

# 言語グリッド: 異文化コラボレーション支援環境の構築

## Language Grid: Development of Support Environment for Intercultural Collaboration

重信 智宏\*1      藤井 薫和\*2\*1      吉野 孝\*3\*1      灘本 明代\*1  
 Tomohiro Shigenobu      Kunikazu Fujii      Takashi Yoshino      Akiyo Nadamoto

\*1 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  
 Language Grid Project, National Institute of Information and Communications Technology

\*2 和歌山大学 システム工学研究科  
 Graduate School of Systems Engineering, Wakayama University

\*3 和歌山大学 システム工学部  
 Faculty of Systems Engineering, Wakayama University

To increase the mutual understanding of different cultures with different languages, it is essential to build a language infrastructure on top of the Internet that improves the accessibility and usability of existing online language services so that users can create new cross-language services for their communities. To realize this infrastructure, we propose the language grid. This approach can facilitate intercultural collaboration through the Internet. By combining several language services, we create a new service to support multilingual communication for various NPO's activity. Moreover, we have developed intercultural collaboration tools named Langrid Input, Langrid Blackboard and Langrid Chat.

### 1. はじめに

近年、多言語コミュニケーションを目的としたシステムの研究開発が行われており、メッセージの多言語化に機械翻訳が用いられている [Flournoy 01]。また、一般的なサービスとして誰でも利用可能な翻訳機能付き電子掲示板サービスも存在している\*1。しかし、それらの種類や言語にも依るが、現時点での翻訳精度は円滑なコミュニケーションに耐える程の品質に達していない。翻訳精度の問題もあるが、機械翻訳で使われる辞書データは、一般的な用語にのみ対応していることが多く、特定のドメインに依存する用語（造語など）が含まれる文章の翻訳精度は不十分である。そのため、あるコミュニティでのみ用いられる様な独自の用語を含む文章が与えられた場合、翻訳に失敗することが多い。このように、機械翻訳を多言語コミュニケーションへ応用するためには、そのドメインを考慮したドメイン指向の翻訳を実現する必要がある。

本稿では、機械翻訳を介した円滑なコミュニケーションを実現させるために、ユーザ自身が作成したコミュニティ辞書と既存の機械翻訳とを連携するドメイン指向翻訳の実現方法、および、それを利用したコラボレーションツールの開発について述べる

### 2. 言語グリッド

我々は、言語の壁を超えることを目的とした、異文化コラボレーションの基盤となる言語グリッドの構築に向けた研究開発を行っている [Ishida 06]。現在、インターネット上には、機械翻訳や形態素解析などの言語処理機能や対訳辞書などの言語資源が蓄積されつつある（以下、言語サービスと呼ぶ）。言語グリッドは、それらの言語サービスを容易に利用可能とし、さらに、ユーザ自らが新たな言語資源や言語処理機能を追加可能であり、それらを組み合わせることで新たな言語サービスを生み出すことが可能な言語基盤である。そこで、コミュニティユーザが独自に作成したドメイン依存の専門用語の対訳辞書（コ

\*1 enjoy Korea: <http://enjoykorea.jp/transboard/>

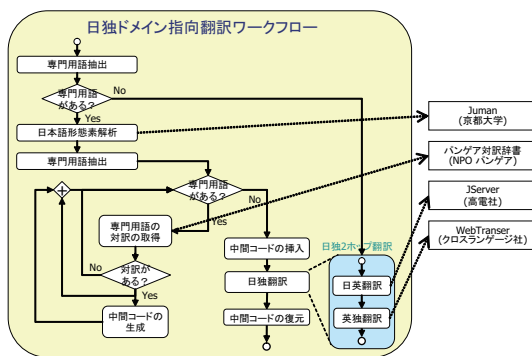


図 1: ドメイン指向翻訳ワークフロー

ミュニティ辞書)と既存の機械翻訳とを連携させることで、ドメイン指向翻訳を実現する。対訳辞書を作り翻訳精度を向上させるといった点では、ユーザ間の協調作業により翻訳精度の向上を可能としている機械翻訳サイト“訳してネット\*2”が存在する [Murata 03]。コミュニティ毎に辞書データの編集が可能であり、このサイトを利用することで、ユーザはより品質の高い翻訳結果を得ることができる。

### 3. ドメイン指向翻訳

ドメイン指向翻訳は、コミュニティ辞書と機械翻訳とを連携させた、そのコミュニティ向けに特化した翻訳サービスである。図 1 に NPO パンゲア向けに作成した日独翻訳に対応するドメイン指向翻訳のワークフローを示す。このワークフローで利用されている各処理コンポーネントは、機械翻訳、形態素解析、対訳辞書、その他の文字列処理の Web サービスからなる。

まず、入力文に含まれる専門用語が抽出される。常に専門用語が含まれているとは限らず、正確な抽出を行うと処理時間が増加するため、単純な文字列マッチングで粗く抽出する。もし、含まれていない場合は、そのまま翻訳される。含まれていた場合は、形態素解析の結果を用いて入力文に含まれる専門用語が正確に抽出する。抽出された専門用語に対応する対訳をコ

