

言語グリッド (Language Grid) の構想 - 異文化コラボレーション基盤の構築を目指して - Language Grid: An Infrastructure for Intercultural Collaboration

アーレム ベン ハッシン*1 藤原 義功 林 良彦*4*1 菱山 玲子*6*1 稲葉 利江子*1
Ahlem Ben Hassine Yoshinori Fujihara Yoshihiko Hayashi Reiko Hishiyama Rieko Inaba

石田 亨*9*1 喜多 千草*7*1 松原 繁夫*2*1 村上 陽平*1 灘本 明代*1 中口 孝雄*3
Toru Ishida Chigusa Kita Shigeo Matsubara Yohei Murakami Akiyo Nadamoto Takao Nakaguchi

中西 英之*8 重信 智宏*1 吉野 孝*5*1
Hideyuki Nakanishi Tomohiro Shigenobu Takashi Yoshino

*1独立行政法人 情報通信研究機構

National Institute of Information and Communications Technology

*2NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories

*3NTT アドバンステクノロジー(株)

NTT Advanced Technology Corporation

*4大阪大学言語文化研究科

Graduate School of Language and Culture, Osaka University

*5和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

*6京都女子大学現代社会学部

Faculty for the Study of Contemporary Society, Kyoto Women's University

*7関西大学総合情報学部

Faculty of Informatics, Kansai University

*8大阪大学工学研究科

Department of Adaptive Machine Systems, Osaka University

*9京都大学情報学研究科

Department of Social Informatics, Kyoto University

Although English has become the standard language in various areas, people do not use it in local activities. To increase mutual understanding between different cultures and opinions in different languages, it is essential to build a language infrastructure on top of the Internet. To increase the accessibility and usability of online language services, this paper proposes the language grid to create composite language services for various communities. The language grid is called "horizontal," when the grid connects the standard languages of nations, or "vertical," when the grid combines the language services generated by communities. The language grid is an infrastructure on the top of Internet that aims to improve accessibility and usability of existing language services so that users can create new language services according to their needs.

1. はじめに

インターネットは世界の人々を繋いだと言われるが、言語の壁は依然として存在する。インターネット上の言語人口は多様化し、標準言語がない状態ともいえる。Global Reach*1の調査によれば、インターネット上の英語人口は35.2%、アジア言語人口が26%、欧州言語人口が28%である。Webで様々な情報が共有されているが、それらの情報を知るためには、多くの言語の理解が必要となり非常に困難である。これまでに日中韓馬の異文化コラボレーション実験ICE2002(Intercultural Collaboration Experiment 2002)を実施した[野村 03]。この実験では、40名を超える日中韓馬の学生、教員が、母国語を用いてオープンソースソフトウェアの共同開発を行った。コミュニケーションを仲介したのは、機械翻訳システムである。以降、日中間のブロードバンドネットワークを用いたICE2003、日中韓馬泰で被験者を共有したICE2005と、実験は続いている。これらの経験より、異文化コラボレーションにおける機械翻訳は、コラボレーションを促進するには十分とはいえない翻訳品質であるものの、母国語で他国の議論を表示できること

連絡先: 灘本 明代, 独立行政法人 情報通信研究機構, 京都府相楽郡精華町光台 3-5, TEL:0774-98-6875, FAX:0774-98-6960, nadamoto@nict.go.jp

*1 <http://global-reach.biz/globstats/>

で、異文化コラボレーションの Awareness 支援としては有効であった。

すでにインターネットには、このような機械翻訳などの言語処理機能や電子辞書などの言語資源が蓄積され始めている(以降、総じて言語サービスと呼ぶ)。しかし、国際的な活動を行っている NPO が、それらを利用して自らが望むサービスを構築しようとする時、様々な問題が生じる。まずユーザが、その機能や品質、それに伴うコストを判断することは困難である。さらに、契約・知財の複雑さ、利用方法が標準化されていないといった問題がある。ユーザが必要なサービスを容易に構築することを可能とするためには、言語資源や言語処理機能のユーザビリティおよびアクセシビリティを向上させる必要がある。インターネットや Web が世界の人々の異文化コラボレーション基盤となるためには、言語の壁を越える仕組みが必要となる。そこで、本研究ではインターネット上の既存の言語サービスを容易に利用でき、ユーザ自らが新たな言語資源や言語処理機能を追加し、さらに組み合わせが可能な仕組みを持つ言語グリッド(Language Grid)について提案する [Ishida 06]。

