

日本語 WikiTalk の開発

Nihongo WikiTalk: Wikipedia-based information access dialogue

小此木 謙一
Kenichi OkonogiGraham Wilcock†
Graham Wilcock山本誠一†
Seiichi Yamamoto

1. まえがき

本稿では、多言語 WikiTalk の日本語対応について述べる。WikiTalk は、Wikipedia からテキストを取得し音声に変換しユーザに伝える対話型情報提供システムである。WikiTalk は、英語を対象として 2012 年に作られ、その多言語化は英語とフィンランド語を対象に 2014 年に開発された。今般、日本語を対象とした多言語化を日本語 Wikipedia と Nao ロボットの日本語の音声合成と音声認識システムを利用して開発を行ったので、この開発について述べる。

以下、2 では、英語 WikiTalk について 3 では多言語 WikiTalk について説明する。英語 WikiTalk と多言語 WikiTalk については、[1]を基に説明すると共に日本語 WikiTalk の開発について述べる。4 は本稿のまとめと今後の課題である。

2. WikiTalk

WikiTalk[2][3]は、Wikipedia に記述されているトピックに関して、音声対話により情報提供を行う対話型情報検索である。WikiTalk は、4 つの機能(Contact, Perception, Understanding, and Reaction)を利用してトピック管理と現在の情報、会話中のシェアされた知識の構成を統合する対話モデルフレームワークに基づいて構成されている[4]。検索されたトピックから、①リンクが貼られている関連トピックに飛ぶ、②発話中のトピックの前の段落に戻る、③発話し終えた段落を繰り返す、ことによりトピックについてより知る機能を有している。ユーザ発信のトピックはもちろん、WikiTalk が Wikipedia の”Did you know?”を利用して新しいトピックを提案する機能も有している。

初期 WikiTalk は、PyRobot (Python Robotics)ソフトウェアと Windows Speech Engine を利用して 2011 年に G.Wilcock によって開発された[3][5]。更に、Nao ロボットを利用したマルチモーダルな振る舞いが出来るようにを拡張された形態の WikiTalk[1][6]は、2012 年にフランスのメッツで行われた 8th International Summer Workshop on Multimodal Interfaces (eINTERFACE 2012)で公開された。

WikiTalk のマルチモーダルな振る舞いは、ロボットの発話に特別なコミュニケーションファンクションのためのジェスチャを統合、同期することで構成されている[7]。また、画像認識でジェスチャを認識し、ソナーセンサを利用して近接検出することでノンバーバル情報を認識する様に設定されている[8]。

マルチモーダル対話のための Nao への対話モデルの実装を Figure1 に示す。

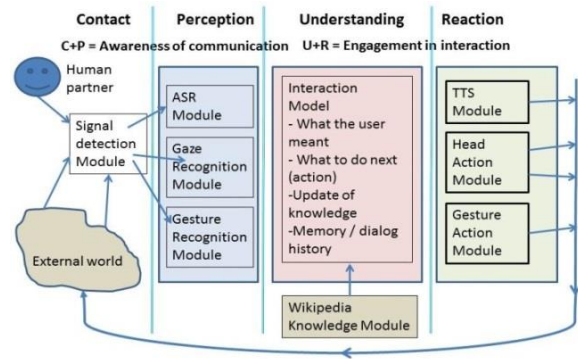


Figure1: Constructive Dialogue Modelling in human-robot interaction, from [9].

2.1 スムーズトピックシフト

WikiTalk にインタラクティブな対話を可能とする一つの要素としてスムーズトピックシフトがある。ユーザが次に何を話したいかを予測してスムーズにトピックをシフトさせる。スムーズトピックシフトをサポートするメカニズムは[2]で詳述されているが、WikiTalk での対話は、状態遷移ネットワークで管理されている。すなわち、①新しいトピックを選択、②Wikipedia の記事をダウンロードして段落ごとに区切る、③記事内のハイパーリンクを抜き出し保存し、スムーズトピックシフトのための情報として利用される。抜き出されたリンクは NewInfos に記憶され、WikiTalk は、ユーザがトピックシフトする遷移先の予測情報として利用される。

最初の段落が選択されると、一つもしくはそれ以上の段落が continueTopic 状態として設定され発話される。continueTopic 状態の制御によってユーザは、①同じトピックについてより情報を得る、②関連するトピックにシフト、③全く新しいトピックに移行を選択したり、又は対話を終わらせることができる。最新の NewInfos の中の一つを話題を選択することによって、ユーザはいつでもスムーズトピックシフトをすることができる。その場合、選ばれた新しい NewInfo は即座に新しいトピックに変更、若しくは、startNewTopic ステイトに戻り、新しいトピック記事をダウンロードして、新しい Newinfos を構築する。そして新しい記事を読み始める。

何回でも繰り返しスムーズトピックシフトは行える。そのため、対話はトピックのつながりの長さだけ行うことができる。また、ユーザは Wikipedia の記事を自由に探索できる。

3 WikiTalk の多言語化

Wikipedia には多言語の記事が存在する。それは、多言語 Wikipedia システムの開発の可能性を意味している。ただし、音声合成や音声認識等のシステムは前もって準備する必要がある。Nao Robot は、Aldebaran によってサポートされている言語を利用でき、その言語は当初 8 種から現在

†同志社大学

‡University of Helsinki と CDM Interact

19種に増加した。(英語, フランス語, 日本語, スペイン語など)

