

Dr Paula BATTERY

University Senior Lecturer

I am a computational linguist with research interests in both computer applications (Natural Language Processing) and language cognition (Computational Psycho-linguistics). My research focuses on building Natural Language Processing tools that work with non-canonical forms of natural language (spoken language, learners, aphasics) and also with low resource languages (endangered languages, dialects). I am interested in both the automatic machine processing of non-canonical language and the cognitive processes underlying that language. Understanding the cognitive aspects of language processing is essential if we are to build language tools that can be intuitive to users. I am currently a Principal Investigator in the Automated Language Teaching and Assessment (ALTA) project. My work within ALTA focuses on the spoken language of learners of English; but I am also funded to work on under-resourced endangered languages as part of the Cambridge Africa programme which partners Cambridge researchers with colleagues in Uganda and Ghana. Other research interests include the computational modelling of first and second language acquisition and language evolution. For full details see research contributions from page 5.

I won a Cambridge University Pilkington Prize for teaching in 2015 and was short listed for the Cambridge University Student Union Student-Led Teaching Awards 2015 in both the Lecturer and Supervisor categories. For teaching contributions see page 2.

Contact Details

Address: Gonville and Caius College, Trinity Street, Cambridge, CB2 1TA, UK

Email: paula.battery@cl.cam.ac.uk

GENERAL INFORMATION

Current and Recent Academic Positions

University Senior Lecturer **Computer Laboratory**, University of Cambridge, 2016+

Senior Lecturer **Department of Theoretical and Applied Linguistics**, University of Cambridge, 2014–2016

Lecturer **Department of Theoretical and Applied Linguistics**, University of Cambridge, 2011–2014

Permanent Senior Research Associate (with contracted teaching) **Research Centre for English and Applied Linguistics**, University of Cambridge, 2007–2011

Research Associate (with contracted teaching) **Research Centre for English and Applied Linguistics**, University of Cambridge, 2005–2007

Associated Researcher **Department of Theoretical and Applied Linguistics**, Modern and Medieval Languages, University of Cambridge, 2016+; **Natural Language and Information Processing Group**, **Computer Laboratory**, University of Cambridge, 2006–2016

Fellow: College Lecturer in Computational Science and Linguistics, **Gonville and Caius College**, University of Cambridge.

Director of Studies **Director of Studies in Linguistics** Gonville and Caius, April 2013+; Sidney Sussex, Oct 2013+; Wolfson 2010–2012, Hughes Hall 2010+; **Director of Studies in Computer Science** Hughes Hall, 2009–2015

Other Associate Positions Teaching Associate, **Queens' College** 2012; Research Associate, Department of Linguistics, **University of California**, Davis Campus, 2009–2010

Industry Positions and Employment

Senior Technical Officer Information Extraction & Data Mining Engineer **European Bioinformatics Institute**, European Molecular Biology Laboratory Outstation, Wellcome Trust Genome Campus, Hinxton, 2010–2013

Enterprise Positions Enterprise Champion, **Cambridge Enterprise**, University of Cambridge, 2008+; Business Fellow, **London Technology Network**, 2007–2009

University Spin-out **iLexIR**, Natural Language Processing consultancy, 2006+

Internships Research Engineer **Google**, Mountain View, CA, USA, 2005; Research Assistant **Dragon Systems**, Cheltenham, UK, 2000; Research Analyst, **Intel Corporation**, Swindon, UK, 1999

Education

PhD **Computer Science (Computational Linguistics)** under supervision of Ted Briscoe, Computer Laboratory, Churchill College, University of Cambridge

MPhil **Advanced Computer Science** (Computer Speech, Text and Internet Technology), Department of Engineering & Computer Laboratory, Churchill College, University of Cambridge

BA (hons) **Computer Science**, Churchill College, University of Cambridge

Awards and Scholarships

2015 Pilkington Teaching Prize, University of Cambridge

2015 Cambridge University Student Union Student-Led Teaching Awards 2015 (shortlisted)

2002–2005 Engineering and Physical Sciences Research Council Award

2000–2001 Engineering and Physical Sciences Research Council Award

1998–2000 Intel Corporation Studentship

1997–2000 National Electronics Council Scholarship

TEACHING

Faculty/Department Teaching Administrative Duties

Directory of Undergraduate Teaching Undergraduate Teaching Co-ordinator for the Linguistics Tripos in the Department of Theoretical & Applied Linguistics, Modern and Medieval Languages.

This is a non-stipendiary position which involves **overseeing all department and College teaching** for papers in the Linguistics Tripos (because **supervisions for Linguistics are organised at department level**). The Linguistics department has circa 90 students. Linguistics papers are also borrowable by students in ASNaC, AMES, Classics, MML, PBS as well as Erasmus students. 2012–2016.

Member of the Undergraduate Studies Committee Sole representative from Department of Theoretical & Applied Linguistics on a Faculty committee which oversees all teaching standards and teaching strategy in Modern and Medieval Languages. 2012–2016

Member of the Faculty Board Member of a Faculty committee which oversees the running (including teaching) of all of the Departments in Modern and Medieval Languages. 2013–2016

Co-ordinator of Li18 Computational Linguistics Undergraduate Paper Co-ordinator in the Department of Theoretical & Applied Linguistics. I created this paper in 2012. It is attended by both undergraduate and MPhil students. The co-ordinator roll involves the organisation of all supervisions, practical sessions and examining for this year-long undergraduate paper. 2012+

Co-ordinator of MPhil Research Methods Seminars Postgraduate Course Co-ordinator in the Department of Theoretical & Applied Linguistics. Involved co-ordinating 32 hours of research methods training for postgraduate students. 2012–2016

Co-ordinator of Computational Linguistics for MPhils Postgraduate Course Co-ordinator in the Department of Theoretical & Applied Linguistics, Modern and Medieval. This is an advanced course that follows on from the undergraduate course in Computational Linguistics. Involved co-ordinating 16 hours of lecture based tutorials for postgraduate students. 2011–2015

Examiner for Linguistics Tripos One of two overseeing Examiners for the Linguistics Tripos, responsibilities include: proof reading Linguistics Tripos examination papers (15 papers per year); attending Examiners meetings. 2015–2016

Faculty Teaching

Undergraduate Teaching

Paper Co-ordinator and Lecturer, Li18 Computational Linguistics: 16 lectures plus 4 practicals, **Part II Linguistics Tripos**. I developed this new paper in 2012. I deliver all contact hours for this paper.

