

# 知能ロボットにおける挨拶文の自動生成方式

## The Automatic Generation System of the Greeting Sentence in Intelligent Robots

吉村 枝里子\*1  
Yoshimura Eriko

渡部 広一\*1  
Watabe Hirokazu

河岡 司\*1  
Kawaoka Tsukasa

\*1 同志社大学大学院工学研究科

Department of Knowledge Engineering and Computer Sciences, Graduate School of Engineering, Doshisha University

**Abstract :** Recently, the computer is spreads in general, but it requires practicing its interface (key board or mouse), when we treat for the first time. In order to use it more easily, the easy communication system is required. In order to realize these kinds of computer, it is needed to make the system to response naturally conversation. This paper proposes a method of the Automatic Generation System of the Greeting Sentence in Conversation System.

### 1. はじめに

コンピュータは大規模かつ高速な計算が可能であり、上手く活用できれば人間の暮らしを豊かに、また、高度に発展させることができる。しかし、活用するためにはコンピュータのインターフェースに人間が合わせその使い方を学習しなければならない。そのため、現在のコンピュータは誰でも簡単に操作できるとはいえない。人間がコンピュータをより容易に利用するためには、コンピュータとの簡単なコミュニケーションが必要である。そこで、実際に人間同士が行うような会話処理をモデル化し、コンピュータのインターフェースに取り入れることが望まれる。

人間同士の会話では、一般的に最初に挨拶が行われる。挨拶はコミュニケーションを円滑にする一端を担っているのは確かなことである。コンピュータ及びロボットが人間のように自然で円滑な会話・応答を行うことを目標としているため、本稿では会話処理の中でも特に挨拶処理についての仕組みを提案する。

### 2. 会話処理メカニズム

本稿における会話処理とは入力文の意図を理解し、その意図に基づいて意味理解を行うことで応答・発話処理を行うものである。

まず入力文を意図理解システムによって挨拶・呼掛、命令・依頼、質問、情報の4つに分類する。これにより、話し手が何を期待しているかという意図を判断する。次にその意図に基づいて詳細な意味を解析する意味理解システムに通し、その内容から応答処理を行う。

この会話処理の流れを図1に示す。このうち本稿では意図理解システムにより挨拶

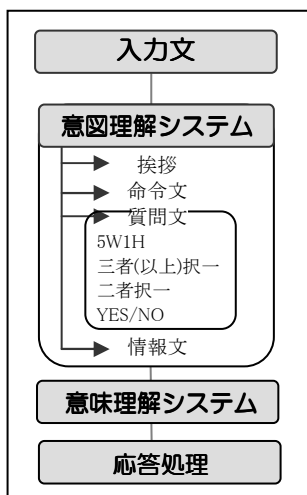


図1 会話処理の流れ

定した応答処理の仕組みを提案する。

### 3. 挨拶処理メカニズム

会話するロボットに代表される一般的な挨拶応答処理は作成者が用意したいくつかのテンプレートをそのまま使っている。そのため、毎回同じ応答を再現しパターン化されてしまい、発展性がない。この問題を解消するには無数のテンプレートを用意しなくてはならない。これには非常に多くの手間と労力が費やされるという新たな問題が発生する。

そこで、本稿においては少ないテンプレートからその時の状況に応じた多種多様な応答文を返すことを目的とする。これを実現するために、本研究では概念ベースと呼ばれる知識ベースを使用する。また、定型的な挨拶語を一言返すだけでなく、会話を広げるための追加語を生成することに注目する。例えば、「こんにちは」という入力語に対してその時の状況から「今日は晴れですね」「気分はどうですか」「夜空が美しいですね」などの言葉を返すものである。

挨拶処理メカニズムではまず入力文が挨拶文かを判断する。挨拶ならば、出会い・別れ・呼掛などの状況に分ける。次に基本応答語を選択し、一文目にその応答文を返す。この挨拶処理の流れを図2に示す。

会話を発展させるのは人と別れる場合ではなく、出会うときだと考えられる。そこで、状況が「出会い」の時に追加する文を作成し、これを二文目として基本応答語に追加し、挨拶の応答とする。

