# Investigation of Statistical Machine Translation Applied to Answer Generation for a Speech-Oriented Guidance System

Kazuma Nishimura, Hiromichi Kawanami, Hiroshi Saruwatari and Kiyohiro Shikano*
* Graduate School of Information Science, Nara Institute of science and Technology, Japan
E-mail: {kazuma-n,kawanami,sawatari,shikano}@is.naist.jp

***Abstract*—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*.**

## I. Introduction

Automatic speech recognition (ASR) has been widely applied to dictation, Voice Search, and car navigation, to name a few. In this paper, we describe a speech-oriented information guidance system, *Takemaru-kun*, which aims to realize a natural speech interface using ASR [1].

*Takemaru-kun* is a real-environment speech-oriented information guidance system whose task domain is not given beforehand. It is an example-based question answering system, which is flexible to respond to user's questions on demand.

An answer to a user's question is selected by referring to a question and answer database (QADB), which can be easily maintained without paying particular attention to the scope of the system.

One of the problems in an example-based system such as *Takemaru-kun* is that the system cannot respond to unexpected user's utterances, if a similar example of a question-answer pair does not exist in the QADB. A secondary problem is that the answer sentences cannot reflect differences in nuance as the set of answer sentences are fixed beforehand. For example, the system would respond the same sentence to different questions such as “Would you tell me how to get to the nearest bus stop?” and “Where can I ride on a bus?” To realize a familiar speech interface, it is preferable to arrange answers using more appropriate phrases.

An analytical result of user's utterances shows that about 5% of user's questions are not included in the QADB. To treat this problem, research on out-of-task utterances (OOT) detection using SVM (Support Vector Machine) and BOW (Bag-of-Word) has been conducted [2]. A detected OOT can be sent to an Internet Voice Search engine as a query.

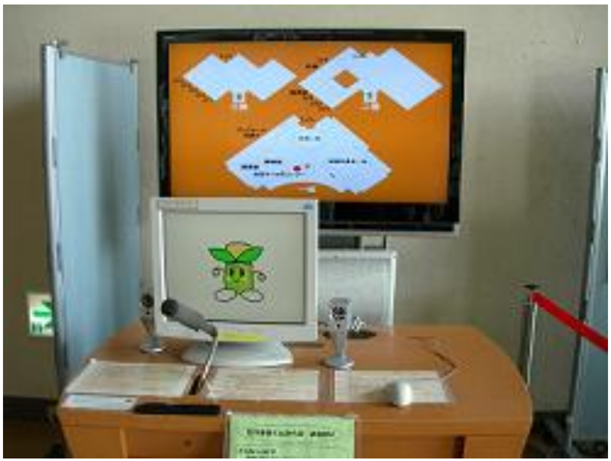

Fig. 1. Speech-oriented guidance system *Takemaru-kun*.

In this paper, we propose an approach to deal with unknown questions by introducing a Statistical Machine Translation (SMT) technique. In the proposed paradigm, we treat the question set and the answer set as different languages. That is to say, a question to the system is translated to the corresponding system answer by SMT models. By introducing SMT technique to question answering (QA), differences of nuance are also expected to be reflected in an answer sentence.

In the following sections, the system overview of *Takemaru-kun*, the concept of SMT and how we introduce it to QA are described. They are followed by an experimental evaluation using transcriptions of real user's utterances received by *Takemaru-kun* and discussion and conclusion of the paper.

## II. Speech-Oriented Guidance System Takemaru-kun

### A. *Takemaru-kun System Overview*

*Takemaru-kun* is a speech-oriented information guidance system that has been in operation since Nov. 2002 at the entrance hall of *Ikoma City North Community Center* (Fig. 1) [1]. The system answers user's questions about the center facilities, services, neighboring sightseeing, agent profile and so on. The system employs a one-question-one-answer strategy. This approach is simple, but it achieves robust answer generation. A system answer is provided by synthetic speech, Web browser and CG agent animation.

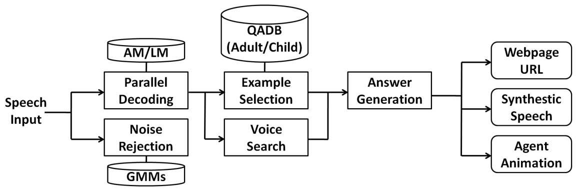

Fig. 2. Processing flow of *Takemaru-kun*.