Guest Lecturer, Natural Language Processing: guest lecture in Computational Psycholinguistics within the Natural Language Processing course, **Part II Computer Science Tripos**.

Organiser, Dissertation Lectures: I organise 5/6 research training lectures for **Part II Linguistics Tripos**. Introduced to the Tripos in 2012. Topics also lectured by me: Corpus Linguistics; How to Choose a Dissertation Topic.

Organiser and Lecturer, Information Lectures: Organiser of 5 information lectures, **Parts I & II, Linguistics Tripos**. Lectures include: Welcome Lectures; Paper Choices Lectures; ‘Switching to Linguistics Lecture’ for students in MML. All lectures delivered by me in the role of Director of Undergraduate Teaching since 2012.

Postgraduate Teaching

MPhil Course Co-ordinator, Research Methods for Linguists: Co-ordinator of 32 hours of research training lectures, **MPhil in Theoretical and Applied Linguistics**. Topics also lectured by me: How to Read a Paper; Presenting Data; Corpus Linguistics. Co-ordinator since 2012.

MPhil Course Co-ordinator and Lecturer, Advanced Computational Linguistics: Lecturer of 16 hours of lecture based tutorials tailored to suit student interests, **MPhil in Theoretical and Applied Linguistics**. Topics selected by students including: Information Retrieval; Computational Models of Language Acquisition; Advanced Corpus Linguistics; Machine Translation; Computational Psycholinguistics. Developed as a new course in 2011.

Seminar Convenor, Computational Linguistics Research Cluster: Organiser of 16 hours of talks and tutorials, **PhD in Theoretical and Applied Linguistics**. Topics also delivered by me: Programming in R. Research cluster introduced in 2011.

Postgraduate Lecturer, Occasional Skills Lectures: MPhil General Seminars, **MPhil in Theoretical and Applied Linguistics**, topic lecturer: ‘When is a Linguistic Model good enough?’. PhD-training lectures, **PhD in Theoretical and Applied Linguistics**, **PhD in Applied Linguistics**, **PhD in Computation Cognition and Language**, topics lectured: ‘Programming in R’; ‘How to Draw a Graph’; ‘The Data Pipeline: Collection, Storage and Processing’; ‘How to give a Lecture’.

Course Co-ordinator and Lecturer, Advanced courses in Computational Linguistics: three 16 hour courses run in tri-yearly rotation, **MPhil in Applied Linguistics**, **PhD in Applied Linguistics**. Courses included Introduction to Computational Linguistics; Computational Language Learning; Computational Psycho-linguistics. Developed as new courses from 2005. Ran until 2010.

MPhil Lecturer, Corpus Linguistics Lecturer of 6 hours of training lectures (part of Research Methods Training), **MPhil in Applied Linguistics**. Developed as a new lectures in 2005. Ran until 2010.

Summary of Courses Developed

Developed Courses: Li18 Computational Linguistics **New Undergraduate Paper**, 2012; Introduction to Computational Linguistics **New Postgraduate Course**, 2005; Computational Language Learning **New Postgraduate Course**, 2006; Computational Psycho-linguistics **New Postgraduate Course**, 2007; Advanced Computational Linguistics **New Postgraduate Course**, 2011

Redesigned Course: Research Methods for Linguists **Redesigned Postgraduate Course**, 2012

Examining

Examining has been undertaken for the Computer Laboratory; Department of Theoretical and Applied Linguistics (DTAL); and the former Research Centre for English and Applied Linguistics (RCEAL), which had a taught MPhil with written examinations.

PhD Dissertations **PhD in Computer Science**: 4 Internal (Computer Laboratory), 1 external (University of Limerick). 2011+

PhD 1st Year Vivas **PhD in Computer Science** Computer Laboratory, 2012+

MPhil Dissertations **MPhil in Theoretical and Applied Linguistics**, Department of Theoretical and Applied Linguistics, 2011+; **MPhil in Advanced Computer Science**, Computer Laboratory, 2010+; **MPhil in Computer Speech Text and Internet Technology**, Computer Laboratory, 2006–2010

MPhil Long Essays **MPhil in Theoretical and Applied Linguistics**, Department of Theoretical and Applied Linguistics, 2011+; **MPhil in Applied Linguistics**, Research Centre for English and Applied Linguistics, 2005–2010

MPhil Written Exams in Semantics and Language Acquisition, **MPhil in Applied Linguistics**, Research Centre for English and Applied Linguistics, 2005–2010

Undergraduate Written Exams in Li18 Computational Linguistics, **Linguistics Tripos**, Department of Theoretical and Applied Linguistics, 2012+ (includes setting the examination paper)

Postgraduate Supervision

i) Supervision of Research Projects in Linguistics

2012–2015 Adaptive Grammar: Towards a quantitative model of language evolution and change.

Supervisor of Ph.D. thesis by *Christian Bentz*

2013 Systematic translation mismatches and semantic transfer in Japanese-English machine translation.

Supervisor of MPhil thesis in Linguistics by *Alicja Piotrkowicz*.

2009–2010 Methods for the extraction and evaluation of Criterial Features for English.

Supervisor of Ph.D thesis by *Stephen Spencer*

2006–2007 An automatic text classification approach to developmental profiling in second language acquisition.

Supervisor of Ph.D thesis by *Caroline Williams*.

2006–2010 ‘You talking to me?’ Zero auxiliary constructions in British English.

Supervisor of Ph.D thesis by *Andrew Caines*

ii) Supervision of Research Projects in Computer Science

Supervision of MPhil Projects in Computer Science involves weekly supervisions of 1 hour.

2016- Adaptive Learning Strategies for Language Learning

Supervisor of Ph.D. thesis by *Russell Moore*

2010 Language Models for Language Learning on Mobile Devices.

Supervisor of MPhil thesis in Advanced Computer Science by *Duncan Roberts*.

