

言語グリッドの世界展開に向けた連邦制運営モデル

村上 陽平[†] 石田 亨[‡] 稲葉 利江子[‡]

(独) 情報通信研究機構 言語グリッドプロジェクト[†]

京都大学大学院情報学研究科 社会情報学専攻[‡]

1. はじめに

インターネット上には多くの言語資源 (データやプログラムの両方) が存在するにも関わらず、異文化コラボレーション活動の現場ではそれらを組み合わせて、多言語サービスとして活用することができていない。UIMA[2] など既に言語資源や言語処理ツールを連携させる試みは行われているが、これらの既存の言語資源連携フレームワークでは、言語資源に絡む複雑な知財の問題は利用者が自分で解決しなければならないため、ほとんどの利用者には、異文化コラボレーションの活動に必要な多言語サポートを得ることは困難であった。

言語グリッド[3]は、このような知財の問題に対応するために、言語資源を連携させるだけでなく、提供者自身のポリシーに従ってサービスを制御可能にする言語サービス管理アーキテクチャを構築している。さらに、システム開発だけでなく、サービス提供者とサービス利用者の両者のインセンティブを合致させる制度設計も行い、2007 年 12 月より京都大学が非営利限定で運営を開始している[4]。

しかしながら、2 年間の運営を通して、いくつかの課題に直面している。一つは運営組織の地理的な原因で、言語グリッドの利用組織のほとんどが日本ユーザであることである。もう一方は、運営組織が大学であり非営利限定にしているため、利用組織の大半が大学であることである。今後の言語グリッドの世界展開に向けて、上記の課題に対処するために、本研究では、複数の運営組織が各々の運営ポリシーを定め、連携して運営していく言語グリッドの連邦制運営モデルを提案する。

2. 言語グリッド

言語グリッドは、機械翻訳や形態素解析などの高度に専門的な言語資源や多文化共生の現場で生まれた言語資源をインターネット上で共有し、それらを自由に組み合わせて利用できる仕

Federated Operation Model for Globally Spreading the Language Grid

[†]Yohei Murakami, Toru Ishida, Language Grid Project, National Institute of Information and Communications Technology (NICT)

[‡]Rieko Inaba, Department of Social Informatics, Kyoto University

組みである。各言語資源は Web サービス化され、WS-BPEL で記述されたワークフローを用いて連携され、言語サービスとして登録される。この言語サービスも Web サービスとして公開されるため、あらゆる異文化コラボレーションツールから言語サービスを利用することが可能である。

言語グリッドの特徴は、世界中の語彙データベースの統合を目指した Global WordNet Grid[1]と異なり、語義に基づく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。したがって、これまでの機械翻訳サービスに対して、専門辞書の辞書引きサービスや用例対訳検索サービスを連携させることで、利用者は利用現場に特化した精度の高い翻訳サービスを受けることが可能である。

3. 中央集権型運営モデル

言語グリッドの運営を開始するにあたり、ステークホルダーであるサービス提供者、サービス利用者の要件を収集し、非営利運営に限定した中央集権型運営モデルを構築した。

中央集権型運営モデルでは、単一組織で言語グリッドを運営し、全ての利用組織はこの運営組織と覚書を締結する。サービス提供者は提供サービスへのアクセスを完全に制御可能である。たとえば、サービス利用者の利用統計情報 (アクセス回数、アクセスログ、データ転送量等) をモニタリングでき、利用可能なユーザの選択や利用制約の設定も可能である。さらに、不正アクセスを防ぐために、サービス利用者はツールを公開して不特定多数のエンドユーザからのアクセスを可能にしてはならない。つまり、ツールを公開する場合は、各エンドユーザをサー