世の中には、専門家が作成した言語サービスが数多く存在する。それらをインターネット上に移し、さらにユーザが作成した言語サービスも追加する。このようにインターネット上に言語サービスを蓄積させることで、ユーザはそれらを連携させ、必要とする言語サービスを新たに構築することが可能となる。

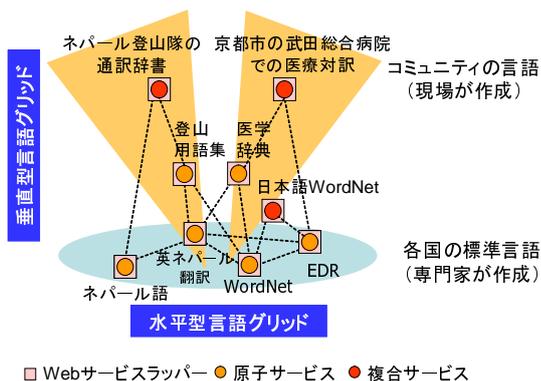


図 1: 言語グリッド

つまり、言語グリッドは、言語サービスの容易な利用と追加の仕組みを提供することで、専門家の力だけではなく、人々の力で言語の壁を越えることが目的であり、Web のアプローチと似ている。

本論文では、言語グリッドの構想について述べる。その中で、これまでの活動成果および今後の研究課題について述べる。

2. 言語グリッド

言語グリッド (Language Grid) は、インターネット上の既存のサービスを自由に組み合わせ、新たな言語サービスを生み出せる枠組みである。言語グリッドの特徴は、ヨーロッパ各国の語彙データベースの統合を目指した EuroWordNet[Vossen 04] と異なり、データの統合ではなくサービスの統合を目指している点である。グリッドとは、「分散した資源を連携させ、オープンスタンダードなプロトコルが利用でき、質の高いサービスを提供できるもの」と定義される。言語グリッドには、対訳翻訳や機械翻訳などを縦横に組み合わせる「水平型言語グリッド」と、応用に特化しコミュニティが必要とする言語サービスを生み出す「垂直型言語グリッド」の 2 種類の機能がある。各言語グリッドがどのような言語サービスから構成されるのかを図 1 に示す。

水平型言語グリッドは、国の標準言語に関する言語資源や言語処理機能をワークフローで接続するものである。アジアの約 10 言語、世界全体で 20 言語程度の対訳辞書や機械翻訳を利用する予定である。一方、垂直型言語グリッドは、異文化コラボレーション活動の現場で使われる、医療通訳、絵文字コミュニケーションなどのコミュニティ固有の辞書や用例対訳を利用可能としていく。

Global な水平型言語グリッドと Local な垂直型言語グリッドとを組み合わせることにより、グローバルな言語インフラを構築することができる。我々は、誰でもが、「コミュニティ固有の言語資源を作成し、言語グリッドに登録できる」さらに、「言語グリッド上に登録されている言語サービスを簡単に接続し、利用できる」仕組みを構築することを目的としている。

3. システムアーキテクチャ

前章で述べたように、言語グリッドは、インターネット上に個別に分散した言語資源や言語処理機能を、仮想的に一つの言語サービス群として統合して利用可能にする技術である。これにより、言語サービスへのアクセシビリティやユーザビリティが向上することで、国際的な活動を行っている人々自らが望む