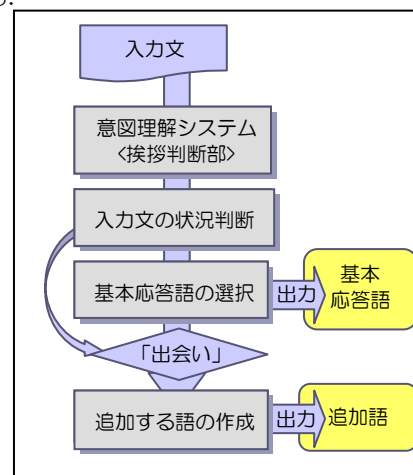


図2 挨拶処理の流れ

## 4. 概念ベースと関連度計算

### 4.1 概念ベース

概念ベース(CB)とは、複数の国語辞書や新聞等から機械的に自動構築した、語(概念)とその意味を表す単語集合(属性)からなる知識ベースのことである。本研究で使用した概念ベースには概念と属性のセットが約9万語登録されている。また、この概念と属性のセットにはその重要性をあらわす重みが付与される。ある「概念 A」は、その概念と関連が深く、その概念の意味となると考えられる「属性  $a_i$ 」と、その属性の重要性を表す「 $w_i$ 」の組の集合<sup>\*式(1)</sup>で表される。この重みは情報量を利用して付与されている。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_i, w_i), \dots, (a_n, w_n)\} \dots \dots (1)$$

属性  $a_i$  を概念 A の一次属性と呼ぶ。これに対し、 $a_i$  を概念とした場合の属性を A の二次属性と呼ぶ。展開していけば一つの概念は N 次までその属性を持つことができる。[大井 2002]

### 4.2 関連度計算

二つの概念 A, B の関連の深さを、関連度という方法を用いて定量的に評価する。この関連度を求める計算方式は、それぞれの概念を二次属性まで展開し、重みを利用した計算によって最適な一次属性の組み合わせを求め、それらの一致する属性個数を評価することで算出する。

この関連度の値は 0~1 の実数値をとり、値が高いほど関連の深い語であることを意味する。属性が全て一致する(完全な同義語及び同じ語)場合、1 となる。[大井 2002]

## 5. 挨拶判断

相手が入力した文は、まず意図理解システムで、その意図を「挨拶・呼掛」、「命令・依頼」、「質問(5W1H, 三者択一文, 二者択一文, YES/NO)」、「情報」に分類する。この意図理解システムで、挨拶と判断するために、必要な知識ベースを作成した。

ここで作成した知識ベースが、基本挨拶語データベースである。この基本挨拶語データベースには基本的な挨拶語と状況を約 200 セット格納している。この基本挨拶語データベースの一部を表 1 に示す。意図理解システムでは、このデータベースと一致する入力文を挨拶と判断する。

表 1 基本挨拶語データベースの一部

挨拶語	状況
あけましておめでとう	年始の挨拶
おはよう	朝の挨拶
おはようございます	朝の挨拶
こんにちは	昼の挨拶
こんばんは	夜の挨拶
さようなら	別れの挨拶 1
さよなら	別れの挨拶 1
:	:

## 6. 第一応答文処理(基本応答語)

挨拶判断で作成し利用した基本挨拶語データベースに、基本的な応答も格納した。これらは状況によって分類し、その状況はそれぞれ挨拶語の状況と対応している。

このデータベースを使用して第一応答文処理を行う。入力文の状況から基本的な応答語を取り出して一語を選択し、これを挨拶に対する一文目の応答として返す。

例。「別れの挨拶」→さようなら、バイバイ

## 7. 第二応答文処理(追加語)

会話を発展させるために基本応答語に追加する語を同義語辞書・反対語辞書・概念ベース、また関連度計算などを利用して作成する。その流れはテンプレート選択、単語抽出、分類判断、多種類増殖、現在状態とのマッチング、単語変換及びテンプレートへの返戻である。これらについて、次小節で述べる。

### 7.1 テンプレート選択

まず、基本的なパターンとなるテンプレートを選択する。テンプレート選択は、何を話題に選ぶかが問題となるが話題判定には判断要素が存在する。そこで、判断要素を“入力文に依る話題設定”、“現在状態に依る話題設定”、“その他”と分けた。

“入力文に依る話題設定”では、入力文が疑問文のときは[聞き返し]、天候に関する話題のときは[天候]、初対面の会話であれば[初対面]、長く会っていない場合の挨拶には[久しぶり]の話題を設定する。

入力文が話題に関係しなければ“現在状態に依る話題設定”に移る。普段と比べてうるさい場所や寒い場所であれば、人間はその話題を口にしやすい。そこで[天候][温度][環境音][相手との再会時間]等の現在情報から普段とどのくらい違うかという異常さを点数化する。この点数が高いものほど重要度の高い話題と判断する。