2008 Integrating simulation and experiment: A study in Japanese Second Language Acquisition.

Supervisor of M.Phil thesis by *Lu Gram*.

2008 Automatic analysis of child language.

Co-supervisor of M.Phil thesis by *Andrew Thwaites*

2008 Features detection for automated poetry analysis.

Co-supervisor of M.Phil thesis by *Marcus Duyzend*.

2006 Acquiring Grammatical Fingerprints for genres.

Co-supervisor of M.Phil thesis by *Zheng Mang*.

iii) *Advisor / Research Committee Member for PhDs in Linguistics*

2013+ Computational Models of Implicit Language Learning.

1st Advisor for Ph.D thesis by *Dimitrios Alikaniotis*

2010–2011 An emergent perspective on L1 influence in the L2 acquisition of English grammatical morphemes.

Research committee member for Ph.D. thesis by *Akira Murakami*

2009–2011 Evaluation of syntactic complexity metrics: diachronic analysis of the dative shift.

Research committee member of Ph.D thesis by *Chris Veysey*

2008–2009 Development of discourse features by Czech and Hungarian learners of English.

Research committee member of Ph.D thesis by *Nobert Vanek*

2008–2009 The acquisition of numerically quantified expressions.

Research committee member of Ph.D thesis by *Chris Cummins*

2006–2007 The acquisition of tense and agreement morphology in L2 English.

Research committee member of Ph.D thesis by *Tae Yamazaki-Hasegawa*

2005–2006 The assignment of stress in English noun-noun combinations.

Research committee member of Ph.D thesis by *Melanie Bell*

Cambridge College Teaching

College Lecturer in Computer Science and Linguistics **Gonville and Caius College**, April 2013+

Director of Studies in Linguistics **Gonville and Caius**, April 2013+; **Sidney Sussex**, Oct 2013+; **Wolfson** 2010–2012;
Hughes Hall 2010+

Director of Studies in Computer Science **Hughes Hall** 2009–2015

Teaching Associate **Queens' College** 2011–2012

Undergraduate Supervisor Computational Linguistics (Linguistics Tripos); Probability (Computer Science Tripos);
Regular Languages (Computer Science Tripos); Discrete Mathematics (Computer Science Tripos); Information
Retrieval (Computer Science Tripos); Natural Language Processing (Computer Science Tripos) 2005+

Part II (final year) Project Supervisor Computer Science 2012+; Linguistics 2013+

Other University Teaching

Teaching Demonstrator Automated Question Answering Practical (MPhil in Computer Speech Text & Internet
Technology) 2003–2004

Practical Class Examiner (Ticker) Assessed Mathematics Exercise (Natural Sciences Tripos); Programming in Java
(Computer Science Tripos) Programming in ML (Computer Science Tripos) 2002–2005

RESEARCH CONTRIBUTIONS

Scholarly Publications and Output

i) Peer Reviewed Publications

Journal Articles

2016 Bentz, C., Alikaniotis, D., Samardzic, T. and Buttery, P. (accepted). *Variation in word frequency distributions: Definitions, measures and implications for a corpus-based language typology*. **Journal of Quantitative Linguistics**. (49 pages)

2016 Katshemererwe, F., Caines, A. and Buttery, P. (accepted). *Building natural language processing tools for Runyakitara*. **Applied Linguistics Review**. (29 pages)

- 2015 Bentz, C., Verkerk, A., Kiela, D., Hill, F. and Buttery, P. *Adaptive communication: Languages with more non-native speakers tend to have fewer word forms* **PLOS ONE**.
<http://dx.doi.org/10.1371/journal.pone.0128254> (23 pages)
- 2015 Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, R., Buttery, P. and Marslen-Wilson, W. *Tracking cortical entrainment in neural activity: Auditory processes in human temporal cortex*. In **Frontiers in Computational Neuroscience**. 9:5
<http://dx.doi.org/10.3389/fncom.2015.00005> (14 pages)
- 2014 Bentz, C., Kiela, D., Hill, F. and Buttery, P. *Zipf's law and the grammar of languages: A quantitative study of Old and Modern English parallel texts*. 10:2, p175—211. **Corpus Linguistics and Linguistic Theory**. (37 pages)
- 2010 McEntyre, J.M., Ananiadou, S., Andrews, S., Black, W.J., Boulderstone, R., Buttery, P., Chaplin, D., Chevuru, S., Cobley, N., Coleman, L., Davey, P., Gupta, B., Haji-Gholam, L., Hawkins, C., Horne, A., Kim, J., Lewin, I., Lyte, V., MacIntyre, R., Mansoor, S., Mason, L., McNaught, J., Newbold, E., Nobata, C., Ong, E., Pillai, S., Rebholz-Schuhmann, D., Rosie, H., Rowbotham, R., Rupp, C.J., Stoehr, P., and Vaughan, P. *UKPMC: a full text article resource for the life sciences*. **Nucleic Acids Research**, 39: D58-D65 (8 pages).
- 2010 Hawkins, J. and Buttery P. *Criterion Features in Learner Corpora: Theory and Illustrations*. **English Profile Journal**, 1: e5 (23 pages).

