

## 言語グリッドによる Wikipedia の多言語議論支援

高崎俊之<sup>†</sup> 石松昌展<sup>†</sup> 門脇恒平<sup>†</sup> 森田大翼<sup>†</sup> 嶋田雅彦<sup>†</sup> 松野淳<sup>‡</sup> 里田旭彦<sup>†</sup> 夏琳思<sup>†</sup> 石田亨<sup>†</sup>  
 京都大学大学院 情報学研究科 社会情報学専攻<sup>†</sup> 京都大学 工学部情報学科<sup>‡</sup>

### 1. はじめに

非営利団体 Wikimedia 財団<sup>(1)</sup>のオンライン百科辞典 Wikipedia<sup>(2)</sup>は、英語版や日本語版など言語毎にシステムが独立して用意され、それぞれのコミュニティでの議論を通して記事生成がなされている。そこでは、言語を超えて議論されるべきグローバルなトピックも多々あるが、非英語話者の参加は難しい現状があり、特に日英語話者の多いアジア圏で課題になっている。例えば日本語版 Wikipedia には、英語版 Wikipedia の記事の誤りについて指摘したいが、言葉の壁で悩んでいるユーザの書き込みが散見され、多言語支援を求めるニーズが見受けられる。一方システムの観点では、ポリシーとして中立性を保つために特定のベンダーに依存することは避けており、Wikipedia ではオープンソースソフトウェアの MediaWiki<sup>(3)</sup>を採用している。

### 2. 言語グリッドの MediaWiki への適用

#### 2.1 言語グリッド

言語グリッド<sup>(4)</sup>は、辞書、対訳、機械翻訳等の各種言語資源を登録し共有することができる Web サービス技術による多言語サービス基盤である [1]。(独)情報通信研究機構により基盤ソフトウェアが開発されている。認証による各言語資源へのアクセスコントロールも備える。既存の単体の機械翻訳システムと異なり、利用者が作成した辞書や対訳集を登録・連携可能なため、利用者コミュニティに応じた、より精度の高い翻訳結果が得られるという特徴がある。言語グリッドの基盤ソフトウェアはオープンソースで提供されている。特定のプロプライエタリな言語サービスに依存することなく、複数の企業・団体・個人から提供された多様な言語サービスを組み合わせ、同一のインタフェース(API)でアクセスすることができるため、特定の企業・団体への依存を避けたい、という Wikimedia のポリシーと合致している。また言語グリッドは、スケーラビリティを意識したサービスコンピューティング技術であるため、大規模な Wikipedia システムとも親和性がある。

Multilingual support in Wikipedia discussion by Language Grid  
 Toshiyuki Takasaki, Masanobu Ishimatsu, Kohei Kadowaki,  
 Daisuke Morita, Masahiko Shimada, Jun Matsuno, Akihiko  
 Satoda, Xia Linsi, and Toru Ishida,  
 Department of Social Informatics, Kyoto University

#### 2.2 言語グリッドエクステンション

言語グリッドによる Wikipedia の多言語化基盤を構築するために、言語グリッドエクステンションが(独)情報通信研究機構により実装されている。これは MediaWiki システムの Extension という拡張方式に従ったソフトウェアモジュールで、言語グリッドに登録された様々な機械翻訳や辞書を自由に選択し、組み合わせて利用することができる。言語グリッドエクステンションは GPL ライセンスで提供され、MediaWiki から言語グリッドへアクセスするための手段を API として提供している。言語グリッドエクステンションの翻訳設定や辞書アクセスは、言語グリッド Toolbox[2]の成果を利用している。MediaWiki 関連の開発者は本 API を利用することで、ゼロからコードを書くことなく、MediaWiki から容易に言語グリッドへアクセスすることができる。

言語グリッドエクステンションは、MediaWiki に特化した付加機能である、PageDict と Service Settings も提供する。PageDict 機能は、記事に現れる専門用語やコミュニティ特有の語の訳語を登録可能な、記事毎に存在する辞書である。Wikimedia プロジェクトの指針に従い、PageDict は誰でも編集が可能としている。言語グリッドエクステンションが提供する API を用いて翻訳する際に、PageDict を機械翻訳と連携させることができ、そこに登録されている訳語を使用して翻訳することで、翻訳品質の改善が期待できる。一方 Service Settings 機能は機械翻訳や辞書をページ毎に設定するユーザインタフェースで、言語対単位で機械翻訳サービスを選択でき、機械翻訳ごとに利用したい辞書を設定できる。現在はテストフェーズだが、将来的なシステム構成は図 1 を想定している。Wikimedia 財団側に独自にオープン言語グリッドとして言語グリッドサービスを配備し、自前で調達した言語資源を独自に登録する。また言語グリッドの連邦制により、詳細な使用条件や免責などを含む覚書を交わせれば、より多くの言語サービスを持つパブリック言語グリッドへの相互アクセスも可能となる。