入力文からも現在状態からも話題が設定されなければ“その他”とし、いつでも使用できる[天候][時間][相手]内の話題を設定する。

このように、話題を設定した後、その話題に関するテンプレートを追加語テンプレート集から選択する。テンプレート文章と話題のセットをデータベース化したものが追加語テンプレート集である。“入力文に依る話題設定”のテンプレート選択の例を図 3 に示す。

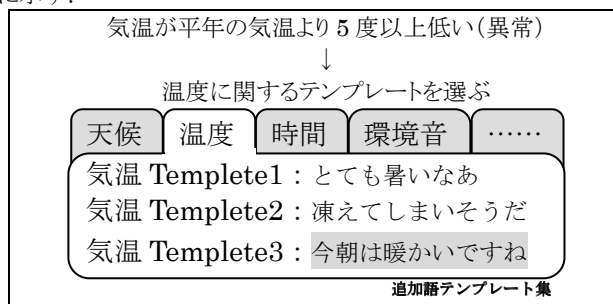


図 3 テンプレート選択例

### 7.2 単語抽出

7.1節で設定したテンプレート文から形態素解析を行い、名詞と形容詞を抽出する。この例を図 4 に示す。

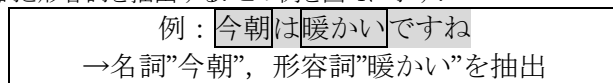


図 4 単語抽出例

### 7.3 分類判断

7.2節で抽出した各々の単語が変換元の単語となる。この単語を[時間][天候][気温][場所]に分類する。

単語の分類方法は変換元単語と分類項目間で関連度を計算し、それらのうちで最高関連度が得られたものとする。これによって変換元の単語はそれぞれ分類される。但し、全ての単語間で低関連度であることを想定し、閾値を実験により設定した。関連度が低く、閾値処理により分類されない単語は7.5節“現在状態とのマッチング”に進む。

変換元単語と分類との関連度の例を図5に示す。図5の場合、変換元単語の「今朝」と最も関連度の高いものは[時間]となる。このため、「今朝」は[時間]に分類する。

変換元単語	分類名	関連度(0~1)
今朝	時間	0.13
	天候	0.0057
	気温	0.0034
	場所	0.0038

図5 分類との関連度計算例

### 7.4 多種類増殖

分類された語を、反対語辞書を用いて増やす。反対語は意味的に逆転するが、種類としては同じである。直接求められる反対語を一次展開の反対語とし、この一次展開された語の反対語を二次展開の反対語とする。ここで一次展開した語群に二次展開した語群を追加し、同じ語を削除する。この語群を多種類増殖における候補群とし、出てきた単語は分類語との関連度計算により雑音を排除する。この例を図6に示す。

暖かい→寒い, 涼しい, 暑い, 暖かい  
 一次展開: 寒い, 涼しい 二次展開: 暑い, 暖かい

図6 多種類増殖例

### 7.5 現在状態とのマッチング

7.4節の単語から現在状態を用いてどの単語を選ぶのかを決める。分類には各々、現在状態から基本的な単語を付与する。その単語と7.4節の単語群との関連度計算を行い、最も現在状態に近い単語を決定する。これを図7に示す。

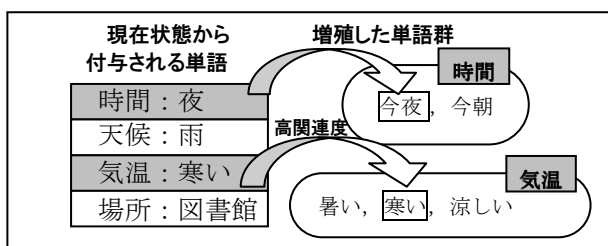


図7 現在状態とのマッチング

### 7.6 単語変換

7.4節及び7.5節で決定した単語を同様の意味の単語に拡張する。に利用する辞書には同義語・類義語辞書・概念ベースの比較実験を行い、適切だと結論付けられた同義語辞書を使用した(8.1節)。こうして得られた単語群が単語変換の候補となる。この中の一語をランダムに選択し、テンプレート内の語と置換することによって単語変換とする。単語変換の例を図8に示す。

