

## 統計的機械翻訳の手法を用いた 音声情報案内システムのための応答文生成手法の検討

西村 一馬<sup>†1</sup> 川波 弘道<sup>†1</sup>  
猿渡 洋<sup>†1</sup> 鹿野 清宏<sup>†1</sup>

実環境下で運用されている音声情報案内システムの応答文生成手法として、質問応答データベースを参照して応答文を決定する、用例ベース応答生成の研究が行われている。しかし、これは入力質問文に一番近い質問文を選択し、その質問文に対応付けられている応答文を出力する手法であるため、データベースにないような想定外の質問が入力された場合には適切な応答を返すことができない。また、応答文はデータベース中の事前に決められた定型文を出力するため、入力質問文の表現に応じた柔軟な応答文を出力することができない。これらの問題点を解消するため、統計的機械翻訳の手法を応用した応答文生成を提案する。本稿では、統計的機械翻訳を用いた応答文生成の実現可能性と課題点を、たけまるくんのユーザ発話をを用いて実験調査した。

### Answer Sentence Generation Method Using Statistical Machine Translation in a Speech-Oriented Guidance System

KAZUMA NISHIMURA,<sup>†1</sup> HIROMICHI KAWANAMI,<sup>†1</sup>  
HIROSHI SARUWATARI<sup>†1</sup> and KIYOHITO SHIKANO<sup>†1</sup>

An example-based question answering (QA) is a robust and practical approach for a real-environment information guidance system. However, it cannot appropriately respond to unexpected user's utterances if a similar example of a question-answer pair does not exist in the QA database; in addition, the answer sentences cannot reflect differences in nuance, because the set of answer sentences are fixed beforehand. To deal with these problems, we propose a new method, which introduces statistical machine translation training to answer sentence generation. In the proposed method, we treat questions and answer sentences as different languages. In this paper, we investigate a feasibility of translation from question into answer using real user utterances for *Takemaru-kun*.

### 1. 背景

近年、音声認識を利用したインターフェースの研究開発が行われている。ディクテーションや音声検索やカーナビや音声対話システムなどが挙げられるが、中でも音声対話システムは人と機械が音声による対話をしながら、情報案内など何らかのタスクを達成するシステムであり、幅広く需要が期待されている。

筆者らは音声対話システムとして、音声情報案内システム「たけまるくん」を開発・運用している<sup>1)</sup>。対話システムの設計の際にはタスクを絞って目的に応じたシステムを構築するのが一般的である。しかし、「たけまるくん」は用例ベース方式で質問応答を行うシステムとなっており、ユーザの発話を反映してデータベースを拡張することも容易であり、ユーザの質問に対して幅広いタスクの応答を返すことができる。応答文の生成は質問応答データベース (QADB) を参照することで行われる。入力質問文と一番類似した質問例をデータベース中から選択し、その質問例に対応付けられた応答文を出力する。質問応答データベースは対話制御に関する特別な知識がなくても簡単に管理することができる設計になっている。

しかし、この方式のシステムにも問題点が存在する。まず第一に、もし QADB 中に類似した質問例がないような想定外の質問文が入力された場合には、適切な応答を返すことができない。ユーザ発話を集計したところによると、発話全体の約 5% は QADB に類似の文が見付からない想定外の発話であった。この問題を扱うために、サポートベクトルマシン (SVM) と Bag-of-Words (BOW) を用いたタスク外発話の検出の研究が行われている<sup>2)</sup>。この研究においては、検出されたタスク外発話はクエリとしてインターネット検索にかけるシステム構成となっている。

第二の問題点は、応答文は事前に QADB に登録された定型文であるので、それぞれの質問発話の表現ニュアンスを反映させた応答文を出力することができないという点である。例えば、「一番近いバス停はどこですか?」という質問と「バスにはどこで乗れるの?」という質問の応答文として、「バス停は、玄関を出て、道路を渡ってください。」という同じ文で答える。これを、例えば「バス停は～を渡ったところです。」と「バスには～で乗れるよ。」など、質問文に応じた表現で柔軟に回答できるようにすると、ユーザからの親しみやすさと

<sup>†1</sup> 奈良先端科学技術大学院大学

Nara Institute of Science and Technology



図1 音声情報案内システム「たけまるくん」  
Fig. 1 Speech-oriented guidance system *Takemaru-kun*.

