

Curriculum Vitae — Paula Buttery

DEPARTMENT :

Computer Science & Technology

EDUCATION :

Ph.D. Computer Science, Computer Laboratory, University of Cambridge

M.Phil. Computer Speech Text & Internet Technology, Department of Engineering & Computer Laboratory, University of Cambridge (now called Part III / MPhil Advanced Computer Science)

B.A. (Hons) Computer Science, Churchill College, University of Cambridge

PROFESSIONAL HISTORY :

- 2019+ **Professor of Language and Machine Learning**, Dept of Computer Science & Technology, Uni. of Cam.
- 2018–2023 **Director**, Cambridge Institute for Automated Language Teaching and Assessment
- 2017–2018 **Acting Director**, Cambridge Institute for Automated Language Teaching and Assessment
- 2017+ **Reader in Computing and Language**, Dept of Computer Science & Technology, University of Cambridge
- 2016–2017 **University Senior Lecturer**, Dept of Computer Science & Technology, University of Cambridge
- 2014–2016 **Senior Lecturer**, Dept of Theoretical and Applied Linguistics, University of Cambridge
- 2013+ **Fellow**, Gonville and Caius College, University of Cambridge
- 2011–2014 **Lecturer**, Dept of Theoretical and Applied Linguistics, University of Cambridge
- 2010–2013 **Senior Technical Officer**, Information Extraction & Data Mining Engineer
European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton
- 2007–2011 **Permanent Senior Research Associate** (independent researcher, with contracted teaching)
Research Centre for English and Applied Linguistics, University of Cambridge
- 2005–2007 **Research Associate** (career position for an independent researcher, with contracted teaching)
Research Centre for English and Applied Linguistics, University of Cambridge

OTHER ACADEMIC POSITIONS :

Visiting Professor, Stanford Computer Science, Stanford University, 2017;

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;

Research Associate Department of Linguistics, University of California, Davis Campus, 2009–2010;

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

Director of Studies in Linguistics, Gonville and Caius, 2013+; Sidney Sussex, 2013+; Wolfson 2010–2012, Hughes Hall 2010–2017;

Director of Studies in Computer Science, Hughes Hall, 2009–2015;

Teaching Associate, Queens' College, 2012.

INDUSTRY AND ENTERPRISE POSITIONS :

Enterprise Champion, Cambridge Enterprise, University of Cambridge, 2008+;

Business Fellow, London Technology Network, 2007–2009;

Member of University Spin-out iLexIR, Natural Language Processing consultancy, 2006+;

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

TEACHING AWARDS :

Pilkington Prize, for excellence in teaching across the collegiate University, 2015;

Cambridge University Student Union Student-Led Teaching Awards, shortlisted in Lecturer and Supervisor categories, 2015

Grants and Funded Projects

- 2018–2021 **Director**, *The Cambridge 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). Renewed funding from **Cambridge Assessment English**; the ALTA Institute is intended to be a rolling research programme.
- 2018–2019 **Principal Investigator**, *Data Science Approaches to Understanding Key Actors on Online Hacking Forums—Follow-up funding*. In collaboration with Alice Hutchings. Cybersecurity and Social Data Science call, **Alan Turing Institute**
- 2017–2018 **Principal Investigator**, *Data Science Approaches to Understanding Key Actors on Online Hacking Forums*. In collaboration with Alice Hutchings. Cybersecurity and Social Data Science call, **Alan Turing Institute**
- 2017–2018 **Principal Investigator**, *Computational Linguistic skills training at Makerere University*. In collaboration with Andrew Caines and Marek Rei. **Cambridge-Africa Alborada Fund**.
- 2017–2018 **Principal Investigator**, *Annotating Spoken Language Corpora* **Cambridge Assessment English**
- 2016–2018 Various projects in collaboration with Andrew Caines, Ioanna Sitaridou, Russell Moore, Andrew Rice and Ianthi Tsimpli funded by **Cambridge Language Sciences IRC Incubator Fund**.
- 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 **Principal 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 Assessment English**.
- 2012–2013 **Principal Investigator**, *Understanding the Effects of Topic on ‘Opportunity of Use’*. Funded by **Cambridge Assessment 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-Principal 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 Assessment English and iLexIR.

Scholarly Publications and Output

i) Peer Reviewed Publications

Journal Articles

- 2018 Caines, A., Pastrana, S., Hutchings, A. and Buttery, P. *Automatically identifying the function and intent of posts in underground forums*. **Crime Science**. Accepted.
- 2017 Bentz, C., Alikaniotis, D., Samardzic, T. and Buttery, P. *Variation in word frequency distributions: Definitions, measures and implications for a corpus-based language typology*. **Journal of Quantitative Linguistics**. 24:2–3

- 2017 Katushemerewe, F., Caines, A. and Buttery, P. *Building natural language processing tools for Runyakitara*. **Applied Linguistics Review**.
- 2016 Bentz, C., Alikaniotis, D., Samardzic, T. and Buttery, P. *Variation in word frequency distributions: Definitions, measures and implications for a corpus-based language typology*. **Journal of Quantitative Linguistics**.
- 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>
- 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*. **Frontiers in Computational Neuroscience**. 9:5 <http://dx.doi.org/10.3389/fncom.2015.00005>
- 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**.
- 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
- 2010 Hawkins, J. and Buttery P. *Criterial Features in Learner Corpora: Theory and Illustrations*. **English Profile Journal**, 1: e5

Books and Book Chapters

- 2018 Caines, A., Buttery, P. and McCarthy, M., 'You still talking to me?' The zero auxiliary progressive in spoken British English, twenty years on. In: Brezina, Love & Aijmer (eds.), **Sociolinguistic variation in contemporary British English: exploration of the Spoken BNC2014**. London: Routledge.
- 2017 Caines, A. and Buttery, P. *The effect of topic on language use in the Cambridge Learner Corpus*. In: Vaclav Brezine and Lynne Flowerdew (eds.), **Written and spoken learner corpora and their use in different contexts**. London: Bloomsbury.
- 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.
- 2012 Hawkins, J. and Filipovic, L. with Buttery, P., Capel, A., Hawkey, R., Salamoura, A., Saville, N., and Trim, J. *Criterial 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.
- 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.
- 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.
- 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.
- 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.

