



# 自然言語処理研究の厳しさと楽しさ

【受賞タイトル】 自然言語の深い構文・意味解析の研究とその応用

宮尾祐介 国立情報学研究所

このたび長尾真記念特別賞の受賞対象となったのは、「深い構文解析 (Deep Parsing)」とよばれる自然言語処理技術の研究である。構文解析は自然言語処理の根幹であり、長年さかんに研究されてきた。リアルテキストを高精度で解析できる技術はなかなか実現されなかったが、1990年代半ばから確率モデルや機械学習を応用することで高精度な曖昧性解消が可能となり、構文解析技術が急速に進歩した。しかし、このときの主な研究対象は、確率・統計モデルを適用しやすい文脈自由文法や係り受けといった比較的単純な構造の解析である。それ以前は、構文解析はより「深い」構造である意味表現 (述語項構造や論理表現) をターゲットとしていた。深い構文解析の研究は、この古の構文解析の枠組みと、現在の機械学習主導の構文解析を両立したものである。

そもそも自然言語処理の研究を志したのは、機械翻訳を実用化して英語を学ばなくてよい世界を実現したい、という浅薄な動機による。その後、構文解析すらまともにできない現実に直面し、やや挫折しながらも英語の構文解析の研究を始めることとなった。その当時の構文解析や機械翻訳はまったく未成熟で、機械翻訳は私の研究人生のうちにはまともに動かないかもしれないと感じていた。

しかしその後の自然言語処理の進歩はすさまじく、私が研究・開発してきた深い構文解析器も2000年代前半には実用レベルの精度を達成し、情報抽出やテキストマイニングに応用してその有効性を実証するまでになった。また、この構文解析器は英日機械翻訳に応用され、大幅な精度向上に貢献することができた。研究を始めた当初は10年ほどでここまで構文解析や機械翻訳が進展するとは予想しておらず、うれしい誤算である。ただし、現在でも機械翻訳は自然な翻訳

とは言えず、構文解析も機械学習頼りなので学習データと異なるタイプのテキストやデータが整備されていない言語では精度が著しく低下する。今までの目覚ましい進歩に微力ながら貢献できたことは幸運だが、残された問題の方が遥かに大きいのも現実である。

現在はテキスト間含意関係認識という意味解析の研究を進めているが、意味解析の世界は構文解析よりはるかに混沌としている。関連する研究分野は、論理的側面を対象とする形式意味論、単語やフレーズが持つ豊かな意味関係についての語彙意味論や知識表現、文脈やさまざまな要素によって変化するパラフレーズ認識などなど、多岐にわたる。さらに、構文解析よりも複雑な構造的曖昧性の問題があり、これらを解決して実用レベルの意味解析を実現できるかどうか、今後の研究に期待いただきたい。

自然言語処理は、哲学から機械学習まで毛色の異なるさまざまな理論・技術が複雑に絡み合った応用技術であると同時に、人間の知的活動の基盤である言語に対する計算機科学的探求という基礎学問の側面も持つ。その一面だけを切り取って研究してもなかなか革新的な技術に結びつかないという厳しさがある反面、その絡み合った糸をきれいにほぐして言語の美しい一面が明らかになったときの感動は格別である。

最後になるが、今回受賞対象となった研究成果は、国内外の共同研究者・研究仲間の協力や彼らとのさまざまな議論によるものである。この場を借りてあらためて感謝の気持ちを表したい。

(2014年5月15日受付)

宮尾祐介 (正会員) yusuke@nii.ac.jp

1998年東京大学理学部情報科学科卒業。2006年同大学院にて博士号(情報理工学)取得。2001年より同大にて助手、のち助教。2010年より国立情報学研究所准教授。