

PAULA J BUTTERY

LECTURER & SENIOR TECHNICAL OFFICER

CONTACT DETAILS

Address: Department of Theoretical and Applied Linguistics
University of Cambridge,
9 West Road, Cambridge, CB3 9DP, UK

Phone: +44 7989 965107

Fax: +44 1223 767398

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

CURRENT POSITIONS

Lecturer: **Computational Linguistics Cluster**
Department of Theoretical and Applied Linguistics
University of Cambridge.

Senior Technical Officer: **Information Extraction & Data Mining Engineer**
Literature Services
European BioInformatics Institute
European Molecular Biology Laboratory Outstation,
Wellcome Trust Genome Campus, Hinxton

Associated Researcher: **Natural Language and Information Processing Group**
Computer Laboratory
University of Cambridge

Director of Studies: **Hughes Hall, University of Cambridge:**
Computer Science and Linguistics
Wolfson College, University of Cambridge:
Linguistics.

Software Consultant: **iLexIR**

Enterprise Champion: **Cambridge Enterprise**

RECENT AND FUTURE FUNDED PROJECTS

2008—2011 **Computational Natural Language Processing and the Neuro-Cognition of Language.** Co-PI in collaboration with Lorraine K. Tyler, William Marslen-Wilson & Anna Korhonen. Co-funded by the EPSRC, ESRC and MRC (Cognitive Systems Foresight project).

2008—2010 **Developing Multilingual Technologies for Automatic Lexical Acquisition (TAACL).** In collaboration with Anna Korhonen & Thierry Poibeau. Funded by the British Council (Alliance Programme).

2005— **English Profile: Reference Level Descriptions for English.**
In collaboration with Cambridge University Press, University of
Cambridge Local Examinations Syndicate and iLexIR. RCEAL com-
ponent funded by Cambridge ESOL.

EDUCATION

2002–2005 **PH.D. Computer Science**
under supervision of Ted Briscoe, Computer Laboratory
Churchill College, University of Cambridge;

2000–2001 **M.PHIL. Computer Speech, Text and Internet Technology**
Department of Engineering & Computer Laboratory
Churchill College, University of Cambridge;

1997–2000 **B.A. (HONS) Computer Science**
Churchill College, University of Cambridge.

PREVIOUS POSITIONS AND EMPLOYMENT

2009-2010 **University of California:** Research Associate.
Department of Linguistics, Davis Campus.

2006-2009 **London Technology Network:** Business Fellowship.

2005 **Google:** Internship, Research Engineer.
Chinese to English machine translation.

2000 **Dragon Systems:** Internship, Research Assistant.
Speech data analysis; software error and efficiency analysis.

1999 **Intel Corporation:** Internship, Market Research Analyst
Business-client sales data analysis.

AWARDS AND SCHOLARSHIPS

2002–2005 University of Cambridge Domestic Research Studentship

2000–2001 Engineering and Physical Sciences Research Council Award

1998–2000 Intel Corporation Studentship

1997–2000 National Electronics Council Scholarship

PAPERS AND PUBLICATIONS

2011 Devereux, B., Tyler L.K., Buttery, P. & Korhonen, A. *The role of verb subcategorization frames and selectional preferences in sentence processing: an investigation using corpus-derived measures*
Paper presentation at Multidisciplinary Workshop on the Mental Representation of Verbal Argument Structure. (Struc-Arg 2011). Paris, France.

- 2011 Bozic, M., Fonteneau, E., Tyler L.K., Devereux, B., Buttery, P. and Marslen-Wilson W. *Grammatical categories in the fronto-temporal language network*. Paper presentation at 11th International Conference on Cognitive Neuroscience (ICON XI). Mallorca, Spain.
- 2011 Buttery, P. and McCarthy, M. *Lexis in Spoken Discourse*. In the Routledge Handbook of Discourse Analysis, Routledge, Gee, J., Handford, M. (eds.) 2011 (plus paper presentation at Inter-Varietal Applied Corpus Studies Symposium, Newcastle Jan 2011).
- 2011 Caines, A., Buttery, P. and McCarthy, M. *Turn taking in child-caregiver interaction*. Paper presentation at Inter-Varietal Applied Corpus Studies Symposium (IVACS Symposium 2011), Newcastle, UK.
- 2010 McEntyre, J.M., Ananiadou, S., Andrews, S., Black, WJ., 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*. 2010, doi:10.1093/nar/gkq1063
- 2010 Caines, A. and Buttery, P. *You Talking to Me? A predictive model for zero-auxiliary constructions*. NLP and Linguistics Workshop, ACL-2010, Uppsala.
- 2010 Hawkins, J. and Buttery P. *Criterion Features in Learner Corpora: Theory and Illustrations*. English Profile Journal, Vol 1, Cambridge University Press.
- 2010 Thwaites, A., Geertzen, J., Marslen-Wilson, W., Buttery, P. *Lexical Isolation Point Software*. Poster presentation at LREC 2010, Malta.
- 2010 Williams, C., Thwaites, A., Buttery, P., Geertzen, J., Randell, B., Shafto, M., Devereux, B., Tyler, L. *The Cambridge Cookie-Theft Corpus: A Corpus of Spontaneous and Directed Speech of Brain-Damaged Patients and Healthy Individuals*. LREC 2010, Malta.
- 2009 Buttery, P. *Using large-scale corpora within the English Profile program; computational methods for constructing reference levels descriptors*. Paper presentation at AAAL 2009, Denver, Colorado.
- 2009 Vlachos, A., Buttery, P., Saghdha, D., Briscoe, T. *Biomedical Event Extraction without Training Data*. BioNLP 2009 at NAACL, Boulder, CO.
- 2009 Rice, A., Buttery, P., Rai, I. and Beresford, A., *Language learning on a next-generation service platform for Africa*. Africa Perspective on the Role of Mobile Technologies in Fostering Social and Economic Development, W3C, April 2009