この状況を利用して最初の WikiTalk の多言語化はフィンランド語に対して行われた。

3.1 言語間の差異の解決

WikiTalk の多言語化には音声合成や音声認識機能に加えて全く異なる2つの処理が必要となる。まず、文と段落を Wikipedia 記事から抽出する必要があるが、これらの文や段落の構成法は、それぞれの言語で異なるため、それぞれ対応した言語の記事をダウンロードする。それぞれの言語で文の書き方が異なるため複雑さがある。しかし、ほとんどの問題は、ユニコードをエンコードすることで解決した。しかし、①言語の違い②文字の違い③記事の内容の違い④文書の構成やフォーマットの違いが存在し、多言語 WikiTalk を開発する際の問題となる。

次に、それぞれの言語で固有の言い回しを対話の流れの中で使う。ユーザコマンドで使う "continue", "repeat", "enough" や Nao の発話を遮ってしまった時の "Oh, sorry!", 複数文から構成される "I cannot access Wikipedia. Please check the Internet connection." など、それぞれの言語に対応させる必要がある。

3.2 日本語 WikiTalk

日本語 WikiTalk は、日本語 Wikipedia と Nao ロボットの日本語の音声合成と音声認識システムを利用して開発を行った。

WikiTalk ではロボットが発話する時、一文ごと音声に変換し発話する。英語やフィンランド語の場合、Natural Language Toolkit(以下 NLTK)を利用して文単位で区切っている。日本語 WikiTalk でも NLTK を利用して、句点ごとに文を区切るように開発した。また、Wikipedia の記事から関連するトピックへのハイパーリンクの情報を抽出する際、英語では、ハイパーリンクがホワイトスペースの間にあるため単語毎に区切って探索するが、日本語の場合、連続する文字列の中にハイパーリンクが存在するので、文毎に探索を行っている。

日本語 WikiTalk におけるユーザコマンドは、英語 WikiTalk におけるユーザコマンドを直訳するのではなく、口語表現になるように訳した。例えば、"continue" を直訳すると「続ける」となってしまう口語表現として違和感があるので「続けて」と訳した。また、多言語に対応させるため、ユーザが使用したい WikiTalk の言語を Nao に発話することによって WikiTalk の使用言語を日本語と英語で変更できるようにした。日本語から英語に変更する時は "English", 英語から日本語に変換する時は「日本語」と発話することで言語を切り替える。

WikiTalk は Wikipedia の "Did you know?" を利用して新しいトピックを提案する機能も有しているが、日本語 Wikipedia には、"Did you know?" という項目が存在していない。そこで日本語 WikiTalk の場合は、「新しい記事」という項目を利用した。

4 まとめと今後の課題

多言語 WikiTalk の日本語対応を目的として日本語 Wikipedia からの情報を取得し、ヒト型ロボット Nao とインタラクティブに対話しながら情報検索を行う WikiTalk システムを開発した。

今後の課題として以下の2つが想定される。Wikipedia は、文語表現であるため、Nao が発話した際、発話の区切りが

おかしく違和感がある場合がある。また、WikiTalk では、Wikipedia のテキストを発話しているため、同音異義語や、人物の漢字表記の説明などで上手く伝わらないことが想定される。

形態素解析を行い発話の区切りを適切に行うこと、漢字の説明を行えるようにすることが必要である。

今後、開発した日本語 WikiTalk を利用して情報検索を行った際のユーザの理解に関する評価実験を行う予定である。

5 参考文献

- [1] G. Wilcock and K. Jokinen (2014). Advances in Wikipedia-based Interaction with Robots. Proceedings of the Workshop on Multimodal Multiparty Human-Robot Interaction, International Conference on Multimodal Interaction (ICMI 2014), Istanbul, 2014.
- [2] G. Wilcock (2012). WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains, pages 57-69, International Conference on Computational Linguistics (COLING 2012), Mumbai, 2012.
- [3] K. Jokinen (2009). Constructive Dialogue Modelling: Speech Interaction and Rational Agents. John Wiley & Sons, 2009.
- [4] K. Jokinen and G. Wilcock (2012). Constructive interaction for talking about interesting topics. Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, 2012.
- [5] G. Wilcock and K. Jokinen (2011). Adding Speech to a Robotics Simulator. Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011), Budapest, 2011.
- [6] A. Csapo, E. Gilmartin, J. Grizou, J. Han, R. Meena, D. Anastasiou, K. Jokinen, and G. Wilcock (2012). Multimodal conversational interaction with a humanoid robot. Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), pages 667-672, Kosice, 2012.
- [7] R. Meena, K. Jokinen, and G. Wilcock (2012). Integration of gestures and speech in human-robot interaction. Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), pages 673-678, Kosice, 2012.
- [8] J. Han, N. Campbell, K. Jokinen, and G. Wilcock (2012). Investigating the use of non-verbal cues in human-robot interaction with a Nao robot. Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), pages 679-683, Kosice, 2012.
- [9] K. Jokinen and G. Wilcock. Multimodal open-domain conversations with the Nao robot. In J. Mariani, S. Rosset, M. Garnier-Rizet, and L. Devillers (editors). Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice, pages 213-224. Springer, 2014.
- [10] N. Laxström, K. Jokinen and G. Wilcock (2014). Situated Interaction with Multilingual WikiTalk. Proceedings of Fifth International Workshop on Spoken Dialogue Systems (IWSDS 2014), Napa, California, 2014.