

# 自然言語処理コミュニティにおける言語資源構築・共有のあり方

松林 優一郎

東北大学

{y-matsu}@ecei.tohoku.ac.jp

電子化された文章の蓄積により、自然言語処理技術に対する需要はますます高まっていると言える。Web 上で行われる文章を通じた社会的コミュニケーション量の爆発的増大により、その需要はこれまで以上に現実的なものとなっており、文章に人間がエンコードしている情報を極力精密に取り出すことの出来る高度な技術の実現が期待されている。これらの中には、談話構造解析、因果関係・目的手段関係・原因理由関係などの関係抽出、言い換え・含意関係理解、意見抽出、センチメント解析など、テキストが伝達する意味を対象とした解析のほか、機械翻訳などコミュニケーション自体を促進するための技術も含まれる。

一方で、これらの技術が、言語の様々な現象に対して包括的に対処しなければ十分な精度が得られない、言ってみれば「総合技術」であることもエラー分析などから明確になりつつある。また、文の基本構造に対する解析技術について残された課題をいくつか例に挙げてみても、述語項構造解析、照応・共参照解析、モダリティ解析、時間関係認識など、内在する問題自体が未だ十分に把握されていないものが多く、実用レベルの解析精度に達していない。

このような一筋縄ではいかない課題に対して、当然ながら我々はそれぞれの課題で問題となる現象とその分布を把握・整理する必要がある。しかしながら、そのような土台の準備を我々はどのように行っているだろうか。近年の言語処理技術の高精度化はコーパスに付与された正解分析例にもとづいた機械学習によって実現されているが、この正解例こそが、言語の構造分析のために入念に練られた理論の結晶であり、現状のほとんどの技術では人手で与えられた構造にもとづいて解析を行わなければ十分な精度を得ることはできない。一方で、我々が今後取り組まなければならない課題では、どのような構造を仮定して解析を進めるべきかという問いにさえ答えを見いだせていない問題のほうが多い。このような状況において、我々はどのような方法論を必要とするだろうか。

本発表では、自然言語処理コミュニティにおいて、それぞれの解析技術の発展に必要な言語現象分析と言語資源構築・共有のあり方について、過去の事例や海外コミュニティにおける動向などを例に出しながら議論を行う。