The system structure is illustrated in Fig. 2. Speech/Noise discrimination using Gaussian Mixture Models (GMM) is executed in parallel with ASR. MFCCs of the GMMs are constructed from five kinds of real input to the system, which are adult speech, child speech, laughing, coughing and other noise. If the likelihood in any of the last three is the highest, the input is rejected as a noise.

Adult and child classification is also conducted during parallel decoding using acoustic likelihoods. The N-Best ASR result is used to calculate a similarity score with each example question in the QADB, prepared for adult and child separately. The nearest neighbor approach is employed for example selection using equation (1) [3]. The example question with the highest score is regarded as the user input and the corresponding answer is used as the output message.

Similarity score =

$$\frac{\text{number of word coincidences in } S_I \text{ and } S_E}{\max\,(\text{number of words in } S_I,\ \text{number of words in } S_E\ )} \quad (1)$$

$$S_I \in \ \{ \text{ Input utterances } \}$$

$$S_E \in \ \{ \text{ Example utterances } \}$$

All system input have been collected from the start of operation. The data for the first two years were manually transcribed with tags concerning noise and labels about age-group and gender. The tags and labels were given by hearing of four trained labelers. These data were used to construct the GMMs and to adapt the acoustic models and language models used in the daily operation. The transcription data are used in the experiment section.

### B. Answer Sentence Extension in Takemaru-kun

In the system, the tasks are not limited beforehand. As the system offers example-based QA, the domains of answer have been extended on demand of users. The QADB in the system consists of example questions and corresponding answer pairs. As the QADB can be simply updated by adding question-answer pairs, to deal with a variety of phrases that appear in spontaneous speech, transcriptions of user's utterances have been added to the example questions. Introducing ASR results as question examples is also effective to update QADB [3]. However, the addition of new answer sentences must be conducted manually with the preparation of the corresponding example questions.

## III. Statistical Machine Translation Method

SMT builds statistical translation models from the analysis of bilingual corpora and makes possible to translate a language into another language automatically. This technique makes it possible to build quickly and easily translation systems for various languages or systems for special fields, for example, patent documents.

### A. Fundamentals of SMT

Here, we explain the fundamentals of SMT. Suppose you want to translate a sentence from source language $\mathbf{f}$ into a sentence from target language $\mathbf{e}$. There are innumerable choices of translated results $\mathbf{e}$. The decoder in SMT system calculates $P(\mathbf{e}|\mathbf{f})$, the probability that $\mathbf{e}$ is the translation result of $\mathbf{f}$, for all pairs of $(\mathbf{e}, \mathbf{f})$. The system outputs the sentence $\hat{\mathbf{e}}$ for which $P(\mathbf{e}|\mathbf{f})$ is the greatest (Fig. 3). Using Bayes' theorem, we can express this problem as below,

$$\begin{aligned} \hat{\mathbf{e}} &= \arg \max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) \\ &= \arg \max_{\mathbf{e}} P(\mathbf{f}|\mathbf{e}) P(\mathbf{e}) \end{aligned} \quad (2)$$

In this formula, $P(\mathbf{f}|\mathbf{e})$ expresses the translation model and $P(\mathbf{e})$ expresses the language model. The translation model is the probability of translating, and the language model is the expression of fluency of the sentence. The translation model is built from the analysis of bilingual corpora and the language model is built from the analysis of the corpus of the target language. The IBM model [4], built by learning the alignment of words, was used originally as the translation model. Recently, phrase-based translation model has been proposed. In the phrase-based model, a phrase is used as the alignment unit, instead of a word. Now, "phrase" means simply a sequence of words, not a linguistic unit, for example, verb phrase or noun phrase.

### B. Phrase-based SMT

In this paper we apply the phrase-based SMT. Koehn's method [5] is popular in phrase-based SMT methods. In this method, the translation model is formulated as below.

$$P(\mathbf{f}|\mathbf{e}) = P(\overline{f}_1^I|\overline{e}_1^I) = \prod_{i=1}^{I} \phi(\overline{f}_i|\overline{e}_i) d(a_i - b_{i-1}) \quad (3)$$

The input sentence of source language $\mathbf{f}$ is segmented into a sequence of $I$ phrases $\overline{f}_1^I$. Each phrase $\overline{f}_i$ in $\overline{f}_1^I$ is translated into a phrase of target language $\overline{e}_i$. The phrases $\overline{e}_i$ may be reordered. $\phi(\overline{f}_i|\overline{e}_i)$ is the phrase translation probability, $d(a_i - b_{i-1})$ is the phrase distortion probability. $a_i$ is the start position of the phrase of source language that was translated into the $i$-th phrase of target language. $b_{i-1}$ is the end position of the phrase of source language that was translated into the $(i-1)$-th phrase of target language.