今夜→今夜, 今夜, 今日, 夜中...  
 寒い→寒い, 寒々しい, 冷たい...  
 今朝は暖かいですね → 今夜は寒い ですね

図8 単語変換例

## 8. 実験評価と考察

### 8.1 単語変換方式における拡張方法の検討

次の五つの辞書を使い、7.6節の単語変換方式の単語拡張を行った。各辞書の平均正答数(各々の出現単語中正答単語数の平均)と正答率(全出現単語中の正答単語数の割合)を図9に示す。

- ①概念ベース
- ②同義語+類義語
- ③同義語
- ④同義語∧概念ベース
- ⑤(同義語+類義語)∧概念ベース

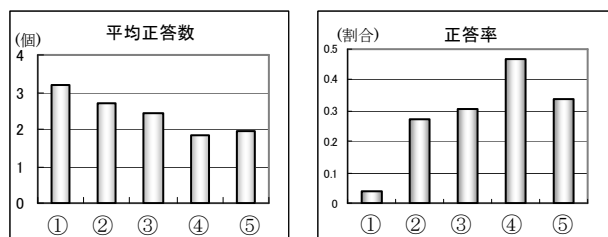


図9 各辞書の平均正答数と正答率

ここでは、次のテストデータを用意した。“TOEIC 会話問題集”・“英語で手紙を書く例文集”から300文の文章を用意し、これらの文章から形態素解析を行って名詞と形容詞を抜き出す。その単語を出現頻度の多い並べ替え、その上位25個の単語をこの単語変換方式のテストデータとして用いる。

平均正答数と正答率を検証した結果から、③同義語辞書を採用するのが適切と考えられた。同義語辞書を用いた場合正答率は30.5%、一単語辺りに出現する平均正答数は2.44個であった。

### 8.2 単語変換方式の雑音除去・精度向上

8.1節の単語変換方式ではあまり精度が高くない。検証した結果、次のような問題があることがわかった。

- 同義語辞書 DB に多義性のある言葉が含まれている。
- 同義語辞書 DB に同義の意味が少ない言葉が含まれている。
- 日常会話にあまり使用しない単語が含まれる。

まず、多義性の問題がある。言葉には多義性を含むものが存在する。同じ読み、漢字でありながら、複数の意味を持つものである。多義性を含む単語を同義語辞書データベースで検索すると、その異なる意味に関する同義語が出てくることがある。

また、同義語 DB 精練に問題がある。語「朝」を膨らませると「あした」「翌朝」など、関係はあるが同義とはいえない言葉が出現する。これは、同義語辞書データベースが機械的に自動構築されたため、精練されていないためだと考えられる。

三つ目の問題点に挨拶に使用する日常性がある。同義語 DB の中には同義ではあるが、日常会話にあまり使用しない語が出現する。挨拶に使用する単語としては日常的に用いられる単語が必要のため、これらの単語は不適切だと考えられる。

これらの問題点を解決するために関連度と WEB 検索 HIT 数を利用する2つの手法を試行した。

(1) 関連度計算による雑音除去・精度向上

7.6節の単語拡張方式で拡張された単語群と分類項目との間で関連度計算を利用し、雑音除去を図った。拡張された単語群と単語の分類語との関連度を計算し、ある閾値以下のものを切り捨てることで関係の低い雑音を除去する。閾値を高く設定すれば低関連の語が除去され、関連の高い語のみが残ることになる。しかし、その閾値が高すぎると、その一単語に対して変換する候補が極端に減少し、言い換える候補が無くなることになる。そのため閾値を設定するための実験を行った。この閾値を様々な値に設定した単語変換候補の正答率と平均正答数の結果を図 10 に示す。

結果、平均正答数と正答率から閾値を 0.1 に設定することが妥当と考えられた。これは閾値が 0.13 以上になると平均正答数が極端に低くなり、閾値が 0.1 の場合と比較して正答率が大きく変化しないからである。これにより、精度が約 10% 向上した。

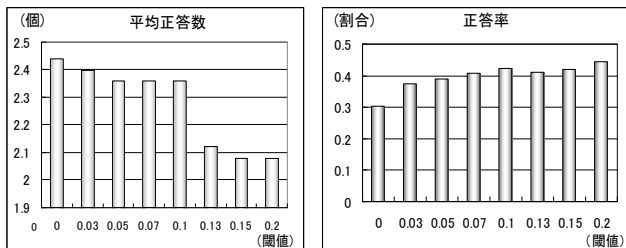


図 10 閾値を設定した正答率と平均正答数

(2) Web 検索数を利用した雑音除去・精度向上

日常会話にあまり使用されない単語を除去するため WEB 検索 HIT 数を利用する。

WEB 空間には様々な情報が存在するが、その情報は言語によって書かれている。日常生活で用いられやすい語ほど検索 HIT 数は高い。例えば、検索エンジンに「私」という単語を入力した場合は約 9,800,000 件の HIT 数が得られる。しかし、単語「下拙」を入力すると検索 HIT 数は約 70 件となる。これらの中には有意な差が見られる。

