# 日本語会話文を用いたキャラクタエージェントのための ジェスチャ自動生成

An Automatic Gesture Generation System for Character Agents by Using Japanese Scripts

滝 由貴　Werner Breitfuss　石塚 満
Yuki Taki　Mitsuru Ishizuka

東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

In this paper, we introduce an automatic gesture generation system. This system gives different types of non-verbal behaviors to an input Japanese dialogue script for two virtual character agents. Non-verbal behaviors are generated based of the analysis of linguistic and contextual information from the input Japanese text. The result of generated gestures is presented as the Multimodal Presentation Markup Language for 3D agents(MPML3D). This system allows us to make it very easy to create appropriate non-verbal behaviors like eye gaze and conversational gesture behavior.

## 1. はじめに

現在，オンライン上では Second Life[*1] やニコッとタウン[*2] を始めとする，様々なバーチャルワールドが提供されている．これらのバーチャルワールドでは，ユーザは自分の分身となるアバターと呼ばれるキャラクタエージェントを用いて，他のユーザとコミュニケーションを行う．その他にも，オンラインゲームや Web ページにおいて，ユーザ同士のコミュニケーションやユーザの活動を支援する役割を持つキャラクタエージェントが活躍している．

キャラクタエージェントを用いたコミュニケーションを円滑に行う上で，キャラクタエージェントによるジェスチャや顔表情などの非言語情報の表現が重要である．[中野 03] しかし，キャラクタエージェントに対して非言語情報を人手で付加することは，コストがかかる．多くのアバターサービスにおいては，ユーザが手動でジェスチャーを選択しなければならない．

この問題を解決するため，キャラクタエージェントへジェスチャを自動的に付加する研究が行われている．The Behavior Expression Animation Toolkit（BEAT）[Cassell 01] は，入力された英語文章からジェスチャを生成する．しかし，このシステムにおいてジェスチャ生成の対象となるキャラクタエージェントは１体のみである．また，他のシステムでは，提示されたジェスチャの中から，最適なジェスチャを人手で選択しなくてはならないものもある．

本研究では，２体のキャラクタエージェントが音声のスピーチおよびジェスチャを用いて，視聴者であるユーザに案内，または説明等を行う場面を想定し，日本語で入力された会話テキストから，２体のキャラクタエージェントに対してジェスチャを生成し付与するシステムを提案する．キャラクタエージェントは，話し手と聞き手に別れ，それぞれの役割に最適なジェスチャを行う．この時，聞き手のジェスチャは話し手のキャラクタエージェントに与えられた台詞から自動的に生成される．２体のキャラクタエージェントが会話を行い，ユーザの支援や説明を行うことにより，ユーザが聞き手のキャラクタエージェントに共感し，また，２体のキャラクタエージェントが交互に話すことにより，ユーザの注意を常に引きつけておくことを狙う．

連絡先: 滝由貴，東京大学情報理工学系研究科創造情報学専攻，〒 113-8656 東京都文京区本郷 7-3-1，03-5841-6774，e-mail: yuki@mi.ci.i.u-tokyo.ac.jp

∗1 http://jp.secondlife.com/

∗2 http://www.nicotto.jp/

## 2. 関連研究

BEAT システムは，入力された英語文章にタグ付けを行い，それを基にジェスチャを生成し，ユーザへ提示するシステムである．このシステムでは，キャラクタエージェントのしゃべる音声とジェスチャが同期する．また，The Conversational Agent System for neTwork applications（CAST）[Nakano 03] は BEAT システムを基にしており，入力された日本語文章から the RIStex animated Agent system（RISA）のためのジェスチャと表情を生成する．[中野 04] では，ジェスチャの出現に関係する日本語の言語特性を実際のプレゼンテーションデータから収集し，分析を行っている．さらに，分析結果を CAST システムに組み込み，ジェスチャの生成を行っている．ゲームキャラクタに対するジェスチャ自動生成として，ポーズ，姿勢，およびしぐさを合成し会話に組み込み，ゲームキャラクタの感情などの心理状態をプレイヤに伝達するという取り組みも行われている．[Nakano 06]