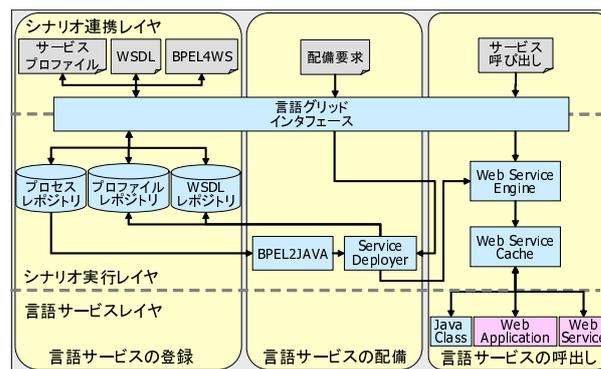


図 2: 言語グリッドシステムのアーキテクチャ

サービスの構築が可能になる。

この言語グリッドの実現には、作成者や提供者が異なる言語サービス群の連携が必要となる。それぞれが個別に実装した言語サービスはインタフェースや実装言語が異なるため、実装方法に依存しない手法で連携しなければならない。このような相互運用の問題を解決する一つの手法として、Web サービスが注目されている。Web サービスとは、XML に基づく通信プロトコル SOAP(Simple Object Access Protocol) やサービスのインタフェース記述言語 WSDL(Web Service Description Language) を用いて、プラットフォームに非依存なアプリケーション連携を可能にする技術である。言語サービスに関しては、三省堂デイリーコンサイズ^{*2}などが体験版ではあるが既に提供されている。また、近年このような Web サービスを連携するためのビジネスプロセスの標準化も活発に行われており、BPEL4WS(Business Process Execution Language for Web Services) などが提案されている [Andrews 03]。例えば、既存の仏英翻訳サービスや英和翻訳サービスといった Web アプリケーションを Web サービス化できれば、その二つを順に接続するワークフローを BPEL4WS で記述することで、仏和翻訳の Web サービスを提供することが可能になる。

我々は、この Web サービス技術に基づいて言語グリッドシステムの開発を行った。開発した言語グリッドシステムのアーキテクチャを図 2 に示す。言語グリッドシステムは、Web サービス化された言語資源・言語処理機能からなる言語サービスレイヤと、連携プロセスである BPEL4WS を実行するシナリオ実行レイヤ、そして、連携プロセスを作成するシナリオ連携レイヤの 3 つのレイヤから構成される。以下では各レイヤについてより詳しく説明する。

3.1 言語サービスレイヤ

言語グリッドに統合される言語サービスは、その提供方法により、実行プログラムとして提供されているもの、Web アプリケーションとして提供されているもの、Web サービスとして提供されているものの 3 種類に分けられる。一つめの言語サービスは言語グリッドシステム上の Web サービス実行エンジンに直接配備することで言語グリッドに統合可能である。一方、残りの二つの言語サービスは言語グリッドシステムの外部にサービスの実体が存在するため、言語グリッドに統合するには Web サービス用のインタフェースが必要である。三つめは既に Web サービスとして公開されているため、そのサービスインタフェース記述である WSDL と、サービスのプロフィール情報を言語グリッドシステム上の WSDL リポジトリ、プロ

*2 <http://www.btonic.com/ws/>

フィールリポジトリにそれぞれ登録することで言語グリッドに統合可能である。一方、二つめの言語サービスは、各言語サービスの種類に応じたラッパーを構築し、Web サービス化しなければならない。サービス間の連携を容易にするために、このラッパーは言語サービスの種類ごとに標準化されたインタフェースが必要である。現在、我々は Web 上に公開されている Web アプリケーションや言語資源を収集し、それぞれのラッパーを構築しながら標準インタフェースを整理している。