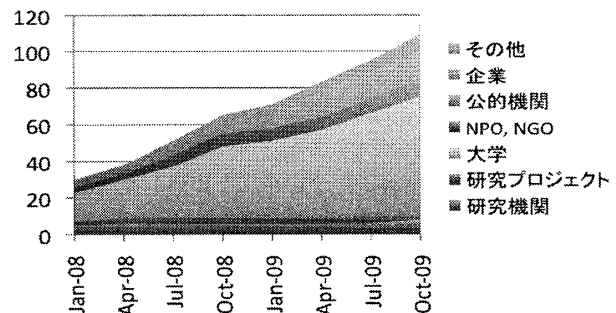


図1 言語グリッドの利用組織の推移

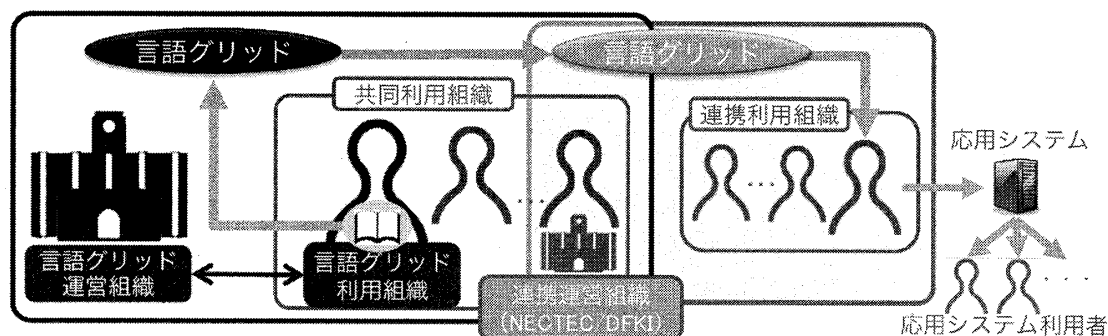


図 2: 連邦制の運営モデル

ビス利用者が同定できる必要がある。

この結果、現場で異文化コラボレーションに直面している NPO や NGO, 自治体での利用は予想に反して少なく、大学での研究利用が多くなっている。図 1 に言語グリッドの利用組織の推移を示す。

4. 連邦制運営モデル

既存の中央集権的な非営利運営モデルでは、二つの課題がある。一つ目は、言語グリッドの利用組織のうち、国内の利用組織が 70%以上占めていることである。二つ目は、さらにその利用組織の大半が大学であることである。この課題に対処するために、複数の運営組織による連邦制での運営モデルの設計を行った(図 2)。連邦制運営モデルでは、単一の運営組織だけではなく、複数の運営組織が連携して運営を行うために、連携運営組織と連携利用組織というロールを導入している。連携運営組織とは、同一の覚書を用いて自らの言語グリッドを別途運営している共同利用組織のことをいう。また、連携利用組織とは、連携運営組織と覚書を締結した利用組織のことを言い、運営組織とは直接覚書を締結していない利用組織である。この連携利用組織によるサービスの利用を許可することで、単一の運営組織では連絡をとることができなかった利用組織や、またその利用組織によって提供される言語資源を収集することが可能になる。

もう一方の課題に対しては、言語グリッドの利用目的や利用方法の範囲を拡張し、選択式にしている。提供者はサービス登録時に、許可する利用目的と利用形態を設定し、利用者は利用時に利用目的や利用形態を送信する。システム上でお互いのマッチングをとりアクセス制御を行っている。これにより、営利企業の研究利用やサーバ型でのツール公開が可能になり、利用方法が多様化することで利用組織の種類が増加が期待される。

5. おわりに

本稿では、複数の運営組織が連携しながら言語グリッドの運営を行う、連邦制運営モデルについて述べた。連邦制運営モデルでは、連携運営組織と連携利用組織という二つのロールを導入することで、連携利用組織は連携運営組織を介して、別の言語グリッドの言語サービスの利用が可能になる。また、提供側は同一の言語グリッドの利用組織以外にも、提供サービスの利用機会を増やすことが可能である。

今後は、2010年4月からタイの NECTEC と、連邦制による言語グリッドの運営を開始することで合意している。これにより、東南アジアでの言語資源獲得が促進され、日本における東南アジアからの労働者との異言語コミュニケーション支援に活用されることが期待される。

謝辞

本研究は、総務省戦略的情報通信研究開発推進制度 (SCOPE) の補助を受けた。

参考文献

- [1] C. Fellbaum and P. Vossen. Connecting the Universal to the Specific: Towards the Global Grid. *Intercultural Collaboration, LNCS 4568*, pp. 1-16, 2007.
- [2] D. Ferrucci, and A. Lally. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Journal of Natural Language Engineering*, Vol. 10, pp.327-348, 2004.
- [3] T. Ishida. Language Grid: An infrastructure for intercultural collaboration. *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96-100, keynote address, 2006.
- [4] T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa. A Non-Profit Operation Model for the Language Grid. *The First International Conference on Global Interoperability for Language Resources (ICGL-08)*, pp. 114-121, 2008