いう観点から望ましい。

本稿では、上記のような問題点を解決するために統計的機械翻訳 (SMT) の手法を応用して応答文生成を行うことを提案する。提案手法においては、質問文と応答文とを別の言語として扱う。すなわち、システムへの質問発話を SMT モデルによって適切な応答文に翻訳すると考える。統計的機械翻訳の手法を質問応答に導入することで、質問文のニュアンスの違いを反映させた応答文生成が可能になることが期待できる。また、想定外の発話に対しても何らかの応答を返すことができる。

以下、第2節では従来の「たけまるくん」システムの概要説明を行う。第3節では提案手法に用いられている統計的機械翻訳について説明する。第4節では統計的機械翻訳を用いた応答文生成について述べる。本稿では、統計的機械翻訳を用いた応答文生成の実現可能性と課題点を、たけまるくんのユーザ発話を用いて実験調査した。その結果を第5節で報告し、第6節で今後の展望を述べる。

## 2. 音声情報案内システム「たけまるくん」

### 2.1 「たけまるくん」システムの概要

筆者らが開発・運用している音声情報案内システム「たけまるくん」は生駒市の北コミュ

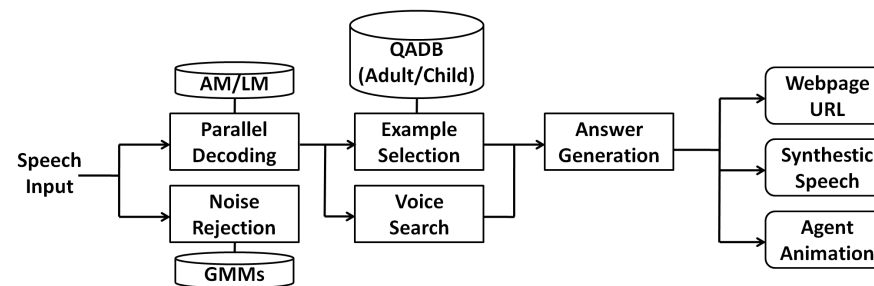


図2 「たけまるくん」のシステム構成図  
Fig. 2 Processing flow of *Takemaru-kun*.

ニティセンターに設置されている<sup>1)</sup>(図1)。2002年11月から運用を開始し、現在まで約8年強の期間に渡って運用を継続している。「たけまるくん」が応答できる内容としては、センターの施設案内、センターのサービス案内、周辺の観光案内などの他、ニュースや天気予報の情報などがある。また、情報案内を期待しない挨拶や、「たけまるくん」自身のプロフィールに関する質問など雑談にも答えられるなど、幅広いタスクに対応している。

システムの構成図を図2に示す。システムはGMMによって雑音と音声とを区別することができ、雑音を応答処理にかけることなく棄却できる。音声は大人話者のモデル、子供話者のモデルをそれぞれ用いて並列デコーディングされ、大人/子供の判別がされて音声認識結果が出力される。その後、音声認識結果を用いて応答文が生成され、ウェブブラウザと合成音声とエージェントアニメを用いた応答提示が行われる。ユーザとシステムは基本的に一問一答で対話を行う。なお、これまで「たけまるくん」は長期にわたって運用されているが、その間に集められたユーザ音声のうち最初の2年5か月間の全発話については人手により雑音性・年齢層・書き起こし・正解応答などの発話ラベルが付与されて音声データベースとして整備されている。

### 2.2 用例ベース応答文生成

「たけまるくん」の応答文生成について説明する。「たけまるくん」は幅広いタスクに対して応答が可能であるが、それを可能としているのはQADBを用いた用例ベースの応答文生成方式である(図3)。「たけまるくん」は質問例と応答文のペア(QAペア)を集めた質問応答データベース(QADB)をシステム内に保持している。ユーザ発話が音声認識されると、その認識結果  $I$  の  $N$ -best を用いた QADB 内の質問例  $E$  との類似度計算が行われ、類

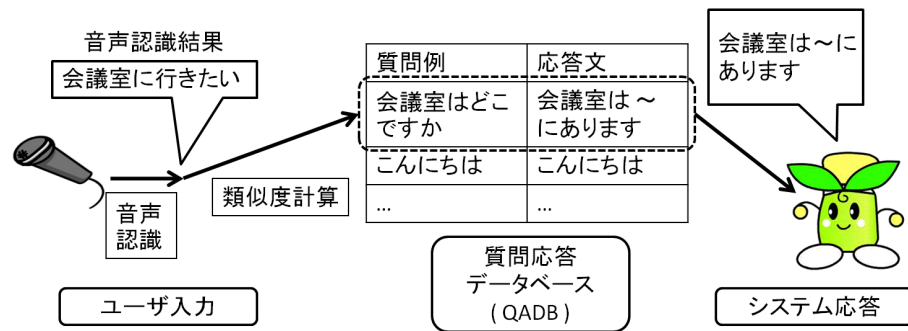


図3 「たけまるくん」の応答文生成  
Fig.3 Answer generation of Takemaru-kun.

