

ドメインごとにエージェントを導入した音声ガイドシステム

酒井 桂一 八木沢 津義 藤田 稔

キヤノン (株) 情報メディア研究所

〒211 神奈川県川崎市幸区鹿島田 890-12

Tel: (044)549-5111 E-mail: keiichi@cis.canon.co.jp

概要: 本稿では、音声ガイドシステムにおいて、ドメインごとにエージェントを導入した対話モデルを提案する。複数のドメインを扱える対話システムでは、ユーザは、異なるドメインにまたがる知識をシステムが持っているとは錯覚し、システムの能力を越えた質問をすることがある。そのような誤解を避けるために、複数のエージェントを用意し、ドメインごとに切り替える。さらに、一つのドメインに対しても異なる対話戦略を持つ複数のエージェントを切り替えることで、ユーザの知識レベルや好みの差異にも対応できる。また、ユーザと各エージェント間に加え、内線案内のようなドメインでは検索対象者を含めた待遇判定を行い、さまざまな人間関係に対して適切な応答を生成することで、より自然な対話の実現を可能とした。

和文キーワード: 音声対話システム, ガイダンス, エージェント指向, 待遇表現, 語調, 呼称

A Speech Guidance System Assigning an Agent to Each Domain

Keiichi SAKAI Tsuyoshi YAGISAWA Minoru FUJITA

Media Technology Laboratory, **Canon Inc.**

890-12, Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 211, Japan

Phone: +81-44-549-5111

Abstract: A new conversational model, which assigns an agent to each domain in a multiple domain speech dialog system, is proposed. A dialog system which deals with multiple domains often misleads the users that the system has a united knowledge across domains. Thus, the users sometimes ask questions beyond each domain knowledge. To avoid such misleading, we introduce the agent for each specific domain knowledge, and exchange them according to the exchanged domains. We also characterize the agents with different strategies in order to adapt the agent to the knowledge levels or the strategic tastes of the users. Furthermore, this paper discusses a realization of a natural conversation through a proper reply generation according to the various human relations. When the domain is a phone number guidance, the "treatment expression judge" utilizes not only the relation among the user and the agents, but also the person to be guided.

keywords: spoken dialog system, guidance, agent-oriented, treatment expression, politeness, how-to-call

1 はじめに

我々は、CD-ROMを始めたとする実用規模の「準構造化された文書(モデル化テキストと呼ぶ)」を検索対象とする音声ガイドシステムの研究を行っている。[7]では、システム主導で対話を進める手順とする「対話項目」を「ジャンル」ごとに宣言的に記述することにより、互いに関連しない複数の「ドメイン」を扱うことができる対話システム Multi TARSAN を作成した。Multi TARSAN では「ジャンル」および「ドメイン」を以下のように定

義する。

ジャンル: 「温泉」、「ゴルフ場」や「洋画」、「邦画」など、モデル化テキストを定義する書式を作成する単位。

ドメイン: ドメイン知識(メタ知識および「ドメイン専門用語」[8])を共有する単位。複数のジャンルから構成される。[7]では「観光案内」、「プロ野球情報」、「映画情報」の3つのドメインに対応し、今回「内線案内¹」を追加した。

¹内線、メールアドレスなどを人手で入力したモデル化テキスト

Multi TARSAN では、ユーザの発話中にドメイン専門用語が現れた場合に、そのドメインに切り替える²。しかしながら、システムが複数のドメインを扱えるようにすると、ユーザは、異なるドメインにまたがる知識をもシステムが持っている」と錯覚し、システムの能力を越えた質問をすることがある。

一方、最近注目を浴びているユーザインタフェースのパラダイムにエージェント指向 [1] がある。これは、計算機を機械としてではなく、擬人化されたエージェントとして、能動的にユーザとやりとりを行うというもので、さらに、複数のエージェント間のコミュニケーションやコラボレーションについての研究もマルチエージェントシステムとして盛んに行われている [2]。

このようなエージェントシステムにはユーザと自然言語で対話を行うものもあるが、ユーザとの窓口として立つエージェントは通常ひとつである。また、エージェント間のやりとりも研究段階であり、汎用性を持つには至っていない [3]。

