

Language Sciences Annual Symposium

Thursday 17 November 2016, Emmanuel College

University of Cambridge

13:15-13:50 Registration and coffee; poster exhibition (Old Library)

Talks will take place in the main lecture theatre, Queen's Building

14:00-15:00 Introduction by **Professor Robert Foley**, Co-Chair, Cambridge Language Sciences and Leverhulme Professor of Human Evolution

Professor Lorraine Tyler, Dept. of Psychology (Chair: Professor Ianthi Tsimpli, Dept. of Theoretical & Applied Linguistics)

Language dynamics: a neurocognitive approach to incremental interpretation

15:00-15:30 **Dr Nigel Collier**, Dept. of Theoretical & Applied Linguistics (Chair: Professor Ianthi Tsimpli)

Natural Language Processing and online health reports (or OMG U got flu?)

15:30-16:00 Group session by the poster presenters (Chair: Guy Emerson, Computer Laboratory)

16:00-16:30 Refreshments; poster session (Old Library)

16:30-17:00 **Professor Ann Copestake**, Computer Laboratory (Chair: Professor Ted Briscoe, Computer Laboratory)

Does natural language understanding have anything to do with understanding natural language?

17:00-18:00 **Professor Simon Fisher**, Max Planck Institute for Psycholinguistics, Nijmegen (Chair: Professor Robert Foley)

A molecular genetic perspective on speech and language

18:00-19:30 Wine and canapés Offered with the generous support of Cambridge University Press

Language dynamics: a neurocognitive approach to incremental interpretation

Professor Lorraine K. Tyler

Professor of Cognitive Neuroscience, Dept. of Psychology, University of Cambridge



Lorraine Tyler heads the *Centre for Speech, Language and the Brain*, an interdisciplinary research group which combines neuroimaging neuropsychological and behavioural methods to reveal how the human brain is organised to support language, perception and meaning. She also leads the *Cambridge Centre for Ageing and Neuroscience (Cam-CAN)*, a University-wide consortium to study how age-related changes in brain structure and function relate to patterns of preserved and declining cognitive functions with age.

Abstract: Understanding spoken language involves a complex set of processes that transform the auditory input into a meaningful interpretation. Our percept is not of acoustic-phonetic detail but of the speaker's intended meaning. This transition occurs on millisecond timescales, with remarkable speed and accuracy, and with no awareness of the complex computations on which it depends. How is this achieved? What are the processes and representations that support the transition from sound to meaning, and what are the neurobiological systems in which they are instantiated? Surprisingly little is known about the specific spatio-temporal patterning and the specific neuro-computational properties of this complex dynamic system. In current research we address these issues by combining advanced techniques from neuroimaging, multivariate statistics and computational linguistics to probe the dynamic patterns of neural activity that are elicited by spoken words and the incremental processes that combine them into syntactically and semantically coherent sentences. Computational linguistic analyses of language corpora enable us to build quantifiable models of different dimensions of language interpretation – from phonetics and phonology to argument structure and semantic integration - and we test for their presence using multivariate methods on combined electro- and magneto encephalography (EMEG) data, as the utterance unfolds in real time. In this talk, I will present the novel account of speech comprehension that is emerging from this research.

Natural Language processing and online health reports (or OMG U got flu?)

Dr Nigel Collier

Language Technology Laboratory, Dept. of Theoretical & Applied Linguistics



Nigel is PI and Director of Research in Computational Linguistics at the Department of Theoretical and Applied Linguistics. He was awarded a PhD in computational linguistics from the University of Manchester (UMIST) in 1996 for his work on *Lexical Transfer using a Hopfield Neural Network*. He was subsequently awarded a Toshiba Fellowship to continue his research on neural networks for machine translation and then joined the NLP group at the University of Tokyo where he coordinated the GENIA text mining project. After becoming faculty at the National Institute of Informatics (NII) in 2000, Nigel led the BioCaster research programme (2006 to 2012) for multilingual news surveillance and served as technical advisor to the G7's Global Health Security Action Group's working group on Risk Management and Communication. He was awarded a Marie Curie fellowship at the European Bioinformatics Institute from 2012 to 2014 where he continued his investigation into biomedical text mining for scientific texts.

Nigel's research interests are in information extraction and biomedical knowledge discovery with a focus on machine learning approaches for representation learning of concepts. He is the author of over 110 peer-reviewed articles and conference papers on biomedical NLP. He currently leads the EPSRC-funded *Semantic Interpretation of Personal Health messages* (SIPHS) project which investigates biomedical concept encoding of laymen's terms in the social media for real world applications such as digital disease surveillance.