Books and Book Chapters

- 2016 Caines, A. and Buttery, P. (in press). *The effect of examination topic on opportunity of use in learner language*. In: Vaclav Brezine and Lynne Flowerdew (eds.), **Written and spoken learner corpora and their use in different contexts**. London: Bloomsbury. (34 pages)
- 2013 Buttery P, McCarthy M, Carter R. *Chatting in the academy: informality in spoken academic discourse*. In: Suganthi, C., Groom, N. and Charles, M. (eds), **Corpus and Academic Discourse**. Amsterdam: John Benjamins. (33 pages)
- 2012 Hawkins, J. and Filipovic, L. with Buttery, P., Capel, A., Hawkey, R., Salamoura, A., Saville, N., and Trim, J. *Criterion features in L2 English: specifying the reference levels of the Common European Framework*. In: Millanovic, M. and Saville, N. (eds), **English Profile Studies**, Volume 1, Cambridge: Cambridge University Press. (183 pages)
- 2012 Buttery, P. and Caines, A. *Normalising Frequency Counts to Account for 'opportunity of use' in Learner Corpora*. In: Tono, Y., Kawaguchi, Y., and Minegishi, M. (eds.), **Developmental and Crosslinguistic Perspectives in Learner Corpus Research**. Amsterdam: John Benjamins. p187—204 (18 pages).
- 2011 Buttery, P. and McCarthy, M. *Lexis in Spoken Discourse*. In: Gee, J., and Handford, M. (eds.), **The Routledge Handbook of Discourse Analysis**. London: Routledge. p285—300 (16 pages).
- 2009 Hawkins, J. and Buttery, P. *Using Learner Language from Corpora to Profile Levels of Proficiency: Insights from the English Profile Programme*. In: Taylor, L. and Weir, C. (eds.), **Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment**. Cambridge: Cambridge University Press. p158—175 (18 pages).
- 2008 Briscoe, E. J. and Buttery, P. *Linguistic adaptations for resolving ambiguity*. In: Smith, A. D. M., Smith, K., and Ferrer i Cancho, R. (eds.), **The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)**. Barcelona, Spain. Singapore: World Scientific Press. p51—58 (8 pages).

Published Conference Proceedings

- 2016 Moore, R., Caines, A., Graham, C. and Buttery, P. *Automated speech-unit delimitation in spoken learner English*. Proceedings of the **26th International Conference on Computational Linguistics (COLING 2016)**, Osaka, Japan. (12 pages)
- 2016 Graham, C., Buttery, P. and Nolan, F. *Vowel characteristics in the assessment of L2*. Proceedings of **INTER-SPEECH**, San Francisco, USA. (5 pages)
- 2016 Caines, A., Bentz, C., Alikaniotis, D., Katshemererwe, F. and Buttery, P. *The Glottolog data explorer: Mapping the world's languages*. Proceedings of the **Tenth International Conference on Language Resources and Evaluation (LREC'16)**, Portoroz, Slovenia, 2016. (8 pages)

- 2016 Caines, A., Bentz, C., Graham, C., Polzehl, T. and Buttery, P. *Crowdsourcing a multilingual speech corpus: recording, transcription and annotation of the CROWDED CORPUS*. Proceedings of the **Tenth International Conference on Language Resources and Evaluation (LREC'16)**, Portoroz, Slovenia, 2016. (8 pages)
- 2016 Zhang, W., Caines, A., Alikaniotis, D and Buttery, P. *Predicting author age from Weibo microblog posts*. Proceedings of the **Tenth International Conference on Language Resources and Evaluation (LREC'16)**. Portoroz, Slovenia, 2016. (8 pages)
- 2015 Moore, R., Caines, A., Buttery, P. and Calbert G. *Incremental dependency parsing and disfluency detection in spoken learner English*. Proceedings of **Text, Speech, and Dialogue (TSD)** Pilsen, Czech Republic. (9 pages)
- 2015 Graham, C., Buttery, P., Caines, A. and Nolan, F. *Automated Assessment of non-native speech using spectral features of vowels*. Proceedings of **INTERSPEECH**, Dresden, Germany. (5 pages)
- 2015 Caines, A., Bentz, C., Graham, C., Polzehl, T. and Buttery, P. *Crowdsourcing a multi-lingual speech corpus: recording, transcription, and natural language processing*. Proceedings of **INTERSPEECH** Dresden, Germany. (5 pages)
- 2014 Caines, A. and Buttery, P. *The effect of disfluencies and learner errors on the parsing of spoken learner language*. Proceedings of the **First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages** p74—81. **Computational Linguistics (CoLing)**, Dublin, Ireland. (8 pages)
- 2014 Bentz, C. and Buttery, P. *Towards a computational model of grammaticalization and lexical diversity*. Proceedings of the **5th Workshop on Cognitive Aspects of Computational Language Learning (CogACLl)**. Gothenburg, Sweden. p38—42 (5 pages)
- 2012 Buttery, P. and Caines, A. *Reclassifying subcategorization frames for experimental analysis and stimulus generation*. Proceedings of the **Eighth International Conference on Language Resources and Evaluation (LREC'12)**. Istanbul, Turkey. p1694—1698 (5 pages)
- 2012 Caines, A. and Buttery, P. *Annotating progressive aspect constructions in the spoken section of the British National Corpus*. Proceedings of the **Eighth International Conference on Language Resources and Evaluation (LREC'12)**. Istanbul, Turkey. p1699—1703 (5 pages)
- 2010 Caines, A. and Buttery, P. *'You Talking to Me?' A predictive model for zero-auxiliary constructions*. Proceedings of the **Workshop on NLP and Linguistics, Finding the Common Ground, Association for Computational Linguistics (ACL-2010)**. Uppsala, Sweden. p43—51 (9 pages).
- 2010 Thwaites, A., Geertzen, J., Marslen-Wilson, W., and Buttery, P. *Lexical Isolation Point Software*. Proceedings of the **Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Malta. p3727—p3731 (5 pages)
- 2010 Williams, C., Thwaites, A., Buttery, P., Geertzen, J., Randall, B., Shafto, M., Devereux, B., and Tyler, L. *The Cambridge Cookie-Theft Corpus: A Corpus of Spontaneous and Directed Speech of Brain-Damaged Patients and Healthy Individuals*. Proceedings of the **Seventh International Conference on Language Resources and Evaluation (LREC'10)**. Malta. p2824—2830 (7 pages)
- 2009 Vlachos, A., Buttery, P., Ó Séaghdha, D., and Briscoe, E. J. *Biomedical Event Extraction without Training Data*. Proceedings of **BioNLP, North American Chapter of the Association for Computational Linguistics (NAACL 2009)**, Boulder, Colorado. p37—40 (4 pages)
- 2009 Rice, A., Buttery, P., Rai, I., and Beresford, A. *Language learning on a next-generation service platform for Africa*. Proceedings of **Africa Perspective on the Role of Mobile Technologies in Fostering Social and Economic Development, W3C Workshop**. Maputo, Mozambique. (7 pages)
- 2007 Buttery, P. and Korhonen, A. *'I will shoot your shopping down and you can shoot all my tins'; Automatic lexical acquisition from the CHILDES Database*. Proceedings of the **Workshop on Cognitive Aspects of Computational Language Acquisition**, Association for Computational Linguistics (ACL-2007). Prague, Czech Republic. p33—40 (8 pages).
- 2007 Buttery, P., Villavicencio, A., and Korhonen A. (eds.). Proceedings of the **Workshop on Cognitive Aspects of Computational Language Acquisition**. Prague, Czech Republic. Association for Computational Linguistics. (97 pages)