似度  $s(I, E)$  が一番大きい質問例が QADB 中から最近傍法によって選択される<sup>3)</sup>。

$$s(I, E) = w(I, E) / \max(g_I, g_E) \quad (1)$$

$$w(I, E) = \sum_{k \in I \cup E} \min(w_I(k), w_E(k))$$

ここで、 $w_I(k), w_E(k)$  はそれぞれ単語  $k$  の文  $I, E$  の一文あたりの平均出現数であり、 $g_I, g_E$  はそれぞれ  $I, E$  の一文あたりの単語数である。選択された質問例に対応付けされた応答文が出力応答文として決定され、ユーザに提示される。

### 2.3 「たけまるくん」における応答文の拡張

用例ベースの応答文生成方式では、未知の質問に対する応答文を用例データベースに追加することで応答精度の向上を図る。「たけまるくん」においては、新たな質問例とそれに対応する応答文のペアを QADB に追記することで新たな質問文に対して答えることができる。QADB へのペアの追加は対話制御に関する知識を持たずとも簡単に行える設計になっている。従来の「たけまるくん」は発話の書き起こしが質問例として登録されているが、音声認識結果も質問例に登録しておくことで応答精度が向上したという研究もある<sup>3)</sup>。ただしいずれにせよ、質問例ひとつひとつに対して応答文を付与しなければならない。より多くの未知質問文に対応しようとする、QADB へのペア追加が簡単な設計であるとはいえ、システム管理者の負担は大きい。また、同内容の表現が異なる入力質問文に対して、それぞれの質問文に応じた表現バリエーションの応答文を出力したいと考えたとき、現在の QADB 方式

では応答文の用意が大変な負担になるという問題もある。

次節以降で説明する統計的機械翻訳を応用した応答文生成の手法はそれらの問題点を解消しようとするものである。

### 3. 統計的機械翻訳

統計的機械翻訳は対訳コーパスを分析して翻訳規則や対訳辞書にあたる統計モデルを自動学習し、ある言語から異なる言語への変換を行えるようにする技術である。モデル構築が自動で行えることにより、様々な言語の翻訳システムや、ある特定の分野に特化した翻訳システムが速く容易に構築可能となる。

#### 3.1 統計的機械翻訳の概要

本節では、統計的機械翻訳の概要について説明する。

原言語の文  $f$  を目的言語の文  $e$  に翻訳したいとする。このとき、翻訳結果  $e$  の候補は無数に存在する。翻訳デコーダは全てのペア  $(e, f)$  に対して  $f$  が  $e$  に翻訳される確率  $P(e|f)$  を計算し、 $P(e|f)$  を最大化する  $\hat{e}$  を探索する。ベイズの定理により、この問題は以下のよう表すことができる。

$$\hat{e} = \arg \max_e P(e|f) = \arg \max_e P(f|e)P(e) \quad (2)$$

この式において、 $P(f|e)$  は翻訳モデルを表し、 $P(e)$  は言語モデルを表す。翻訳モデルは翻訳の可能性を表し、言語モデルはその文の言語としての流暢さを示す。言語モデルは目的言語のコーパスから学習される。元々は単語アライメントを学習して作成される IBM 翻訳モデル<sup>4)</sup> が翻訳モデルとして使用されていたが、後に、フレーズベースの翻訳モデルが提案された。フレーズベース翻訳モデルにおいては、単語ではなく句がアライメントの単位として用いられる。ここで、「句」とは任意の単語列を指し、名詞句や動詞句といった言語学的なまとまりを指すものではない。

#### 3.2 フレーズベース翻訳モデル

本稿ではフレーズベース翻訳モデルを用いて応答文生成を行う。フレーズベース翻訳モデルの研究としては Koehn らの手法<sup>5)</sup> が有名である。この手法では、翻訳モデルは以下のよう定式化される。

$$P(f|e) = \prod_{i=1}^I \phi(\bar{f}_i | \bar{e}_i) d(a_i - b_{i-1}) \quad (3)$$

まず原言語文  $f$  を  $I$  個の句  $\bar{f}_1 \bar{f}_2 \dots \bar{f}_I$  に分割し,  $f$  中のそれぞれの句  $\bar{f}_i$  を目的言語の句  $\bar{e}_i$  に翻訳する. そして句  $\bar{e}_i$  の順序を入れ替える.  $\phi(\bar{f}_i | \bar{e}_i)$  は句翻訳確率であり,  $d(a_i - b_{i-1})$  は相対的な句歪み確率である.  $a_i$  は目的言語の  $i$  番目の句に訳される原言語句の開始位置である.  $b_{i-1}$  は目的言語の  $(i-1)$  番目の句に訳される原言語句の終端位置である.

