

## 対話エージェントを用いた異言語間コミュニケーションシステム

高橋大和\* 片岡明\* 麻野間直樹\* 山田節夫\* 古瀬蔵\*

日本電信電話株式会社 NTT サイバースペース研究所

{ takahashi.yamato, kataoka.akira, asanoma.naoki, yamada.setsuo,  
furuse.osamu}@lab.ntt.co.jp

### 概要

音声によるタスク達成型コミュニケーションを異言語間で実現する方式を提案する。タスク達成対話では、必要な情報だけをキーワードとして伝達できればコミュニケーションを成立させることが可能である。そこで、キーワードの獲得を行うために発話を制限・誘導する技術である対話エージェントを利用する。対話エージェントとのやりとりを通してキーワードを確認することで、正確なキーワードの獲得が行える。また、キーワードを翻訳して定型文に埋め込み、対話エージェントがガイダンスとしてもう一方の話者に発話することで、タスク達成対話を対象とする異言語間コミュニケーションを成立させることができる。

## A System with Spoken-Dialog Agents for Goal-oriented Communication across Languages

Yamato Takahashi<sup>†</sup>, Akira Kataoka<sup>†</sup>, Naoki Asanoma<sup>†</sup>, Setsuo Yamada<sup>†</sup>, and Osamu Furuse<sup>†</sup>

NTT Cyber Space Laboratories, NTT Corporation

{ takahashi.yamato, kataoka.akira, asanoma.naoki, yamada.setsuo,  
furuse.osamu}@lab.ntt.co.jp

### Abstract

This paper proposes a novel architecture for goal-oriented communication across languages. Goal-oriented dialog makes it possible for communication partners to catch each other's intention by simply exchanging a few keywords; this advances the achievement of their goals. In order to acquire such keywords, the technique of the spoken-dialog agent, which controls utterances, is introduced to a speech translation system. The spoken-dialog agents in this system lead the users' utterances toward the goal and confirm the results of speech recognition. By embedding the translated keywords into sentence templates, the system's utterances to each user can be generated in the user's language.

---

\* 日本電信電話株式会社 NTT サイバースペース研究所

† NTT Cyber Space Laboratories, NTT Corporation

## 1. はじめに

音声翻訳は、母国語の音声で外国人とのコミュニケーションを実現させるための技術である。しかし、例えば、対象分野を旅行会話に限定しても[1]、現在の音声認識と機械翻訳技術では、それぞれの処理で常に正しい結果が得られるとは限らず、これらを単純に結合させただけで実用的な音声翻訳システムを構築することは困難である。そこで、対象分野を限定することに加え、音声認識誤りを手動で修正する[2]、システム主導の対話を通して、認識誤りを修正する[3]などの方法や、認識誤りの修正に加え、複数の訳語候補をあげ、翻訳誤りも手動で修正を行う[4]といった方法も提案されている。これらの方法は、発話された内容を全て認識し、全ての内容を正確に翻訳することに目標を置いている。

しかし、タスクを限定した外国人とのコミュニケーションでは、そのタスクに重要なキーワードが相手に理解されることが必要で、必ずしもすべての発話内容が正確に理解される必要はない。例えば、予約のようなタスクでは、日時、人数、連絡先などの重要なキーワードが正確に理解されるように翻訳できればよい。異言語間でのタスク達成型コミュニケーションでは、重要なキーワードを発話から正確に獲得し、相手言語で理解できるように翻訳することが必要である。

タスク達成に重要なキーワードの獲得に有用な方法として、対話エージェント[5]の導入がある。対話エージェントに発話の制限・誘導を行わせることで、キーワードの獲得ができる。また、聞き返しなどの対話制御[6]を行わせることで、音声認識誤りが起きてもキーワードが正しいかどうかの確認ができる。

一方、タスクに重要なキーワードが理解できるように、翻訳されたキーワードが埋め込まれた定型文を対話エージェントに提示させる。ここで、タスクに重要なキーワードは、変換辞書を利用することによって翻訳される。