- 2006 Buttery, P. *Computational models for first language acquisition*. **Ph.D. thesis published as a technical report UCAM-CL-TR-675**. Computer Laboratory, University of Cambridge. (176 pages)
- 2005 Buttery, P. and Korhonen, A. *Large-scale analysis of verb subcategorization differences between child directed speech and adult speech*. Proceedings of the **Verb Workshop 2005: Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes**. Saarland University, Germany. (6 pages)
- 2004 Buttery, P. *A quantitative evaluation of naturalistic models of language acquisition; the efficiency of the Triggering Learning Algorithm compared to a Categorical Grammar Learner*. Proceedings of the **Workshop on Psychocomputational Models of Human Language Acquisition. COLING-2004**, Association for Computational Linguistics. Geneva, Switzerland. p1—8 (8 pages).
- 2004 Buttery, P. and Briscoe, E. J. *The significance of errors to parametric models of language acquisition*. In: Cohen, P., Clark, A., Hovy, E., Oates, T., and Witbrock, M. (eds.), **Language Learning: An Interdisciplinary Perspective. Papers from the AAI Spring Symposium**. Association for the Advancement of Artificial Intelligence. Stanford, CA. p15—20 (6 pages).
- 2003 Buttery, P. *A computational model of first language acquisition*. Proceedings of the **6th Annual CLUK Research Colloquium (CLUK-6)**. Edinburgh. p1—8 (8 pages).

Other Refereed Conference Presentations and Posters

- 2015 Caines, A., Graham, C., Buttery, P. and McCarthy, M. *Crowdsourcing error annotations in a corpus of learner spoken English*, **EUROCALL 2015**, Padova, Italy. Aug 2015.
- 2015 Buttery, P., Caines A., and McCarthy, M. *Automated Processing, Grading and Correction of Spontaneous Spoken Learner Data*, **Corpus Linguistics 2015**, Lancaster University, Lancaster. July 2015.
- 2015 Katushemererwe, F., Caines, A. and Buttery, P. *Building Natural Language Processing tools for Runyakitara*, **British Association of Applied Linguistics; Language in Africa Special Interest Group Annual Meeting**, Aston University. May 2015.
- 2015 Caines, A., McCarthy, M. and Buttery, P. *Working in shades of grey: error analysis in spoken learner language*, **British Association of Applied Linguistics; Testing, Evaluation and Assessment Special Interest Group Conference**, University of Cambridge. May 2015.
- 2014 Caines, A. and Buttery, P. *The effect of topic on opportunity of use in the Cambridge Learner Corpus*, **Teaching and Language Corpora (TaLC) Conference**, Lancaster. July 2014.
- 2014 Buttery, P., Caines, A. and McCarthy, M. *Infinite shades of grey: what constitutes an error?*, **Inter-Varietal Applied Corpus Studies (IVACS) Conference**, Newcastle. June 2014.
- 2012 Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, Roy D., Buttery, P. and Marslen-Wilson, W.D. *Locating computable functions using signal correlation techniques in magneto-encephalography*, **19th Annual meeting of the Cognitive Neuroscience Society**, Chicago, USA, March 2012.
- 2012 Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, R.D., Buttery, P. and Marslen-Wilson, W.D. *Dissecting neuro-computational functions in EMEG source space*, **British Association for Cognitive Neuroscience Conference and Annual Meeting**, Newcastle, UK, April 2012.
- 2011 Devereux, B., Tyler L.K., Buttery, P. and Korhonen, A. *The role of verb subcategorization frames and selectional preferences in sentence processing: an investigation using corpus-derived measures*, **Multidisciplinary Workshop on the Mental Representation of Verbal Argument Structure. (Struc-Arg 2011)**. Paris, France, Sept 2011.
- 2011 Bozic, M., Fonteneau, E., Tyler L.K., Devereux, B., Buttery, P. and Marslen-Wilson W. *Grammatical categories in the fronto-temporal language network*, **11th International Conference on Cognitive Neuroscience (ICON XI)**. Mallorca, Spain, Sept 2011.
- 2011 Caines, A., Buttery, P. and McCarthy, M. *Turn taking in child-caregiver interaction*, **Inter-Varietal Applied Corpus Studies Symposium (IVACS Symposium 2011)**, Newcastle, UK, Jan 2011.
- 2011 Buttery, P. and McCarthy, M. *Lexis in Spoken Discourse*, **Inter-Varietal Applied Corpus Studies Symposium (IVACS Symposium 2011)**, Newcastle, UK, Jan 2011.

- 2009 Buttery, P. *Using large-scale corpora within the English Profile program; computational methods for constructing reference levels descriptors*, **American Association for Applied Linguistics 2009**, Denver, Colorado, USA, March 2009.
- 2008 Gram, L. and Buttery, P. *Integrating simulation and experiment: A study of Japanese second language acquisition* (poster presentation), **Architectures and Mechanisms for Language Processing 2008**, Cambridge, UK, Sept 2008.
- 2008 Blevins, J.P., Ackerman, F., Buttery, P. and Malouf, R. *An entropy-based measure of morphological information* **Proceeding of Quantitative Investigations In Theoretical Linguistics (QITL3)**. Helsinki, Finland, July 2008.
- 2007 Briscoe, E. J. and Buttery, P. *The influence of prosody and ambiguity on English relativization strategies*, **Interdisciplinary Approaches to Relative Clauses**, Cambridge, UK, Sept 2007.
- 2003 Buttery, P. and Villavicencio, A. *Language acquisition and the universal grammar* (poster presentation), **Architectures and Mechanisms for Language Processing 2003**, Glasgow, UK, Aug 2003.

