

言語グリッドを用いた留学生向け教材多言語化支援システムの提案

北川 大輔[†] 木野村 晋吾[‡] 桑原 和宏[†]

[†]立命館大学 情報理工学部

[‡]立命館大学大学院 理工学研究科

1 はじめに

近年の国際化の流れによって、日本へ来る留学生が増えている。それに従い、外国語での講義を行う大学も増えている。授業で配付するレジュメや表示するスライドも日本語ではなく外国語で行う場合もあり得る。

日本語で作成した講義資料を外国語へ翻訳する際、手軽に低コストで翻訳するには、機械翻訳が適している。しかし、機械翻訳を利用すると、専門用語が正しく訳せなかったり、意味の通らない文になってしまう可能性があるため、受講生の学習を妨げる可能性がある。予め訳しておいた対訳を用いて翻訳する、用例対訳を用いた翻訳を行えば、機械翻訳を上回る精度で翻訳が可能である。しかし、用例対訳を人手で作る手間がかかり、かつ用例対訳を検索する際に完全一致を条件に検索した場合、言い換えなどの表記揺れに対応できない。このため、用例対訳を用いた翻訳を積極的に利用する場合、訳文の準備を行う手間や費用がかかってしまう。

本稿では、HTML で記述された教材を対象として、翻訳する部分を簡単に指定できる仕組みを導入するとともに、ユーザが入力した修正訳文から用例対訳を自動的に集めることができる拡張機能を提案する。対訳の用例を集めることにより翻訳の精度を高めつつ、教材の多言語化を容易にすることが本システムの目的である。

2 提案システム

2.1 機能

多言語化を意識した教材であると、翻訳してほしい部分とそうでない部分がある。そのため、教材の文章すべてを翻訳するのではなく、指示された部分のみを翻訳するような仕組みを導入した。具体的な手法としては、HTML 化された教材を前提に、HTML の class

“Language Grid enabled multilingualization support of educational materials for international students”
Daisuke KITAGAWA[†], Shingo KINOMURA[‡], and Kazuhiro KUWABARA[†]

[†]College of Information Science and Engineering, Ritsumeikan University

[‡]Graduate School of Science and Engineering, Ritsumeikan University

HTML の各要素に対して、イベントリスナーを設定することができる。以下にイベントとイベントリスナーの動作確認用のエリアを設けた。この例ではイベントの種類とイベントに含まれるマウスの座標を、イベントリスナーが表示する。ここでは、window と canvas 要素にそれぞれイベントリスナーを設定している。canvas タグはHTML5で導入される図形描画用のタグである。 [翻訳](#)

可以对HTML的各要素设定监听项。下面设置了事件与事件监听项的动作确认框。在这个例子里，事件监听项显示事件的种类和事件中包含的鼠标坐标。在这里 window 和 canvas 中分别设定了事件监听项。canvas 标记是用HTML5导入的描画图形用标记。

[翻訳の間違いを報告](#)

図 1: 翻訳例

属性として“translatable”を指定した部分を翻訳の対象と識別する。HTML の class 属性として記述することにより、HTML 要素との親和性が高まり、教材閲覧者に対して、「翻訳ボタン」「翻訳の間違いを報告ボタン」などを配置することも容易になるという副次的な効果も生まれる。

本システムは JavaScript のライブラリとして提供される。これにより、教材作成者は、教材の HTML 文章中に、JavaScript のライブラリへのリンクを記述するだけで本システムの機能を利用することができる。また、教材閲覧者はシステムが文末に自動的に付加したボタンをクリックすることにより、図 1 のように文章の翻訳や、原文と機械翻訳文を見比べながら適切な訳を入力し、送信することで図 2 のように誤訳の報告ができる機能を実装した。この機能を提供することにより、本システムでは、ユーザから修正訳文を集める。本システムではユーザが報告した正しい訳を蓄積し、次回以降の翻訳において報告された訳を用いることで、より精度の高い翻訳文をユーザに提供する。

2.2 システム構成

本システムは「翻訳サーバ」と「用例対訳サービス」「辞書サービス」から成る(図 3)。また、機械翻訳のサービスとして言語グリッド [1] が提供するサービスを使用している。言語グリッドの翻訳サービスには、辞書サービスが提供されている URI を指定して翻訳を行うと、原文中の該当単語について自動的に辞書サービスを用いて翻訳を行う機能がある。この機能を使用すると、辞書サービスを用いない場合と比べ精度の高い翻訳文 [2] を生成することができ、専門用語を含む文でも、辞書を用いない場合に比べ品質の高い翻訳を行