本稿では、キーワードの獲得に対話エージェントを利用することで、タスク達成を目的とした異言語間の音声コミュニケーションが実現できる手法を提案する。提案手法では、異なる言語を発話する話者間のコミュニケーションに対話エージェントを介在させる。対話エージェントは、対話シナリオにしたがって、話者の発話からタスク達成に重要なキーワードを獲得し、このキーワードを確認するための発話を行う。また、このキーワードを理解させるための発話を、翻訳されたキーワードを定型文に埋め込むことで生成し、相手に提示する。すなわち、異なる言語を発話する話者の間でキーワードのやりとりにより意思の疎通ができるよう対話シナリオを用意し、対話エージェントが対話シナリオにそって対話を制御し、タスク達成を成功へと導く。

この手法の利点として、

1. キーワードが正しく獲得されたかどうかを話者に確認することによって、音声認識誤りを回避
2. 話者の意図を、もう一方の話者の言語で伝達するために、定型文にキーワードを埋め込んだ発話を提示することにより機械翻訳誤りを回避

があげられる。

## 2. 対話エージェントを使った異言語間コミュニケーション

本節では、提案手法を実装した異言語間コミュニケーションシステムの構成と動作について説明する。なお、本実装では、甘粕ら[5]の対話エージェントを利用している。

### 2.1. 異言語間コミュニケーションシステムの構成

提案手法に基づく異言語間コミュニケーションシステムでは、異なる言語を使用する二人の話者が直接対話するのではなく、対話エージェントが二人の話者の間に介在することによって意思の疎通を確実に行わせる。単一の対話エージェントが話者ごとに言語を切り替えて対話するというシステム構成も可能ではあるが、本稿では、話者ごとに対話エージェントを用意したシステム構成について説明を行う。

図1に、システムの全体構成を示す。以下、話者Aは対話エージェントAと日本語で対話し、話者Bは対話エージェントBと英語で対話を行うものとする。対話エージェントは対話シナリオによって制御され、次の機能を有する。

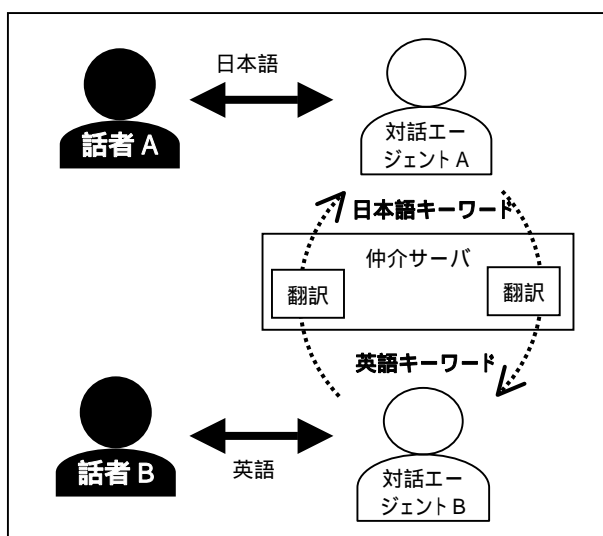


図 1. 全体構成

(1)話者の発話の音声認識、(2)タスク達成のためのキーワードをその認識結果から獲得、(3)伝達されたキーワードを定型文に埋め込んで提示、(4)次の動作を決定。次の動作としては、話者と対話を行うか、あるいは、獲得したキーワードを、仲介サーバを経由してもう一方の対話エージェントに伝達することをを行う。

### 2.2. 対話エージェントと話者のやりとり

図2に対話エージェントの構成を示す。対話エージェントは話者に対して、発話開始のガイダンス、キーワードを確認するためのガイダンス、もう一方の話者からのキーワードを埋め込んだガイダンス、のいずれかを発話する。一方、話者は対話エージェントに対して、ガイダンスにしたがって応答の発話をする。