そこで我々は、自然言語による対話機能を持った複数のエージェントを用意し、ドメインごとにエージェントを切り替えるモデルを提案する。エージェントが切り替わることにより、ユーザはドメインが変わったことを自然に認識し、システムの能力を越えるような質問をすることがなくなる。また、エージェントに対話戦略を持たせることによって、ドメインごとに異なるユーザの知識レベルや好みによって対話戦略を制御することが容易になる。

さらに、エージェントが切り替わったことを強調し、より自然な対話を実現するために、ユーザとエージェント、さらには内線案内の場合に検索対象者を含めた三者間の待遇判定処理を作成した。この処理では、ユーザが用いた待遇/語調/呼称を抽出し、三者の上下関係/親疎関係から、エージェントが応答文で用いる待遇/語調/呼称を決定する。

本稿では、2章でMulti TARSAN に導入したエージェントのモデルについて述べ、3章では、待遇判定処理について述べる。また4章で、その効果を検証するための実験システムについて述べる。

2 エージェントによる対話モデル

2.1 ドメインごとのエージェントの切替

図 1 に示すように、ドメインごとにその知識を持つ「エージェント」を割り当てる。そして、エージェントごとに音声では声の高さや発声スピード、文字列ではフォ

²ただし、複数のドメインに関連する知識を表現する枠組は作成できていない。

ントや音色などの属性を与え、ドメインが変わった場合にエージェントを切り替え、ユーザにドメインが変わったことを認識させる。すなわち、ユーザは、エージェントが変わったことで、複数のドメインにまたがる質問が困難であることがわかる。

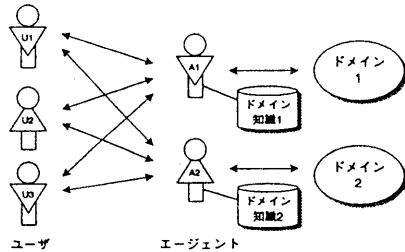


図 1: ドメインごとのエージェントの切替

2.2 同一のドメインでのエージェントの切替

また、図 2 に示すように、同一のドメインに対して、異なる対話戦略を持つ複数のエージェントを用意する。これは、例えば、ユーザの知識レベルによって、初心者用のエージェントや熟練者用のエージェントを切り替えたり、ユーザの好みによって、例示解答方略あるいは促進方略 [4] を多用するエージェントを選んだりするといった、対話戦略を制御するのに有効である。

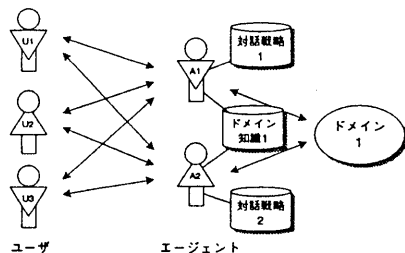


図 2: 同一ドメインでのエージェントの切替

2.3 エージェントを強調する待遇判定

このように、ユーザと同等の個人情報を持つエージェントを設定し、それを自然な形で強調する手段として、ユーザ-エージェント間の待遇を判定し、応答文に反映させることが有効であると考えられる。また、「内線案内」というドメインでは、検索対象者をさらに含んだ三者間の待遇を考慮することで、より自然な対話を実現できる。

エージェントを導入した内線案内のモデルを図 3 に示す。個人情報 DB には、氏名、所属、役職、入社年度など、個人に固有の情報が格納されており、その個人情報

DB からロードしたユーザ/エージェント/検索対象者の個人情報テーブルと、「対人関係シソーラス³」を参照して、三者間の上下関係、親疎関係を求め、待遇表現(待遇/語調/呼称)を判定する。

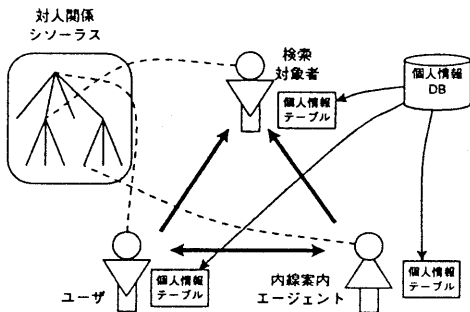


図3: エージェントを導入した内線案内のモデル

3 待遇判定処理

菊地は、待遇表現を判定するファクターを以下のように分類しているが、その判定方法は複雑で、人間ですら実際に使いこなせている人が少ないことをも指摘している [5]。

1. 社会的ファクター

(a) 場および話題

- i. その場の構成者: 話し手、聞き手、第三者のこと
- ii. 場面の性質など: 話し言葉か書き言葉か、手紙かメモかなど
- iii. 話題: その場の構成者のことかその場にはない第三者のことか