これらの取り組みでは，全て１体のキャラクタエージェントが対象になっており，話を聞く側のジェスチャについては考慮されていない．また，[Nakano 06] を除く上記のジェスチャ生成システムでは，文章中に登場する指示動作や強調語などに呼応してジェスチャを決定しているため，機械的な印象を与えやすく，同じ文章を入力すると，毎回同じジェスチャが出力され得る．日本語入力を基に話し手，および聞き手のジェスチャを同時に生成する点で，本研究は新しいと言える．

## 3. ジェスチャ自動生成システム

本システムは，本研究室の先行研究となる [Breitfuss 08] を基に構成されている．[Breitfuss 08] は，Language Tagging モジュール，Non-Verbal Behavior Generation モジュール，および Transformation to simple script or MPML3D モジュールからなる．MPML3D[Nischt 06] とは，“Multimodal Presentation Markup Language” の略であり，3D キャラクタエージェントのジェスチャを記述するための XML ベースの記述言語である．また，Language Tagging モジュールは，BEAT システムを基に作られている．

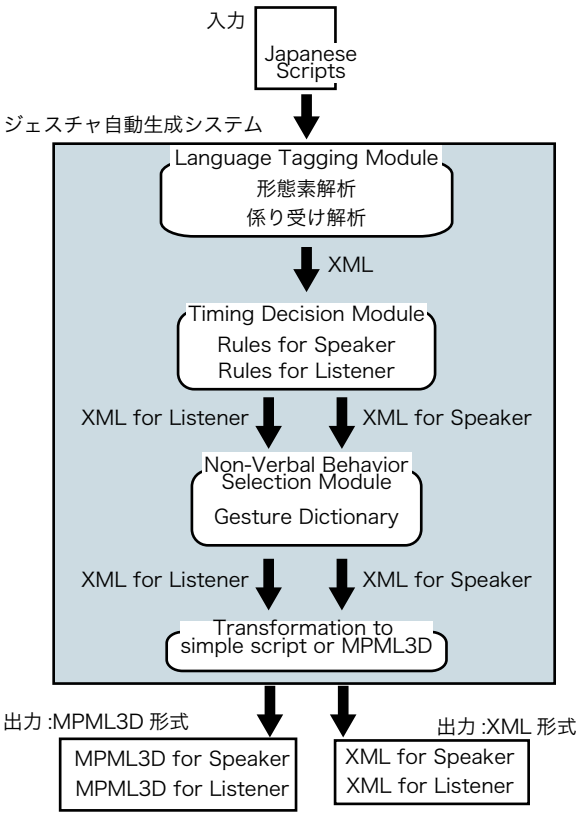


図 1: システム構成図

### 3.1 システム構成

図 1 に，本システムの構成図を示す．本システムは，Language Tagging モジュール，Timing Decision モジュール，Non-Verbal Behavior Selection モジュール，および Transformation to simple script or MPML3D モジュールの４つのモジュールから成る．

入力された日本語会話文は，まず Language Tagging モジュールにおいて，形態素解析，および係り受け解析され，その結果を基にタグ付けされる．次に Timing Decision モジュールにおいて，あらかじめ決められたルールとタグ情報に基づき，話し手と聞き手の役割毎にジェスチャを付与する文節（タイミング）を決定する．さらに，Non-Verbal Behavior Selection モジュールにおいて，役割毎に付与すべきジェスチャを決定する．最後に，Transformation to simple script or MPML3D モジュールにおいて，ジェスチャを行うキャラクタエージェントの環境に即した形式の XML 文，または MPML3D 形式の出力を生成する．システム内部のデータは，全て XML 形式で扱われる．

ジェスチャのタイミングを決定するモジュールと，ジェスチャの種類を決定するモジュールを分離することによって，キャラクタエージェントのパーソナリティやその場の雰囲気などを考慮しジェスチャの種類を選択することを可能とする．

### 3.2 Language Tagging

Language Tagging モジュールでは，入力された日本語の会話文にタグ付けを行う．語彙的・統語的情報が，ジェスチャを付与すべき場所を識別する上で効果的である [中野 04] ことから，入力文に対して形態素解析，および係り受け解析を行う．図 2 に，Language Tagging モジュールによる出力木を示す．まず，形態素解析によって得られた結果から構文木を生成する．