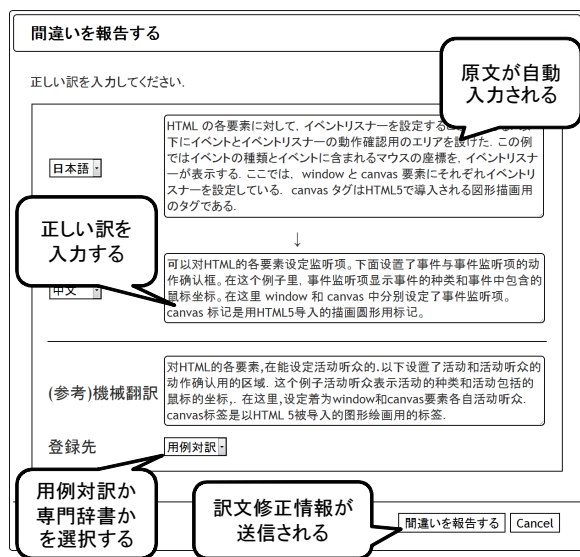


図 2: 訳文修正例

うことが可能となる。

2.3 処理の流れ

「翻訳サーバ」には、クライアントから送信された原文を翻訳し、訳文として返す機能と、クライアントから送られる訳文修正情報を処理する機能の2つがある。

訳文の生成は以下のように行われる。クライアントの JavaScript プログラムが HTML 文章中の translatable クラス属性の文を翻訳サーバに送信する。翻訳サーバは、原文を用例対訳サーバに送信し、用例対訳データベース中に原文が存在するかどうかを検索する。用例対訳が見つければ、翻訳サーバはクライアントに対して用例対訳を訳文として返す。用例対訳が見つからなければ、翻訳サーバは言語グリッド上の翻訳サービスを利用して辞書サーバを用いて機械翻訳を行い、生成された文章を訳文として返す。

また、クライアントから送られてくる訳文修正情報は次のように処理される。まず、ユーザが翻訳結果に不備を見つけた際に、「翻訳の間違いを報告ボタン」を押して表示されるダイアログに正しい訳を入力すると、原文と訳文及びそれぞれの言語、登録対象が文章であることをセットとした情報が作成される。これを本稿では訳文修正情報と呼ぶ。また、ユーザが専門用語を登録する際は、登録対象が単語である情報と前述の情報を訳文修正情報として扱う。ユーザが翻訳サーバに訳文修正情報を送信すると、翻訳サーバは受け取った登録対象が単語であるなら専門辞書データベースへ、文章であれば用例対訳データベースへ送信する。

辞書サーバが単語単位での訳を保存、利用するためのサーバであるのに対し、用例対訳サーバは文節や文単位での訳を保存、利用するためのサーバである。ク

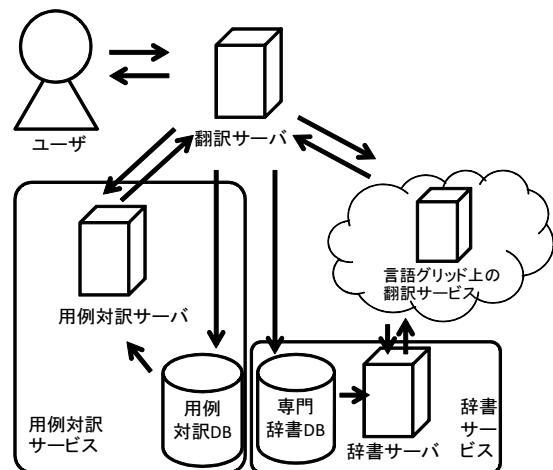


図 3: 本システム構成図

ライアントからの訳文修正情報のうち、専門辞書データベースに登録された単語は辞書サーバを通じて機械翻訳時に利用され、用例対訳データベースに登録された用例対訳は用例対訳サーバを通じて検索要求があった際に使用される。

翻訳の際には、翻訳サーバがクライアントに対する代表となって、用例対訳サービスや翻訳サービスにアクセスして最も適切な訳を返すことで、精度の高い翻訳文を得ることが可能である。

3 まとめ

本稿では、言語グリッドを用いて HTML 化された講義資料の多言語対応を支援するシステムを提案した。本システムでは、JavaScript ファイルをインポートし、特定の class 属性を与えることにより、講義資料を多言語化できる。また、ユーザから報告された修正訳を用例対訳データベース等に登録する機能を設けることで、より翻訳精度を高めることを可能としている。

本提案手法では、教材が HTML 文書の場合に多言語化が可能である。留学生の学習支援という観点から、教材の多言語化のみならず、例えば学習内容について討議する掲示板の多言語化にも適用していきたい。

参考文献

- [1] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *Proc. IEEE/IPSJ Symposium on Applications and the Internet(SAINT-06)*, pp.96-100 (2006).
- [2] 西山七絵, 村上仁一, 徳久雅人, 池原悟: 単文文型パターン辞書の構築, *言語処理学会第 11 回年次大会発表論文集*, Vol. 11, pp. 372 - 375 (2005).