

多言語対訳用語集における付加情報の活用

村上 知弘[†] 油井 貴之[†] 桑原 和宏[‡]

[†]立命館大学大学院 理工学研究科

[‡]立命館大学 情報理工学部

1 はじめに

Google 翻訳などのインターネット機械翻訳サービスには専門用語などの日常に利用されない用語については必ずしも適切な訳語が選択されないという問題がある。これに対して多言語対訳用語集を用いた機械翻訳の精度向上の試みがされている [1]。多言語対訳用語集には機械翻訳ではカバーできない専門的な用語の対訳が入っており、これを用いることでよりの確な訳語を選択することができる。

この多言語対訳用語集には対訳の他に、言語間の違いを説明する為に多言語対訳用語集を作成した専門家が必要と判断して付けた付加情報が存在することがある。これは人手による翻訳においては有用であるものの、インターネット機械翻訳サービスにおいては利用されておらず付加情報が活かされていない。

そこで本論文では付加情報を機械翻訳サービスと組み合わせるよう、付加情報付き多言語対訳用語集の Web サービスを提案するとともに、多言語文章作成における訳語の選択支援への適用例を述べる。

2 付加情報

2.1 付加情報とは

多言語対訳用語集の一例を図 1 に示す。この多言語対訳用語集は多言語・学校プロジェクト (<http://www.tagengo-gakko.jp/>) が教育現場で使用される用語についてまとめたものである。図 1 において付加情報とは、縁なしやスペイン、北部などの大括弧で囲まれた箇所を指す。付加情報は同じ言語で同じ意味を持つが、国や地域などにより異なる 2 用語が存在する時にその 2 用語を区別する為に利用される。例えば、図 1 の「crayolas {ペルー} /crayon {スペイン}」は、日本語のクレヨンに対する訳語としてペルーでは crayolas、スペインでは crayon が用いられることを示している。人

手による翻訳では、ペルー向けの翻訳とスペイン向けの翻訳を区別して作成でき、より適切な翻訳を実現できる。

日本語	英語	スペイン語	ベトナム語
帽子	hat	gorro {縁なし} /sombbrero {縁あり}	mũ {北部} / nón {南部}
新入生	bagong mag-aaral	alumnos nuevos {小・中} /estudiantes nuevos {高}	học sinh mới nhập học
クレヨン	crayons	crayolas {ペルー} /crayón {スペイン}	sáp màu

図 1: 多言語対訳用語集の一例

2.2 機械翻訳との連携の課題

付加情報は人手による翻訳には有効だがインターネット機械翻訳サービスには利用されていない。その理由として機械翻訳では付加情報の種類を判断できないことが挙げられる。

例えば、付加情報を持つ用語を翻訳して表示させた場合、表示させる用語は、必ずしも一つだけに絞ればよいわけではない。前節で述べた用語集を用い、「新入生歓迎会を開きます」という連絡を翻訳したい時に、クラスにスペイン語を用いるペルー出身の子とスペイン出身の子がいる場合は、図 1 の新入生のスペイン語訳からわかるように小・中と高では用いる用語が異なるため、表示させる用語を状況によって一つに絞る必要がある。しかし、「クレヨンを持ってきてください」という連絡を翻訳したい時には、クレヨンのスペイン語訳からわかるように、クラスにペルー、スペイン出身の子がいる場合ではどちらも表示させる必要があるのである。つまり、付加情報の活用には次の 2 つが考えられる。

- 複数の用語のどちらか一つに絞らせる場合
- 複数の用語をすべて表示させる可能性もある場合

これらを区別するには付加情報の中身の意味を知る必要があるため人手による翻訳であれば翻訳家が付加

“Use of additional information in a multilingual glossary”

Tomohiro MURAKAMI[†], Takayuki YUI[†], and Kazuhiro KUWABARA[‡]

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

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

情報から判断できるが、機械翻訳においてはその判断は難しく適切な翻訳文の表示ができない。

3 付加情報の利用手段の提案

3.1 多言語対訳用語集の RDF 表現

付加情報の種類を区別するために RDF(Resource Description Framework) を使用して用語集を表現する。それぞれの言語の付加情報を持つ用語には Select, All という属性を定義した。Select 属性は 2.2 節で述べたどちらか一つに絞らせる付加情報, All 属性はすべて表示させる可能性もある付加情報を持つことを示す。これらを踏まえて構築した用語集の一例を図 2 に示す。

