

## Twitter のリプライ文と役割語変換を利用した会話文の生成

木下 翔央<sup>†</sup> 福地 隆<sup>†</sup> 幸崎 剣<sup>†</sup> 韓 東力<sup>†</sup>日本大学文理学部 情報システム解析学科<sup>†</sup>

## 1. はじめに

人間とコンピュータに会話を行わせようとする試みは、1966年のELIZA[1]に始まり、現在に至るまで、様々な方法でアプローチが続けられている。これまでの研究や既存システムでは、返答文を生成する技法として、データベース型や n-gram マルコフ連鎖型などの方法が用いられてきた。本研究では既存の考え方をもとに、「キーワード型双方向マルコフ連鎖辞書」と Twitter の利用を新たに考案する。また、システムを人間と錯覚させることを目的のひとつとしているため、役割語によるキャラクター性の付与を行う。

## 2. システムの構成

本研究では、人工無脳研究サイト「人工無脳は考える」[2]に基づき、システムを三つのブロック: Planning、Generating、Encoding に振り分け、独立して稼働する各ブロックを統合することにより全体の目標を実現する。

全体のフローとしては、ユーザーの入力を Planning Block で分析し、その結果を Generating Block に渡して会話文を生成する。生成された会話文を Encoding Block に渡してキャラクター性を付与し、Encoding Block の出力が最終的に提示される出力となる。また、P->G->E のサイクルとは別に、各 Block に形態素解析やデータ構造の提供を行う Analyzing Block を設けている。実際のシステムでは、形態素解析の手法として、全て MeCab[3]を使用している。

## 3. Planning Block

Planning Block では入力文に対して返答と成り得る「関連度の高い単語」（以下「キーワード」）を Web 検索を利用して抽出し、抽出されたキーワード群を Generating Block へ渡す。検索は GoogleAPI[4]を使用する。図1は処理の流れを示している。

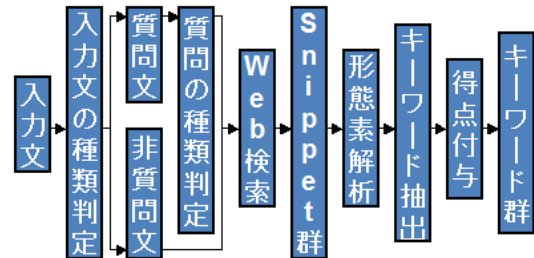


図1 Planning Block の処理の流れ

## 4. Generating Block

ここでは、前節の Planning Block から渡された「キーワード」を元に、マルコフ連鎖を使用して会話文を生成する。マルコフ連鎖の辞書に記録させる文章は、Twitter[5]の Public\_Timeline から日本語のつぶやきをランダムに取得し、そのなかでも特に Mention 属性を持ったものを使用している。これは、Twitter という Web サービスが、会話文的な性格を持ったつぶやきを投稿するサービスであること、そして Mention 属性を持ったつぶやきは、誰かに言及(Reply)することから、より会話文らしい文章になっているであろうという予想に基づくものである。また、各々のつぶやきからは、本来会話文に必要な記号(@や URL、RT、ハッシュ等)を除去し、文末判定を用いて文章を区切り、辞書に記録する。

マルコフ連鎖辞書の実装には、従来の「先頭から順に連鎖する」タイプの順方向に加え、「文末から順に連鎖する」タイプの逆方向のもの、「キーワード」と成り得る形態素のつながりのみを記録したキーワード辞書の二つを加えた、双方向型マルコフ連鎖辞書(Bidirectional Markov Dictionary: 以下「BMD」)を考案した。キーワード辞書とは、各方向のマルコフ連鎖辞書に記録される文章の中から、特にキーワードと成り得る「名詞、動詞、形容詞、副詞」のみに注目し、その品詞の前後に出現する形態素のみを記録した辞書である。BMD では、Planning Block から渡された「キーワード」がキーワード辞書に登録されているかの問い合わせを行い、存在する場合はその「キーワード」の前後に出現した形態素を元に、順方向と逆方向のマルコフ連鎖を、それぞれ文末と文頭に向かい行う。

高橋らの研究[6]でも双方向マルコフ辞書が使用されているが、特定の「キーワード」から文頭と文末への連鎖ではなく、本研究と異なっている。BMD を使用することにより、マルコフ連鎖の辞書を再構築することなく必要なキーワードを含んだ会話文を生成できることや、大量に記録された文章の中から効率的にキーワードを含んだ会話文を生成できるなどのメリットがある。

