

## 理解しやすさとユーモアを考慮した Web コンテンツの対話番組変換

蓬 菜 博 哉<sup>†</sup> 灘 本 明 代<sup>††</sup> 田 中 克 己<sup>†</sup>

これまで我々は「容易に」「楽しく」「片手間に」Web コンテンツを取得するために、Web コンテンツを放送型コンテンツに変換してきた。本論文では、この放送型コンテンツをより親しみのあるコンテンツに変換することを目的とし、Web コンテンツを対話型番組に変換することを行う。本論文で提案する対話型番組とは漫才を用いた放送型コンテンツの事である。Web コンテンツを漫才に変換することにより、幅広い年齢のユーザ層にとって内容の理解を容易にするとともに、これまで Web コンテンツ取得時に必要であった能動的操作を軽減する事が可能となる。そこで、本論文ではまず漫才の形式化に取り組み、Web コンテンツから漫才形式に基づいた対話型番組に自動変換する機構を提案する。

### Transformation from Web contents into humorous user-friendly talk show type contents.

HIROYA HOURAI,<sup>†</sup> AKIYO NADAMOTO<sup>††</sup> and KATSUMI TANAKA<sup>†</sup>

We propose a framework for transforming Web contents into humorous conversation that aims at exploring easy-to-understand and enjoyable ways to access Web contents. In this paper, we especially focus on transforming multiple Web pages into a Japanese traditional humorous talk show (Manzai). By watching and listening to a transformed Manzai content instead of reading original Web contents, users of all ages can understand the contents easily. Furthermore, since the transformed contents can be viewed in a less-clicking-more-watching style, it can reduce the positive operations (scrolling, clicking). In this paper, first, we present our framework to analyze and formulate Manzai contents. Next, we propose a way to automatically transform multiple Web pages collected under a theme into Manzai-type contents.

#### 1. はじめに

日本にはユーモアのある独自のコンテンツとして古くは狂言に始まり、川柳、落語、漫才、コントと多くの種類のコンテンツがある。我々日本人は日ごろからこのようなユーモアのあるコンテンツを親しみ、楽しんでいる。特に、漫才はひとつのコンテンツで子供からお年寄りまで容易に理解し笑いを誘うものであり、最も理解しやすく親しみのあるコンテンツの一つであるといえる。しかしながら、狂言や落語に比べて明確な型がないことから計算機上で漫才を生成することは困難であると言われている。そこで我々は、ある程度

漫才を形式化することにより計算機上で漫才を自動生成することができるのではないかと考えた。

一方、これまで我々は、利用者に「読む」「スクロールする」「クリックする」等の能動的な操作を要求している Web コンテンツを、キャラクターアニメーションや音声合成を用いて「見る」「聞く」といった受動的な操作により「容易に」「片手間に」情報を取得できる受動的視聴方式<sup>6)</sup>を提案してきた。これまでの研究では、Web コンテンツをそのまま受動的視聴方式である放送型コンテンツに変換しているため、本当に親しみのあるコンテンツに変換しているとはいいがたく、ユーザは楽しみながらコンテンツを取得しているとは限らない。

そこで我々は、Web コンテンツから対話を生成することにより、より親しみのあるコンテンツに変換することを考えた。我々日本人にとって年齢に関係なく、最も親しみのあるコンテンツとはユーモアを考慮したコンテンツである。そこで本論文では、計算機上で漫才を自動生成する第一段階として、Web コンテンツをより親しみのあるコンテンツに変換することを目的

<sup>†</sup> 京都大学大学院 情報学専攻 社会情報学専攻

Division of Social Informatics,  
Graduate School of Informatics, Kyoto University

<sup>††</sup> 独立行政法人通信総合研究所 けいはんな情報通信融合研究センター メディアインタラクティブグループ

Interactive Communication Media Contents Group  
Keihanna Human Info-Communication Research Center  
Communications Research Laboratory

とし、Web コンテンツを漫才形式の対話型番組に変換することを行う。本論文ではこの漫才形式に基づいた対話型番組を漫才型コンテンツと呼ぶ。Web コンテンツを漫才型コンテンツに変換して視聴する利点は以下のとおりである。