(b) 人間関係

- i. 上下の関係: 社会通念上の身分や地位の上下
- ii. 立場の関係: 恩恵を与える/受ける立場など立場の強弱
- iii. 親疎の関係: 親しい間柄とそうではない間柄
- iv. 内/外の関係: 聞き手から見て話し手の身内か否か

2. 心理的ファクター: 話し手の心理状態など

この分類に従うと、ユーザが要求する情報をエージェントがユーザに提示する対話では、1.(a) ii. および 1.(b) ii. が固定されている。また、内線案内では、個々のユーザ

³いわゆる組織図としての上下関係以外に、例えば、社内サークルなど、会社組織以外のリンクを張ることができる。

に関する情報を把握するだけでなく、ユーザとエージェント、さらには情報を提示する検索対象者との三者間について、ドメインに応じて、1.(b)の残りファクターを決定する。具体的には、「対人関係シソーラス」を用いて、三者間の人間関係のうち、上下関係(1.(b) i.)、親疎関係(1.(b) ii. と 1.(b) iii. を合わせたもの)を求める。

さらに、2.の心理的ファクターを上下関係、親疎関係から待遇表現を求める関数と考え、「丁寧型」/「標準型」/「ぞんざい型」といった複数のテーブルを用意し、それらをエージェントごとに「待遇タイプ」として割り付けることで、応答に用いるさまざまな待遇表現(待遇/語調/呼称)を生成する。

このような考え方に基づいて、待遇判定処理を作成した。待遇判定処理部とその他のモジュールとの情報の流れは以下ようになる(図4)。

1. Default DataBase⁴から、ユーザとエージェントのIDを得る。
2. 入力文から抽出された待遇表現(待遇/語調/呼称)を発話行為抽出部から得る。
3. ユーザ、エージェント、検索対象者について個人に固有の情報を個人情報DBから得る。
4. ユーザ、エージェント、検索対象者間の上下関係、親疎関係を判定するための情報を対人関係シソーラスから得る。
5. 上記の情報から決定された待遇/語調/呼称を文生成に送る。
6. エージェントごとに設定した声質情報(声立ての高さ/スピード)を音声規則合成に送る。

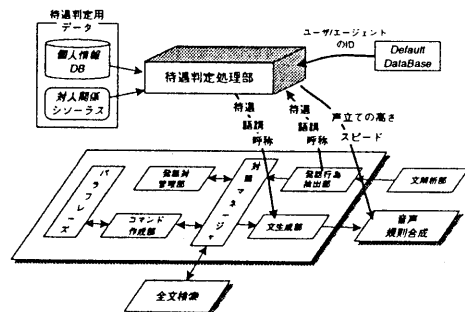


図4: 待遇判定処理部を含んだ対話管理部の構成

⁴NEXTSTEPに用意されている環境変数のデータベース

3.1 発話行為抽出部での処理

3.1.1 待遇の抽出

文解析用辞書の助動詞に、値を“尊敬”/“謙譲”/“丁寧”/“ぞんざい”とする「待遇項目」を付与することにより、その項目を持つ語が現れた場合、ユーザの待遇をその値とする。現れない場合には“丁寧”とする。また、複数個現れた場合には高い方(尊敬寄り)を採用する。

3.1.2 語調の抽出

文解析用辞書の助動詞に、値を“尊敬”/“丁寧”/“対等”/“ぞんざい”とする「語調項目」を付与することにより、その項目を持つ語が現れた場合、ユーザの語調をその値とする。現れない場合には、“対等”とする。また、複数個現れた場合には、値の高い方(尊敬寄り)を採用する。