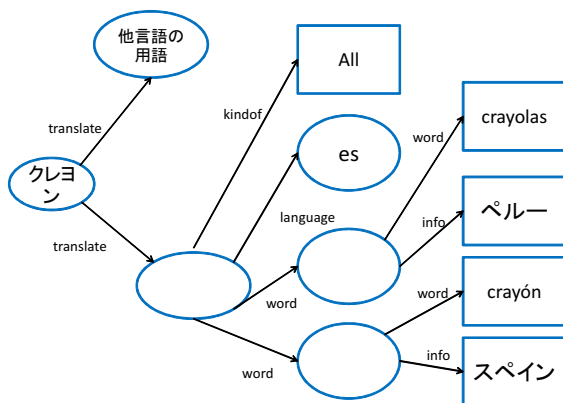


図 2: 付加情報を持つ用語の表現例

3.2 用語集の Web サービス化

用語集は翻訳の補助となるものであり、他の翻訳、多言語システムと組み合わせて利用されるものである。そこで用語集を Web サービスとして提供する。利用者が入力パラメータとして検索ワード、検索元言語、翻訳先言語などを付与した Web サービスへのリクエストを送る。サーバが入力パラメータを基に用語集検索を行い、レスポンスとして JSON で表現された検索ワードに対する付加情報を含む対訳を返却する。

3.3 付加情報の利用例

付加情報の利用例として、2.1 節で示した教育関連の付加情報付き多言語対訳用語集を機械翻訳と組み合わせることにより、多言語文章作成において訳語の選択支援に適用した。

このサービスは、Web ブラウザに日本語で入力された文章を機械翻訳により指定の言語に翻訳して表示するものである。システム構成図を図 3 に示す。処理の流れとしてはユーザーの文章翻訳要求に対してアプリケーションサーバは多言語サービスの基盤である言語グリッド [2] と 3.2 節で構築した Web サービスを用いて文章の翻訳を行う。アプリケーションサーバは付加情報を含んだ翻訳文を JSON で表現し、レスポンスデー

タとしてユーザー側に返す。ユーザー側はレスポンスデータを解析し翻訳された文章をユーザーに表示する。その際、付加情報を持たない用語については、そのまま用語集の対訳が使われて翻訳される。付加情報を持つ用語が含まれている場合は、用語部分が通常と異なる色と下線付きで表示される。ここで 3.1 節で述べた Select 属性を持つ用語は、その部分にカーソルを合わせると訳語の選択肢が表示される。例を図 3 の吹き出し部に示す。ユーザーは提示された訳語から自分の意図する意味の用語を選択する。選択結果は、翻訳文に自動的に反映される。一方、All 属性を持つ用語は、訳語を選択する必要がないので、カーソルを合わせると付加情報が表示される。

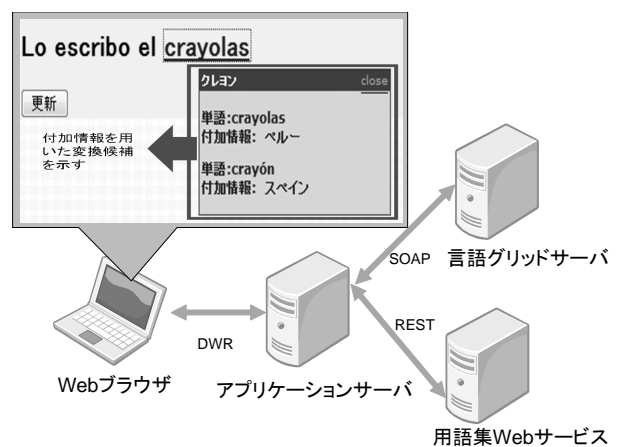


図 3: 付加情報付き用語集サービスの利用例

4 まとめと今後の展望

本論文では対訳用例集において訳語に付加された情報を活用する手法を提案し、多言語文章作成における訳語の選択支援に適用した。

今後、現在は単文ごとの翻訳になっているので、文脈情報を活用した複数文の文章翻訳を行い、より実用的な多言語文章支援サービスへ発展させていきたい。

謝辞

多言語対訳用語集を提供していただいた関西大学総合情報学部喜多千草教授に深く感謝いたします。

参考文献

- [1] Murata, T., Kitamura, M., Fukui, T., and Sukehiro, T.: Implementation of Collaborative Translation Environment 'Yakushite Net', *Proceedings of the 9th Machine Translation Summit System Presentation*, pp. 479-482 (2003).
- [2] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *Proceedings of the International Symposium on Applications on Internet (SAINT-2006)*, pp. 96-100 (2006).