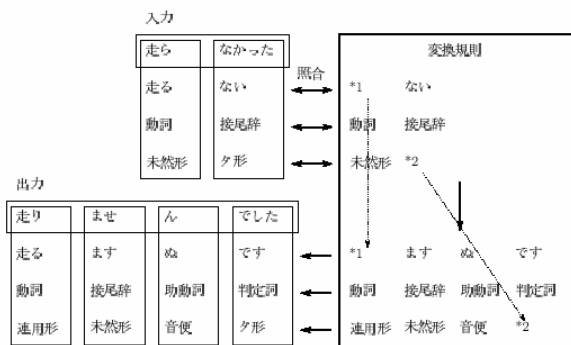


図3：変換規則の適用例

わが国における地層処分概念は、安定な状態にガラス固化した廃棄物を、地下深部の安定した地質環境中に、人工的な安全対策を施したうえで埋設することにより、長期にわたって人間の生活環境から隔離し、将来の世代にも廃棄物による影響が及ばないようにすることである。ここで、地質環境とは、地層処分場としての地下の環境を指し、具体的には岩盤とそこに含まれる地下水からなる。空間的には、深度数百mから1,000m程度で、数km四方の水平的広がりを想定している。

図4：普通体から丁寧体への変換例

て話し言葉に変換したといえるようなレベルではない。また、使用したコーパスは比較的閉じた分野のもので、このルールが他分野の書き言葉にも適用できるかどうか疑問である。今後は、より一般的な規則を構築できるように広範なコーパスを用いたり、変換結果を評価する基準を明確にする必要がある。

5. 自然言語テキストからのプレゼンテーション補助資料の生成

数文から数十文の意味的まとまりを持つテキスト(以降、これを知識カードとよぶ)を対象として、要約スライドを自動生成する手法[9]について報告する。知識カードの文章構造の解析、主題や重要な説明表現の抽出、それらの適切配置によって要約スライドを作成する。知識カードの例と、そこから報告する方法によって自動生成された要約スライドの例を図5に示す。

5.1 文章構造解析

スライドを作成するには、知識カードの文章構造を捉える必要がある。ここで、入力文はあらかじめJUMANで形態素に分け、KNPで構文解析しておく。文章構造のモデルとして、一文をノードとし、それらがある結束関係を持ってリンクされた木構造[4]を採用した。文と文の結束関係としては、次のものを考えた。

エージェントプラットフォームはエージェントを構築実行するためのソフトウェアプラットフォームである。いくつかのエージェントプラットフォームはモバイルエージェント機能を提供している。モバイルエージェントはコンピュータ間を移動可能な能動的なプログラムである。モバイルエージェントはネットワーク上のホストからホストへと移動して処理を行なう。モバイルエージェントが移動するとき、プログラムのコードだけでなくプログラム内の変数内容といった実行状態も移動先に転送され、移動先コンピュータでは移動前の実行状態から継続した処理が行なわれる。

エージェントプラットフォーム

- エージェントを構築実行するためのソフトウェアプラットフォーム
- モバイルエージェント機能を提供
 - モバイルエージェント
 - コンピュータ間を移動可能な能動的なプログラム
 - ネットワーク上のホストからホストへと移動して処理を行なう

図5：知識カードとスライドの例

- #0
 - └ (初期化) #1 エージェントプラットフォームはエージェントを構築実行するためのソフトウェアプラットフォームである。
 - └ (主題連鎖) #2 いくつかのエージェントプラットフォームはモバイルエージェント機能を提供している。
 - └ (焦点主題連鎖) #3 モバイルエージェントはコンピュータ間を移動可能な能動的なプログラムである。
 - └ (主題連鎖) #4 モバイルエージェントはネットワーク上のホストからホストへと移動して処理を行なう。
 - └ (焦点主題連鎖) #5 モバイルエージェントが移動するとき、プログラムのコードだけでなくプログラム内の変数内容といった実行状態も移動先に転送され、移動先コンピュータでは移動前の実行状態から継続した処理が行なわれる。

図6：図5の入力文に対する文章構造

並列、対比、主題連鎖、焦点主題連鎖、詳細化、理由、原因結果、変化、例提示、例説明、質問応答

例えば、図5の例文の文章構造は図6のようになる。

文章構造解析は、種々の表層の手がかりをもとに、各入力文に対して、以前の文(接続文)との間の結束関係を次に示す基準によって求め、最終的に最も高い合計点を得たものを採用した。