3.1.3 呼称の抽出

文解析用辞書の「所長」、「部長」などの役職名を表す名詞、および「さん」、「くん」などの人稱名詞に接続する接尾語に、値をそれぞれ「役職」/「人稱」とする「呼称項目」を付与することにより、その項目を持つ語が現れた場合、検索対象者に対する呼称をその「見出し項目」の値とする。

3.2 待遇判定処理部での処理

ユーザと各エージェント間の待遇表現は、システム起動時に、Default DataBaseによって設定された値に基づいて、個人情報DBからロードした個人情報テーブルと対人関係シソーラスを参照して決定する。

また、内線案内の場合、検索対象者とユーザおよびエージェントの間の待遇表現は、検索対象者の待遇判定用データをロード後に随時決定する。

3.2.1 上下関係の判定

既知のユーザの場合には、待遇表現用データから求められる対人関係を優先する⁵。両者の個人情報テーブルに「役職項目」の値がある場合には、その値を比較して、上下関係(“上位”/“対等”/“下位”)を決定する。片方の個人情報テーブルのみに「役職項目」の値がある場合には、値がある方を“上位”とする。両者の個人情報テーブルとも値がない場合には、「入社年度項目」の値を比較して上下関係を決定する。

⁵多くの対話例の収集実験(たとえば[6])で報告されているように、計算機を相手にした対話の場合、ユーザは丁寧な表現を用いがちなため。

未知のユーザの場合には、発話行為抽出の語調の値が“ぞんざい”の場合ユーザが“上位”，“対等”or“丁寧”の場合“対等”，“尊敬”の場合エージェントが“上位”とする。

3.2.2 親疎関係の判定

既知のユーザの場合には、待遇表現用データから求められる対人関係を優先する⁵。両者の個人情報テーブルの所属項目の値から、対人関係シソーラスを参照して対人関係シソーラス上のノードの距離を求める。エージェントごとに設定する親疎閾値とノードの距離を比較し、それ以下であれば、親疎関係を“親近”とする。それ以外の場合、入社年度項目の値が同じであれば、親疎関係を“親近”とする。それ以外の場合、親疎関係を“疎遠”とする。

未知のユーザの場合には、“疎遠”とする。

3.2.3 待遇の決定

3.2.1, 3.2.2により求められた上下関係・親疎関係の値から、表1に示すような待遇テーブルを参照して、エージェントがエージェント自身、ユーザ、検索対象者を扱う待遇を決定する。なお、ユーザおよびエージェントの自分に対する親疎関係の値はそれぞれ“親近”，上下関係の値はそれぞれ“対等”を用いる。

表 1: 待遇テーブルの例

親疎関係		エージェントとの上下関係		
ユーザ	エージェント	上位	対等	下位
親近	親近	丁寧	ぞんざい	ぞんざい
親近	疎遠	尊敬	丁寧	丁寧
疎遠	親近	謙譲	謙譲	謙譲
疎遠	疎遠	尊敬	丁寧	丁寧

たとえば、未知のユーザに対して、エージェント自身の上司の情報を提示する、すなわち、ユーザとエージェント、ユーザと検索対象者⁶の親疎関係が“疎遠”で、エージェントと検索対象者の親疎関係が“親近”，エージェントに対してユーザが“上位”，検索対象者が“上位”の場合、エージェントがそれぞれに対する待遇は、ユーザ:“尊敬”，エージェントおよび検索対象者:“謙譲”となる。

3.2.4 語調の決定

3.2.2で決定した親疎関係、3.2.3で決定した待遇の値から、表2に示すような語調テーブルを参照して、エージェントのユーザに対する語調を決定する。

⁶他のエージェントに対する待遇は検索対象者に同じ。

表 2: 語調テーブルの例

ユーザ-エージェント 間の親疎関係	待遇		
	尊敬 or 謙譲	対等	ぞんざい
親近	丁寧	ぞんざい	対等
疎遠	尊敬	丁寧	丁寧

3.2.5 呼称の決定

3.2.3と同様に、表3に示すような呼称テーブルを参照して検索対象者及びエージェント自身以外のエージェントの呼称を決定する。なお、表中の「優先」項目は、個人情報テーブルにその値があれば優先して用いる呼称である。

