

## 本ワークショップの趣旨

計算心理言語学は、人間の言語処理・獲得の仕組みの解明を目指す心理言語学的問題に対して、計算論的アプローチで取り組む学問である。数理モデルを用いた明確な仮説の構築と、計算機を用いた定量的な仮説の検証を一貫して行うという特徴を持ち、これまで、言語獲得、文・意味処理、音声知覚などの様々な問題に関する新たな知見が提案されてきた。昨今の統計的学習モデルや言語資源の急速な進歩に伴い、さらなる展開が期待される分野である。

しかし、この分野は比較的歴史が浅く、とりわけ国内の言語学コミュニティの中で十分に認知・理解されているとは言い難い。本ワークショップは、形態、統語、意味、談話の諸問題に取り組む計算心理言語学的研究を概観し、隣接分野のコーパス言語学、実験言語学、自然言語処理との関連と違いを明確にしながら、計算心理言語学についての理解を深めることを目的とする。

## 本ワークショップの構成

担当者	発表題目
折田 奈甫, 大関 洋平	イントロダクション
大関 洋平	形態: レキシコンと形態処理の計算モデル
Douglas Roland	統語: Computational Models of Sentence-level Language Comprehension
窪田 悠介	意味: カテゴリ文法・意味計算・文処理
折田 奈甫	談話: 代名詞選択の形式化から計算心理言語学を考える
松林 優一郎	自然言語処理: 「柔らかく」なった自然言語処理
川原 繁人 (コメンテーター)	各発表に対するコメント
全員	全体ディスカッション
企画・司会: 折田 奈甫, 大関 洋平	

## 各発表の要旨

## 形態: レキシコンと形態処理の計算モデル

大関 洋平

計算形態論および自然言語処理における形態素解析では、文字列レベルの変換を仮定する有限状態オートマトンが専ら主流であった (Kaplan & Kay, 1994; Beesley & Karttunen, 2003; Roark & Sproat, 2007)。しかしながら、単純な語が大半を占める有限のレキシコンを効率的に近似できるという計算論的な結果は、人間の形態処理における有限状態オートマトンの心理的実在を必ずしも意味しない。本発表では、文字列だけでなく形態素・品詞・統語構造など様々なレベルを対象とした形態処理の計算モデルを構築し、認知・脳科学で得られる行動・神経活動データに基づいて検証する計算心理言語学のアプローチを概観する。そして、工学的な目的で構築された計算モデルを、人間の言語処理モデルとして科学的な目的のために援用する可能性について議論する。

## 統語: Computational Models of Sentence-level Language Comprehension

Douglas Roland

This presentation will summarize the current generation of computational models of sentence-level language comprehension, addressing the strengths and weaknesses of each model. Particular attention will be paid to data such as that presented in Yun et al. (2012) and Roland et al. (2012), and the difficulties the neighborhood effects found in this data poses for many of the models. The implications of such data on both theories of language comprehension and computational models of sentence-level language comprehension will be discussed.

## 意味: カテゴリ文法・意味計算・文処理

窪田 悠介

カテゴリ文法の一つである CCG が 80 年代に統語変形を用いない文法理論として提案された際、初期の「売り文句」の一つとして、文処理に伴う逐次的な意味構造の構築が可能であるという点が強調された (Ades & Steedman 1982, Altmann & Steedman 1988 など)。心理言語学における計算モデルに関する理論的研究 (Hale 2001, Levy 2008 など) が大きく進展し、統語コーパスや計算機に実装されたパーザなどモデリングに利用可能な言語資源の整備が整ってきた現在、80 年代に提案され、その後様々な歴史的事情によりほぼ全く顧みられることのなかった、カテゴリ文法の心理言語学的モデルとしての可能性を現代的観点から再考察すべき時期が来ているように思われる。本発表では、発表者の観点から、そのような再考察の下準備としていくつかの論点を整理することを試みる。

## 談話: 代名詞選択の形式化から計算心理言語学を考える

折田 奈南

計算心理言語学では、顕著性や新旧情報などの情報を定量化し数理モデルに組み込むことにより、談話レベルの問題の具体的なメカニズムの提案・検証に歩を進めている。本発表では、最初に、談話や語用論レベルの問題を対象とした計算心理言語学的研究を概観する。次に、話者の代名詞選択の形式化を例に、計算心理言語学の主要な理論である Uniform Information Density Hypothesis (Levy and Jaeger 2007) と Rational Speech Act model (Frank and Goodman 2012) ではこの現象を捉えられない可能性を議論する。また、心理言語学の実験で得られた知見を把握した上で数理モデルの妥当性を評価することの重要性についても論ずる。

## 自然言語処理: 自然言語処理: 「柔らかく」なった自然言語処理

松林 優一郎

1990 年代に Penn Treebank が登場して以来、言語処理はコーパスからの統計的な学習手法にもとづいた計算モデル構築により飛躍的な発展を遂げてきた。2010 年代には、数十億文規模の超大規模コーパスとニューラルネットによる表現学習により、言語処理は記号や規則の世界から、「単語ベクトル」「単語系列の構造を計算する行列」といった実数値を使った「柔らかい」表現の上で言語を扱えるように変化し、さらなる飛躍を遂げつつある。本発表では、述語項構造解析の技術を中心に、こうした自然言語処理における歴史的変化や最新の話題に触れつつ、今後の課題点を紹介し、急速に変化を遂げつつある「柔らかい」言語処理と、将来のコーパス言語学や計算心理言語学との接点について議論する。

## 謝辞

本ワークショップ/発表は国立国語研究所共同研究プロジェクト「大規模コーパスを利用した言語処理の計算心理言語学的研究」、「対照言語学的観点から見た日本語の音声と文法」、および JSPS 科研費 JP18K00523, JP18K12357 の研究成果の一部である。