句歪み確率は句 (もしくは単語) の翻訳前後の位置の違いで与えられるペナルティである. 句翻訳確率は以下のような相対頻度で与えられる.

$$\phi(\bar{f} | \bar{e}) = \frac{\text{count}(\bar{f}, \bar{e})}{\sum_{\bar{f}'} \text{count}(\bar{f}', \bar{e})} \quad (4)$$

Koehn らによって開発されたフレーズベースの統計的機械翻訳のツールキットとして Moses<sup>\*1</sup>がある. Moses では対訳コーパスから IBM 翻訳モデルの単語アライメントを基にしたヒューリスティックを用いてフレーズ抽出を行う.

#### 4. 統計的機械翻訳を用いた質問応答

統計的機械翻訳は元々は異言語間の翻訳を行う技術である. ここで質問文と応答文を別言語とみなして “翻訳” を行うことで, 質問文から応答文の生成ができると考えられる. 図 4 は本来の統計的機械翻訳の構成を表したものである. 対して, 図 5 は統計的機械翻訳の手法を用いた応答文生成手法の構成を表したものである. 言語間翻訳においては, 翻訳モデルは互いに異なる言語 (例えば英語と仏語など) の対訳コーパスから学習される. 応答文生成においては, 翻訳モデルは質問応答ペアの集合から学習される. 翻訳モデルを学習した後, 入力質問文をデコーダにかけることで応答文が生成できる.

#### 5. 実験

本実験では統計的機械翻訳を用いた質問文からの応答文生成の実現可能性を調査した. 異言語間翻訳を行う元々の統計的機械翻訳においては, 単語もしくは句の対は同じ意味を持っているが, 質問応答においては, それぞれのペアは異なる意味を持っているという差がある. また, 異言語間翻訳においては短い入力文は短く, 長い入力文は長く翻訳されたりするものであるが, 質問応答においては質問文の長さで応答文の長さの間に関係性があるかどうかは不明である. この実験の目的は, 上記のような違いがある中で, 統計的機械翻訳を利用

\*1 <http://www.statmt.org/moses/>

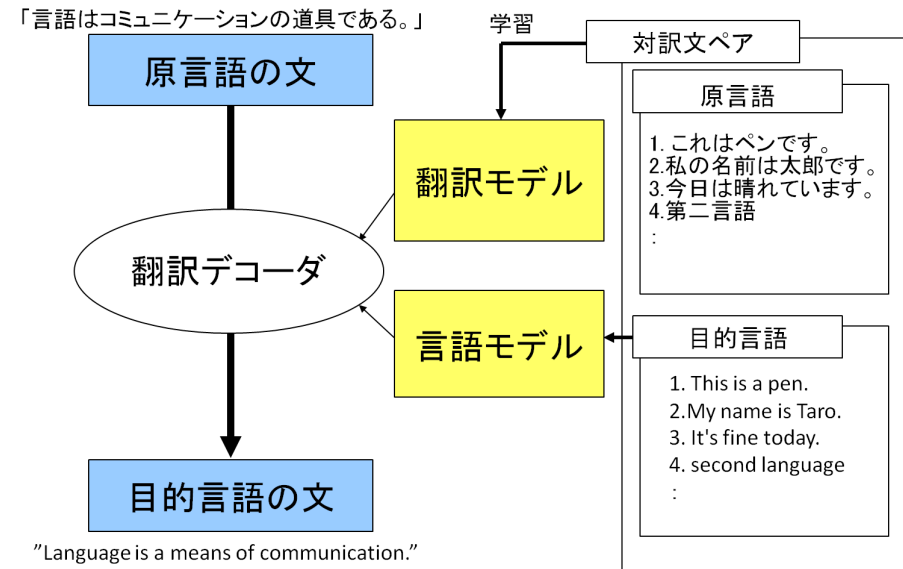


図 4 統計的機械翻訳の概要図  
Fig. 4 Statistical Machine Translation system.