The phrase distortion probability is a penalty for the difference of the position between phrases (or words) before translation and after translation.

Phrase translation probability is given by relative frequency.

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

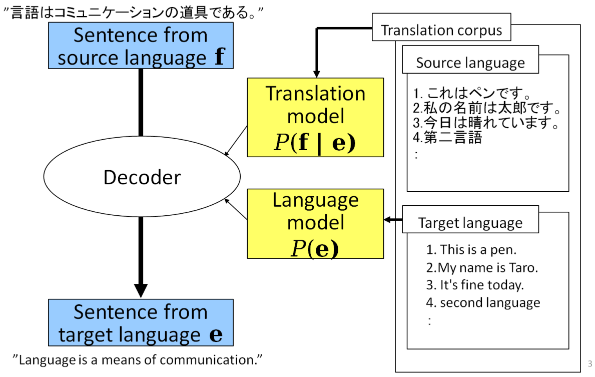
”言語はコミュニケーションの道具である。”
Sentence from source language $\mathbf{f}$
Translation corpus
Source language
1. これはペンです。
2.私の名前は太郎です。
3.今日は晴れています。
4.第二言語
:
Translation model $P(\mathbf{f} \mid \mathbf{e})$
Decoder
Language model $P(\mathbf{e})$
Target language
1. This is a pen.
2.My name is Taro.
3. It's fine today.
4. second language
:
Sentence from target language $\mathbf{e}$
”Language is a means of communication.”

Fig. 3. Statistical Machine Translation system.

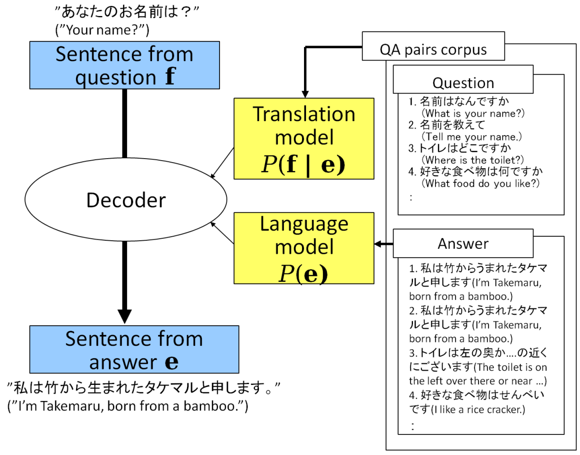
”あなたのお名前は？”
(”Your name?”)
Sentence from question $\mathbf{f}$
QA pairs corpus
Question
1. 名前はなんですか
(What is your name?)
2. 名前を教えて
(Tell me your name.)
3. トイレはどこですか
(Where is the toilet?)
4. 好きな食べ物は何ですか
(What food do you like?)
:
Translation model $P(\mathbf{f} \mid \mathbf{e})$
Decoder
Language model $P(\mathbf{e})$
Answer
1. 私は竹からうまれたタケマルと申します(I'm Takemaru, born from a bamboo.)
2. 私は竹からうまれたタケマルと申します(I'm Takemaru, born from a bamboo.)
3.トイレは左の奥か....の近くにございます(The toilet is on the left over there or near ...)
4. 好きな食べ物はせんべいです(I like a rice cracker.)
:
Sentence from answer $\mathbf{e}$
”私は竹から生まれたタケマルと申します。”
(”I'm Takemaru, born from a bamboo.”)

Fig. 4. SMT for QA system.

Moses a the phrase-based SMT toolkit developed by Koehn, et al.(http://www.statmt.org/moses/). Moses extracts phrases from the bilingual corpora using the heuristics based on the word alignment of the IBM model.

## IV. Answer generation by using SMT

SMT is originally a technique that makes possible automated translation between different languages.