<https://sites.google.com/site/nhcollier/>

Abstract: Online health information is widely reported by individuals in social media, chat rooms, discussion boards and also by the news media. These informal sources of evidence about our individual health, attitudes and behaviours are driving the development of new techniques for studying human health in areas ranging from real-time disease surveillance, to understanding mental illness, to providing evidence for new applications of drugs. Informal patient data on the Web is increasing, accessible, low cost and seems likely to cover a greatly expanded population compared to traditional survey methods. However in order to understand and integrate this data researchers in Natural Language Processing (NLP) must grapple with theoretical, practical and ethical challenges. For example: How can machines achieve fine-grained analysis and understanding of laymen's health language? How does online health report data complement traditional survey data? How can we integrate online health data with other data sources such as databases and ontologies? What benefit could there be in a longitudinal analysis of an individual's online health reports over a period of time? My talk will illustrate answers to these questions in the light of our recent attempts to harness social and news media as a new type of signal for understanding human health.

Does natural language understanding have anything to do with understanding natural language?

Professor Ann Copestake

Professor of Computational Linguistics, Computer Laboratory

Ann Copestake's research is in Computational Linguistics/Natural Language Processing - developing computer models of human languages (or, more precisely, models of some aspects of human languages). In conjunction with DELPH-IN, an informal international consortium, she has developed software which has been used to develop formal computational accounts of the syntax and compositional semantics of many different languages. Her current research mainly concerns the development of models of compositional and lexical semantics which are compatible with broad-coverage computational processing (parsing and generation). She has worked on a variety of application areas including scientific text processing, information extraction, augmentative and alternative communication (AAC), machine translation, Natural Language Interfaces, lexical acquisition and on tools for lexicographers.



Abstract: Natural language understanding (NLU) and natural language generation (NLG) are the main general goals of computational work on human language. Does such work have anything to tell us about the scientific/linguistic goal of understanding how natural languages behave? Can better understanding of linguistics help the development of practical computational techniques? It is far from obvious that the answer to these questions is yes, since successful computational modelling does not necessarily imply any real understanding. For instance, trajectories of physical objects can be modelled without understanding the underlying physics. Fred Jelinek, in a talk given at an award ceremony in 2004, admitted making the notorious comment 'every time I fire a linguist, our system performance improves' but argued that the goals of practical speech recognition simply did not coincide with the interests of the linguists. However, in this talk, I will suggest some more positive answers to these questions. I will describe DELPH-IN, a long-standing international collaboration involving researchers who are explicitly addressing both linguistic and computational goals, and discuss some of its successes. I'll outline how the development of precision grammars for various languages allows us to investigate language scientifically, and also to build practical systems for end users. I'll conclude by speculating how this sort of work can further progress in the brave new computational world of deep learning.

A molecular genetic perspective on speech and language

Professor Simon Fisher

Director, Max Planck Institute for Psycholinguistics, Nijmegen

Professor of Language and Genetics, Donders Institute for Brain, Cognition and Behaviour, Nijmegen



Simon Fisher obtained his Natural Sciences degree at Trinity Hall, Cambridge followed by a DPhil at the Genetics Unit of the Biochemistry Department, Oxford University. For his postdoctoral research he joined Prof. Anthony Monaco's group at the Wellcome Trust Centre for Human Genetics (WTCHG) in Oxford, and worked on identifying genetic factors that contribute to developmental disorders such as dyslexia and speech and language impairments. During this time he and his colleagues discovered FOXP2, the first case of a gene mutated in speech and language impairment. In 2002, Simon was awarded a Royal Society Research Fellowship and became head of his own laboratory at the WTCHG, where he used state-of-the-art methods to uncover how language-related genes influence the brain. From 2007-10 Simon was also the Isobel Laing Fellow in Biomedical Sciences at Oriel College, Oxford, where he taught Biochemistry and Medical Genetics. In 2010 he was appointed Director of a new department specifically devoted to "Language and Genetics" at the Max Planck Institute in Nijmegen.