キーワード辞書には、Planning Block から渡された「キーワード」が登録されていない場合も考えられる。その場合には、BMD はそのキーワードをクエリにして、Twitter API を使用して検索を行い、その結果を BMD に記録させる。このように BMD を用いることにより、必ず「キーワード」を含んだ会話文の生成が可能となる。

### 5. Encoding Block

このブロックは金水が考案した役割語[7]から派生された「キャラ助詞」と「キャラコピュラ」の概念[8]に基づき、会話文に汎用的なキャラクターを付与することを目的としている。

Mecab の解析結果で「終助詞」とされる品詞をはじめとした、文章の終端につく品詞を、本研究では「語尾」と呼ぶ。文章中において語尾は「文の意味を大きく変化させない」、「性別、年齢、丁寧さの3つの印象に強く関わる」という特徴を持つ。

本研究ではキャラ助詞の分類方法を元に、日常会話で見られるような各々のキャラクターの口調を「汎用役割語」として設定し、発話文に反映させることを提案する。なお、キャラクターは口調に関わるもののみで、会話の癖・話題の傾向などの部分には関係しない。

表1 汎用的なキャラクター

男	少年	青年	高年	老年
乱暴	男, 乱, 少	男, 乱, 青	男, 乱, 青	男, 乱, 老
普通	男, 普, 少	男, 普, 青	男, 普, 青	男, 普, 老
丁寧	男, 丁, 少	男, 丁, 青	男, 丁, 青	男, 丁, 老
女	少年	青年	高年	老年
乱暴	女, 乱, 少	女, 乱, 青	女, 乱, 青	女, 乱, 老
普通	女, 普, 少	女, 普, 青	女, 普, 青	女, 普, 老
丁寧	女, 丁, 少	女, 丁, 青	女, 丁, 青	女, 丁, 老

終助詞の特徴から「性別」、「年齢」と「丁寧さ」の3つの「印象値」を設定し、それぞれに対してパラメータを設定した、24通りのキャラクターに対して汎用役割語を表1のように設定し、役割語変換動作を以下4つに細分類する。

- 【非丁寧語変換】：キャラクターの印象値に「乱暴」が含まれる場合行われる。
- 【丁寧語変換】：キャラクターの印象値に「丁寧」が含まれる場合行われる。
- 【語尾変換】：キャラクターごとに変換の規則が設定されており、全ての発話文に対して行われる。
- 【人称変換】：文中の「一人称代名詞」と「二人称代名詞」を探し、適切な形に変換する。キャラクターごとに変換の規則が設定されており、全ての発話文に対して行われる。

役割語変換によるキャラクター表現の例を以下に示す。「断定」の男口調を表す終助詞「ぜ」を付与することで、「乱暴さ」と「男らしさ」を表すことができる。

(例文)

「また、つまらぬものを斬ってしまった」

(男, 乱暴, 少年)

「また、つまらぬものを斬ってしまったぜ」

### 6. まとめ

本研究では既存研究で行われてきた会話返答文の生成方法をもとに、Twitter のリプライ文と役割語変換の利用を新たに取り入れる手法を提案している。今後は構築したシステムを実際に利用してもらい、手法全体および各モジュールの機能評価を行いながらより人間らしい会話文生成を目指していく。

### 参考文献

1. Joseph Weizenbaum: "ELIZA—a computer program for the study of natural language communication between man and machine", Communications of the ACM, Vol. 9, No. 1, pp36-45. 1966.
2. <http://www.ycf.nanet.co.jp/~skato/muno/>
3. <http://mecab.sourceforge.net/>
4. <http://code.google.com/intl/ja/more/>
5. <http://twitter.com/>
6. 高橋瑞希, Rafal Rzepka, 荒木健治: "Web検索と単語 n-gram モデルを用いた文生成手法の性能評価": 言語処理学会 第 16 年次大会論文集 pp. 391-394. 2010.
7. 金水敏 『ヴァーチャル日本語 役割後の謎』, 第 5 刷, 岩波書店. 2003.
8. 定延利之 「キャラ助詞が現れる環境」, 編: 金水敏 『役割語研究の地平』, 第 1 刷, pp27-48, くろしお出版. 2007.