また、発話行為抽出で求められた呼称と、待遇判定処理で判定されたユーザが用いるべき検索対象者の呼称とが一致しないものを検索対象から外す。

表 3: 呼称テーブルの例

親疎関係		上下関係			
ユーザ	エージェント	優先	上位	対等	下位
親近	親近	(なし)	さん	(nil)	くん
親近	疎遠	役職	さん	さん	くん
疎遠	親近	役職	(nil)	(nil)	(nil)
疎遠	疎遠	役職	さま	さま	さん

3.3 応答文の例

3.2で決定した待遇表現(待遇/語調/呼称)に基づき、応答文を生成する。以下に、内線案内の応答文の例を示す。

「酒井」と同等の個人情報を持つエージェントに対して、

1. 未知のユーザが「田村所長(エージェントの上司)のメールアドレスを示せ。」と入力した場合

上下関係: ユーザ, 検索対象者(田村所長)ともエージェントに対して“上位”

親疎関係: ユーザはエージェントに対して“疎遠”, 検索対象者はエージェントに対して“親近”

待遇: 検索対象者に対して“謙譲”, ユーザに対して“尊敬”, エージェント自身に対して“謙譲”

語調: “尊敬”

検索対象者の呼称: 所長

となり、「田村所長のメールアドレスは tamura@canon.co.jp でございます。」という応答文を生成する。

2. “親近”で“上位”のユーザが「相澤くん(エージェントの後輩)の内線を知りたい。」と入力した場合

上下関係: ユーザはエージェントに対して“上位”,

検索対象者(相澤)はエージェントに対して“下位”

親疎関係: ユーザ, 検索対象者ともエージェントに対して“親近”

待遇: 検索対象者に対して“ぞんざい”, ユーザに対して“丁寧”, エージェント自身に対して“ぞんざい”

語調: 丁寧

検索対象者の呼称: くん

となり、「相澤くんの内線は 3556 です。」という応答文を生成する。

3. “親近”で“対等”な三者間の場合に「酒井(エージェントと対等)のアドレスは。」と入力した場合

上下関係 三者とも“対等”

親疎関係 三者とも“親近”

待遇: 三者とも“ぞんざい”,

語調: ぞんざい

検索対象者の呼称: nil(呼び捨て)

となり、「酒井のアドレスは keiichi@cis.canon.co.jp だよ。」という応答文を生成する。

4 実験システム

今回提案したエージェントのモデルと待遇判定処理の効果を検証するための実験システムを作成した。今回提案したモデルでは、ドメインごとに設定したエージェントの切替をいかにユーザに提示するかによって、その効果を調査する。また、今後の評価の基準とするための対話例も同時に収集する。

実験システムでは、NEXTSTEPのGUIを用いて、プリファレンスでエージェントの属性や待遇判定用のテーブルの値を切り替えることができる。また、各エージェントに対して、状態ごとにアイコンを割り当て、切り替えることによって、エージェントの切替を強調する。

4.1 プリファレンスによるエージェントの設定

個人情報設定プリファレンスパネルを図5に示す。プリファレンスパネルでは、エージェントの待遇判定情報、応答文で用いる色とフォント、および合成音声情報(声立ての高さとスピード)を入力し、「待遇タイプ⁶」をメニューから選択することによって、待遇テーブル(表1)、語調テーブル(表2)、呼称テーブル(表3)の既定値を設定することができる。さらにそれぞれのテーブルの値は、メニューから微調整することができる。