- Web 閲覧の際の能動的な操作から開放され、「容易に」「片手間に」情報を取得できる。
- 難しかったり、固い内容のコンテンツでもわかりやすくかつ親しみやすく視聴する事ができる。日本の漫才台本作家の第一人者である秋田実氏は「洒落やおどけた言い廻しを混ぜて雑談風の対話にすると、硬い時事解説の内容でもびっくりするくらい柔らかくなって、風変わりな滑稽読物になった」と著作の中で述べている<sup>2)</sup>。
- 漫才は一つのテーマについて、それに関連する出来事や意見などをおもしろおかしく味付けしながら喋るため、特定のテーマに関連する複数の Web コンテンツを表層的に紹介することが可能である。
- 会話は、人間にとってなじみの深い情報収集手段であり、書き言葉によるテキスト的表現に比べ、情報受信者の理解を促進すると言われており<sup>5)</sup>、Web コンテンツを理解しやすいコンテンツに変換することが可能である。

本論文では、漫才型コンテンツのスタイルを決定することにより漫才をある程度形式化し、Web コンテンツを漫才形式のコンテンツへ変換する。

2 節では本研究に関連する事項の説明を行い、3 節では漫才を Web コンテンツからの自動変換で生成する事を念頭においた視点によって分析を行い実現可能な領域を模索する。4 節では3 節を踏まえた上で自動変換機構を提唱する。5 節では結論と今後の課題について述べる。

## 2. 基本事項と関連研究

### 2.1 Web ページの受動的視聴機構

これまで、我々は Web ページを放送型コンテンツに変換する「Web ページの受動的視聴機構」<sup>6)</sup>を提案してきた。これらの研究では S-XML という Web コンテンツと放送型コンテンツを同時に生成できるマークアップ言語を提唱した。自動変換機構では、HTML において、画像とテキストとの同期化領域を発見すると共に演出を付加して S-XML を生成する。そして S-XML から TVML<sup>8)9)</sup>に代表される放送型コンテンツ記述言語に変換する事によって受動型視聴を実現していた。本研究では S-XML のような中間言語を用いずに直接放送型コンテンツに変換する。これは人手で

中間言語の編集などを行う事を念頭に置かずに、自動変換にこだわったからである。また、これまでの研究では、Web コンテンツの内容そのものには手を加えずに放送型コンテンツに変換していたが、本研究においては、ユーモアを考慮した上で Web コンテンツをわかりやすく親しみのある放送型コンテンツに変換する事を目指している。

### 2.2 漫才の歴史

現在知られている漫才とは、2 人(ないし 3 人)の芸人が、互いに滑稽な掛け合いを演ずるもので、今日寄席演芸として人気を高め、大阪では落語をしのいで主流の地位を占めるに至っている。現在の漫才はいわゆる「しゃべくり漫才」と言われるもので、昭和初期にエンタツ・アチャコによって確立されたものである。それ以前のマンザイは万歳と呼ばれる歌や舞いが基本となる物であった。太夫と才蔵と呼ばれる役割の 2 人が 1 組となり、太夫が扇をかざし、いろいろとめでたい寿(ことほ)ぎの詞(ことば)を言い立て、才蔵が小鼓を打ち囃して合の手を入れるという、掛け合いで進行するものである。しかし、昭和 5 年にコンビを組んだ横山エンタツと花菱アチャコは従来の万歳のユニフォームであった和服と鼓や張り扇をかなぐり捨てて、背広姿で現れ、歌も歌わずに「しゃべくり」だけで押し通した。当時としてはかなり冒険であったはずだが、それだけに観客にとっては新鮮であり、語り合う内容も世相のスケッチやスポーツの話題など、日常生活に題材を求め、自然な会話の妙味にその芸を見出す、今日の漫才の先駆けをなしたのである。なお、現在の漫才に見られる「ボケ」と「ツッコミ」という役割分担の原型は、当時の万歳の「才蔵」と「太夫」に見る事ができる。

## 3. 漫才の形式化

Web コンテンツを漫才型コンテンツに自動変換するために、漫才をある程度形式化し漫才型コンテンツのスタイルを決定する。図 1 に漫才型コンテンツ台本の構成を、以下に我々の行った漫才の形式化について述べる。

### 3.1 前処理

漫才とは、決められた形式に基づいてキャラクターに台本を演じさせる事により実現するものであり、漫才の形式と台本とキャラクターは漫才を形作る三大要素である。台本とキャラクターは独立したものは無く、キャラクターにあった台本を生成しなければならない。漫才台本作家秋田実氏は「ヘドが出る位漫才をみよ、コンビの特徴が判らんと漫才は書けん」<sup>4)</sup>を持論