3.2 シナリオ実行レイヤ

ビジネスプロセスの記述言語として開発された BPEL4WS はロールバック等の複雑な処理が可能である反面、実行速度に問題がある。言語サービスがしばしば人と人の共同作業の場で用いられることを考慮すると、複雑な処理よりも実時間性の方が重視される。そこで、シナリオ実行レイヤでは二種類の高速化手法を利用し、連携プロセスの処理を効率化させている。一つは BPEL4WS から Java へのトランスレータ (BPEL2JAVA) であり、もう一つは Web サービスの実行結果のキャッシュである。前者は特に複数クライアントから呼び出される場合に効果的であり、後者はあいさつなどの定型文が頻繁に発生するチャットなどで利用される言語サービスに効果的である。したがって、多人数での多言語チャットで呼び出される多国語翻訳サービスの実行に適している。

3.3 シナリオ連携レイヤ

言語サービスは、Web サービス化され言語グリッドに統合されると、シナリオ連携レイヤで作成する連携プロセスのコンポーネントとして利用することが可能になる。言語グリッドシステムは、連携プロセスの作成を支援するためのコマンド群も提供しており、ユーザは直接 BPEL4WS を記述しなくても、選択した言語サービスをワークフローへ追加・削除することで、自分たちの活動に適した連携プロセスを作成できる。作成されたワークフローから自動的に BPEL4WS が生成され、ユーザからの配備要求によって、新しい言語サービスが生成される。このような言語サービスの連携により構築される言語グリッドには、水平型言語グリッドと呼ばれるものと垂直型言語グリッドと呼ばれるものが存在する。

水平型言語グリッドは、国の標準言語を取り扱う汎用的な言語サービスを連携するため、グリッドを構成するコンポーネントの数は比較的少なく、連携のパターンもマルチホップ翻訳のように単純である。今後、言語資源や言語処理機能の数が増加するにつれて、水平型言語グリッドも容易に拡大していく。

一方、垂直型言語グリッドは、各コミュニティの活動に特化した言語サービスの構築を目的とするため、その連携プロセスは複雑かつ多様である。例えば、医療現場の通訳支援では、正確な翻訳が求められるため、医療用の対訳辞書が必要となる。ただし、病院ごとの対訳辞書を通訳支援団体がそれぞれ作成しては、その労力は少なくない。したがって、既に電子化された一般的な医療対訳サービスと、各病院に特化した医療対訳集を電子化したサービスを連携させることが必要となってくる。図 3 にその連携プロセスを表すワークフローを示す。入力となる日本語の文は初めに病院に特化した対訳サービス