した応答文生成の実現可能性や課題点を調査することである.

##### 5.1 実験条件

実験には「たけまるくん」で収集された大人ユーザ発話の人手による書き起こし文のデータセットを用いた. 発話それぞれに対して 276 種類の応答文のうち一つが付与されている. データ収集時期は 2002 年 11 月から 2004 年 10 月である. この期間から 2003 年 7,8 月を除いた期間の質問応答ペアにより翻訳モデルを作成した. 質問応答ペアのうち, 応答文から言語モデルを作成した. 2003 年 8 月のデータはテストデータとして, 2003 年 7 月のデータは開発データとして用いた. 開発データを用いて, BLEU スコアが最大となるようにパラメータチューニングを行った. テストデータの質問文を応答文に変換したあと, テストデータの応答文 (以下, 正解応答文と呼ぶ) と比較して性能評価を行った. 単語アライメ

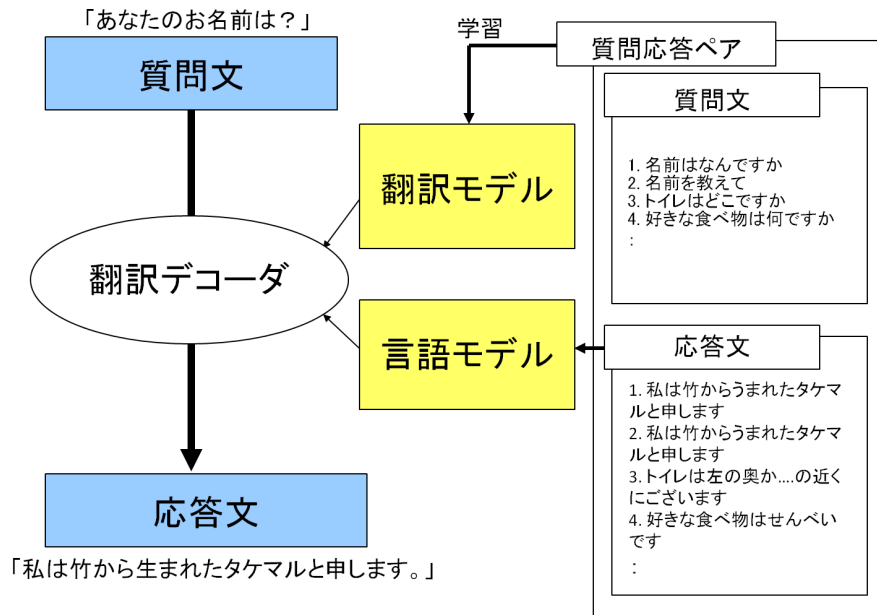


図 5 統計的機械翻訳を用いた質問応答の概要図  
Fig. 5 SMT for QA system.

ントを GIZA++<sup>\*1</sup>により学習し、フレーズ抽出して句翻訳モデルを求めた。言語モデルは SRILM を用いて求めた。

## 5.2 結果

生成された応答文から BLEU スコアを求めると、0.660 であった。しかし、BLEU スコアは応答としての正しさを直接反映するとはいえないので、一名の評価者による主観評価によっても性能を調査した。主観評価の基準としては「応答として適切か否か」という 2 択で評価を行った。その結果、生成された応答文が適切な応答文であるとして評価されたのはテストデータ 1053 文のうち 592 文であった。このうち 543 文は正解応答文そのものに変換さ

表 1 実験データの諸元  
Table 1 Dataset features.

学習データ	期間	2002 年 11 月-2004 年 10 月 (2003 年 7, 8 月を除く)
	データ数	18509 ペア
開発データ	期間	2003 年 7 月
	データ数	872 ペア
テストデータ	期間	2003 年 8 月
	データ数	1053 ペア

表 2 実験結果の評価  
Table 2 Evaluation of result.