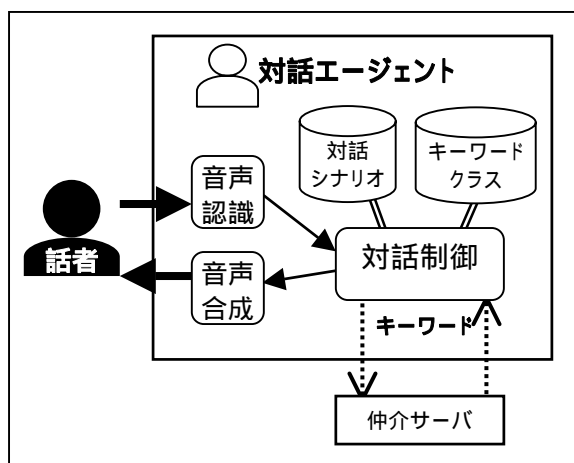


図 2. 対話エージェントの構成

対話エージェントは、音声認識、対話制御、音声合成の処理により話者との対話を行う。話者の発話に対して音声認識を行い、その認識結果を対話制御に送る。対話制御は認識結果からキーワードを獲得する。また、獲得したキーワードに対し、対話シナリオとキーワードクラスを参照して、音声合成に送るガイダンスを決定する、もしくは、キーワードを仲介サーバに送る。前者の場合、音声合成されたガイダンス

を対話エージェントが話者に発話する。

キーワードクラスは、音声認識した単語のうち、キーワードとして扱う単語を意味や役割の面から分類したデータベースである。例えば、[肯定]というクラスに、「はい」「ええ」「そうです」といったキーワードを列挙しておくことで、音声認識結果に「はい」といったキーワードが含まれた場合は、[肯定]の発話がされた、と認識する。

対話シナリオの例として、表 1 にタクシー手配手続きの一部を示す。対話シナリオは、対話状況に固有なラベルが付与されている。これを対話状況ラベルと呼ぶ。各対話状況について、対話エージェントが発話すべきガイダンス、獲得したキーワードクラスに応じて遷移すべき対話状況、そして、仲介サーバへ送るべきキーワードクラス、その具体的な値、および相手先の対話シナリオにおける対話状況、が規定されている。

表 1. 対話シナリオ例  
(タクシー手配手続きの一部)

対話状況ラベル	ガイダンス	キーワードクラス	遷移先の対話状況ラベル	仲介サーバへ送信する情報
手続き開始	タクシーの手配をいたしますか？	[肯定]	台数確認	
		[台数]	台数確認	
		[否定]	手続き終了	
			手続き開始	
台数確認	[台数]でよろしいでしょうか？	[肯定]	(待ち状態)	[台数], 台数提示
		[否定]	台数確認	
		[台数]	台数確認	
			台数確認	

対話シナリオの基本的な流れは、話者にガイダンスを発話し、その発話に応答した話者の発話を音声認識し、その結果に含まれているキーワードクラスの組合せを獲得する。さらに、獲得したキーワードの確認を行う。この確認により、重要なキーワードを間違えて伝えることがない。

### 2.3. 対話エージェント間のやりとり

話者が発話したキーワードを、もう一方の話者に伝達する対話状況では、対話シナリオにしたがって、対話エージェント間でキーワードと対話状況に関する情報のやりとりが、仲介サーバを通じて行われる。

図 3 に仲介サーバの構成を示す。仲介サーバは、対話エージェントから送られた情報を、もう一方の対話エージェントへ伝達する。仲介サーバは、エージェント間制御と翻訳エンジンからなる。翻訳エンジンは、日英翻訳と英日翻訳のように双方向用意する。

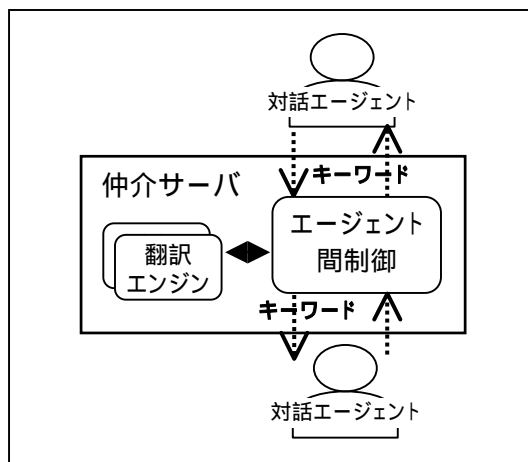


図 3. 仲介サーバの構成

### 3. 対話例: タクシー手配手続き

図 4 に示す対話例で、タクシー手配手続きについて、タスク達成までの流れを説明する。話者 A が話者 B に、配車して欲しい台数を伝え、話者 B が話者 A に、どのくらいで到着するかという情報を伝えられればタスクを達成でき