iii) Peer Tested Research Software, Technical Manuals

- 2010 Stoehr, P., Buttery, P., Cobley, N., Horne, A., Pillai, S. **Research Software: UKPMC—webservice 2: An information processing and extraction system supporting UKPMC** (<http://ukpmc.ac.uk>)
- 2009 Buttery, P., and iLexIR . **Research Software: Corpus Collection Portal: Website for corpus data collection for the English Profile Programme** (see <http://epp.ilexir.co.uk/>)
- 2009 Thwaites, A. and Buttery, P. **Research Software: Lexical Isolation Point Software**. Software for cohort modelling in speech.
- 2008 Gram, L. and Buttery, P. **Research Software Update & Technical Manual: iLexIR Search**. Manual for the Cambridge Learner Corpus search tool for researchers in the English Profile Programme.
- 2007 Buttery, P. and Briscoe, E.J. **Research Software: Rule-defined pattern matcher over grammatical relations**, part of the Robust Accurate Statistical Parser toolkit.
- 2006 Buttery, P. **Research Software: Indexer and syntax-based search tool—iLexIR Search**. Cambridge Learner Corpus search tool for researchers in the English Profile Programme.
- 2004 Korhonen, A. and Buttery, P. **Research Data: Lexical-semantic database**.
- 2004 Buttery, P. **Research Data: Large-scale Annotated Corpus of Child Directed Speech**. Original data from the CHILDES database.
- 2004 Buttery, P. **Research Data: Large-Scale Annotated Corpus of Spoken Interactions between Families and Friends**. Original data from the British National Corpus.

Invited Talks

i) Keynote and Plenary Talks

- 2016 'The English Profile: Individualised Multidimensional Error Profiling.' Plenary at **The Kaken Project International Conference on the CEFR-J (Common European Framework of Reference—Japan)**. Tokyo, Japan, March 2016
- 2016 'Automated Marking for Learner Language.' Plenary at **The Kaken Project International Conference on the CEFR-J (Common European Framework of Reference—Japan)**. Tokyo, Japan, March 2016
- 2015 'Automated Assessment and Feedback.' Joint keynote address at **Mock IELTS Launch Event at the British Embassy**, Beijing, China,
- 2013 'The English Profile: L2 criterial features of English.' Keynote at **The Kaken Project International Conference on the CEFR-J (Common European Framework of Reference—Japan): reference level descriptions using textbook & learner corpora**. University of Aizu, Aizuwakamatsu, Fukushima, Japan, Sept 2013.
- 2012 'Normalising frequency counts to account for "opportunity of use" in learner corpora'. Plenary talk at **Biennial IVACS (Inter-Varietal Applied Corpus Studies) Conference**, Leeds Metropolitan University, UK, June 2012.

- 2011 'Extracting and classifying excerpts from learner corpora.' Keynote at **The TUFUS G-COE (Tokyo University of Foreign Studies, Global Centres of Excellence) International Symposium on Learner Corpora**. Tokyo University of Foreign Studies, Japan, March 2011.
- 2009 'Natural Language Processing for the Analysis of Spoken Corpora.' Keynote at **The IVACS (Inter-Varietal Applied Corpus Studies) Symposium**, University of Edinburgh, UK, January 2009.
- 2008 'Using Learner Language from Corpora to Profile Levels of Proficiency (CEFR)—Insights from the English Profile Project.' Joint keynote with John Hawkins at **ALTE (Association of Language Testers in Europe), 3rd International Conference**, University of Cambridge, Cambridge, UK, April 2008.

ii) *Other Invited Talks*

- 2016 'Developments in Learner Feedback'. **Automated Language Teaching and Assessment Annual Review (Year 3)**, Cambridge Assessment, Cambridge. Oct 2016.
- 2015 'Automated Assessment of Spoken Language'. **National Education Examinations Authority (NEEA)**, Beijing, China, Oct 2015.
- 2015 'Automated Teaching of Spoken Language'. **Beijing Normal University**, Beijing, China, Oct 2015.
- 2015 'Finding utterance boundaries in learners' speech' **Automated Language Teaching and Assessment Annual Review (Year 2)**, Cambridge Assessment, Cambridge. Oct 2015.
- 2015 'Automated Processing, Grading and Correction of Spontaneous Spoken Learner Data'. With Russell Moore and Andrew Caines. **Automated Language Teaching and Assessment Institute Seminar**, Cambridge, June 2015.
- 2015 'Disfluency detection in spoken learner English', With Russell Moore and Andrew Caines. **Natural Language and Information Processing Group Seminar**, University of Cambridge. May 2015.
- 2015 'The automatic assessment of spoken language'. With Andrew Caines, Calbert Graham and Michael McCarthy. **English Profile Seminar**, Cambridge University Press. Feb 2015.
- 2015 'Natural Language Processing'. **The Vocabulary of Big Data, Cambridge Big Data Strategic Initiative Event**, Cambridge, Jan 2015.
- 2015 'Challenges for data driven approaches to Automated Language Teaching and Assessment.' **Machine Learning Day**. Computer Laboratory, Cambridge Jan 2015
- 2014 'Spoken corpus annotation' with Andrew Caines. **Automated Language Teaching and Assessment Institute Seminar**, Cambridge. Oct 2014.
- 2013 'The English Profile.' **Tokyo University of Foreign Studies** Toyko, Japan, Sept 2013.
- 2013 'Text mining: challenges and applications.' **BCS (British Computer Society) UK Symposium on Knowledge Discovery and Data Mining 2013** London, March 2013.
- 2012 'An introduction to language processing concentrating on techniques that have made it into deployed systems.' **BCS (British Computer Society) AI-2012, Thirty-second International Conference on Artificial Intelligence**, Cambridge, Dec 2012.
- 2011 'Individualised Models for Language Learning.' **Google Tech Talk**. Google, Mountain View, California, USA, July 2011.
- 2009 '21st century linguistics: integration with neuroscience and impact on social/economic development.' **The Cambridge Network Open Meeting: 'Cambridge ideas change the world'**, Cambridge, UK, October 2009.
- 2009 'Language learning on a next-generation service platform for Africa.' With Andrew Rice (Computer Laboratory), **English Profile Network Seminar Series 2009 Criterial Features towards Reference Level Descriptions of English**, Cambridge, UK.
- 2009 'Criterial features in learner corpora: Theory and illustrations.' with John Hawkins, **English Profile Network Seminar Series 2009 Criterial Features towards Reference Level Descriptions of English**, Cambridge, UK.
- 2008 'Automated Syntactic Analysis of the Cambridge Learner Corpus.' **Data is Second Language Research, CILR (Cambridge Institute of Language Research) Graduate Seminar**, Cambridge, Dec 2008.