テストデータ総数	1053 文
適切な応答	592 sentences
正解応答文と一致	543 文
正解応答文と一致していないが適切	49 文
不適切な応答	461 文
BLEU スコア	0.660

れたものであった。残りの 49 文は正解応答文そのものではないが応答として適切であると主観的に評価されたものである。適切な応答はテストデータ全体の 56.2%にあたる。提案手法による応答文生成の実現可能性という観点においては、過半数のテストデータが適切な応答に変換されたので、実現可能性はあると判断した。

ちなみに、従来システムの正解応答率は 79.9%<sup>6)</sup>である。提案手法をシステムに導入するためには、適切な応答が生成される割合を 79.9%に近づけることが必要である。

## 5.3 考察

ここからは、提案手法の課題点を探るべく、生成された応答文の例を挙げ、それぞれについて考察していく。変換された応答文の例を表 3 に示す。例 1 は、変換された応答文が正解応答文と完全に一致した例である。例 2 は、正解応答文とは異なるが、応答として適切な文が生成された例である。表 3 に挙げた残りの例は応答として不適切な文が生成されたものである。例 3 は単語の接続が誤りを含む例である。具体的には「～うまれたのホームページ～」の「の」という助詞が直前の「た」という助動詞から接続されているのが誤りである。例 4 や例 5 は文が途中から始まったり途中で切れてしまっている例である。このような例に対しては、単語の品詞から判断して直前の単語と接続可能かどうか、また、文の開始・終了位置に存在できるかどうかという制限をかけることで対応できると考えられる。

\*1 <http://code.google.com/p/giza-pp/>



表 3 生成された応答文の例  
Table 3 Examples of result.

例 1	Q 「あなたのお名前は」 T 「私は竹からうまれたタケマルと申します」 A 「私は竹からうまれたタケマルと申します」
例 2	Q 「大阪に行きたいんですけど」 T 「大阪方面に行くには難波行きの電車に乗ってください」 A 「大阪方面に行くには難波行きの電車をご利用ください」
例 3	Q 「名前を教えてください」 T 「私は竹からうまれたのホームページにアクセスします」 A 「私は竹からうまれたタケマルと申します」
例 4	Q 「市役所はどこですか」 T 「は近鉄生駒駅の南側にあります」 A 「生駒市市役所は近鉄生駒駅の南側にあります」
例 5	Q 「あなたの名前は何ですか」 T 「私は竹からうまれたタケマルと」 A 「私は竹からうまれたタケマルと申します」
例 6	Q 「お名前を教えてください」 T 「うまれたタケマルと申します」 A 「私は竹からうまれたタケマルと申します」

Q:入力質問文 T:生成された応答文 A:正解応答文

また、翻訳モデルの学習の結果、そのモデルからどのような文が生成されるのか、実際に生成してみるまで正確に分からないという難しさもある。例 3 と例 6 を比べると、入力質問文は僅かな違いしかないが、生成された応答文は大きく異なる。また、本稿では書き起こし文を用いた実験のみを行っているが、実際にシステムで運用する際には音声認識結果テキストを入力とする。音声認識結果を入力としたときの性能も調査する必要がある。

## 6. 結 論

本稿では、音声情報案内システムへの想定外の発話に対応する手法として統計的機械翻訳を用いることを提案し、その実現可能性と課題点を調査した。提案手法では、翻訳モデルと言語モデルを質問応答ペアから学習し、質問文から応答文を生成する。実験の結果から、この手法の実現可能性はあると思われる。しかしながら、不適切な応答が生成されるケースも多数見受けられた。今後は、第 5 節で述べたような制約をかけて性能向上をはかる。さらに、実際の運用を想定して、書き起こし文ではなく音声認識結果を入力としたときの性能調査も行う予定である。

謝辞 本研究の一部は、科学技術振興事業団・戦略的基礎研究推進事業 (CREST) の支援

を受けて実施された。

## 参 考 文 献

- 1) Nishimura, R., Lee, A., Saruwatari, H. and Shikano, K.: Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2004)*, pp.433–436 (2004).
- 2) 藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡 洋, 鹿野清宏: 音声情報案内システムにおける SVM を用いたタスク外発話の検出, 情報処理学会研究報告, Vol.2009-SLP-77, No.14, pp.1–6 (2009).
- 3) Takeuchi, S., Cincarek, T., Kawanami, H., Saruwatari, H. and Shikano, K.: Question and Answer Database Optimization Using Speech Recognition Results, *INTERSPEECH 2008*, pp.451–454 (2008).
- 4) Brown, P.F., Pietra, V. J.D., Pietra, S. A.D. and Mercer, R.L.: The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, Vol.19, No.2, pp.263–311 (1993).
- 5) Koehn, P., Och, F.J. and Marcu, D.: Statistical Phrase-Based Translation, *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Vol.1, pp.48–54 (2003).
- 6) Takeuchi, S., Kawanami, H., Saruwatari, H. and Shikano, K.: Unknown Example Detection for Example-based Spoken Dialog System, *Oriental COCODA 2009*, O3-5, Urumqi, China (2009).