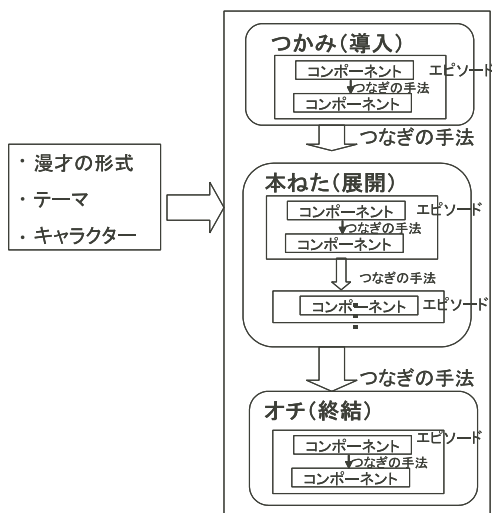


図 1 漫才型コンテンツの構成

Fig. 1 Composition of scenario for Manzai-like content

としていたそうである．そのため，台本を生成する前に前処理として，漫才の形式やテーマ，キャラクターを決定する事が重要である．

#### 漫才の形式

本研究では，現代一般的となっている「しゃべくり漫才」を用いる．「しゃべくり漫才」の形式としては「ぼやき漫才」「時事漫才」等がある．「ぼやき漫才」とは，身近な事，世の中のことなどを大げさに述べぼやきながら面白おかしく伝える漫才である．「時事漫才」とは，いくつかの時事ニュースを面白おかしく伝える漫才である．

#### テーマ

漫才にはテーマがあり，テーマに則した話題を話すのが普通である．本論文における漫才型コンテンツでは，複数のニュースを対象としているため，「明るいニュース」や「悲しいニュース」などの印象語をテーマとし，そのテーマにそったニュースをいくつか選びだして漫才型コンテンツの生成に用いる．

#### キャラクター

先に述べたようにキャラクターは漫才の構成において非常に重要な要素である．キャラクターが変われば台本も変わってくるため，まずこれを決定しなければいけない．一般の漫才は，2ないし3人で演じられる．2人で行う漫才はコンビと呼ばれ，「夫婦コンビ」「兄弟コンビ」「姉妹コンビ」「友達コンビ」などがある．通常は「夫婦コンビ」の場合女性が強く，「兄弟・姉妹コンビ」では弟や妹が強いというように，一般常識から考えると逆の性質の設定となっており，それが意外性をまねいて笑いを誘う基礎となっていると考えられ

る．このように，各々のコンビは独自の性質を持ち，それが漫才のコンテンツを生み出す重要な役割を担っている．今回はPC上で架空のキャラクターを用いるので実際にはありえないような「子供とロボット」などといったコンビを選択する事も可能である．

### 3.2 台本

漫才を自動生成することの大部分は，台本を自動生成することに要約される．以下，台本の生成に必要な技術について述べる．

#### 基礎技術

漫才の最大の特徴は「ボケ」と「ツッコミ」があることである．ボケとツッコミが一組となって一つの笑いを構成している．漫才台本には多様なパターンの笑いが含まれているが，それらの笑いはボケとツッコミを基盤として構成されているのである．

#### ● ボケとツッコミ

ボケは非常識的な発言や行動であり，ツッコミはそのおかしなところに対する指摘や修正である．漫才コンビは，ボケを専門的に行う「ボケ役」，ツッコミを専門的に行う「ツッコミ役」という風に役割分担されているのが一般的である．ツッコミが漫才の進行役になっているところが多い．

#### ● 間

漫才特有の「間」というものがあり，この間があるから笑える場合もある．しかしながら，適切な間を自動で生成するのは難しいため，今後の課題である．

#### 台本の構成

漫才は一つのテーマに基づいた複数の話題から構成されている．本論文ではこの漫才を構成する各々の話題をエピソードと呼び，それを構成する会話の意味的な一つのまとまりをサブエピソードと呼ぶ．

漫才の構成手法の定義も行った．構成手法に基づいて台本を分類すると，台本はつかみ(導入)，本ねた(展開)，オチ(終結)の3つの部分に分かれている．構成手法ではそれらを構成する要素である，エピソード間のつなぎの手法，エピソードを構成するサブエピソード間のつなぎの手法それら全てを規定する．