- 2008 'Using Electronic Corpora to Profile Levels of Proficiency in Second Language Acquisition: the Cambridge English Profile Project.' with John Hawkins Nick Saville (Cambridge English) and Mike McCarthy (Cambridge University Press), **University of California, Davis Campus** California, USA, March 2008.
- 2008 'English Profile—Reference Level Descriptions for English; annotating and searching the learner corpora.' **CALPER (Center for Advanced Language Proficiency Education and Research)** Pennsylvania State University, Pennsylvania, USA, March 2008.
- 2007 'Searching the Cambridge Learner Corpus.' **English Profile Academic Network Seminar** Cambridge, UK, Sept 2007.
- 2007 'English Profile: Reference Level Descriptions for English.' **Linguistic Society of America Summer Institute, Stanford University**, Stanford University, California, USA, July 2007.
- 2007 'Automatic analysis of verb constructions in child speech and the possible implications for first language acquisition models.' **Journée "Acquisition du langage : observations et modèles"**, **Université Charles de Gaulle - Lille 3**, Lille, France, June 2007.
- 2007 'Automatic lexical acquisition from CHILDES.' **The Tuesday Colloquium**, Research Centre for English and Applied Linguistics, University of Cambridge, UK, April 2007.
- 2007 'RASP and the Cambridge Learner Corpus.' **English Profile Project January Seminar, Hughes Hall**, January 2007, also **University of Cambridge ESOL Examinations**, Cambridge, UK, May 2007.
- 2006 'Searching the Cambridge Learner Corpus.' **Natural Language and Information Processing Group Seminars**, Computer Laboratory, University of Cambridge, UK, July 2006.
- 2004 'Large scale analysis of verb subcategorization differences between child-directed speech and adult speech.' **Information Technology Research Institute**, University of Brighton, UK, June 2004.
- 2004 'Reliable Language Acquisition from Real Data.' **Columbia University**, also **City University of New York**, New York, USA, March 2004.
- 2003 'A Computational Model for First Language Acquisition.' **Research Centre for English and Applied Linguistics** University of Cambridge, UK, January 2003.

Grants and Funded Projects

- 2015–2016 *Computer-Assisted Language Learning (CALL) for the revitalization of endangered languages: The case of Runyakitara* In collaboration with Fridah Katushemererwe (Makerere University, Uganda) and Andrew Caines (University of Cambridge). Funded by **Cambridge-Africa Alborada Research Fund** and the **Cambridge Africa Partnership for Research Excellence**.
- 2014–2015 *Building Research Capacity in Computational Linguistics* In collaboration with Dr Fridah Katushemererwe (Makerere University, Uganda) and Andrew Caines (University of Cambridge). Funded by **Cambridge-Africa Alborada Research Fund**.
- 2013–2018 **Co-Principle Investigator**, *The Institute for Research in Automated Language Teaching and Assessment (ALTA)*. In collaboration with Ted Briscoe (Computer Laboratory), Mark Gales (Engineering) and Andy Rice (Computer Laboratory). Funded by **Cambridge English** for 5 years in the first instance, the ALTA Institute is intended to be a rolling research programme with funding renewed annually at the end of the first phase.
- 2012–2013 **Principle Investigator**, *Understanding the Effects of Topic on 'Opportunity of Use'*. Funded by **Cambridge English**.
- 2011–2016 Named as **Leading Developer**, *UK version of PubMed Central (UKPMC)*. In collaboration with Johanna McEntyre (Project Lead), Peter Stoehr (Technical Architect), Dietrich Rebholz-Schuhmann (Textmining), **EMBL-European Bioinformatics Institute**, **NacTeM & MIMAS**, **University of Manchester**, and the **British Library**. Funded at the European Bioinformatics Institute by the **Wellcome Trust** and the UKPMC funders group.
- 2008–2011 **Co-Principle Investigator**, *Computational Natural Language Processing and the Neuro-Cognition of Language*. In collaboration with Lorraine K. Tyler (Psychology), William Marslen-Wilson (MRC-CBU) and Anna Korhonen (Computer Laboratory). Funded by the **Engineering and Physical Sciences Research Council** (as a **Cognitive Systems Foresight** Project).

2005—2009 **Co-Investigator and Research Associate** *English Profile: Reference Level Descriptions for English*. In collaboration with Cambridge University Press, Cambridge English and iLexIR. RCEAL component funded by **Cambridge English**.

Academic Esteem Indicators

Committee Member of Professional Bodies Committee Member for the **British Computer Society** Special Interest Group in Artificial Intelligence, 2013–2015;

Reviewer for National Funding Agencies **Economic and Social Research Council**, 2012+; **Leverhulme Trust**, 2007+

Editorial Board of International Journal **International Journal of Learner Corpus Research**, 2013+

Reviewer for International Journals **PLOS one**, 2016+; **Journal of Child Language**, 2009+; **Computer Speech and Language**, 2006+; **Natural Language Engineering**, 2005

Programme Committees / Reviewer for International Conferences **LCR**, Learner Corpus Research, 2015+; **ACL**, Association of Computational Linguistics, 2012+ and **ACL-workshops**, Cognitive Aspects of Computational Language Acquisition, 2007+ ; **CoLing**, International Conference on Computational Linguistics, 2012+; **CoNLL**, Conference on Computational Natural Language Learning, 2009+;

Conference Chairs International Conference Workshop Chair, **ACL 2007—workshops**, 45th Annual Meeting of the Association for Computational Linguistics workshop, ‘Cognitive Aspects of Computational Language Acquisition’. Prague, 2007; National Conference Chair, **IVACS 2011** Annual Inter-Varietal Applied Corpus Studies Symposium, ‘Speech is where it’s at...’. Cambridge, 2011; Conference Session Chair, **Cambridge Language Sciences 2013**, “Language Sciences in the 21st Century: the interdisciplinary challenge”. Cambridge, 2013