(1) 表層的な手がかり

表層的な手がかりを元にし、ルールを作成した。談話構造解析のルールの例を表2に示す。適用範囲とは、入力文と前の文がどれだけ離れた文までを考えるか、

接続可能文，入力文のパターンは，表層表現のパターンであり，これらのルールとマッチすると，接続文との間の結束関係にスコアを与える．

(2) 語連鎖

各文において，焦点と主題を考える．2文間に，主題と主題，焦点と主題に連鎖があるときは，それぞれ，主題連鎖，焦点主題連鎖という結束関係とする．語の連鎖は，完全一致と部分一致を考え，完全一致の時は部分一致の時よりも高いスコアを与える．

(3) 2文間の類似度

2文が並列または対比の関係にあるときは，それらの文の間には一定の類似性があると考え，類似度が高いものには，対比および並列にスコアを与えた．品詞の一致，語の一致，シソーラス辞書上での語の近さなどによって計算された文節間の類似度を組み合わせることにより，文節列間の類似度を計算する．

文章構造解析の単位(ノード)は1文を基本とするが，1文内に大きな並列があるときは，分割してそれぞれ別のノードとして扱う．

5.2 見出し語，重要説明表現の抽出

各文から主題(「見出し語」)と主題を説明するような表現(「重要説明表現」)を取り出す．

見出し語の抽出については，「～とは」，「～という」，「～は」，「...する{の/とき}は～{だ/になる}」などの手がかり表現に着目した処理を行う．1文中にこれらの手がかり表現が複数存在する場合は，そのうち一番前にあるものを抽出する．ただし，一文目は特に重要であると考え，「～では」，「～について」，「」でくられた語などを見出し語として抽出する．

重要説明表現の抽出を行うためには，次のようなパターンに着目した処理を行う．(二重線が見出し語，一重線が重要説明表現)

- (a) 見出し語{は/とは/では/というの}... x
- (b) 見出し語...， x: 主節
- (c) x する{の/とき}は 見出し語{だ/になる}．
で囲まれた部分の節を重要説明表現とする．
- (d) x が重要だ．

5.3 スライドへの出力

得られたカードの文章構造と抽出した見出し語・重要説明表現をもとに次のような基準でスライドに配置する．

1. 文章中の一番最初の見出し語をスライドのタイトルとして出力する．
2. 見出し語があれば出力し，次の行にインデントを一つ下げて重要説明表現を出力する．見出し語がなければ，重要説明表現だけを出力する．
3. 1枚のスライドに表示される量が多くなりすぎないように，文章構造木の深さ2までの文を出力対象とする(一般に，文章構造の根に近い方が重要であると考えられることができるからである)．
4. 親との結束関係に応じて，適宜インデント処理を行う．

6. 言語・非言語モダリティの同期の強調

比喩的な意味の可視化と，重要語のスーパーインポーズによって，言語モダリティと非言語モダリティの同期を強調する手法[10]について述べる．

6.1 ジェスチャーの分析

ジェスチャーを，具体的なイベントを表す象徴的ジェスチャー，抽象的な概念や関係比喩的を表すジェスチャー，指示的な動作による直示的ジェスチャー，特に独立した意味をもたない反復動作であるビートジェスチャー，対照を表す対照ジェスチャーに分類し，日本人のPowerPoint スライドを使ったプレゼンテーションをビデオに記録して，講師の視線，音声波形，発話内容，ジェスチャー，談話構造を分析した．

その結果，最も使われているジェスチャーの種類はビートジェスチャー，直示的ジェスチャー，比喩的ジェスチャーであり，ビートジェスチャーが行われるときは，発表者が強調をするときであることなどがわかった．

6.2 マルチモーダルプレゼンテーションの同期の強調

(1) ビートジェスチャーと発話の同期の強調

ビートジェスチャーに伴う句に応じて，表5に示すルールに基づいてスーパーインポーズを用いてビートジェスチャーと発話の間の同期を取る．出力イメージを図7に示す．

(2) 比喩的ジェスチャーと発話の間の同期の強調

表5: ビートジェスチャと発話の同期の強調

ビートジェスチャに伴う句	強調	例
名詞	修飾部分を除いた名詞部分をスーパーインポーズする。	ビートジェスチャと「報告する」という句に対しては、「報告」のところだけをスーパーインポーズする。
名詞句（複合名詞を含む）または接続詞（例えば、「議論の順序」）	全体をスーパーインポーズする。	
形容詞的ないしは副詞的な修飾語	修飾語（あるいは、修飾語の後に非修飾語が続くときは修飾語とその非修飾語）をスーパーインポーズする。	例えば、「残った」という修飾語と「もの」という修飾語のペアをスーパーインポーズする。
疑問詞	疑問詞にマッチするものをスーパーインポーズする。	例えば、「どうやって」と「いつ」がスーパーインポーズされる。

