

# ドメイン指向翻訳を備えた 言語グリッドコミュニケーションツール

重信 智宏<sup>1</sup> 藤原 義功<sup>2</sup> 村上 陽平<sup>3</sup> 吉野 孝<sup>4</sup> 石田 亨<sup>5</sup>  
 情報通信研究機構<sup>1</sup> 情報通信研究機構<sup>3</sup> 和歌山大学<sup>4</sup> 京都大学<sup>5</sup>

## 1. はじめに

これまでに機械翻訳を利用した多言語コミュニケーションツールの開発が行われている[1]。しかし、機械翻訳の種類や利用言語にも依るが、現状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。単純な例としては、人名の翻訳は高い割合で失敗するため、英字で入力するユーザが多い。このように、機械翻訳を人間同士のコミュニケーションへ応用するためには、ドメイン指向の翻訳を実現する必要がある。

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[2]。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”<sup>1</sup>が存在する[3]。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。

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

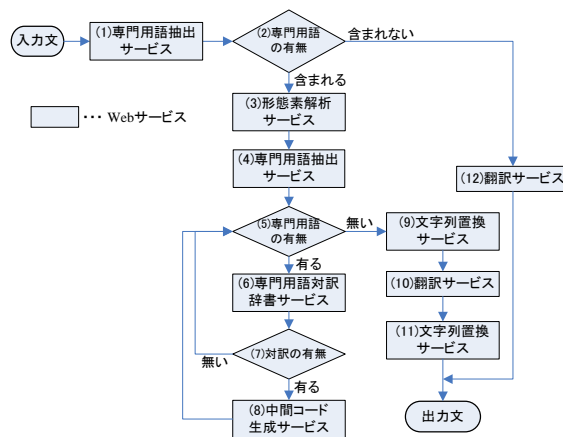


図1 ドメイン指向翻訳ワークフロー

訳の実現方法、および、それを利用した言語グリッドコミュニケーションツールの開発について述べる。

## 2. 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。

入力文に含まれる専門用語が抽出される(1)。常に専門用語が含まれているとは限らず、形態素解析の結果を用いて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(12)。含まれていた場合は、形態素解析による形態素と(1)の結果から入力文に含まれる専門用語が正確に抽出される(3, 4)。

抽出された専門用語に対応する対訳を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(6, 8)。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(9)。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

### Language Grid Communication Tools with Domain Oriented Translation

1. Tomohiro Shigenobu, National Institute of Information and Communications Technology.
2. Yoshinori Fujihara.
3. Yohei Murakami, National Institute of Information and Communications Technology.
4. Takashi Yoshino, Faculty of Systems Engineering, Wakayama University.
5. Toru Ishida, Department of Social Informatics, Kyoto University.

1 <http://yakushite.net/>  
 2 <http://www.pangaeon.org/>  
 3 <http://www.jearn.jp/japan/>

した翻訳が実現する(10, 11). 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. つまり, コミュニティ辞書を切り替えることで, 他のコミュニティへの応用も容易に可能であり, 現在, NPOパンゲア<sup>2</sup>およびNPO JEARN<sup>3</sup>において利用されている.

実際に, コミュニティ辞書を作成しドメイン指向翻訳と通常の翻訳との比較評価を行った. 評価方法はTIDES[4]を参考とし, 主観的評価による5段階評価により, 翻訳結果の正確さを確認した. 日英翻訳および日韓翻訳の評価を行った結果, どちらも理解度の改善が確認できた. 日英翻訳については平均3.0から平均3.8へ, 日韓翻訳については平均3.4から平均4.5へと改善した.

### 3. コミュニケーションツール

ドメイン指向翻訳を備えた言語グリッドコミュニケ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**: 入力文の作成に特化したツールであり, Langrid Blackboard および Langrid Chat の入力インタフェースとして共通利用される(図2). 入力支援機能として母国語で翻訳精度を確認することができる“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, 別アプリケーションの編集エリアのカーソル位置へ翻訳結果を挿入することが可能であり, JEARN において英語のBBSへ日本文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. このツールは現在, 日本語の分からない中国人留学生に対して, 所属研究室で行われる研究会の日本語の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. ユーザが母国語で入力したメッセージは,



図2 Langrid Input

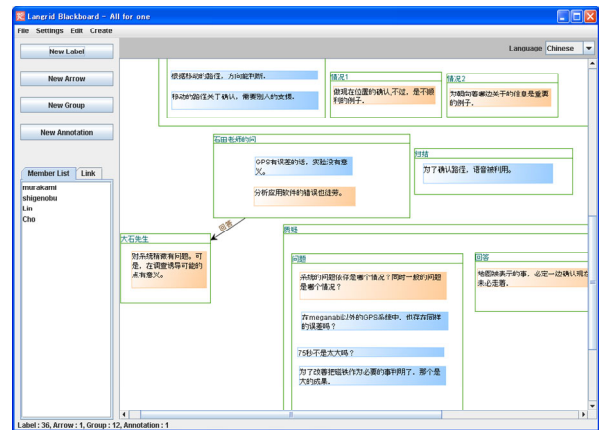


図3 Langrid Blackboard

各ユーザの利用言語に翻訳され表示される.

### 4. おわりに

コミュニティ辞書を連携させたドメイン指向翻訳を備えた言語グリッドコミュニケーションツールを開発した. ドメイン指向翻訳の評価により, 理解度の向上を確認でき有用であることが分かった. 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る. この問題に対応すると共に, 現在は名詞のみの対応であるため, 今後は動詞や形容詞などへの対応も検討する.

JEARN の利用状況を観察していると, Langrid Input を入力環境として利用するだけでなく, BBS の英文の翻訳にも利用していることが分かった. それらの英文には省略語(例えば, YOU を U と省略)が多数存在し, そのまま翻訳することはできないが, ドメイン指向翻訳に表記の揺れを校正する処理を追加することで対応するなど, 今後は, 様々な状況下において翻訳精度の向上が可能な仕組みが必要となる. また, バイリンガル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作成を可能とする支援を行う予定である.

### 参考文献

- [1]R. S. Flournoy, and C. Callison-Burch. 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.
- [2]T. Ishida, “Language Grid: An Infrastructure for Intercultural Collaboration”, in proc. of IEEE / IPSJ Symposium on Applications and the Internet (SAINT06), pp.96-100, 2006.
- [3]T. Murata, M. Kitamura, T. Fukui and T. Sukehiro: “Implementation of Collaborative Translation Environment: Yakushite Net”, the 9<sup>th</sup> Machine Translation Summit System Presentation, pp.479-482, 2003.
- [4]Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations, 2002. <http://www ldc.upenn.edu/Projects/TIDES/Translation/TransAssess02.pdf>