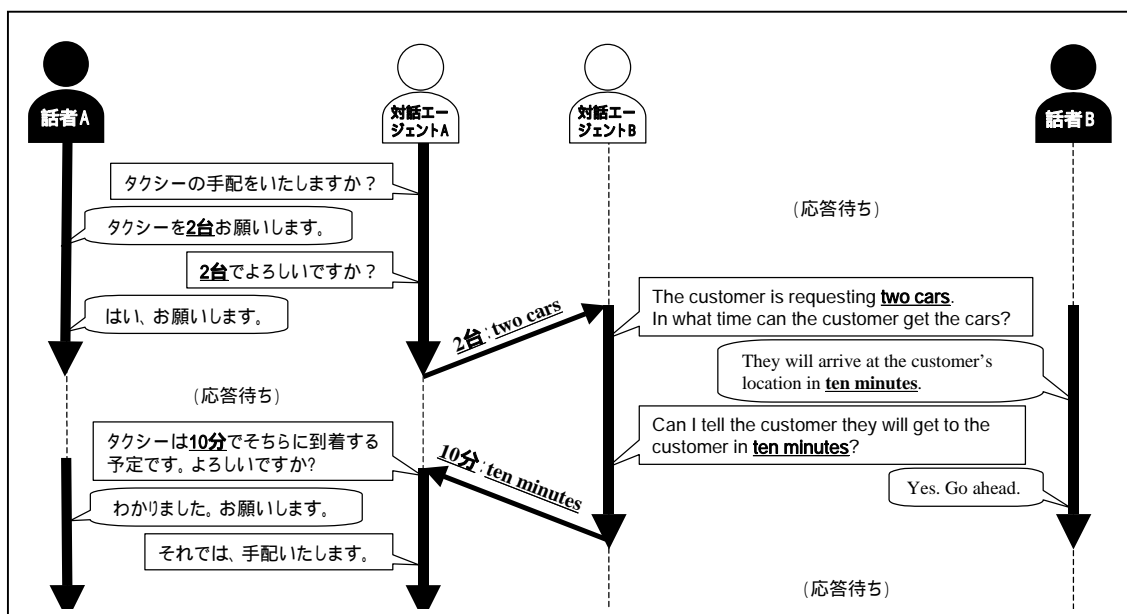


図 4. 対話例(タクシー配車依頼の一部)

たこととする。

話者 A と対話エージェント A とのやりとりで、対話エージェント A が「2 台」と確認した[台数]の情報が、仲介サーバを経て、「two cars」という値で対話エージェント B に伝達される。対話エージェント B は、「two cars」でタクシー依頼について[到着時間]の値を尋ねるガイダンスを話者 B に発話する。それに対し、話者 B は[到着時間]として「ten minutes」を応答し、対話エージェント B とのやりとり、対話エージェント B から仲介サーバを経て対話エージェント A への情報伝達により、対話エージェント A から話者 A に[到着時間]が「10 分」であることを提示するガイダンスが発話される。話者 A がその提示内容を了承することで、「10分後にタクシー2台が到着する」というタクシー予約のタスクが達成される。

具体的には、対話は以下のように進められる。対話の開始状態において対話状況ラベルには「手続き開始」が与えられているものとする。対話エージェント A は『タクシーの手配をいたしますか?』というガイダンスを話者 A に発話す

る。「タクシーを 2 台 お願いします」と話者 A が応答し、対話エージェント A は、この発話の音声認識結果から[台数]のキーワードクラスに含まれる「2 台」を具体的な値として獲得して、対話状況「台数確認」に遷移する。対話エージェント A は、対話状況「台数確認」におけるガイダンスの定型文に「2 台」を埋め込み、『2 台でよろしいですか?』と話者 A に発話する。