(1) Wikimedia 財団: <http://www.wikimedia.org>

(2) Wikipedia: <http://www.wikipedia.org>

(3) 言語グリッド: <http://www.langrid.org>

(4) MediaWiki: <http://www.mediawiki.org>

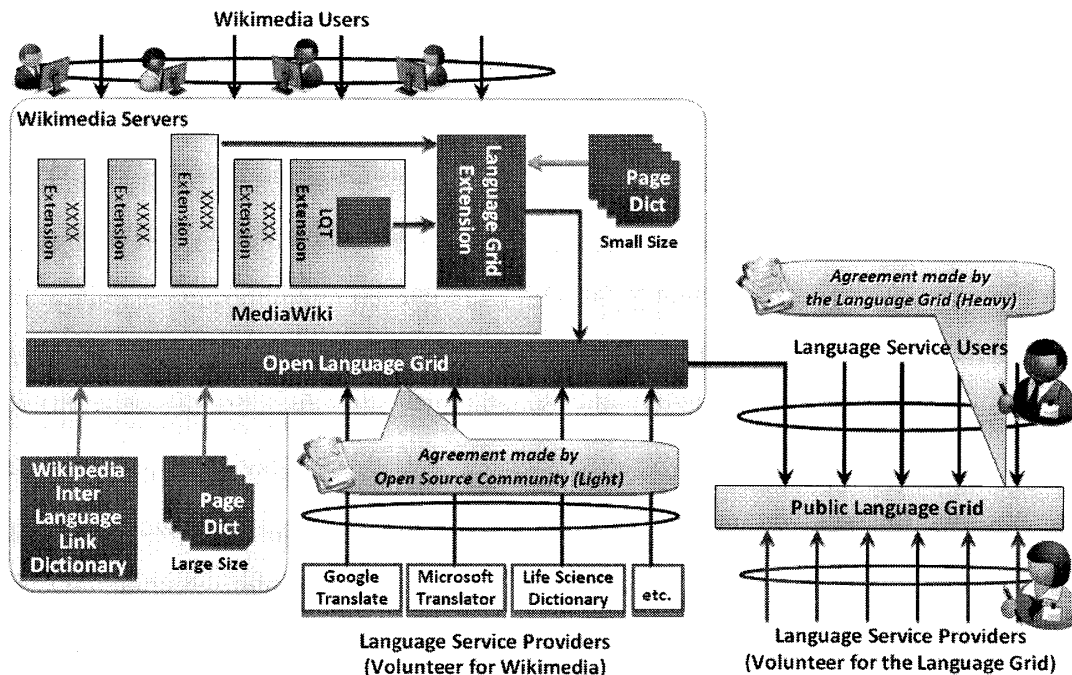


図 1. MediaWiki と言語グリッドエクステンションのシステム構成イメージ

### 3. 多言語議論支援

Wikipedia では、記事内容についての議論は各記事ページに用意されている「ノート」(英語では“discussion”)で行われている。前述のように Wikipedia 内での議論の多言語化ニーズがあるため、我々は言語グリッドエクステンションを介し、ノート機能の多言語化を行なった。Wikimedia 財団がユーザビリティ向上のため推進しているノート機能の GUI 強化版である Extension:LiquidThreads を多言語化した。これにより、ユーザはノートに書かれている議論を機械翻訳を通して母国語で閲覧することができ、また母国語で書き込みをすることができる。なお言語グリッドに現在登録されている言語資源では、52 言語の機械翻訳が可能である。

本研究開発を通して様々な課題が見えてきたがここでは 2 点挙げる。一つ目は「”翻訳に関する議論”の翻訳」の問題である。例えば、”おはよう”は”Good morning”という意味ですか?という日本語文を英訳すると、Does “Good morning” mean “Good morning?”となり意図が伝わらない。この問題を解決するため文章中の特定の部分を機械翻訳対象から除くと共に、原文語の意味を理解するのに参考となるような情報を付加する仕組みを言語サービスとして用意することを進めている。次に、機械翻訳の品質の課題も挙げられる。これについては、ユーザによる PageDict と他の辞書等とを連携した翻訳によって翻訳品質の向上を目指すと共に、パイリンガ

ルなユーザなどが機械翻訳結果を人的に修正する、というコミュニティによる支援ができる機能の追加を検討している。これは実際の NGO コミュニティで行なわれており [3], Wikipedia コミュニティの特性を考慮して検討していきたい。

### 4. おわりに

MediaWiki システムや Wikimedia コミュニティが、特定の言語資源ベンダーに深く依存することなく、多言語サービスを楽しむことができるオープンな言語グリッドエクステンションを紹介し、それを利用した多言語議論支援システムの開発について述べた。システムの土台は完成したので、今後はフィールドである実際の Wikimedia コミュニティからのフィードバックを得ながら課題を整理して改良を進めていきたい。

#### 謝辞

本研究は、戦略的情報通信研究開発推進制度 (SCOPE), および、京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」から助成を受けた。

#### 参考文献

- 1) Ishida, T: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96-100, keynote address (2006).
- 2) 田仲正弘, 村上陽平, 稲葉利江子ほか: Language Grid Toolbox: 多言語コミュニティ支援のためのオープンソースソフトウェア. 情報処理学会第 72 回全国大会 (2010).
- 3) 高崎俊之: 言語グリッドを活用した NGO 向け多言語コミュニティサイトの開発とフィールド実践, 信学技報, vol. 108, no. 441, AI2008-46, pp. 7-12 (2009).