比喩的なジェスチャは、象徴的なジェスチャに類似しているが、メタファーの介入によって象徴性が生じる。

比喩的なジェスチャに伴う句が動的意味であれば、これらの句の意味の暗示する空間的な方向に従って、矢印のある線のアニメーションを行う。例えば、「増加」という語については、矢印のある上向きの線を表示し（図8）、「減少」については下向きの矢印のついた線を表示する。

比喩的なジェスチャに伴う句が静的な意味のときは、これらの形状の出現のタイミングを考慮し、句の表す比喩的な意味に従う幾何学的な形状を表示する。例えば、「長い」という語に対する比喩的なジェスチャのストロークの間は、ジェスチャ空間の近くに矢印を表示する。「トップ」という語とジェスチャの同期を取るためには、比喩的なジェスチャのストロークが終わろうとするときに矢印を表示する。

(3) 直示的なジェスチャと発話の同期の強調

直示的なジェスチャに関する強調法は採用されるインタフェースとシステムがそのレイアウトと内容についてどれくらい詳細に感知しているかに依存する。インタフェースがスライドを含んでいるとき、句がスライドの一部または全てを参照しているとき、スライドのなかに示された重要語を参照しているときは図9に示すように、当該の句を囲む線や下線などを表示する。

7. まとめ

自然言語テキストからの談話情報抽出に基づく韻律生成手法を中心に、言語・非言語モダリティを統合



図7:(3)におけるビートジェスチャの同期のハイライト



図8:「増加」という発話とそれに対応する比喩的なジェスチャの同期の可視化

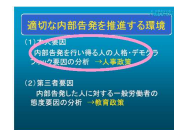


図9: 直示的なジェスチャに対する同期のハイライト

した会話エージェントのプレゼンテーション行動自動生成方式の研究開発に取り組み、会話エージェントの発話の韻律情報と非言語行動の自動生成手法、書き言葉を話し言葉に変換する手法、プレゼンテーション補助資料を生成する手法を研究開発した。

参考文献

- [1] Yoshihiro Tooyama, Li Qing, Toyooki Nishida, Sadao Kurohashi, Conversational speech synthesis model for virtualized-egos, 人工知能学会 第 16 回全国大会, 1E1-07 (2002.5) .
- [2] 石崎雅人, 伝康晴, “言語と計算 3 談話と対話,” 東京大学出版会, 2001.
- [3] 福島孝博, 江原暉将, 白井克彦, “短文分割の自動要約への効果,” 自然言語処理, Vol.6, No.6, pp.131-147, 1999.
- [4] 松本裕治, 伝康晴, “話し言葉の形態素解析,” 音声言語情報処理, No.036-002, 2001.
- [5] Nakano, Y. I., Murayama, T., and Nishida, T.: Multimodal Story-based Communication: Integrating a Movie and a Conversational Agent. Trans. IEICE, Vol.E87-D No.6 pp.1338-1346 2004/6.

- [6] 黒橋禎夫, 長尾真. 表層表現中の情報に基づく文章構造の自動抽出, 自然言語処理, Vol. 1, No. 1, pp. 3-20, 1994.
- [7] 大泉敏貴, 鍛冶伸裕, 河原大輔, 岡本雅史, 黒橋禎夫, 西田豊明: 書きことばから話しことばへの変換, 言語処理学会 第9回年次大会 (2003.3)
- [8] 文化庁編:「ことば」シリーズ12, 話し言葉, 1970.
- [9] 柴田知秀, 河原大輔, 黒橋禎夫: 主題と文章構造の解析に基づくスライドの自動生成, 言語処理学会 第9回年次大会, pp.597-600 (2003.3).
- [10] Li, Q., Nakano, Y., Okamoto, M., and Nishida, T. "Highlighting Multimodal Synchronization for Embodied Conversational Agent", Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004), pp. 50-55, (2004-1).