Other Reviewing Activities Various local conferences organised by the **Cambridge Language Sciences Strategic Research Initiative**, 2012+; **ESSLI** European Summer School in Logic Language and Information, 2006+; **CamLing** Postgraduate Conference in Language Research, 2006+, **CLUK** Postgraduate Conference in Computational Linguistics, 2006+

Ph.D. Thesis Examiner Examiner of 5 PhD theses (4 Internal, 1 External), 2011+

Research Co-workers—Line-managed Research Associates/Assistants

2016 **Undergraduate Research Opportunities Placements** Emma Flint, Elliot Ford, Walter Myer, Olivia Thomas, Gladys Tyen, Natasha Voake
Students in Computational Linguistics funded by *Africa Voices, Cambridge Assessment, The Automated Language Teaching and Assessment Institute*

2014+ **Research Associate** Calbert Graham
Project: *The Automated Language Teaching and Assessment Institute*

2013+ **Research Associate** Andrew Caines
Project: *The Automated Language Teaching and Assessment Institute*

2012–2013 **Research Assistant** Andrew Caines
Project: *Understanding the Effects of Topic on ‘Opportunity of Use’*

2008–2010 **Research Associate** Jeroen Geertzen
Computational Natural Language Processing and the Neuro-Cognition of Language

2008–2009 **Research Assistants** Andrew Thwaites / Caroline Williams
Project: *Computational Natural Language Processing and the Neuro-Cognition of Language*

Membership of Cambridge University Strategic Initiatives

- Member of **Cambridge Language Sciences**
- Member of **Cambridge Neuroscience**
- Member of **Cambridge Big Data**

OTHER CONTRIBUTIONS

Faculty / Departmental Contributions

- Directory of Undergraduate Teaching** This is a non-stipendiary position which involves overseeing all teaching and administrative duties relating to undergraduates in the Linguistics Tripos, Department of Theoretical & Applied Linguistics. 2012–2016.
- Member of the Undergraduate Studies Committee** Sole representative from Department of Theoretical & Applied Linguistics on a Faculty committee which oversees all teaching standards and teaching strategy in Modern and Medieval Languages. 2012–2016
- Member of the Faculty Board** Member of a Faculty committee which oversees the running of all of the Departments in Modern and Medieval Languages. Membership of this board carried no reduction in teaching stint. 2013–2016
- Member of Joint-Tripes Working Party** Member of Faculty committee to oversee the introduction of a joint-Tripes between Modern and Medieval Language and the Department of Theoretical & Applied Linguistics. Membership of this committee carried no reduction in teaching stint. 2014–2016
- Ethics Officer** Position responsible for ensuring that University ethics policy is observed by all students and staff in the Department of Theoretical & Applied Linguistics who run participant based experiments. 2011–2016
- Technology Strategy Officer (Research)** Position that involved developing a strategic plan for building the technical resources required for research activity in the Research Centre for English & Applied Linguistics, Faculty of English. 2009–2010;
- Secretary of IT Committee** Member and secretary of committee responsible for IT provision in the Research Centre for English & Applied Linguistics, Faculty of English. 2007–2009

Cambridge College Contributions

- Fellow College Lecturer in Computer Science and Linguistics**, Gonville and Caius College, April 2013+
- College Committees** **Examiner of Accounts**, position involves auditing of College procedures, Gonville and Caius College, 2013+;
- Research Fellowship Committee: Joint Arts/Science Sub-Committee**, position involves assessment of interdisciplinary applications to the Junior Research Fellowship competition at Gonville and Caius College, 2013+;
- Education Committee**, member of committee responsible for overseeing education procedure and strategy at Churchill College, 2004;
- University Admissions** Admissions interviewing for **Computer Science and Linguistics** at Queens', Wolfson, Hughes Hall, 2009+; and Sidney Sussex, Gonville and Caius 2013+. ~two weeks of interviews.
- Director of Studies in Linguistics** **Gonville and Caius**, April 2013+; **Sidney Sussex**, Oct 2013+; **Wolfson** 2010–2012; **Hughes Hall** 2010+.
- Director of Studies in Computer Science** **Hughes Hall** .2009–2015

Outreach

i) Talks at Schools Events

- Why study Linguistics?** Talks to sixth form students during **Arts, Humanities and Social Sciences summer schools**, Sidney Sussex College 2013+, Gonville and Caius College 2015+
- How does Google work?** Talk to sixth form students for **Social Mobility Foundation**, Gonville and Caius College, 2013+
- Modelling Languages using Finite-state machines** Talk to year 8 and 9 students for the **Oxbridge Academic Programmes**, The Cambridge Prep Experience, University of Cambridge, June 2008.

ii) *Talks at Events for Professional Bodies*

Automated Assessment of Spoken Language **National Education Examinations Authority (NEEA)**, Beijing, China, Oct 2015.

Text mining: challenges and applications Invited talk at **British Computer Society (BCS)**, UK Symposium on Knowledge Discovery and Data Mining 2013, London, March 2013.

An introduction to language processing: concentrating on techniques that have made it into deployed systems Invited talk at **British Computer Society (BCS)**, AI-2012 Thirty-second International Conference on Artificial Intelligence, Cambridge, Dec 2012.

iii) *Talks at Public Events*

Automated Assessment and Feedback Joint keynote address at Mock IELTS Launch Event. **British Embassy**, Beijing, China, Oct 2015

Natural Language Processing The Vocabulary of Big Data, **Cambridge Big Data Strategic Initiative Event**, Cambridge, Jan 2015.

Automated Teaching of Spoken Language **Beijing Normal University**, Beijing, China, Oct 2015.

Individualised Models for Language Learning **Google Tech Talk**, Google, Mountain View, California, USA, July 2011.

21st century linguistics: integration with neuroscience and impact on social/economic development Talk at **The Cambridge Network Open Meeting: 'Cambridge ideas change the world'**, Cambridge, Oct 2009.

Introducing the English Profile Programme Talk at **Festival of Ideas**, University of Cambridge, Oct 2008.