

言語グリッドを用いた医療用例対訳の収集

岩部 正明¹ 田淵 裕章² 村上 陽平³ 重野 亜久里⁴ 石田 亨¹
北村 泰彦² 河原 達也⁵ 吉野 孝⁶ 津村 宏⁷

京都大学大学院 情報学研究科 社会情報学専攻¹ 関西学院大学 理工学部 情報科学科²
独立行政法人 情報通信研究機構 言語グリッドプロジェクト³
特定非営利活動法人 多文化共生センター・きょうと⁴
京都大学 学術情報メディアセンター⁵
和歌山大学 システム工学部 デザイン情報学科⁶
東京医療保健大学 医療保健学部 医療情報学科⁷

1. はじめに

医療においては些細なミスが生命に影響を及ぼすため、外国人患者の受診の際には精度の高い翻訳が必要である。現在は、NPO などにより派遣される医療通訳ボランティアが外国人患者の支援を行っているが、人材不足から多くの外国人患者を支援するには至っていない。そこで本研究では、機械翻訳よりも精度の保証された用例対訳を蓄積し、それを患者自身が自由に検索して利用できる用例対訳収集システムを開発することで、外国人患者の受診を支援する。しかし、医療分野における専門的かつ多言語の用例対訳を収集・利用する際には、以下の問題を解決しなければならない。

用例対訳の蓄積コスト 医療分野では、専門的かつ多言語の膨大な用例対訳が必要となる。しかし、一部の人間の力のみですべてを蓄積することはできない。

用例対訳の相互運用性 膨大な用例対訳の蓄積には、既存の用例対訳も利用できることが望ましい。しかし、Web 上に様々なフォーマットで分散して存在する用例対訳を統合的に利用することは困難である。

用例対訳の検索コスト 医療用例対訳のユーザは、計算機に関する専門的知識を持たない患者である。このようなユーザにとって、膨大な用例対訳の中から必要なものを検索するコストは非常に大きい。

Collecting Medical Parallel Texts Using Language Grid

^{*1}Masaaki Iwabu, Toru Ishida. Department of Social Informatics, Kyoto University

^{*2}Hiroaki Tabuchi, Yasuhiko Kitamura. Department of Informatics, Kwansai Gakuin University

^{*3}Yohei Murakami. Language Grid Project, National Institute of Information and Communications Technology

^{*4}Aguri Shigeno. Center for Multicultural Information and Assistance, Kyoto

^{*5}Tatsuya Kawahara. Academic Center for Computing and Media Studies, Kyoto University

^{*6}Takashi Yoshino. Department of Design and Information Sciences, Wakayama University

^{*7}Hiroshi Tsumura. Division of Healthcare Informatics, Tokyo Health Care University

本稿では、用例対訳収集システムに求められる要件を定義した後、システムの設計について述べる。本システムでは 20ヶ国語 × 2,000 用例以上を収集し、実際に外国人患者の受診支援につなげることを目標とする。

2. 用例対訳収集システムの要件定義

1. で挙げた 3 つの問題点に対処するために、用例対訳収集システムには以下の 3 点が必要となる。

Web ベースの用例対訳収集機能 用例対訳は極めて膨大であり、その蓄積にあらゆる人が参画できる枠組みが必要である。Web のアプローチは、これに適している。実際、Wikipedia や青空文庫では、Web 上での多数の人々の協創により、膨大な知識体系の構築に成功している。本システムでも同様に、Web ベースのインタフェースを採用する。これにより、用例対訳の蓄積に要するコストが分散され、なおかつ集合知の形成が容易になる。

用例対訳データのラッピング機能 用例対訳は Web 上に多様なフォーマットでデータとして存在するが、本システムで作成される用例対訳も含めて、すべてを統合的に利用できることが望ましい。そこで、用例対訳データを共通のインタフェースに基づいて Web サービス化する。さらに、用例検索 Web サービスの連携基盤として言語グリッド [1] を使い、Web サービス連携を用いてより充実した用例対訳の利用を行う。

音声による検索機能 あらゆるユーザが膨大な用例対訳から必要なものを容易に検索できるように、本システムに音声インタフェースを導入する。入力された音声は音声認識エンジンにより文字列に変換され、その文字列を検索クエリとして利用する。本システムでは、収集された用例対訳を音声認識エンジン Julius^{*8}[2] 用の学習データとして与えるこ