Abstract: The rise of molecular technologies has yielded exciting new routes for studying the biological foundations of language. In particular, researchers have begun to identify genes implicated in neurodevelopmental disorders that disrupt speech and language skills. My talk will illustrate how such work can provide powerful entry points into critical neural pathways, using FOXP2 as an example. Rare mutations of this gene cause problems with learning to sequence mouth movements during speech, accompanied by wide-ranging deficits in language production and comprehension. FOXP2 encodes a regulatory protein, a hub in a network of other genes, several of which have also been associated with language-related impairments. Versions of FOXP2 are found in similar form in many vertebrate species; indeed, studies of animals and birds suggest it has conserved roles in the development and plasticity of certain sets of neural circuits. Thus, the contributions of this gene to human speech and language involve modifications of evolutionarily ancient functions. Searches for additional language-related genes are underway, taking advantage of dramatic advances in genomic methods. Overall, the FOXP2 story illustrates the value of an interdisciplinary approach for unravelling the complicated connections between genes, neurons, brain circuits and language.

Posters

Poster exhibition organised by Guy Emerson (Computer Laboratory) and Jamie Douglas (Dept. of Theoretical & Applied Linguistics). Speaker session chaired by Guy Emerson.

Titles & authors

Acquiring production skills through implicit learning - Giulia Bovolenta, John Williams

Acquisition of quantity, relevance and word learning inferences, and their relationship with Theory of Mind - Elspeth Wilson, Napoleon Katsos

Automated speech-unit delimitation in spoken learner English – Russell Moore, Andrew Caines, Calbert Graham, Paula Buttery

Automatic assessment and error detection of non-native English speech using phone distance features - Konstantinos Kyriakopoulos, M.J.F. Gales, Kate Knill

Constrained multi-task learning for automated essay scoring - Ronan Cummins, Meng Zhang, Ted Briscoe

Cross linguistic influences on the acquisition of metaphorical expressions – Chris Mengying Xia

Crowdsourcing a multilingual speech corpus: recording, transcription and annotation of the CrowdED Corpus - Andrew Caines, Christian Bentz, Calbert Graham, Paula Buttery

Deep density networks with uncertainty for automatic assessment of spontaneous spoken English - Andrey Malinin

Evaluating multi-modal deep learning systems with micro-worlds - Alexander Kuhnle, Ann Copestake

Grammatical constraints on lexical and structural processing strategies: EMEG evidence from Russian morphosyntax - Anastasia Klimovich-Gray, Elisabeth Fonteneau, William Marslen-Wilson

Graph- and surface-level sentence chunking - Ewa Muszyńska

Multi-level representations in speech processing in brain and machine: Evidence from EMEG and RSA - Cai Wingfield, Li Su, Barry Devereux, Xunying Liu, Chao Zhang, Phil Woodland, Elisabeth Fonteneau, Andrew Thwaites, William Marslen-Wilson

Neuro-computational modelling of lexico-syntactic representation and integration during speech comprehension - Hun S. Choi, Billi Randall, Barry J. Devereux, Lorraine K. Tyler

Overnight changes in the neural representation of newly-acquired inflectional affixes: univariate and multivariate fMRI evidence - Lydia Viñals, Jelena Mirković, Gareth Gaskell, Matt Davis

Pitch-interval analysis of 'periodic' and 'aperiodic' Question + Answer pairs - Juan Pablo Robledo del Canto, Sarah Hawkins, Ian Cross, Richard Ogden

Sublexical morpheme stripping - Luca Cilibrasi

The development of svarabhakti vowels in the history of Norwegian - Tam Blaxter

The effect of physical activity on implicit language learning in children and adults - Carla Pastorino Campos

The effects of attention and interference on the neural encoding of continuous speech - Andrea Olguin, Tristan Bekinschtein, Mirjana Bozic

The interaction of syntactic structure and lexical constraints during sentence processing - Barry Devereux, Billi Randall, William Marslen-Wilson, Lorraine K. Tyler

The neuromagnetic time-course of semantic ambiguity resolution in speech comprehension - Lucy J. MacGregor, Jennifer M. Rodd, Olaf Hauk, Matt Davis

Presenters (plenary session)

Tam Blaxter, Giulia Bovolenta, Andrew Caines, Carla Pastorino Campos, Juan Pablo Robledo del Canto, Hun Choi, Luca Cilibrasi, Ronan Cummins, Barry Devereux, Anastasia Klimovich-Gray, Alexander Kuhnle, Konstantinos Kyriakopoulos, Lucy MacGregor, Andrey Malinin, Ewa Muszyńska, Andrea Olguin, Lydia Vinals, Elspeth Wilson, Cai Wingfield, Chris Mengying Xia