⁶表1,2,3と同じ値のセットを持つ「標準型」、それより丁寧な値のセットを持つ「丁寧型」、ぞんざいな値のセットを持つ「ぞんざい型」を用意する。

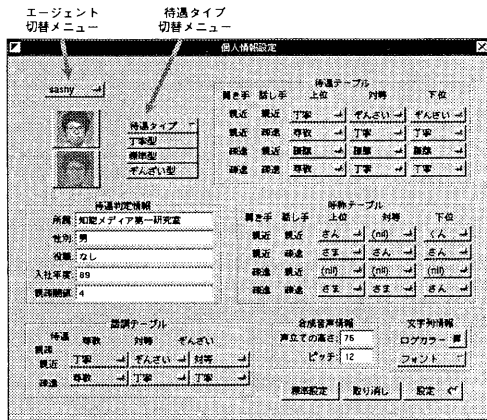


図 5: プリファレンスパネル

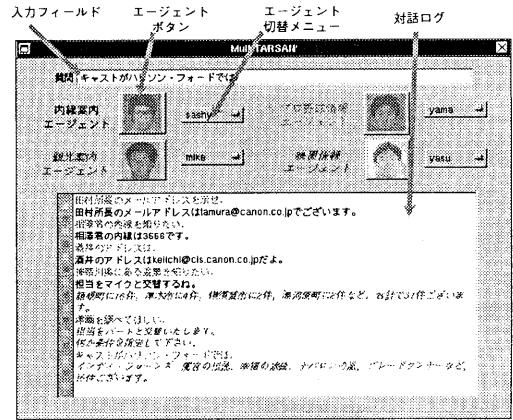


図 6: 実行画面

4.2 実行画面

実行画面 (メインウィンドウ) を図 6 に示す。メインウィンドウは、入力フィールドと対話ログ、4つのドメイン (内線案内、観光案内、プロ野球情報、映画情報) を持ち、各ドメインごとに、エージェントボタン、エージェント切替メニューを持つ。

エージェントには、予め、そのドメインの対話戦略⁸を割り当てている。また、エージェントボタンは、どのエージェントがアクティブになっているかを示す。(図 6 では、映画情報エージェントがアクティブになっている。)

ドメインの切替は、そのドメインを示す語が入力文に現れることにより、システム主導で行われる。ただし、エージェントボタンをクリックすることによっても、ユーザがユーザ主導でドメインを切り替えることができる。

5 おわりに

本稿では、ドメインごとにエージェントを導入した対話モデルを提案した。また、そのモデルに基づいた待遇判定処理を作成し、これらの効果を検証するシステムを作成した。

今後は、実験システムを用いて、対話例を収集するとともに、今回提案したモデルの評価を行う。

評価方法としては、実験システムのプリファレンスを用い、以下に示すような複数の実験環境の下で、エージェントの有無による効果を調査する。たとえば、以下の条件では、 $2^3 = 8$ 通りの実験環境を作成できる。

- (A) エージェントをすべて同じにするものと変えるもの、

⁷待遇判定処理の有無による効果を比較実験するために、すべて同じエージェントにできる。

⁸必須条件項目、シソーラス項目、デフォルト対象項目などの対話項目 [7]。

すなわち、待遇、語調、呼称が一定のものに変えるもの

- (B) すべてのエージェントの応答文のフォントや色を同じにするものとエージェントごとにフォントや色を変えるもの

- (C) すべてのエージェントの声質を同じにするものとエージェントごとに声質を変えるもの

謝辞

本研究を行なう機会を与えて下さいました情報メディア研究所 田村秀行所長に感謝いたします。

参考文献

- [1] Alan H. Bond and Les Gasser, editors: *Readings in Distributed Artificial Intelligence*, Morgan Kaufmann Publishers, Inc., 1988.
- [2] Brenda Laurel: Interface agents: Metaphors with character, In Brenda Laurel, editor, *The Art of Human-Computer Interface Design*, pp. 355-365. Addison-Wesley, 1990.
- [3] 長尾 確: 音声対話とマルチモーダル・インタフェース, O plus E, No.185, pp. 123-140, 1995-4.
- [4] 加藤恒昭, 中川 優: 質問応答における意図の把握と話題の管理, 情処学研報, NL58-6, 1986-11.
- [5] 菊地康人: 敬語, 角川書店, 1994.
- [6] 速水 悟, 伊藤克亘, 田中和世: 音声対話システムの構築とそれを用いた会話音声収集, 信学技報, SP91-101, 1991-12.
- [7] 酒井桂一, 池田裕治, 藤田 稔: ガイダンスシステムのマルチドメイン化の試み, 言語処理学会第 1 回年次大会, C3-3, 1995-3.
- [8] 酒井桂一, 池田裕治, 藤田 稔: パラフレーズ機能を備えた自然言語インタフェースシステム, 情処論, Vol. 34, No. 7, pp. 1518-1526, 1993-7.