- 2009 Hawkins, J. and Buttery, P. *Using Learner Language from Corpora to Profile Levels of Proficiency - Insights from the English Profile Programme*. In: *Studies in Language Testing: The Social and Educational Impact of Language Assessment*, Cambridge University Press, 2009.
- 2008 Gram, L. and Buttery, P. *Integrating simulation and experiment: A study of Japanese second language acquisition*. Poster presentation AMLaP 2008, Cambridge, UK.
- 2008 Briscoe, T. and Buttery, P. *Linguistic adaptations for resolving ambiguity*. In Smith, A. D. M., Smith, K., & Ferrer i Cancho, R. (eds). *The Evolution of Language: Proceedings of the 7th International Conference (EVLANG7)*, Barcelona, 2008. Singapore: World Scientific Press.
- 2007 Briscoe, T. and Buttery, P. *The influence of prosody and ambiguity on English relativization strategies*. *Interdisciplinary Approaches to Relative Clauses*, Cambridge, UK, 2007.
- 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*. ACL-2007 Workshop, Cognitive Aspects of Computational Language Acquisition, Prague, Czech Republic, 2007.
- 2007 Buttery, P., Villavicencio, A., and Korhonen A. (eds.). *Proceedings of the ACL 2007 Workshop on Cognitive Aspects of Computational Language Acquisition*. Prague, Czech Republic. Association of Computational Linguists.
- 2006 Buttery, P. *Computational models for first language acquisition*. PhD thesis published as a technical report, Computer Laboratory, University of Cambridge, 2006
- 2005 Buttery, P. and Korhonen, A. *Large-scale analysis of verb subcategorization differences between child directed speech and adult speech*. Verb Workshop 2005, Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes, Saarland University 2005.
- 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*. In proceedings of COLING-2004 Workshop, Psycho-computational Models of Human Language Acquisition, Geneva 2004.
- 2004 Buttery, P. and Briscoe, E.J. *The significance of errors to parametric models of language acquisition*. In AAI Technical Report SS-04-05, Spring Symposium on Language Learning: An Interdisciplinary Perspective, Stanford, 2004, pages 15–20.

- 2003 Buttery, P. and Villavicencio, A. *Language acquisition and the universal grammar*. Poster presentation AMLaP, Glasgow, 2003, page 119.
- 2003 Buttery, P. *A computational model of first language acquisition*. CLUK6, Edinburgh, 2003, pages 1–8.
- 2001 Buttery, P. *Learning from annotated corpora*. Masters Thesis, University of Cambridge.

BOOKS AND ARTICLES THAT ARE IN PRESS, OR SUBMITTED

- 2012 Buttery, P. and Caines, A. *An Empirical Approach to First Language Acquisition*. Cambridge University Press.
- 2011 Buttery, P. and Caines, A. *Normalising Frequency Counts to Account for opportunity of use in Learner Corpora*. TUFSL vol.4, John Benjamins.
- 2011 Geertzen, J., Randall B., Buttery, P., Tyler L.K. and Shafto M. *Language production across the adult lifespan: An automatic parsing analysis of syntax in spontaneous speech*. Submitted.
- 2011 Thwaites, A., Nimmo-Smith, I., Fonteneau, E., Patterson, R., Buttery, P. and Marslen-Wilson, W. *Locating computable functions in the human brain*. Submitted.
- 2011 Buttery, P. and Caines, A. *Reclassifying subcategorization frames for experimental analysis and stimulus generation*. Submitted.
- 2011 Caines, A. and Buttery, P. *Annotating progressive aspect constructions in the spoken section of the British National Corpus*. Submitted.