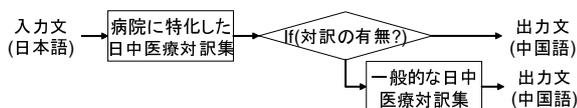


図 3: 医療通訳支援用言語サービスのワークフロー

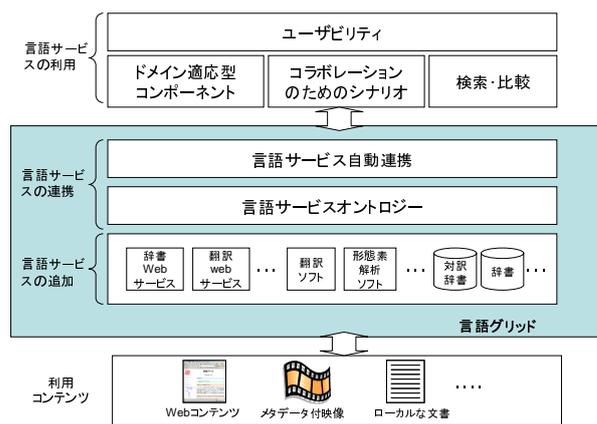


図 4: 異文化コラボレーション支援システムの構成図

に送られ、対訳が存在すれば、対訳の中国語の文を出力する。もし対訳が無ければ、一般的な対訳サービスに送られ、対訳の中国語の文が出力される。

4. 今後の研究構想

ユーザが言語グリッドを用いて異文化コラボレーションに関するアプリケーションを構築するには、ユーザが容易に使用できる異文化コラボレーション支援システムが必要である。実サービス可能な異文化コラボレーション支援システムを構築するためには、(1) 言語グリッドの機能の強化、(2) 異文化コラボレーション支援システム利用のための各種機能が必要である。そこで我々は、言語グリッドにおける言語サービスの追加、連携、利用の視点から研究に取り組んでいる。図 4 に異文化コラボレーション支援システムの構成を示す。本研究では、言語グリッドの強化として言語サービスの追加と連携を、異文化コラボレーション支援システム利用のための各種機能の構築として言語サービスの利用を提案する。

4.1 言語サービスの追加

言語グリッドでは、対訳辞書や機械翻訳、形態素解析といった言語資源や言語処理機能が数多く必要である。あらかじめある程度の言語サービスは言語グリッドに登録されているが、言語グリッドの発展には、Web のように言語グリッドのユーザによる言語サービスの登録が必要不可欠である。したがって、あらゆるユーザが容易にこの言語サービスを追加できる機能を構築しなければならない。追加対象である Web アプリケーションやソフトウェア等の Web サービス化されていないものは、ラッパーを用いて Web サービス化し言語グリッドシステムに登録しなければならない。このラッピング作業を自動化することが今後の研究課題として挙げられる。ただし、言語グリッドでは複数の言語サービスを連携させることが目的であるため、その言語サービスが何であるか各々の言語サービスを事前に定義する必要がある。そのため、言語サービスに対するセマンティックスを用いたセマンティックラッパーを実現する。

4.2 言語サービスの連携

(1) 言語サービスオントロジー

言語グリッドにおいて、複数の言語サービスを連携させるためには、相互運用性を確保・促進する必要があり、そのために、個々の言語サービスに対して適切なメタデータを付与する必要がある。このようなメタデータは、人間、ソフトウェア双方にとって了解可能な言語サービスオントロジーに基づいてい

る必要がある。また、そのオントロジーは言語グリッド上に現在存在する言語サービスを対象とするだけでなく、将来現れる可能性のある新たな言語サービスも対象に含められなければならない。そこで言語サービスの追加時、連携時に有用な標準的な言語グリッドオントロジーを構築する [Hayashi 06]。

(2) 言語サービス自動連携

言語グリッドに登録されている複数の言語サービスから、ユーザ自ら各自の利用形態に合わせて最適な言語サービスの組み合わせを選択するのは困難である。そこで、この言語サービスの連携の自動化を研究開発する。言語グリッドは常に言語サービスが追加・削除されるため、連携可能な言語サービスが動的に変化している。したがって、ワークフローのプランニングとして、このような動的な環境下における Web サービスの自動連携手法を提案する。さらに、実際に言語サービスオントロジーを用いて、この自動選択手法を言語サービスドメインへ適用することで、提案手法の有効性を検証する。

4.3 言語サービスの利用

(1) コンテンツ検索・比較機能

ユーザは言語グリッドを用いて Web コンテンツやメタデータ付映像コンテンツ、ローカルな文書コンテンツ等多メディアなコンテンツを利用する事がある。また、ユーザは会話の中からまたはこの Web に関連する文書といったように、直接キーワードを用いて検索を行うとは限らない。そこで、多言語コンテンツ間・多メディア間を横断検索可能な機能及び Context Aware な機能を用いた検索手法の研究開発を行う。また、異文化間のコラボレーションを行うにあたり、より効率よくコラボレーションを行うために多言語で記述された種々のコンテンツを比較提示する研究開発も行う。