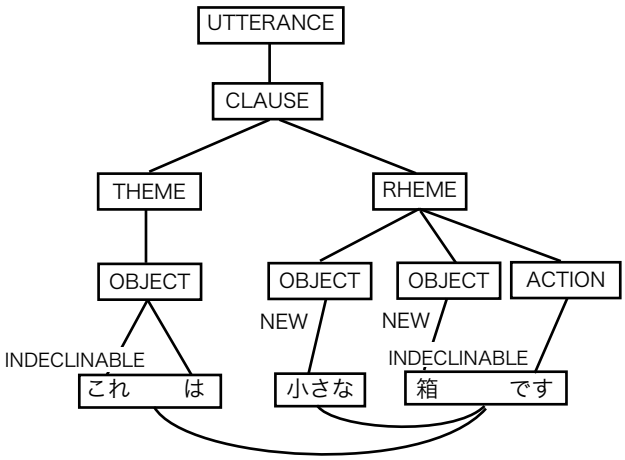


図 2: タグ付け結果

表 1: 付与するタグ一覧

| | |
|---|---|
| 形態素解析によって得られるタグ | THEME |
| | RHEME |
| | ACTION |
| | OBJECT |
| | INDECLINABLE |
| | NUM |
| | ADV |
| | POSTADV |
| | WH |
| | CONJUNCTION |
| 係り受け解析によって得られるタグ | PAR |
| 語の情報によって得られるタグ | NEW |

次に，格助詞 “は” が含まれる文節を主題（THEME），そうでない文節を題述（RHEME）に分類し，タグ付けを行う．さらに，動詞（ACTION）とそれ以外（OBJECT）に分類し，タグ付けを行う．そこへ、品詞情報から体言（INDECLINABLE）である形態素にタグ付けを行う．また，数詞（NUM），副詞（ADV），副詞に後続する文節（POSTADV），疑問詞（WH），接続詞（CONJUNCTION）に，それぞれタグ付けを行う．次に，係り受け解析によって得られた結果から，文節同士が並列句になっている場合は PAR のタグを，どの文節と並列であるのかの情報と共に付与する．

また，ここで，会話中において新出となる語に NEW のタグを付ける．

表 1 に，Language Tagging モジュールによって付与するタグの一覧を示す．

### 3.3 Timing Decision

Language Tagging モジュールでタグ付けされた情報と，あらかじめ定めたルールに従い，ジェスチャを付与する文節を決定する．ルールの作成には，[Cassell 01] および [中野 04] 中の実験結果を参考にしている．このとき，キャラクタエージェントの会話中のジェスチャは，視線と身振りに区別でき，これらは同時に起こり得るものとする．Timing Decision モジュールでは，まず話し手のジェスチャについて，入力された文章からタイミングを決定する．そして，話し手のジェスチャのタイミングを考慮し，聞き手のジェスチャのタイミングを決定する．

```
For each THEME node in the tree
   IF at the beginning of the UTTERANCE
   OR 70% of the time
      Look away from listener
For each RHEME node in the tree
   IF at the end of the UTTERANCE
   OR 73% of the time
      Look at listener
```

図 3: 話し手の視線を動かすルール

```
For each THEME node in the tree
   IF at the beginning of the UTTERANCE
   OR 80% of the time
      Look at speaker
For each RHEME node in the tree
   IF at the end of the UTTERANCE
   OR 47% of the time
      Look at the speaker
```

図 4: 聞き手の視線を動かすルール

図 3 に，話し手が視線を動かすタイミングを決定するためのルールの 1 つを示す．図 3 では，文章の主題では，それが文章の初め，もしくは 70%の確立で聞き手側を見ないという視線の動きを，また，文章の題述では，それが文章の最後，もしくは 73%の確立で聞き手の側を見るという視線の動きを決定する．

図 4 に，聞き手が視線を動かすタイミングを決定するためのルールの 1 つを示す．図 4 では，文章の主題では，それが文章の初め，もしくは 80%の確立で話し手側を見るという視線の動きを，また，文章の題述では，それが文章の最後，もしくは 47%の確立で話し手の側を見るという視線の動きを決定する．この他にも，与えられたタグに応じて，視線と身振りに関するルールを用意している．例えば，並列を表すタグ “PAR” が与えられた語や，接続詞を表すタグ “CONJUNCTION” が与えられた語では，一定の確立でジェスチャが発現するなどである．