#### ● エピソード

エピソードとは一つの話であり，単数または複数のサブエピソードからなる．また，漫才は複数のエピソードからなる．次節で提案するシステムでは，一つのエピソードを一つのWebページのタイトル又は要約から生成した会話とする．

#### ● サブエピソード

サブエピソードとは，漫才的対話を構成する要素

であり、ボケとツッコミ、質問と応答といった切り離すことのできない意味的つながりをもった最小単位であると定義する。そのなかで特に笑いを考慮したサブエピソードは笑いのサブエピソードと呼ぶ事とする。

我々は笑いのサブエピソードを、大きく二種類に分類することによりその性質を明確化する。それは「ありえる笑い」と「ありえない笑い」である<sup>2)</sup>。

－ ありえる笑い

ありえる笑いとは、漫才の視聴者が「あるある」と納得しながら思わず笑ってしまう笑いである。「まね」や「パロディー」などがありえる笑いである。また、「読み間違え」「勘違い」などの間違いを誘うものもありえる笑いと言える。

－ ありえない笑い

ありえない笑いとは、常識ではありえない事を言って笑いを誘うものである。「誇張」「うそ」「ダジャレ」などがある。

● つなぎの手法

サブエピソードをつなぐ構成手法としては、「天井」「かぶせる」「あいづち」「三段オチ」などがある。天井とは、すでにたボケを2度3度ということであり、かぶせるとはボケの上ののっかりさらにボケを言うことである。のせは一つのボケまたはツッコミに対して、ボケたり突っ込んだりすることをいい、三段オチはフリフリオチの形態をとることである。エピソードをつなぐ構成手法としては、「話題転換」「あいづち」がある。不自然な話題転換を行えば笑いを誘う事が可能である。これらは具体的には細かい規則として記述する事が可能である。

● 漫才の流れ

漫才は複数のエピソードから構成されるが、全体としてはつかみ、本ねた、オチの大きな流れがある。つかみではテーマをはっきりとさせ、最初の笑いをとる。本ねたはその漫才のコンテンツの核であり、いろいろな話題をふり複数の笑いをとる部分である。また、最後のオチへのつなぎの意味も持つ。オチは最後の笑いであり、漫才においてはこのオチが重要視されている場合がある。

#### 4. 提案するシステムの概要

本論文で提案するシステムはニュースサイトのニュース記事一覧を入力とし、それらの話題の中から特定の

テーマに沿った話題をいくつか選び、それらの概要を紹介する漫才形式の対話番組に変換してユーザに提示するシステムである。システムの大まかな流れは以下のようなになる。

(1) ユーザが以下の事項を決定する。

● テーマの選択

ユーザはどのテーマの漫才を見たいかを選択する。実際にはユーザは「明るい」「物騒」といった印象語の中から選択する。

● キャラクターの選択

キャラクターの選択とは、どんなキャラクターに漫才を演じさせるかである。プロトタイプシステムでは、スタイルについては「時事漫才」のスタイルのみを採用するためユーザによる選択は必要としない。

(2) システムは、ユーザが選択したテーマに則したニュース記事をリストアップする。

(3) リストアップされたニュース記事数件を元に、それらを表層的に紹介する漫才形式の台本を生成する。

(4) 生成した台本を TVML 形式に変換する。

(5) TVML スクリプトをプレイヤーで再生し、ユーザに漫才型コンテンツとして提示する。

(6) 漫才型コンテンツを提示しているとき、バックグラウンドでは現在演じられている話題の元になった Web ページを Web ブラウザに提示する。これにより、ユーザはニュース記事の表層部分だけでなく、詳細を知りたければそちらを閲覧する事が可能である。

これまでの我々は、1つの Web ページを放送型コンテンツに変える Web コンテンツの受動的視聴方式を提案してきた。それに対し、本論文では、複数の Web ページの内容の概略を放送型コンテンツに変換する方式を提案する。複数ページの内容を一つの放送型コンテンツとして提示する場合、各々のページの内容すべてを放送型コンテンツに変換したのでは、放送型コンテンツが長くなり、退屈なコンテンツになる可能性がある。そこで、我々は、各々のページの概要をリレー形式で一つの放送型コンテンツに変換することを考えた。また、漫才は複数の話題を統合して一つのコンテンツにすることに適していると考え、本論文では、このリレー方式で提示する手法の一つとして漫才型コンテンツに変換することを行う。