\*2 訳してねっと: <http://yakushite.net/>

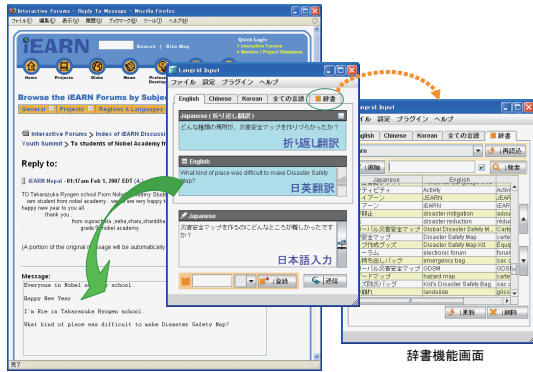


図 2: Langrid Input

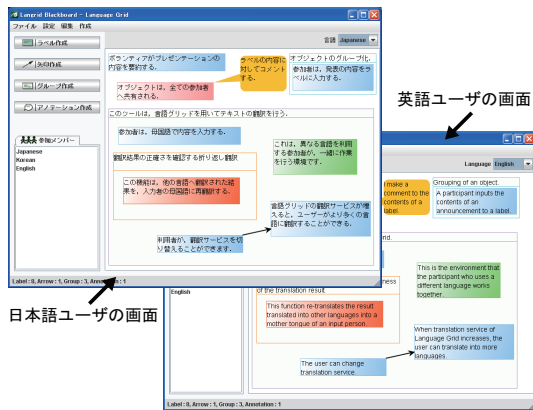


図 3: Langrid Blackboard

コミュニティ辞書より取得し、対訳に対応する中間コードを取得する。中間コードは、専門用語の置き換えに利用される機械翻訳を介しても変化しない文字列である。入力文内の専門用語を中間コードに置き換えてから翻訳を行うことで、翻訳後に対訳を置き換える場所を特定することができる。文字列置換サービスを利用して、入力文内の専門用語を中間コードに置き換える。その中間コードを含む入力文を翻訳し、翻訳結果に含まれる中間コードと対応する対訳とを置き換えることで、コミュニティ辞書を利用した翻訳が実現する。この仕組みを言語グリッドに登録し、ドメイン指向翻訳サービスとして汎用的に利用可能としている。

#### 4. 異文化コラボレーションツール

異文化コラボレーションツールとして、多言語入力支援ツール Langrid Input, 多言語共有黒板ツール Langrid Blackboard, 多言語対話支援ツール Langrid Chat を開発した。各ツールの概要を以下に示す。

Langrid Input: 翻訳文作成に特化したツールであり、Langrid Blackboard および Langrid Chat の共通入力環境でもある (図 2)。入力支援機能として母国語で翻訳精度を確認可能な“折り返し翻訳機能”を備える。また、タブ切り替えにより多言語を扱うことが可能である。単独で利用した場合、他のアプリケーションのテキストエリアのカーソル位置へ翻訳結果を挿入する機能を備える。現在、NPO JEARN において英語の BBS へ日本語から翻訳した英文を投稿する際に用いられている。また、コミュニティ辞書操作機能を備えており、ユーザは、その場で専門用語の新規追加や編集が行え、更新されたデータは即

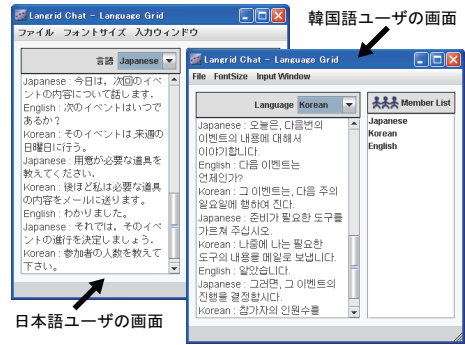


図 4: Langrid Chat

座に翻訳結果へ反映される。

Langrid Blackboard: 多言語のユーザ間で議論の内容を整理する用途などに利用可能なツールである (図 3)。共有ウィンドウを持ち、テキストデータを入力可能なラベルの作成やそれらのグループ化が可能である。共有ウィンドウ内のオブジェクトは、全ユーザ間で共有され、同じ位置に表示される。テキストは各ユーザの利用言語に応じて翻訳される。現在、京都大学の石田研究室で行われる研究会において、日本語の分からない中国人留学生に対して、発表内容を伝えるために利用されている。

Langrid Chat: 多言語対応のチャットツールであり、ドメイン指向翻訳サービスを利用したメッセージの翻訳機能を備える (図 4)。ユーザが母国語で入力したメッセージは、各ユーザの利用言語に翻訳され表示される。

#### 5. おわりに

言語グリッドにより実現するドメイン指向翻訳を備えた異文化コラボレーション支援環境を開発した。ドメイン指向翻訳の翻訳精度を評価したところ、理解度の向上を確認できた。しかし、通常の翻訳でも正確に翻訳される用語が専門用語として登録された場合、未知語である中間コードに置き換えられて翻訳されるため、品詞情報が欠落し前置詞が変化するなど、文法的な誤りが発生することがあった。また、コミュニティ辞書を作成する際、パイリングでなければ正確な対訳の作成は困難であるため、ユーザ間の協調作業による対訳データの作成を可能とする支援を行う予定である。

#### 参考文献

[Flournoy 01] Flournoy, S. R., and Callison-Burch, C.: Secondary Benefits of Feedback and User Interaction in Machine Translation Tools, Workshop paper for "MT2010: Towards a Roadmap for MT" of the MT Summit VIII (2001).

[Ishida 06] Ishida, T.: "Language Grid: An Infrastructure for Intercultural Collaboration", in proc. of IEEE / IPSJ Symposium on Applications and the Internet (SAINT06), pp.96-100, (2006).

[Murata 03] Murata, T., Kitamura, M., Fukui T., and Sukehiro, T.: "Implementation of Collaborative Translation Environment: Yakushite Net", the 9th Machine Translation Summit System Presentation, pp.479-482 (2003).