(2) ドメイン適応型コンポーネント

異文化コラボレーション支援システムにおいて、ドメインが異なれば必要とされる機能やユーザ・インタフェースは大きく異なる。各コミュニティにおいて多種多様なタスクが存在することは容易に想像できるため、あらゆるコラボレーション形態に対応したツールを提供することは非常に困難である。また、コミュニティに適した言語サービスを構築できたとしても、それを利用するためのツールをユーザ自身で開発することは難しい。そこで、言語グリッドと連携する機能を備えた、ユーザ・インタフェースを含む各種機能をコンポーネント群として生成し、ユーザがそれらを容易に組み合わせることで、よりドメインに特化したコラボレーションツールの生成を実現する開発環境を構築する。つまり、ユーザは必要なコンポーネントを組み合わせることで、種々のコミュニティにおけるコラボレーションに適した支援システムを容易に生成することが可能となる。

(3) コラボレーションのためのシナリオ記述

異文化コラボレーションの支援を目的とした言語サービスは、しばしば人と人の共同作業の場で用いられるため、Web サービスのみから構成される複合サービスとは限らない。言語サービスにおける多くの連携プロセスでは、コラボレーションの現場の人のアクティビティへの考慮も必要になる。つまり、人とのインタラクションも含んだ連携プロセスを考えなければならない。このような連携プロセスの実現には、単純なサービス呼出しに基づいたワークフロー記述言語ではなく、各サービスの実行主体をエージェントとみなした、インタラクションに基づくシナリオ記述言語が適切である。このシナリオ記述言語では、コラボレーション支援の中で発生する人とのインタラクションの記述と、言語グリッド上の言語サービスとのインタラクションの記述を可能にしなければならない。

(4) ユーザビリティ

ユーザ中心の異文化コラボレーション支援システムを構築していくためには、専門家主導のシステム構築ではなく、社会の中で実際に行われる多文化・多言語間のコミュニケーション現場を通し、システムを開発していくことが重要である。そこで、「異文化コラボレーションオープンラボ」を構築し、国内外の異文化コラボレーション活動を行っている関連団体・組織などと連携しながら、ユーザビリティ評価モデルを心理的なストレスの測定を行いながら構築し、そこで得られた知見を言語グリッドシステムおよびコラボレーション支援ツールにフィードバックする。

5. おわりに

本論文では、言語サービスのアクセシビリティおよびユーザビリティの向上を目的とした言語グリッドの構想について述べた。異文化コラボレーションの研究には、機械翻訳などの自然言語処理研究のみならず、エージェント、セマンティック Web などの人工知能研究、ヒューマンインタフェースやコラボレーション研究、さらに異文化間コミュニケーションの研究が連携していく必要がある。これらの連携には、異文化コラボレーションにおける会話のデータや統計データの蓄積が重要であり、実験やアンケートによるデータ収集が、これらの異なる研究分野を結びつけていくと思われる。すでに本研究では3つのNPOの協力による活動を始めており、現場のデータを収集していく予定である。このように本研究の特徴は、産官学だけでなく民をも加えた協働体制で研究を進めていくことにある。このために、ユーザビリティテストおよびビジネスモデルの形成を目的とした異文化コラボレーションオープンラボの開設などを予定している。

参考文献

- [Andrews 03] Andrews, T., Curbera, F., Dolakia, H., Golland, J., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S., Trickovic, I., and Weeravarana, S.: Business Process Execution Language for Web Services, 2003.
- [Hayashi 06] Hayashi, Y. and Ishida, T.: A Dictionary Model for Unifying Machine Readable Dictionaries and Computational Concept Lexicons, in Proc. of the International Conference on Language Resources and Evaluation (LREC-06), 2006.
- [Ishida 06] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, in Proc. of IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.96-100, 2006.
- [Vossen 04] Vossen, P.: EuroWordNet: A Multilingual Database of Autonomous and Language-Specific Wordnets Connected via an Inter-Lingual Index, International Journal of Lexicography, Vol.17, No.2, pp.161-173.
- [野村 03] 野村早恵子, 石田 亨, 船越 要, 安岡 美佳, 山下 直美: アジアにおける異文化コラボレーション実験 2002: 機械翻訳を介したソフトウェア開発, 情報処理学会論文誌, Vol.44, No.5, pp.503-511, 2001.