##### 4.1 キャラクターの選択

今回のシステムでは「助手と博士」「子供とロボット」の以下の二つのキャラクターを選択できるように

した。

#### 助手と博士

- 助手 : 基本的にはコンテンツの内容を真面目に紹介する。分からない事に関して博士に意見を求めたり、博士のボケに対してツッコミを入れたりする
- 博士 : 多岐の話題に精通した専門家である。はずであるが、変なコメントなどを挟んだりダジャレを言ったりする、ボケ役の役割も持つ。

#### 子供とロボット

- 子供 : 基本的にコンテンツの内容をロボットに対して教え込もうとするのだが、ロボットのバカさに腹を立てて辛辣なツッコミを入れたりする。
- ロボット : 基本的にあまり賢くなく、言われた事を勘違いしてボケたりトンチンカンな事を言ったりする。

#### 4.2 テーマの選択

本システムではまず、ユーザが自分の見たい漫才のテーマを選択する。ここでテーマとして選択可能なものは以下の印象語とする。

- 明るい
- 物騒
- 情けない
- 悲しい

以下、これらの印象語を  $I_j, j \in (1, \dots, 4)$  と示す。ユーザがテーマとして上記の印象語を選択した後、システムはニュースサイトにおけるニュース記事の見出し一覧からその印象語に分類されるニュース記事を抽出する。ユーザの指定したテーマである印象語からそれに関連するニュース記事を抽出する手法としては、各印象語とニュース記事の見出し文に含まれる単語との共起関係を用いた特徴ベクトルにより決定する。以下にニュース記事の特徴ベクトル算出の具体的なプロセスを示す。

- (1) ニュース記事の見出し文を形態素解析する(形態素解析ツールとしては茶筌<sup>11)</sup>を用いる)。そこで得られた単語列から、記号、非自立語、固有名詞を取り除き、接頭語は直後の名詞に付加する。そうして得られた単語列を  $W_i, i \in (1, \dots, n)$  とする。
- (2)  $W_i$  と各印象語  $I_j, j \in (1, \dots, 4)$  の共起度  $c_{ij}$  を求める。 $U(X)$  は単語集合  $X$  を含む Web ページの数であり、検索エンジン(Google Web API<sup>12)</sup>)を用いて求める。

$$c_{ij} = U(W_i \cap I_j) / U(W_i \cup I_j) \quad (1)$$

これを用いて、単語  $W_i$  の特徴ベクトル  $V_{W_i}$  を

$$V_{W_i} = \{c_{i1}, c_{i2}, c_{i3}, c_{i4}\} \quad (2)$$

と定義する。

- (3) 全ての  $W_i$  について  $V_{W_i}$  を求め合計する。ここで求められた合成ベクトルがニュース記事の特徴ベクトルである。

例えば「離島にこそブロードバンドを」というニュース見出し内に現れている「離島」「ブロードバンド」という単語と「明るい、物騒、情けない、悲しい」に対する共起度を求め、各単語の特徴ベクトルとする。「離島」の特徴ベクトルは(明るい=0.2, 物騒=0.08, 情けない=0.1, 悲しい=0.1)、「ブロードバンド」は(明るい=0.5, 物騒=0.1, 情けない=0.1, 悲しい=0.08)となる。したがって、このニュース記事の特徴ベクトルは(明るい=0.7, 物騒=0.18, 情けない=0.2, 悲しい=0.18)となる。各ニュース記事の特徴ベクトルを求めた後、「明るい話題」を選択するには、特徴ベクトルの「明るい成分」が閾値  $\alpha$  以上のニュース記事を選択すればよい。

このようにして各ニュース記事特徴ベクトルを算出しておき、各印象語に対する値の大小によってニュース記事を選択する。

#### 4.3 台本の生成

漫才の形式、キャラクター、テーマが決定したならば後は台本を生成すれば漫才を記述する事ができる。台本は「つかみ」「本ネタ」「オチ」の部分に分ける事ができ、それぞれにおいて異なる生成規則によって生成するものとする。台本の生成規則は、前節で述べた、サブエピソードを構成手法に乗っ取ってつなぎ合わせていくという方法を取る。しかし計算機で自動生成する事が困難であるサブエピソード、構成手法は用いず、いくつかの限られた物を用いた。

