

言語サービスオントロジーの検討

林 良彦† 梶和 千春‡

大阪大学大学院言語文化研究科／独立行政法人 情報通信機構†

京都大学大学院情報学研究科‡

1 はじめに

近年、さまざまな言語資源や言語処理機能が Web 上で利用可能となっている。テキスト翻訳や辞書アクセスなどの言語サービスを直接に提供するサイトのほか、形態素解析などの各種の自然言語処理ツールも Web 上で公開されている。一方、ドメインに特化した言語資源が NPO などのコミュニティによって独自に開発されている。これらを必要に応じて連携させることができれば、異文化コラボレーションのさまざまな局面において有用な複合的な言語サービスを容易に実現できると期待される。

言語グリッド[1]は、そのような言語基盤を Web 上に構築するためのプロジェクトである。本報告は、言語グリッドのような言語基盤上において言語サービスを記述するための言語サービスオントロジーの最上位階層の構成案を提示する。

2 言語サービスオントロジーの必要性

ユーザの要求に合致する複合的な言語サービスを実現しようとする際は、要素となる言語処理機能や言語資源、および、そのアクセス機能を組み合わせることが必要となるが、このためには、これらの構成要素に対して適切な記述が与えられている必要がある。また、その記述は標準に沿った形式にのっっている必要がある。言語サービスオントロジーは、このような構成要素の記述において、共通理解の得られたセマンティクスを与えるための基盤である。

言語サービスオントロジーは、言語処理機能や言語資源の記述だけではなく、既存の、または、新規の構成要素を言語基盤上で新たに利用可能とする際にも必要となる。すなわち、新たな構成要素を言語基盤上で利用可能とする際には、そこで定められている標準インタフェースによるアクセスが必要であ

An Upper Ontology for Linguistic Web Services

† Yoshihiko Hayashi, Osaka University and NICT

‡ Chiharu Narawa, Kyoto University

り、各要素の個別的な詳細を隠蔽するラッパープログラムを開発する必要があるが、言語サービスオントロジーのタクソノミーにおける主要なタイプごとにラッパープログラムの基本的な構成を対応付けておくことにより、ラッパープログラムの開発を効率化することができる。もしくは、対象の構成要素がすでに Web 上で CGI などによるサービスを提供している場合には、必要な情報を抽出するためのパターンを探索する際の制約として利用することにより、ラッパープログラムの半自動生成が可能となる。

3 言語サービスオントロジーの要件

言語サービスオントロジーが満たさなければならない要件は、次のとおりである。

- 入出力の制約の表現：複合言語サービスを可能にするためには、最初のサービスの出力が次のサービスの入力として引き継がれるように、入出力の互換性をチェックしなければならない。従って、言語サービスオントロジーが満たすべき最も重要な要件は、入出力の制約を十分に記述し得る語彙をもつことである。さらに、将来の自動プランニングによるサービス連携に備えて、言語サービスが呼び出される時の前提条件や、サービスが呼び出されたときに、もたらされる状態変化の効果を適切に定義しておかなければならない。
- 構成要素を記述するに十分な語彙：言語サービスの構成要素、即ち、言語処理機能と言語資源の両方を記述する語彙をもつ必要がある。さらには、言語処理機能によって付加されるさまざまなレベルのアノテーションの記述モデルを提供する必要がある。
- 標準との適合：構成要素に関する記述の再利用性・相互運用性を確保するために、記述の体系は、国際標準として規定されている規格に適合していなければならない。

4 オントロジーの最上位階層

図1は、言語サービスオントロジーの最上位階層を示す¹。長方形の枠はクラスを示し、矢印とその横のラベルはクラス間の関係を示す。‘isa’のラベルは、クラス階層を表わしている。

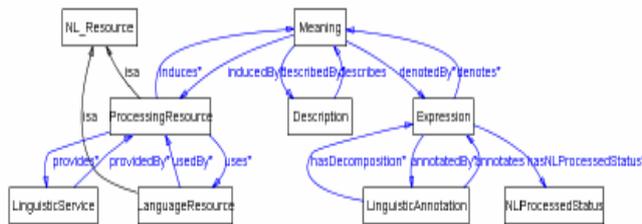


図1 言語サービスオントロジーの最上位階層

まず、言語サービス **LinguisticService** は、言語処理資源 **ProcessingResource** により実現される。これは、静的な言語資源 **LanguageResource** は、言語処理資源に属するなんらかのアクセス機能を介して初めて利用可能となるという制約を示している。図には示されていないが、**LinguisticService** クラスには、非機能的な記述(いわゆるメタデータ)を保持するための属性を持たせる。

言語処理資源、言語資源の二つのクラスの上位クラスが **NL_Resource** である。また、両クラス間に定義される関係‘use’により、言語処理機能は、言語資源機能を利用する(ことがある)という関係を表現している。

さて、図の右上部は、言語表現 **Expression** と意味 **Meaning**、意味に関する記述 **Description** の関係を示している。すなわち、言語表現は何らかの意味を指示(denote)する。ここでは、意味がどのような表現により定義されるかについては規定せず、プレースホルダとして、ある意味が存在することを表現することを可能とする。これにより、翻訳やある種の言い換え処理における入出力の言語表現の意味的等価性を示すことを可能とする。現在のところは、意味クラスの下位分類として、言語資源クラスの下位クラスである辞書クラスのタクソノミーに応じて辞書意味を分類している[2]。また、これらの辞書

の意味に対応して、その意味記述に関する分類も行っている。

上記のように言語表現は意味と関係するが、自然言語処理による結果により注釈付けられる。すなわち、言語処理状態 **NLProcessedStatus** によりどのような言語処理が行われたか、また、**LinguisticAnnotation** によりどのような結果が得られているかが示される。**NLProcessedStatus** は、言語処理機能をパイプライン的に実行する際の入出力条件や前提・結果条件(いわゆる IOPE)を表すために導入されたクラスである。また、言語処理は言語表現に対してアノテーションを付与するものであるという近年の考え方に従い、言語解析の結果は、**LinguisticAnnotation** のインスタンスにより表現され、**Expression** のインスタンスに関係付けられる。

5 おわりに

本報告では、言語サービスオントロジーの最上位階層の構成案を提示し、クラス間の関係を示した。今後の課題としては、実際に言語基盤(言語グリッド)上で利用が検討されている言語資源、及び、言語処理ツールや言語処理システムの記述を行なう作業を通して、さらに考慮すべき技術的要件を抽出し、オントロジーを詳細化する必要がある。

また、本稿で示したように、言語サービスオントロジーは、言語表現に対する言語学的な注釈や、辞書エントリの情報構造を包含するものである。このため、LAF[3]やLMF[4]などの国際標準化の動向を参照していく。

参考文献

1. Ahlem, 他: 言語グリッド (Language Grid) の構想. 人工知能学会第20回全国大会,(2006)
2. Hayashi,Y.: Conceptual Framework of Upper Ontology for Describing Linguistic Services. Proc. of IWIC2007,(2007).
3. Ide, N., Romary, L.: International standard for a linguistic annotation framework. Journal of Natural Language Engineering, Vol.10:3-4, pp.211-225. (2004).
4. Francopoulo, G. et al.: Lexical Markup Framework (LMF). Proc. of LREC2006, (2006).

¹ OWL を作業用の記述言語とし、ツールとして Protégé (<http://protege.stanford.edu/>) を利用している。