### 3.4 Non-Verbal Behavior Selection

ここでは，Language Tagging モジュールによって決定された視線の動きのタイミングに従って，話し手，聞き手それぞれの視線を決定する．次に，タイミング毎の振る舞いを決定する．まず，全てのジェスチャをすべきタイミングにおいて，Beat のジェスチャを適応する．Beat ジェスチャは，全てのジェスチャの基本となる動きである．次に，Gesture Dictionary に登録されているジェスチャを Beat の代わりに置き換える．例えば，“小さい” という単語が登録されていた場合，それに対応する登録されたジェスチャが Beat ジェスチャの代わりとなる．また，全てのジェスチャは，その優先度を数値で持っており，複数のジェスチャが衝突した場合は，優先度の高いジェスチャをその文節のジェスチャとする．

図 5 に，話し手のジェスチャを表す木を示す．まず，視線の

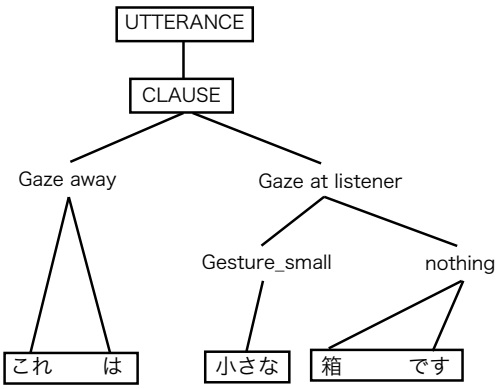


図 5: 話し手のジェスチャ

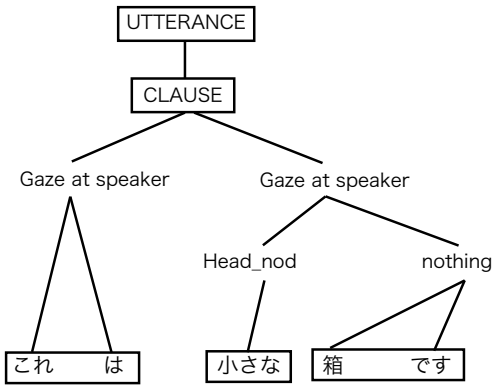


図 6: 聞き手のジェスチャ

動きとして図 3 で示したルールにより決定した動きが “Gaze away” と “Gaze at listener” である．また，語 “小さい” に呼応して，小さいを表すジェスチャ “Gesture$_{s}mall''$が適応されている．

図 6 に，聞き手のジェスチャを表す木を示す．まず，視線の動きとして図 4 で示されたルールにより決定した動きが “Gaze at speaker” である．

### 3.5 Transformation to simple script or MPML3D

Non-Verbal Behavior Selection モジュールで決定されたジェスチャから，キャラクタエージェントが実際にジェスチャを行うための命令となる，MPML3D を生成する．このとき，これまでに生成した聞き手と話し手のジェスチャの木を合成し，一つの時系列上に表現する．図 7 に，MPML3D 形式の出力例を示す．全てのジェスチャは，指定されたタイムラインに従って，並行に行われる．まず，始めの Action タグによって話し手の台詞が指示される．次に，話し手の視線の動きがタイムラインと共に指示される．MPML3D において，視線の動きは，頭の回転の動きで表現されているため，ここでは動く先の座標で指定されている．さらに，聞き手の視線の動きが指示される．また，話し手のジェスチャ，および聞き手のジェスチャが，それぞれタイムラインに従って指示される．また，図 8 に示す通り，単純な XML 形式での出力も同時に行われる．XML 形式の出力では，話し手のキャラクタエージェントの台詞，話し手の状態，およびジェスチャについて記述される．ただし，現在，本システムにおいて話し手の状態は考慮されていない．

```
<Task>
  <Parallel>
    <Action name=" kenspeak" >
    ken.speak(" これは小さな箱です" )
    </Action>                                   Speaker :
                                                Speak
    <Action startOn=" kenspeak[0].begin"
            stopOn=" kenspeak[9].end" >         Speaker :
    ken.turnHead(20, 0.2, 1, 0.2)               Gaze away
    </Action>
    <Action startOn=" kenspeak[10].begin"
             stopOn=" kenspeak[20].end" >       Speaker :
    ken.turnHead(0, 0.2, 5, 0.2)                Gaze at listener
    </Action>
    <Action startOn=" kenspeak[0].begin"
            stopOn=" kenspeak[20].end" >        Listener :
    yuki.turnHead(0, 0.2, 1, 0.2)               Gase at speaker
    </Action>
    <Action startOn=" kenspeak[10].begin" >     Speaker :
    ken.gesture(" showsmallvertical" )          Gesture small
    </Action>
    <Action startOn=" kenspeak[14].begin" >     Listener :
    yuki.gesture(" headnod" )                   Headnod
    </Action>
  </Parallel>
</Task>
```