##### 4.3.1 つかみ

つかみの部分の構成はサブエピソードとしては「状況、話題の説明を行う定型会話文」と「勘違い(印象の取り違い)」をつなぎ合わせる事によって構成する。「状況、話題の説明を行う定型会話文」では、話者の立場から今回の漫才のテーマについての説明を行う。「勘違い(印象の取り違い)」では、最初の話題について軽く触れて、本ネタ部分で具体的な個々の話題(各ニュース記事に対応)に入っていくための導入部分としての役割を持つ。以下にキャラクターに「子供とロボット」、テーマに「情けない」を選択した場合の例を示す。

子供「はい、ども！ちまたで噂の天才少年やすし君や。隣におるんは天才の俺様が発明した割にはポンコ

ツなロボットの・・」

ロボ「マーティーです．よろしく．」

子供「ところでマーティ，最近情けない話が多いよな」

ロボ「情けない？どんな話でしょう？」

子供「ロボットでも新聞くらい読んどきや！こんな話知ってるか？ 痴漢で逮捕の巡査辞職」

ロボ「それは勇ましい話ですね」

子供「なんでやねん！勇ましくないわ！まあちょっと落ち着いて聞けや」

ロボ「はい」

以下，具体的なニュース記事内容について話を展開する本ネタ部分に入る．

（この例では，ニュース記事の特徴ベクトルの“情けない”成分が高いニュース記事として“痴漢で～”のニュース記事があった場合である．）

#### 4.3.2 本ネタ

本ネタ部分では，個々のニュース記事を題材として概要を紹介する．しかし，ニュース記事は平叙文で書かれているのに対して，漫才は会話文であるので平叙文から対話文に変換するテクニックが必要となってくる．そこで漫才形式対話生成の為に，前節で挙げたサブエピソードを用いて対話文を組み立てる．また本ネタ部分は他の「つかみ」「オチ」の部分と違って複数のエピソードで構成される為，エピソード間，サブエピソード間のつなぎの手法が重要となってくる．

##### 4.3.2.1 採用したサブエピソード

今回実現したサブエピソードとしては，あいづち，質問・応答生成<sup>7)</sup>，文の分割，単語の読み間違い，意味の取り違い，誇張がある（初めの3つは対話文の生成のテクニックであるが，後半の3つは笑いを考慮したいいわゆる“ボケとツッコミ”を基本とするテクニックである．）以下に各サブエピソードの詳細を述べる．

あいづち 平叙文を一文読み上げた後に「なるほど」「へー」などといったランダムな相づちを挿入したり，前文の主語を繰り返したりする事（「へー，松井がねえ」等）によって二人の会話にする．

質問・応答生成 1 文の主述関係に注目し，平叙文の読み上げの後に主部を聞き返し，それに対して応答する事で二人の会話にする．

（例）“松井がホームランを打った”という平叙文に対して以下のような会話文を生成する．

A「松井がホームランを打ったらしいで」

B「えっ？誰が？」

A「松井やがな．ちゃんと聞いたかとあかんで」

質問・応答生成 2 平叙文中に現れた珍しい単語や固有名詞などに対して質問を行い，それに対する応答を検索エンジンを用いて発見することができれば，それを用いて会話文を生成する．

（例）“ドリカムの元メンバーだった西川被告に有罪判決が下った”という平叙文に対して以下のような会話文を生成する．

A「ドリカムの元メンバーだった西川被告に有罪判決が下ったらしいで．」

B「ドリカムって何？」

A「ドリームズ・カム・トゥルーのことやがな．そんな事も知らんのかい」 Web 検索によって得られた結果

文の分割 文の主述関係に注目し，主部と述部で二人の会話にする．

（例）“小泉首相がロシア大統領と会談を行った”という平叙文に対して以下のような会話文を生成する．

A「小泉首相がなあ」

B「どうしたん？」

A「ロシア大統領と会談を行ったらしいで」

なお主部と述部の判定は，以下のように行う．

- 対象となる文から，格助詞の「が」と係助詞の「は」を抜き出す．
- 「が」「は」の直前の単語が名詞である場合は，その名詞を含む文節を抜き出しそれを主語とする．文節の解析には日本語係り受け解析器 CaboCha<sup>13)</sup>を用いる．