^{*8}<http://julius.sourceforge.jp/>

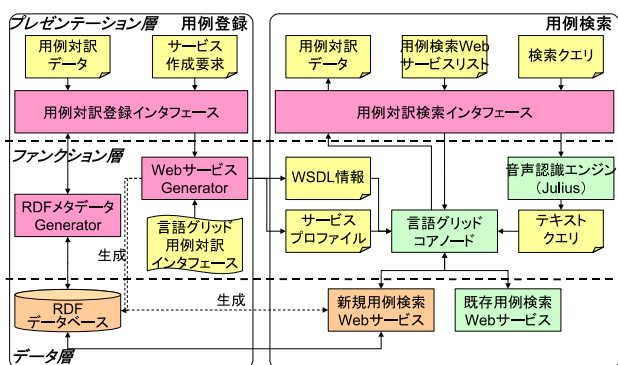


図 1: システムアーキテクチャ

とで、特定のドメインに特化した高性能の音声認識エンジンを実現する。

3. 用例対訳収集システムの設計

2. で言及した要件を踏まえて設計した用例対訳収集システムの構成および動作は、以下のとおりである。

システムアーキテクチャ 本システムの主要な機能である用例対訳の登録と検索は、図 1 に示すデータ層・ファンクション層・プレゼンテーション層の 3 層からなるアーキテクチャで実現する。個々の機能は、以下のとおりである。

1. 用例対訳登録・検索インタフェース：ユーザーに、用例検索 Web サービスの作成および用例対訳の登録・検索に必要なインタフェースを提供する。検索インタフェースとして、テキストと音声の双方を提供する。
2. Web サービス Generator：サービス作成要求に基づいた用例検索 Web サービスと用例対訳を格納する RDF データベースを生成し、Web サービスを言語グリッドに登録する。
3. RDF メタデータ Generator：ユーザーの入力した用例対訳データを Web サービス Generator が作成した RDF データベースに登録できるように、必要なメタデータを自動的に付与する。
4. 言語グリッドコアノード：用例検索 Web サービスに関する情報を管理する。また、用例検索 Web サービスを連携して実行し、結果をまとめて用例対訳検索インタフェースに返す。

動作 本システムにおいて、用例対訳の登録および検索は以下のように実現される。

1. 用例対訳の登録：Web サービスの作成要求に従い、Web サービス Generator が用例検索

Web サービスおよび RDF データベースを作成する。用例を追加すると、RDF メタデータ Generator がメタデータを付与して RDF データベースに格納する。用例は別のユーザーに検索されて対訳が追加され、同様に RDF データベースに格納される。

2. 用例対訳の検索：音声入力されたクエリは音声認識エンジンによってテキストに変換される。変換後のテキストクエリを用いて、言語グリッドコアノードは複数の用例検索 Web サービスを呼び出す。すべての用例検索 Web サービスからクエリに適合する用例対訳が返されると、それらをまとめてユーザーに提示する。

4. まとめ

外国人患者の受診を支援する上で、医療分野の用例対訳を蓄積して利用することが有効である。本稿では、円滑な用例対訳の収集と利用を支援するために、以下の 3 点を実現する用例対訳収集システムを提案した。

Web ベースの用例対訳収集 専門的かつ多言語にわたる膨大な用例対訳の収集に多数の人々が関与できるように、Web ベースのシステムを構築する。

Web サービス技術の適用 用例対訳を Web サービスとして提供する。さらに、言語グリッドを用いて Web サービスを連携させ、Web 上に分散した用例対訳の統合的な利用を実現する。

音声インタフェースの導入 テキストによる用例検索に加えて、音声による用例検索を導入する。

すでに、医療分野に関しては 20ヶ国語 × 2,000 用例の収集に向けて、本システムに日英中韓葡の 5ヶ国語の用例が蓄積され始めている^{*9}。ただし、本システムは医療分野に特化したものではなく、集合知の形成および利用を目的とした汎用的な仕組みを提供するものである。本システムの構築によって、医療に代表されるさまざまな分野で多言語の知識体系の蓄積が進むという社会的インパクトが期待できる。

参考文献

- [1] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. In *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96–100, 2006.
- [2] 河原達也, 李晃伸. 連続音声認識ソフトウェア Julius. *人工知能学会誌*, Vol. 20, No. 1, pp. 41–49, 2005.

*9 <http://medical.langrid.org/collaboration/>