

抽象化を用いた会話機能

平 伸也 三浦 孝夫

法政大学工学部電気電子工学科

塩谷 勇

産能大学経営情報学部

1 前書き

本稿では、自然な会話機能を実現するための方式を論じる。このためには、多種多様な会話文章を生成し選択する機能が必要である。しかし語彙だけから構成される辞書の素朴な利用や、キーワードに固定した返答を対応させるなどの素朴な方式では自然さを得にくく、会話を継続することは困難である。また、会話システムが文章を学習する機構を導入することが難しいことも問題である。

本論文では、辞書に概念階層や各文法に基づく推論を導入し、また各語彙に表情を定めておくことで、上記の問題を改善することができることを示す。

2 会話機能

本章では、はじめに会話生成の基本的な構成を示す。利用者が入力した会話文章を手がかりに、その文章の構造を解析し、構造と辞書を用いて返答を生成するというのが基本的な流れである。

2.1 形態素解析

文章の構造を解析するために形態素解析を行う必要がある。形態素解析とは、自然文を意味のある最小の単位(形態素)に分解し品詞を見分けることをいう。基本的には、辞書をもちいてその語彙が属する品詞を調べていくという方法をとる。日本語では品詞の接続には制限が存在するため、この特徴をうまく利用することによって形態素に分割し品詞を見分けることができる。

例えば「太郎は花子に本を渡した」という文章は「太郎/は/花子/に/本/を/渡し/た」と分割され、「太郎」(名詞 - 固有名詞)、「は」(助詞 - 係助詞)、「花子」(名詞 - 固有名詞)、「に」(助詞 - 格助詞)、「本」(名詞 - 一般)、「を」(助詞 - 格助詞)、「渡し」(動詞 - 自立)、「た」(助動詞)であると判定する。

反面「わかち書き」や「ルーズな文法」などに対する問題点があり、必ず正しい結果が期待できるとは限らない [2]。

2.2 言い換え規則

言い換えとは、同じ内容を別の表現で記述する技術をいう [4]。様々な言い換えがいられているが、本システムでは文章の構造のみで言い換えられるものだけを扱う [2]。これは一般には「N しか V ない」という構造を「N だけ V」という構造に書き換える。ここで N は名詞を、V は動詞を表す。このような規則を用いる事で、文章構造から言い換えを機械的に生成することができる。

本研究で、実際に利用する規則は次のものである。複数の候補がある場合はランダムに決定する。

- (1) 「N しか V しない」を「V するのは N だけだ」に言い換える
- (2) 「V のが N である」を「N は V」に言い換える
- (3) 「(人) は N が V」を「(人) が V なのは N だけ?」および「(人) は N のどんなところが V?」のいずれかに言い換える。

2.3 表情付け

人間同士の場合、互いに相手の顔(の表情)を見ながら会話をするため、状況が強調され意図の伝達が容易になる。このように会話における表情は、自然な会話の促進に重要な役割を果たす効果が期待できる。本研究では、会話に用いられる語彙にパラメタを付すことにより、表情を変化させる機構を実装する。

各文章が表す表情を選択するため、文章に生じる名詞語彙に表情パラメタをあらかじめ定義する。表情パラメタは図 1 のような喜怒哀楽の 4 つの表情と、その表情の度合いを表す数値で構成される。

文章に出てくる語彙の表情パラメタの合計を、その文章の表情とする。ただ、会話の文脈で大きく表情が変わるのは自然ではないため、本研究では入力文章の表情を計算し、これまでの表情パラメタ数値にその数値を加え、最大の表情を選ぶ。

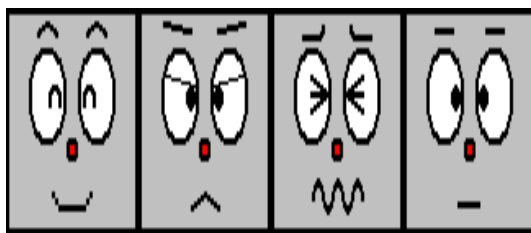


図 1: 表情の設定

3 会話文章の抽象化

会話文章を抽象化するために、本研究では語彙の概念階層を導入し、これを文章生成に利用する。

概念階層とは、図 2 のように、「鳥」は「動物」に含まれ「カナリア」は「鳥」に含まれている、というようにカテゴリ同士の関係を階層的に現す関連をいう。幾つかの階層は事前に辞書に登録しておくが、未知のカテゴリを学習する場合、そのカテゴリが持つ特徴、表情や概念階層の構成を獲得する必要がある。しかしこれらを必要の都度、全てを一度に質問

“Chattering using Abstraction Mechanism” : Shin'ya Taira, Takao Miura : Hosei University, Dept. of Elec. and Elec. Eng.

Kajino-cho 3-7-2, Koganei, Tokyo, JAPAN

Isamu Shioya : SANNO University, Dept. of Management and Informatics

Kamikasuya, Isehara, Kanagawa, JAPAN

するのは自然な会話を得にくいいため、会話中で少しずつ質問し、徐々に学習させていく方法をとる。

概念階層を利用し「鳥」から「動物」に語彙をおきかえることにより、例えば「私は鳥が好きだ」は「私は動物が好きだ」のように、文章の連続性を得る。

また、階層内での左右のつながりを利用し「鳥」の話から「魚」の話に話題を移す事ができる。これにより、「私は鳥が好きだ」という文章から「じゃあ魚は好き？」を生成できる。