Published Conference Proceedings

- 2018 Andrew Caines, Sergio Pastrana, Alice Hutchings and Paula Buttery. *Aggressive language in an online hacking forum*. Proceedings of the 2nd Abusive Language Workshop (ALW 2018) at 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 2018

- 2018 Sergio Pastrana, Alice Hutchings, Andrew Caines and Paula Buttery. *Characterizing Eve: Analysing Cyber-crime Actors in a Large Underground Forum*. Proceedings of the **21st International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2018)**. Crete, Greece, 2018
- 2017 Andrew Caines, Michael McCarthy and Paula Buttery. *Parsing transcripts of speech*, Proceedings of **1st Workshop on Speech-Centric Natural Language Processing (SCNLP 2017) at 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, Copenhagen, Denmark, 2017
- 2017 Andrew Caines, Emma Flint and Paula Buttery. *Collecting fluency corrections for spoken learner English*, Proceedings of the **12th Workshop on Innovative Use of NLP for Building Educational Applications at 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, Copenhagen, Denmark, 2017
- 2017 Emma Flint, Elliot Ford, Olivia Thomas, Andrew Caines and Paula Buttery: *A text normalisation system for non-standard English words*, Proceedings of **The 3rd Workshop on Noisy User-generated Text (W-NUT 2017) at 2017 Conference on Empirical Methods in Natural Language Processing**, Copenhagen, Denmark, 2017
- 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.
- 2016 Graham, C., Buttery, P. and Nolan, F. *Vowel characteristics in the assessment of L2*. Proceedings of **INTER-SPEECH**, San Francisco, USA.
- 2016 Caines, A., Bentz, C., Alikaniotis, D., Katushemererwe, 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.
- 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.
- 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.
- 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.
- 2015 Graham, C., Buttery, P., Caines, A. and Nolan, F. *Automated Assessment of non-native speech using spectral features of vowels*. Proceedings of **INTER-SPEECH**, Dresden, Germany.
- 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 **INTER-SPEECH** Dresden, Germany.
- 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.
- 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
- 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
- 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
- 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

- 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 =
- 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
- 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
- 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.
- 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
- 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.
- 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.
- 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
- 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.
- 2003 Buttery, P. *A computational model of first language acquisition*. Proceedings of the **6th Annual CLUK Research Colloquium (CLUK-6)**. Edinburgh. p1—8.

ii) *Refereed Conference Papers 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

- 2017 Andrew Caines, Diane Nicholls and Paula Buttery. *Annotating errors and disfluencies in transcriptions of speech*. **University of Cambridge Technical Report Number 915**. <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-915.pdf>
- 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* (<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: *Syntactic construction retrieval tool* 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

- 2017 ‘Individualised Language in the Big Data Era’. 1 of 4 headlined talks at **Cambridge Language Sciences Annual Symposium**, Cambridge, Nov 2017
- 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 TUFS 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

- 2018 ‘Director’s Summary’. **Automated Language Teaching and Assessment Annual Review (Year 5)**, Cambridge Assessment, Cambridge, Oct 2018.
- 2018 ‘Data science approaches to understanding key actors on online hacking forums’. **Defence and Security Programme Presentation Day, GCHQ**, Cheltenham, Sept 2018
- 2018 ‘Machine Learning in Education Applications’. **Visit of the Singapore Ministry of Education and the Singapore Examinations and Assessment Board, Department of Computer Science and Technology, Cambridge, Sept 2018**
- 2018 ‘Data science approaches to understanding key actors on online hacking forums’. **Defence and Security Monthly Meet-up, Alan Turing Institute**, March 2018
- 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 Critical Features towards Reference Level Descriptions of English**, Cambridge, UK.
- 2009 'Critical features in learner corpora: Theory and illustrations.' with John Hawkins, **English Profile Network Seminar Series 2009 Critical Features towards Reference Level Descriptions of English**, Cambridge, UK.
- 2008 'Automated Syntactic Analysis of the Cambridge Learner Corpus.' **Data in 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.

Line-managed Research Co-workers

- 2018+ **(Senior) Research Associates:** Helen Yannakoudakis, Marek Rei, Øistein Andersen, Ekaterina Kochmar, Mariano Felice, Zheng Yuan, Christopher Bryant
Project: *The Automated Language Teaching and Assessment Institute*
- 2016–2018 **Undergraduate Research Opportunities Placements:** 20 students in Computational Linguistics funded by *Africa’s Voices, Cambridge Assessment, The Automated Language Teaching and Assessment Institute, Grammarly, Cybershield, Cambridge Language Sciences*
- 2016 **Visiting Academic:** Brendan Flanagan, Kyushu University, Japan
Working within the *Automated Language Teaching and Assessment Institute*
- 2014–2018 **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* (Senior Research Associate, 2016+)
- 2012–2013 **Research Assistant:** Andrew Caines
Project: *Understanding the Effects of Topic on ‘Opportunity of Use’*
- 2008–2010 **Research Associate:** Jeroen Geertzen
Project: *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*