図 7: MPML3D 形式の出力例

```
<utterance>
  <text> これは小さな箱です </text>
  <mood> neutral</mood>
  <gesture> showsmallvertical </gesture>
</utterance>
```

図 8: XML 形式の出力例

## 4. まとめ

入力された日本語会話文から，２体のキャラクタエージェントのためのジェスチャを自動生成するシステムを提案した．キャラクタエージェントは，話し手と聞き手の２つの役割を与えられた会話文に沿って担当する．本システムは，入力された日本語の会話文を解析，タグ付けし，２役それぞれのジェスチャを生成する．

本システムの特徴として，１つの入力から聞き手のジェスチャを自動的に生成するため，聞き手のジェスチャをアニメーション作成者が指定するコストが省ける．また，聞き手のキャラクタエージェントが，話し手のジェスチャや内容に呼応してジェスチャを行うため，ユーザは聞き手のキャラクタエージェントに共感しやすくなるものと考える．さらに，２体のキャラクタエージェントが交互に話をすることによって，ユーザの注意を引きつけやすいと考える．

本システムの構成において，ジェスチャのタイミングを決定するモジュールと，ジェスチャの種類を決定するモジュールをそれぞれ設けたことにより，キャラクタエージェントのパーソナリティや聞き手の雰囲気をジェスチャの種類決定に考慮することが可能となる．また，本システムはジェスチャの種類の決定に辞書を用いているが，これにより，常に特定の単語に呼応してジェスチャが起こるということを防ぐ．例えば，人間は “小さい” という単語が出たら常に “Gesture$_{s}mall$” をするとは考えにくいことが挙げられる．

## 5. 今後の課題

今後の課題として，提案したシステムによって生成されたジェスチャがユーザに与える印象を実験により評価することが挙げられる．評価方法としては，２体のキャラクタエージェントによってスピーチを行い，それを見たユーザに印象と，評価を調査するもの，および，あらかじめユーザにスピーチの原稿とジェスチャーのリストを渡し，どのタイミングにどのジェスチャを付けるかを記入させる．それを正解データとして，本システムの出力結果を比較する２つを考えている．

また，ジェスチャの種類決定において，学習機構を用い，より人間らしいジェスチャや，様々なジェスチャの組み合わせを取り入れることも挙げられる．さらには，ジェスチャのタイミングを決定するルールの拡充も課題となる．

## 参考文献


[Breitfuss 08] Breitfuss, W., Prendinger, H. and Ishizuka, M.: Automatic Generation of Gaze and Gestures for Dialogues between Embodied Conversational Agents, International Journal of Semantic Conputing, Vol. 2, No. 1, pp. 71-90 (2008)

[Cassell 01] Cassell, J., Vilhjalmsson, H. and Bickmore, T.: BEAT: The Behavior Expression Animation Toolkit, Proc. SIGGRAPH’ 01, pp. 477-486 (2001)

[Nakano 06] Nakano, A., Shioiri, K. and Hoshino, J.: Synthesizing Pose, Unconscious Movement, and Gesture for Mental Behavior Expression of Interactive Characters, ACM ACE2006 (2006)

[Nakano 03] Nakano, Y., Murayama, T., Kawahara, D., Kurohashi, S. and Nishida, T.: Embodied Conversational Agents for Presenting Intellectual Multimedia Contents, Proc. KES’ 03, pp. 1030-1036 (2003)

[Nischt 06] Nischt, M., Prendinger, H., Andre, E. and Ishizuka, M.: MPML3D: A Reactive Framework for the Multimodal Presentation Markup Language, Proc. IVA’ 06, pp. 218-229 (2006)

[中野 03] 中野有紀子：知識流通のためのメディア技術 ーインターフェースエージェントの利用ー, 社会技術研究論文集, Vol. 1, pp. 77-84 (2003)

[中野 04] 中野有紀子, 村山敏泰, 西田豊明：会話エージェントによる情報提供 ー非言語情報による重要概念の強調ー, 社会技術研究論文集, Vol. 2, pp. 159-166 (2004)