NAACL HLT 2010

First Workshop on
Computational Neurolinguistics

Proceedings of the Workshop

June 6, 2010
Los Angeles, California

Introduction

Welcome to the NAACL-HLT 2010 Workshop on Computational Neurolinguistics.

This is the first workshop to be held on this emerging topic, which integrates recent advances in computational linguistics and cognitive neuroscience with the latest methods from machine learning. This new field promises to aid in the further development of cognitively plausible theories of language, to provide a third empirical basis as a benchmark for computational linguistics (besides corpora, and data elicited from informants), and to enrich the models of language used in neuroscience with the precision and breadth that computational linguistic methods provide. More ambitious blue-sky applications being pursued include language-based brain-computer interfaces, parsing of sentential structure from recordings of neural activity during language processing, and the derivation of language resources from neuroimaging data.

We hope that this event will provide an interdisciplinary forum for the free exchange of ideas between the participants, whose expertise ranges across computational linguistics, cognitive psychology, brain decoding, psycholinguistics and other areas of cognitive science. In preparation for the workshop we released two neural recording data-sets and corresponding language models (the CMU fMRI set and the Trento EEG set, both on a lexical semantic processing task) to allow researchers from different specialties to contribute. The papers that will be presented at the workshop cover a range of neuroimaging techniques (EEG, fMRI, MEG), models of language phenomena (distributional models of lexical semantics, formal ontologies, word class distinctions, connectionist approaches), and of machine learning and data mining methods (Bayesian learning, source separation models, regression techniques, non-linear classifiers).

In addition to the submitted papers we will have two additional talks. We are very happy to welcome Tom Mitchell to open the event. Over recent years Prof. Mitchell has chosen neuroimaging data, and language, as the two phenomena that he will focus on in his machine learning research. In addition we will give a mini-tutorial: a crash course for computational linguists in the neuroscience of language, and on basic principles of neuroimaging techniques.

Brian Murphy, Kai-min Chang, and Anna Korhonen.

Organizers:

Brian Murphy, Centre for Mind/Brain Studies, University of Trento, Italy
Kai-min Kevin Chang, Language Technologies Institute, Carnegie Mellon University, USA
Anna Korhonen, Computer Laboratory, University of Cambridge, UK

Program Committee:

Afra Alishahi, Saarland University, Germany
Ben Amsel, University of Toronto, Canada
Stefano Anzellotti, Harvard University, USA
Colin Bannard, University of Texas Austin, USA
Marco Baroni, University of Trento, Italy
Gemma Boleda, Universitat Politècnica de Catalunya, Spain
Ina Bornkessel, Max Planck Leipzig, Germany
Augusto Buchweitz, Carnegie Mellon University, USA
George Cree, University of Toronto, Canada
Barry Devereux, University of Cambridge, UK
Katrin Erk, University of Texas Austin, USA
Stefan Evert, Unversity of Osnabrück, Germany
Adele Goldberg, Princeton University, USA
Chu-Ren Huang, Hong Kong Polytechnic University, Hong Kong
Aravind Joshi, University of Pennsylvania, USA
Marcel Just, Carnegie Mellon University, USA
Frank Keller, University of Edinburgh, UK
Charles Kemp, Carnegie Mellon University, USA
Mirella Lapata, University of Edinburgh, UK
Chia-Ying Lee, Academia Sinica, Taiwan
Roger Levy, University of California Sand Diego, USA
Angelika Lingnau, University of Trento, Italy
Brad Mahon, University of Rochester, USA
Robert Mason, Carnegie Mellon University, USA
Diana McCarthy, Lexical Computing Ltd, UK
Ken McRae, University of Western Ontario, Canada
Tom Mitchell, Carnegie Mellon University, USA
Fermin Moscoso del Prado Martin, University of Provence, France
Sebastian Pado, University of Stuttgart, Germany
Francisco Pereira, Princeton University, USA
Massimo Poesio, University of Trento, Italy
Thierry Poibeau, CNRS and Ecole Normale Supérieure, France
Dean Pomerleau, Intel Labs Pittsburgh, USA
Ari Rappoport, Hebrew University of Jerusalem, Israel
Brian Roark, Oregeon Health & Science University, USA

Kenji Sagae, University of Southern California, USA
Hinrich Schuetze, Stuttgart University, Germany
Sabine Schulte im Walde, University of Stuttgart, Germany
Svetlana Shinkareva, University of South Carolina, USA
Nathaniel Smith, University of San Diego, USA
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil
David Vinson, University College London, UK
Yang ChinLung, City University of Hong Kong, China

Invited Speaker:

Tom Mitchell, Carnegie Mellon University, USA

Table of Contents

Workshop Program

Sunday, June 6, 2010

9:00–10:30   **Session I**

                    Invited Talk: Tom Mitchell

                    Learning semantic features for fMRI data from definitional text
                    Francisco Pereira, Matthew Botvinick and Greg Detre

10:30–11:00   **Coffee Break**

11:00–12:30   **Session II**

                    Concept Classification with Bayesian Multi-task Learning
                    Marcel van Gerven and Irina Simanova

                    WordNet Based Features for Predicting Brain Activity associated with meanings of nouns
                    Ahmad Babaeian Jelodar, Mehrdad Alizadeh and Shahram Khadivi

                    Network Analysis of Korean Word Associations
                    Jaeyoung Jung, Na Li and Hiroyuki Akama

12:30–1:30   **Lunch**

1:30–3:00   **Session III**

                    Detecting Semantic Category in Simultaneous EEG/MEG Recordings
                    Brian Murphy and Massimo Poesio

                    Hemispheric processing of Chinese polysemy in the disyllabic verb/ noun compounds: an event-related potential study
                    Chih-Ying Huang and Chia-Ying Lee

                    An Investigation on Polysemy and Lexical Organization of Verbs
                    Daniel Germann, Aline Villavicencio and Maity Siqueira

                    Tutorial (Part 1)

Sunday, June 6, 2010 (continued)

3:00–3:30   **Coffee Break**

3:30–5:00   **Session IV**

Tutorial (Part 2)

Acquiring Human-like Feature-Based Conceptual Representations from Corpora
Colin Kelly, Barry Devereux and Anna Korhonen

Using fMRI activation to conceptual stimuli to evaluate methods for extracting conceptual representations from corpora
Barry Devereux, Colin Kelly and Anna Korhonen

5:00–6:00   **Discussion**