In our approach, we suppose that question sentences could be translated into answer sentences if we consider question and answer sentences as different languages. Fig. 3 shows the original SMT flowchart and Fig. 4 illustrates how the technique is applied to the QA proposed in this paper. In the language translation task, translation models are built from bilingual corpora, for example, English and French. Now, in the QA translation task, translation models are built from QA pairs and language models from answer sentences. The following procedures are the same as the language translation task.

TABLE I
Dataset features

| | | |
|---|---|---|
| Training data | Period | Nov.2002-Oct.2004<br>(excluding Jul.&Aug.2003) |
| | # of data | 18509 pairs |
| Development data | Period | Jul.2003 |
| | # of data | 872 pairs |
| Test data | Period | Aug.2003 |
| | # of data | 1053 pairs |

TABLE II
Evaluation of result

| | |
|---|---|
| Total test data | 1053 sentences |
| Appropriate answers | 592 sentences |
| exactly the same as correct answer sentence | 543 sentences |
| not exactly the same as correct answer sentence | 49 sentences |
| Inappropriate answers | 461 sentences |
| BLEU score | 0.660 |

## V. Experiments

We investigated whether question sentences can be properly translated into corresponding answer sentences in this experiment. In the language translation task, each pair of words or phrases has the same meaning, however, in the QA translation task, each pair has different meanings. Additionally, short sentences tend to be translated into short sentences in the language translation task, but the length of the answer sentence does not directly relate to that of the question sentence generally. The purpose of this experiment is to investigate the possibility of generating answer sentences by SMT in the conditions as mentioned above.

### A. Experimental condition

We employed a dataset that consists of manual transcriptions of adult user's utterances and the answer sentences tagged on them. The amount of the kinds of answer sentences is 276. The dataset was collected with *Takemaru-kun* system from Nov. 2002 to Oct. 2004(Table I). We built the translation model from these QA pairs, and built the language model from answer sentences, excluding the pairs of Jul. and Aug. 2003. The training data consist of 18509 pairs. The data of Jul. 2003 are used as development data, and the feature weights were optimized for BLEU. The data from Aug. 2003 are used as test data. The question sentences of the test data are translated into answer sentences with the translating decoder. We obtained the word alignment by running GIZA++(http://code.google.com/p/giza-pp/), and built the language model by SRILM, and extracted phrases and decoded sentences by Moses.

### B. Results and Discussions

By the translation with the proposed method, BLEU score was 0.660. In the example-based method, in comparison, we didn't use BLEU score. We evaluated the results subjectively from the view of ”the appropriateness as an answer.” The question sentences were translated into exactly the same answer sentences in the QADB for 543 test sentences out of

1053. Additionally, there were 49 cases where the appropriate answers were not exactly the same answer sentences. These 592 cases correspond to 56.2% of the test data, so it shows the feasibility of this method. In comparison, response accuracy of the conventional method using similarity scores described in section 2 is 79.9% [6]. In this experiment, transcription data excluding the test data are used for the QADB. Although the accuracy of the proposed method is lower than the conventional method, the feasibility of the method is illustrated.

Examples of translated sentences are illustrated in Table III. In Ex.1, the translated sentence is exactly the same to the tagged answer, and that in Ex.2 is not exactly the same, however, it is proper as an answer. The rest are examples that have problems in the generated answer sentences. For example, the concatenation in the translated sentence Ex.3 is strange. We think that a restriction of concatenation with POS information may be effective for these cases. Ex.4 and 5 are broken, incomplete sentences. In these cases, complete sentences may be generated restricting first/last words of a sentence as appropriate first/last words. Ex.6 is formed well, however, the content is incorrect. These cases occur because SMT cannot use knowledge of the meaning of the words. This phenomenon should not occur in the guidance task, so we have to think a way to screen sentences with incorrect contents. Additionally, a problem is that the control is difficult. For example, the question sentences of Ex.3 and Ex.7 are almost the same, however, the translated sentences are fairly different. In this paper, transcriptions are used as input question sentences, however, the actual input to *Takemaru-kun* are ASR results, so it is expected that more difficulties will arise. There are many difficulties as mentioned above, however, question sentences can be translated into answer sentences that have entirely different meanings.