ここで、「2 台」を「1 台」と誤って音声認識していたとすると、対話エージェント A は『1 台でよろしいですか?』と発話する。それに対し、話者 A が「いえ、2 台です」と台数を訂正する発話をすれば、対話状況「台数確認」におけるキーワードクラス[台数]の具体的な値が「1 台」から「2 台」に更新される。この更新が正しいかどうかを確認するために、対話状況「台数確認」において再度、『2 台でよろしいですか?』というガイダンスが行われる。このように、話者の発話内容の確認を行うことで、音声認識誤りが発生しても訂正が行え、正しいキーワードの獲得が実現できる。

対話状況「台数確認」に対し、話者 A が「はい、お願いします」と発話し、対話エージェントが[肯定]というキーワードクラスを音声認識結果から獲得できれば、表 1 にしたがって、キーワードクラス[台数]とその具体的な値として「2台」、および、遷移先対話状況「台数提示」という情報が仲介サーバを経て対話エージェント B に伝達される。

仲介サーバでは、対話エージェント A から対話エージェント B へキーワードを伝達する際に、日英翻訳を行い、逆方向では英日翻訳を行う。「2台」というキーワードクラス[台数]の具体的な値は「two cars」と日英翻訳され、対話エージェント B へ伝達される。

表 2 に、対話エージェント B で利用される対話シナリオ例を示す。対話エージェント B は、表 2 の対話シナリオの対話状況「台数提示」の記述にそって、ガイダンスの[台数]に「two cars」を埋め込んで『The customer is requesting **two cars**.』と、話者 B に発話する。この例のように、機械翻訳はキーワードのみに対して行われ、話者への提示は、定型文にキーワードを埋め込んで行われる。これにより、重大な機械翻訳誤りを回避することができる。

表 2. 対話シナリオ例  
(タクシー手配手続きの一部)

対話状況ラベル	ガイダンス	キーワードクラス	遷移先の対話状況ラベル	仲介サーバへ送信する情報
台数提示	The customer is requesting [台数].		時間確認	

#### 4. まとめ

対話エージェントを利用した異言語間コミュニケーションの手法を提案した。提案手法により、タスク達成対話を異言語間で行うことを可

能にするプロトタイプシステムを構築した。提案手法では、対話エージェントによる対話の制御と発話内容の確認により、タスクに重要なキーワードの音声認識誤りが回避できる。また、対話エージェントが、翻訳したキーワードをガイダンスに埋め込んで利用者に発話することにより、翻訳誤りを回避できる。これらにより、異なる言語を使う話者同士のタスク達成における意思疎通を確実にできる。

今後の課題としては、さまざまなタスクへの提案手法の適用、提案手法の評価などが挙げられる。

#### 参考文献

- [1] Eiichiro Sumita, Setsuo Yamada, Kazuhide Yamamoto, Michael Paul, Hideki Kashioka, Kai Ishikawa and Satoshi Shirai. "Solutions to Problems Inherent in Spoken-language Translation: The ATR-MATRIX Approach." In Proceedings of MT Summit '99, pp. 229--235, 1999.
- [2] Kiyoshi Yamabana, Ken Hanazawa, Ryosuke Isotani, Seiya Osada, Akitoshi Okumura, Takao Watanabe. "A Speech Translation System with Mobile Wireless Clients" In ACL-03 Companion Volume to the Proceedings of the Conference, pp.133 136, 2003.
- [3] Chengqing Zong, Bo Xu, Taiyi Huang. "Interactive Chinese-to-English Speech Translation Based on Dialogue Management" In Proceedings of the Workshop on Speech-to-Speech Translation: Algorithms and Systems, pp.61 68, 2002.
- [4] Mark Seligman. "Interactive Real-time Translation via the Internet." In Working Notes, Natural Language Processing for the World Wide Web. AAAI-97 Spring Symposium, vol.5, pp.142--148, 1997.
- [5] 甘粕哲郎, 山本俊一郎, 小川厚徳, 篠崎翼, 平沢純一, 山口義和 "サイバリアテンダント - 自由発話入力に対応したマルチモーダル対話システム", 日本音響学会 春季講演論文集, 3-Q-32, 2004-4.
- [6] 平沢純一, 山本俊一郎, 堀貴明, 大附克年 "CTI 向け自由発話対応 音声対話システム RexDialog", 情報処理学会研究報告, 2003-SLP-47, pp.35--40, 2003.