RESEARCH SOFTWARE, TECHNICAL MANUALS & PUBLICLY AVAILABLE RESEARCH DATA

- 2009 Buttery, P., *iLexIR Corpus Collection Portal*. <http://epp.ilexir.co.uk/> (English Profile Programme)
- 2009 Thwaites, A. and Buttery, P. *Lexical Isolation Point Software*.
- 2008 Gram, L. and Buttery, P. *iLexIR Search*. Technical Manual for Cambridge Learner Corpus search tool (English Profile Programme).
- 2007 Buttery, P. and Briscoe, E.J. *Rule-defined pattern matcher over grammatical relations*, part of the RASP toolkit.
- 2006 Buttery, P. *Indexer and syntax-based search tool—iLexIR Search*. For analysis of the Cambridge Learner Corpus (English Profile Programme)
- 2004 Korhonen, A. and Buttery, P. *Lexical-semantic database*.

- 2004 Buttery, P. *Large-scale Annotated Corpus of Child Directed Speech*. Original data from the CHILDES database.
- 2004 Buttery, P. *Large-Scale Annotated Corpus of Spoken Interactions between Families and Friends*. Original data from the British National Corpus.

INVITED TALKS

- 2012 *Title to be confirmed* Keynote at **Biennial IVACS (Inter-Varietal Applied Corpus Studies) Conference**, Leeds Metropolitan University. June 2012.
- 2011 *Individualised Models for Language Learning*. **Google Tech Talk** Google, Mountain View. July 2011.
- 2011 *Extracting and classifying excerpts from learner corpora*. Keynote at **The TUMS G-COE International Symposium on Learner Corpora**. Tokyo University of Foreign Studies, Japan, March 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 *Computation Cognition and Language at the Research Centre for English and Applied Linguistics*. (with Jim Blevins, Jeroen Geertzen, Andrew Thwaites and Chris Veysey) **RCEAL Seminar Series**, Cambridge, UK.
- 2009 *Language learning on a next-generation service platform for Africa*. (with Andrew Rice) **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.
- 2009 *Natural Language Processing for the Analysis of Spoken Corpora*. Keynote at **The IVACS (Inter-Varietal Applied Corpus Studies) Symposium**, Edinburgh, January 2009.
- 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 Learner Language from Corpora to Profile Levels of Proficiency (CEFR)—Insights from the English Profile Project*. **Association of Language Testers in Europe (ALTE) 3rd International Conference, Cambridge**, Joint keynote presentation with Prof. John Hawkins. April 2008.

- 2008 *Using Electronic Corpora to Profile Levels of Proficiency in Second Language Acquisition: the Cambridge English Profile Project.* **University of California, Davis**, A joint presentation by John A. Hawkins (UC Davis & Cambridge University), Nick Saville (Cambridge Assessment), Paula Buttery (Cambridge University) and Mike McCarthy (Penn State University & Cambridge University Press), 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**, March 2008.
- 2007 *Searching the Cambridge Learner Corpus.* **English Profile Academic Network Seminar, Cambridge**, Sept 2007.
- 2007 *English Profile: Reference Level Descriptions for English.* **Linguistic Society of America Summer Institute, Stanford University**, July 2007.
- 2007 *Automatic analysis of verb constructions in child speech and the possible implications for first language acquisition models.* **Journee “Acquisition du langage : observations et modeles”, Université Charles de Gaulle-Lille3**, June 2007.
- 2007 *Automatic lexical acquisition from CHILDES.* **Research Centre for English and Applied Linguistics, University of Cambridge**, April 2007.
- 2007 *RASP and the Cambridge Learner Corpus.* **English Profile Project January Seminar, Hughes Hall**, January 2007 also presented at **University of Cambridge ESOL Examinations**, May 2007.
- 2006 *Searching the Cambridge Learner Corpus.* **Computer Laboratory, University of Cambridge**, July 2006.
- 2004 *Large scale analysis of verb subcategorization differences between child-directed speech and adult speech.* **Information Technology Research Institute, University of Brighton**, June 2004.
- 2004 *Reliable Language Acquisition from Real Data.* **Columbia University**, March 2004 also presented at **City University of New York**, March 2004.
- 2003 *A Computational Model for First Language Acquisition.* **Research Centre for English and Applied Linguistics, University of Cambridge**, January 2003.

OTHER ACADEMIC ACTIVITIES

Conference Organiser

Inter-Varietal Applied Corpus Studies Symposium **IVACS 2012** (with Andrew Caines).