単語の読み間違い 平叙文中に現れた重要な一般名詞に対して，それを取替えて読み間違い，それを聞いた話者が“ツッコミ”を行う事により会話文を生成する．重要な一般名詞の選択はニュース記事内の単語の出現頻度に重み付けを行う事によって行う．

（例）“台風6は時速15Kmで北北西の方向に進んでいる．”という平叙文に対して以下のような会話文を生成する．

A「大福6号は時速15kmで北北西の方向に進んでいるらしいで」

B「大福？台風ちゃうの？」

A「あ，ホンマや．ごめん」

読み間違いの生成は以下のように行う．

- 読み間違いの単語の通常の読みを茶筌<sup>11)</sup>を用いて調べる（例：「台風」「たいふう」）
- 日本語辞書をもちいてその読みを調べ，その単語の前後の単語でひらがなでも文字数が同

じである単語を取り出す（例：「大福：だいふく」）

誇張 平叙文内に数値を含む表現が合った場合、その数値を大げさな数値に改変し、それを聞いた話者が“ ツッコミ ”を行う事によって会話を生成する。また同様に大きな数値を極端に小さくする場合もある。

（例）“ 地元消防団や一般市民の協力も得て約 3000 人態勢で捜索している ”という平叙文に対して以下のような会話を生成する

A 「地元消防団や一般市民の協力も得て約 3000 万人態勢で捜索しているらしいわ。」

B 「3000 万って多すぎるやろ！」

A 「あ、3000 人やったわ。」

印象の取り違い これは、“ つかみ ”の構成にも用いられているパターンである。これはあらかじめニュース記事の印象が分かっているため、それと逆の感想を述べ、それに対してツッコミを入れる事で会話を生成する。このサブエピソードは他のサブエピソードと違い、一つのエピソードの最後の部分に使われ、次のサブエピソードとのつなぎの役割を果たしているため、それに続くつなぎの生成は行わない（例）ベッカムの来日のニュース記事に対して、特徴ベクトルの「明るい」成分が大きかった場合」

A 「ベッカムは 21 日まで日本に滞在するそうですわ。」（ニュース記事紹介の最終部分）

B 「いやー、悲しい話でしたね。」

A 「どこがやねん！ちゃんと聞いてたか？」

#### 4.3.2.2 採用したつなぎの手法

エピソード間のつなぎ 本ネタ部分では、ユーザの選択したテーマに沿った複数のエピソードを紹介するので、エピソード間のつなぎが重要である。エピソード紹介の順序としては、序盤に視聴者をつかんで、最後に一番盛り上がるような順番で紹介する事が望ましいと考えられる。今回のシステムでは、紹介するエピソードの順番は各ニュース記事の特徴ベクトルの選択要素（例えば「明るい」）の値が、中 小 大の順番になるようにした（以下に例を示す）。他にもテーマにそっていないエピソードを故意に紹介する事によって笑いを取る方法もあるが、ユーザの意図の反映に重点を置いたため今回は採用しなかった。

（最初のエピソード：明るさ = 中）

「明るい話題で結構でしたな。」

「次も明るい話題ですよ。」

（次のエピソード：明るさ = 小）

「今のは普通の話やったな。」

「いやいや、次のは最高に明るい話題ですから。」

（最後のエピソード：明るさ = 大）

（オチ部分に続く）

サブエピソード間のつなぎ エピソードを構成する各サブエピソードは、基本的には対象ニュース記事の平叙文の順番に従ってつなぎ合わせてく。どのサブエピソードをどの文脈で使うかは、キャラクター設定による。ただし、以下のような最低限のルールを設定した。

- 同一エピソード内では最低一度は笑いの要素を考慮したサブエピソードを用いる（後半の 3 つのうちのどれか）
- 「印象の取り違い」のサブエピソードはエピソードの最後に位置する。
- 「単語の読み間違い」のサブエピソードに限っては同一の平叙文に対して二回以上繰り返してもよい。これは「天井」とよばれる同じようなボケを繰り返すつなぎの手法である（ただし読み間違える単語は違う単語にする。）

#### 4.3.3 オチ

オチ部分では、これまでエピソードとして用いたニュース記事に関するダジャレを言ったり、テーマに応じてあらかじめ用意しておいた定型文を述べて終わりにする。ダジャレはあらかじめ用意しておいた辞書からパターンマッチにより探し出す事により生成する。以下示すのは、「子供とロボット」のキャラクターで「明るい」話題として“ 漁をするサルに人気集中 ”等の話題などを紹介した後の生成例である。