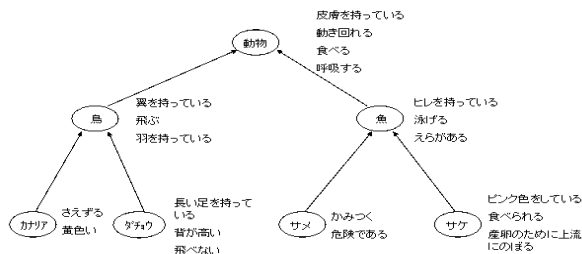


図 2: 概念階層

また、概念階層の属性を継承することにより、新たな文章を生成することができる。例えば「犬」カテゴリが「4つ足をもつ」という属性を有するとき、「土佐犬」が「犬」の部分カテゴリという事実から、「土佐犬」は「4つ足をもつ」という属性を継承する。

文章内に含まれている語彙は、属するカテゴリを検索することで特徴を得る場合が多い。このため、逆に特徴から概念階層内の位置を検索し、辞書内の探索を効率よく処理できることがある。

例えば、未知カテゴリが「鮭」であるとき、「鮭が泳いでいるのを見た」という文章の「泳ぐ」に注目し、カテゴリ「魚」が「泳げる」という特徴を有することから、「鮭」は「魚」の部分カテゴリであると推測できる。

自然な会話を実現させるためには、語彙だけから構成される辞書では、このような文章を生成することは容易ではない。すなわち、語彙間の関連性やこれら利用する状況をできるだけ用い、会話内容の滑らかな移動を可能にする機構が必須のものである。このために概念階層を言い換えに適用する試みる。この結果、入力された文章から複数の返答を生成できると考えられる。

入力文章を概念階層・推論規則を用いることで、様々な返答をつくる事ができる事を以下に示す。

- (1) 抽象化を利用: 「太郎は鳥が好きです」
「太郎は動物が好きなんだ〜」
- (2) 具体化を利用: 「太郎は鳥が好きです」
「太郎はカナリアが好きなの?」
- (3) 概念階層の移動: 「太郎は鳥が好きです」
「太郎は魚は嫌い?」
- (4) カテゴリ属性を利用: 「鳥」カテゴリが属性「翼を持っている」対応させているとき、「鳥って翼を持っているよね」

4 会話システムの実装

4.1 システム動作の流れ

本システムの起動時に、利用者は自分の名前を入力する。図 3 は、会話処理の状況を示す。会話を終わらせる時は「はいはい」などと入力する。

システムは起動されると、まず辞書を読み込む。利用者が自分の名前を入力し、システムが記憶すると会話の準備が完了する。以後のシステムの流れを図 4 に示す。入力された会話文章は、形態素解析により文法構造で表現され、表情が計算

無論、「犬は4つ足をもつ」という文が生成できる。

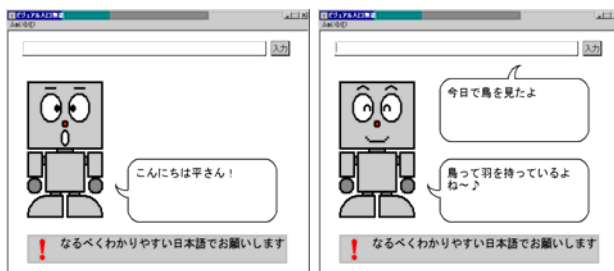


図 3: システムの利用

される。形態素解析には茶筌を用いる [3]。このとき未知のカテゴリの有無を調べ、必要ならば辞書に追加した後、文章生成機能により返答文を生成する。

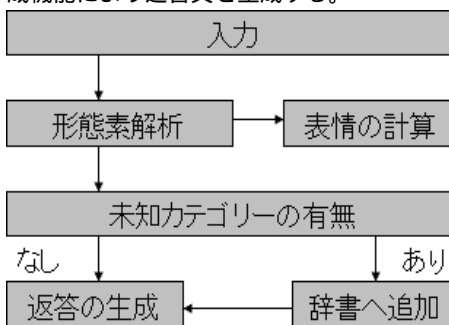


図 4: システム動作の流れ

4.2 辞書

辞書は概念階層に基づいてカテゴリ同士の関係を記憶している。辞書内でカテゴリは以下の 4 つから構成される: (1) カテゴリの特徴とキーワード, (2) 表情パラメータ, (3) 概念階層に基づくカテゴリの位置, (4) 学習途中かどうか。

ここで、キーワードとは、カテゴリの属性に含まれる最も重要な単語（名詞又は動詞の原形）とする。キーワードを確定するために、利用者に質問文章を生成し必要なら記憶する。属性を持たない場合もある。

新しいカテゴリを辞書に追加するとき、学習途中であることを明示する。学習途中のカテゴリは、新しい情報が追加されるたびに上述項目 (1),(2),(3) を更新する。これらが確定した段階で学習済みであるとし、以後変更しない。辞書内の探索方法はカテゴリ名とキーワードの 2 通りありランダムに方式を選択する。

5 結び

「概念階層」「推論規則」「表情」を用いることで自然な日本語会話を実現できた。会話特有の文章の表現方法を考慮すればさらに、自然な会話が可能になることが確認できた。

参考文献

- [1] 森敏昭・井上毅・松井孝雄: グラフィック認知心理学, サイエンス社, 1995
- [2] 乾健太郎: 言語表現を言い換える技術, 言語処理学会第 8 回チュートリアル, 2002
- [3] 松本祐治, 他: 日本語形態素解析システム “茶筌”(V2.2.1) 使用説明書, 2000
- [4] 増山繁, 山本和英: テキスト自動要約における新たな展開, 情報処理学会誌 V43, No.12, 2002