Conference Organiser	New Paradigms in Quantitative Linguistics. Application made for funding to British Academy Conference Grants scheme and CRASSH (with Jim Blevins and Andrew Caines)
Event Organiser	Introducing the English Profile Programme, Festival of Ideas 2008 . University of Cambridge.
Workshop Chair	Cognitive Aspects of Computational Language Acquisition, ACL-2007 workshop (with Anna Korhonen and Aline Villavicencio).
Internal Examiner	Grammatical error prediction. Ph.D. Thesis by Oistein Anderson, University of Cambridge .
External Examiner	Open and Distance Learning Unit, Queen Mary, University of London .
Reviewing Activities	Book reviewer for Natural Language Engineering ; Article reviewer for Computer Science and Language and Journal of Child Language .
Conference Reviewing	Member of programm committees for: ACL Association of Computational Linguists; CoLing , International Conference on Computational Linguistics; ESSLLI , European Summer School in Logic Language and Information; CamLing , Postgraduate Conference in Language Research; CLUK , Postgraduate Conference in Computational Linguistics.

COMMITTEE POSITIONS

2011+	Ethics Officer Department of Theoretical & Applied Linguistics, University of Cambridge
2009-2011	Technology Strategy Officer (Research) Research Centre of English & Applied Linguistics, University of Cambridge
2007-2009	Secretary of IT Committee Research Centre of English & Applied Linguistics, University of Cambridge
2004	Member of Education Committee Churchill College, University of Cambridge
2003–2004	Postgraduate Welfare Officer Churchill College, University of Cambridge

SUPERVISION OF RESEARCH PROJECTS

2010-2011	Language Models for Language Learning on Mobile Devices. Supervisor of MPhil thesis in Advanced Computer Science by <i>Duncan Roberts</i>
-----------	---

2010-2011	An emergent perspective on L1 influence in the L2 acquisition of English grammatical morphemes. Research committee member for Ph.D. thesis by <i>Akira Murakami</i>
2009-2010	Methods for the extraction and evaluation of Criterial Features for English. Supervisor of Ph.D thesis by <i>Stephen Spencer</i>
2009+	Evaluation of syntactic complexity metrics: diachronic analysis of the dative shift. Supervisor of Ph.D thesis by <i>Chris Veysey</i>
2008-2009	Modeling the acquisition and evolution of language. Co-supervisor of Ph.D thesis by <i>Lu Gram</i>
2008-2009	Development of discourse features by Czech and Hungarian learners of English Research committee member for Ph.D thesis by <i>Nobert Vanek</i>
2008-2009	The acquisition of numerically quantified expressions Research committee member for Ph.D thesis by <i>Chris Cummins</i>
2008	Integrating simulation and experiment: A study in Japanese SLA. Supervisor of M.Phil thesis by <i>Lu Gram</i>
2008	Automatic analysis of child language. Co-supervisor of M.Phil thesis by <i>Andrew Thwaites</i>
2008	Features detection for automated poetry analysis. Co-supervisor of M.Phil thesis by <i>Marcus Duyzend</i>
2006+	An automatic text classification approach to developmental profiling in second language acquisition. Supervisor of Ph.D thesis by <i>Caroline Williams</i>
2006-2010	“(Are) you talking to me?” Elision of the auxiliary verb in questions. Co-supervisor of Ph.D thesis by <i>Andrew Caines</i>
2006–2007	The acquisition of tense and agreement morphology in L2 English. Research committee member of Ph.D thesis by <i>Tae Yamazaki-Hasegawa</i>
2006	Acquiring Grammatical Fingerprints for genres. Co-supervisor of M.Phil thesis by <i>Zheng Mang</i>
2005–2006	The assignment of stress in English noun-noun combinations. Research committee member of Ph.D thesis by <i>Melanie Bell</i>

LECTURE COURSES

2012+	Computational Linguistics (Senior year course for BA in Linguistics)
2011+	Computational Corpus Linguistics (option course for M.Phil in Linguistics)
2011+	Corpus Linguistics (research methods course for M.Phil in English and Applied Linguistics)
2008–2009	Computational Language Learning (option course for M.Phil in Linguistics)
2005–2009	Corpus Linguistics (research methods course for M.Phil in English and Applied Linguistics)
2007–2008	Computational Psycho-linguistics (Ph.D seminar series and reading group)
2006–2007	Introduction to Computational Language Learning (Ph.D course with Anna Korhonen)
2005–2006	Introduction to Computational Linguistics (Ph.D course with Anna Korhonen)

TEACHING ASSISTANT POSITIONS

Course Supervisor:

2002–2005	Discrete Mathematics Information Retrieval Natural Language Processing Computer Vision Mathematics for Computation Theory Advanced Graphics and Human Computer Interaction
-----------	---

Teaching Demonstrator:

2003–2004	Automated Question Answering Practical
-----------	--

Practical Class Examiner:

2002–2005	Assessed Mathematics Exercise for Natural Sciences Programming in Java Programming in ML
-----------	--

MODERN LANGUAGES

English:	Native speaker;
French:	Good—gap year in France 2001-2002;
German:	Basic level.