子供「どや、マーティ？明るい話やったやろ？僕は“ 漁をするサルに人気集中 ”の話が好きやなあ。」

ロボ「サルがりハーサル。」

子供「ありゃ、またバグかいな！家帰ってデバッグし直しやな、こりゃ。」

## 5. 結論と今後の課題

本論文では、Web コンテンツを親しみやすく手軽



図 2 システムのスクリーンショット  
Fig.2 Screenshot of the system

に視聴する為の手法として、Web コンテンツを漫才形式のコンテンツに変換する事を提案した。漫才メタファーを用いる事により、複数の話題をとりこんだ一つの親しみやすいコンテンツを生成する事が可能となり、ザッピング視聴のニーズに応える事が可能となった。ただ、漫才の自動生成という視点でとらえると、計算機による自動生成には多くの技術的課題を残す事となった。以下にそれらの技術的課題を示す。

**インスタンスエーション** 一般的な話題から身近な話題に変換する手法である。話者がまるで自分の周囲の出来事のように喋ることにより、ユーザーの興味を引き立てることができる。漫才などでよく取り入れられている手法である。キャラクターの作りこみと膨大な知識データベースを用意すれば将来的には可能である。

**パロディ、嘘などのサブエピソード** これらも自動生成は非常に難しく、膨大な知識データベースを必要とする。

**質問・応答の多様化** 現時点では、“What(Who) is X?”のタイプの質問・応答しか自動生成できていない。これも膨大な知識データベースがあれば可能になる。

**謝辞** 本研究の一部は、平成 15 年度科研費特定領域研究 (2) 「Web の意味構造に基づく新しい Web 検索サービス方式に関する研究」(課題番号:15017249, 代表:田中克己), 平成 15 年度科研費基盤研究 (A)(2) 「モバイル環境におけるコンテンツのマルチモーダル検索・呈示と放送コンテンツ生成」(課題番号:14208036, 代表:田中克己), 平成 15 年度基盤技術研究促進事業 (民間基盤技術研究支援制度) 「クロスメディアコンテンツ基盤技術の研究開発」によるものです。ここに記して謝意を表すものとします。

## 参 考 文 献

- 1) 相羽秋夫, 『上方漫才入門』, 弘文出版, 1995 年
- 2) 秋田実, 『私は漫才作者』, 文芸春秋, 1978 年
- 3) 秋田実, 『名作漫才選集 1, 2』, 日本実業出版社, 1973 年
- 4) 足立克己, 『いいたい放題上方漫才史』, 東方出版, 1994 年
- 5) 西田豊明, 『インタラクションの理解とデザイン』, 岩波書店, 2000 年
- 6) 灘本 明代, 服部 多栄子, 近藤 宏行, 沢中 郁夫, 田中 克己, “ Web コンテンツの受動的視聴のための自動変換とスクリプト作成マークアップ言語 ”, 情報処理学会論文誌:データベース (TOD8) Vol.42No.SIG1, pp.103-116, 2001 年 .
- 7) 蓬萊博哉, 灘本明代, 田中克己, “ トークショーメタファーによる複数 Web ページの受動的視聴 ”, 情報処理学会研究報告 2002-DBS-129 pp.155-162, 2003 年 1 月
- 8) NHK 放送技術研究所 TVML ホームページ  
<http://www.strl.nhk.or.jp/TVML/indexj.html>
- 9) 林 正樹: 番組記述言語 TVML を使った情報の番組化, 情報処理学会 DBS 研究会技術報告, Vol.2000, No.10, 2000-DBS-120-13, pp.91-98, 2000 年 1 月 .
- 10) 久保田秀和, 山下耕二, 福原知宏, 西田豊明 POC Caster: インターネットコミュニティのための会話表現を用いた情報提供エージェント 2002 年 人工知能学会論文誌 17 巻 3 号 SP-H
- 11) 奈良先端科学技術大学松本研究室 茶筌ホームページ  
<http://chasen.aist-nara.ac.jp/index.html.ja>
- 12) Google Web API ホームページ  
<http://www.google.com/apis/>
- 13) 奈良先端科学技術大学松本研究室 南瓜ホームページ  
[http://cl.aist-nara.ac.jp/taku-ku /software/cabocha/](http://cl.aist-nara.ac.jp/taku-ku/software/cabocha/)