## VI. Conclusions

In this paper, we proposed an approach to deal with unknown questions in a speech-oriented information guidance system, by introducing SMT technique. In this method, a translation model and a language model are built from QA pairs, and question sentences are translated into answer sentences. The experimental evaluation shows the feasibility of the proposed approach. However, analytical results show that there are many cases where translation fails. In future work, we will apply the restrictions mentioned in section 5, and we will conduct experiments using ASR results as the actual input, which include recognition errors. Finally, in this paper, we did not convert SMT tools, considering the difference between language translation task and QA translation task. In future work, it is also necessary to research about the difference between the tasks, including problems related to the lack of information on OOT.

## Acknowledgment

This work was partially supported by CREST (Core Research for Evolutional Science and Technology), Japan Science and Technology Agency(JST).

TABLE III
Examples of result

| Index | | Sentence |
|---|---|---|
| Ex.1 | Q | ”あなたのお名前は” (”Your name?”) |
| | T | ”私は竹からうまれたタケマルと申します”<br>(”I’m Takemaru, born from a bamboo.”) |
| | A | ”私は竹からうまれたタケマルと申します”<br>(”I’m Takemaru, born from a bamboo.”) |
| Ex.2 | Q | ”大阪に行きたいんですけど” (”I want to go to Osaka.”) |
| | T | ”大阪方面にいくには難波行きの電車に乗ってください”<br>(”Get on a train bound for Nanba to go to Osaka.”) |
| | A | ”大阪方面にいくには難波行きの電車をご利用ください”<br>(”Take a train bound for Nanba to go to Osaka.”) |
| Ex.3 | Q | ”名前を教えて下さい” (”Please tell me your name.”) |
| | 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”<br>(”I’m accessing born from a bamboo Web page.”) |
| | A | ”私は竹からうまれたタケマルと申します”<br>(”I’m Takemaru, born from a bamboo.”) |
| Ex.4 | Q | ”市役所はどこですか” (”Where is the Ikoma City Hall?”) |
| | T | ”は近鉄生駒駅の南側にあります”<br>(”is on the south of the Ikoma Station.”) |
| | A | ”生駒市市役所は近鉄生駒駅の南側にあります”<br>(”Ikoma City Hall is on the south of the Ikoma Station.”) |
| Ex.5 | Q | ”あなたの名前は何ですか” (”What’s your name?”) |
| | T | ”私は竹からうまれたタケマルと”<br>(”I Takemaru, born from a bamboo.”) |
| | A | ”私は竹からうまれたタケマルと申します”<br>(”I’m Takemaru, born from a bamboo.”) |
| Ex.6 | Q | ”プレイルームはどこですか” (”Where is the playroom?”) |
| | T | ”プレイルームは左の奥のトイレの隣にあります”<br>(”The playroom is next to the toilet further to the left.”) |
| | A | ”プレイルームは左に行った階段の隣です”<br>(”The playroom is next to the stairs to the left.”) |
| Ex.7 | Q | ”お名前を教えて下さい” (”Please tell me your name.”)<br>*more polite expression than Ex.3 |
| | T | ”うまれたタケマルと申します”<br>(”I’m Takemaru, born”) |
| | A | ”私は竹からうまれたタケマルと申します”<br>(”I’m Takemaru, born from a bamboo.”) |

Q:Question sentence
T:Translated sentence from question
A:Tagged answer sentence

## References

[1] R. Nisimura *et al.*, “Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability,” *In Proc. of ICASSP2004*, vol. 1, pp. 433-0436, 2004.

[2] Y. Fujita *et al.*, “Detection of Out-of-Task Utterances Using SVM for Speech-Oriented Guidance Systems,” (in Japanese) *IPSJ Technical report*, 2009-SLP-77(14), pp. 1-6, 2009.

[3] S. Takeuchi *et al.*, “Question and Answer Database Optimization Using Speech Recognition Results,” *INTERSPEECH2008*, pp. 451-454, 2008.

[4] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer, “The Mathematics of Statistical Machine Translation: Parameter Estimation,” *Computational Linguistics*, vol. 19-2, pp. 263-311, 1993

[5] P. Koehn, F. J. Och, D. Marcu, “Statistical Phrase-Based Translation,” *In Proc. of HLT-NAACL*, 2003

[6] S. Takeuchi *et al.*, “Unknown Example Detection for Example-based Spoken Dialog System,” *In Proc. of Oriental COCOSDA2009*