しかし、全ての単語群に同じ閾値を設定すると不都合が生じる。例えば一人称として利用する単語「私」の候補群は、文中で頻繁に利用される可能性が高い。よって、WEB 空間上でその使用頻度の絶対数が多いと考えられる。このように、その単語ごとに他の単語に比べ絶対数が極端に大きい場合や小さい場合がある。このような単語に同じ基準の閾値を設定できない。

そこで、基準とする単語を決め、基準単語の検索 HIT 数から、その語群の検索 HIT 数の閾値を算出することを提案する。この基準単語には、元のテンプレートに使われている変換前の単語を使用する。変換元単語は元来、日常会話に使う正しい語であるからである。

よって、変換元単語の検索 HIT 数を基準とし、変換後の単語の検索 HIT 数との比較を行った。また、(1) で決定した関連度計算の雑音除去を利用する場合と利用しなかった場合についても実験を行った。これらの結果を図 11 に示す。

結果、平均正答数と正答率から、閾値に変換前単語の検索 HIT 数の 1/100 を設定するのが妥当だと考えられた。このことで正答率は 48.6% となり、18.1% の精度向上が見られた。

また、関連度計算による閾値も加味した場合、正答率は 67.1% となり、WEB 検索数のみよりも 18.5% の向上、関連度計算のみよりも 24.8% の向上、何も行わなかった場合より 37.6% の向上がみられた。

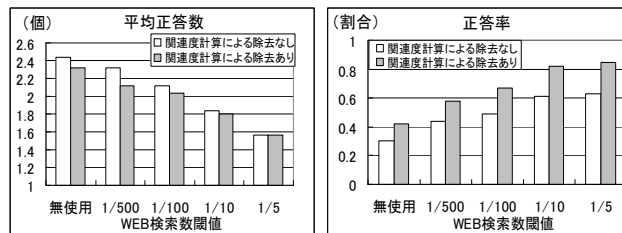


図 11 検索エンジンの HIT 数を利用した単語変換候補の正答率と平均正答数

(3) 応答文全体としての精度と考察

入力した挨拶に対して、応答された文章全体の精度を求めたものを図 12 に示す。

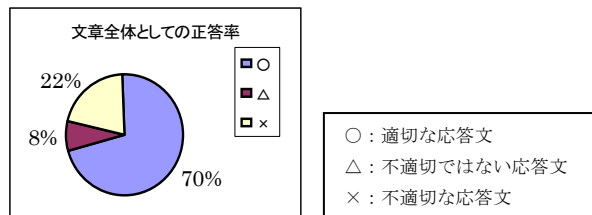


図 12 文章全体としての正答率

応答文の例を図 13 に示す。「不適切な応答文」には単語変換の失敗が多く見られ、「不適切ではない応答文」には状況によって正しくなるものや、誤りではないが日常的に使われない語を含む文章などが存在した。このため単語変換方式をより改良し精度向上させる必要があると考えられる。

適切な応答文(○)の例:  
 入力文「はじめまして」  
 → 応答文「はじめまして。お会いできてうれしいです。」

不適切ではない応答(△)の例:  
 入力文「お元気ですか」  
 → 応答文「おかげさまで。貴女は？」

不適切な応答(×)の例:  
 入力文「おはよう」  
 → 応答文「おはようございます。涼しいあしたですね。」

図 13 応答文の例

9. おわりに

本研究では、入力文が挨拶の場合にその応答文を生成する仕組みを提案した。現在の状態や様々な辞書を用い少ないテンプレートから現在状態に対応した応答文を自動生成した。単語変換方式の改良を行うことで精度を上げ、より自然な会話に近づけることが今後の課題である。

謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「知能情報科学とその応用」における研究の一環として行ったものである。

参考文献

[大井 2002] 大井 健二, 渡部 広一, 河岡 司: “関連度計算を用いた質問応答方式”, 情報処理学会知能と複雑系研究会資料, pp.25-30, 2002 年.