

Dr Sandra Williams

sandrahelenwilliams@gmail.com

PROFILE

Dr Williams is a Research Fellow at The Open University. She has substantial experience in industrial and academic Natural Language Engineering and experience in teaching at post- and undergraduate levels. Recently, she was Principal Investigator of an ESRC Small Grant: Generating intelligent descriptions of numerical quantities for people with different levels of numeracy.

Over the past ten years, Williams has developed intelligent planners for Natural Language Generation (NLG) of text and spoken dialogue for embodied agents, most recently, generating descriptions of classes, individuals and properties in ontologies, numerical expressions, and mathematical word problems.

During her PhD research and the SkillSum project, Williams developed NLG systems that tailored texts for people with limited literacy. Her approach used Constraint Satisfaction Problem Solving with empirical data derived from a linguistic analysis of a corpus annotated with rhetorical relations. During this period, she gained experience in designing, planning and conducting user experiments, statistically analysing results, and developing theoretical and technical insights.

RESEARCH GOALS

Dr Williams' current research goals are to map ontology logics to natural language (NL) for easy human comprehension and to develop models for communicating and understanding numerical information in different contexts and for different audiences.

QUALIFICATIONS

PhD Natural Language Generation, Computing Science, University of Aberdeen, 2004.

MPhil Computer Speech and Language Processing, University of Cambridge, 1991.

BA Hons (upper 2nd) Computing and Artificial Intelligence, University of Sussex, 1988.

EMPLOYMENT HISTORY

2006 – present. Research Fellow, Computing Dept., The Open University, U.K. Recently she took up funding from a Pro Vice Chancellor grant for research development, following employment on EPSRC project SWAT, Semantic Web Authoring Tool, from 2009-2012. Previously, Dr Williams led the NumGen project: Generating intelligent descriptions of numerical quantities for people with different levels of numeracy. Before that, she contributed to the EU-funded Semantic Mining Network of Excellence, generating summaries of patients' medical records.

2003 – 2005. Research Fellow, Computing Science Dept., University of Aberdeen, U.K. SkillSum project: Automatic Generation of Personalised Basic Skills Summary Reports, a PACCIT LINK project with Cambridge Training and Development Ltd. (CTAD) that developed technology to help people assess their basic skills and to automatically generate feedback for people with limited literacy and numeracy.

2000 – 2003. Ph.D. Student, University of Aberdeen, U.K.

1999 – 2000. Technology Consultant, Cambridge Training and Development Ltd., U.K. Williams led a feasibility study funded by a DTI Smart award (Small firms Merit Award for Research and Technology) to explore integrating speech and language technology with e-learning software.

1997 – 1999. Research Scientist (Lecturer B grade), Language Technology Group, Macquarie University, Australia. Williams managed an NLG project to describe routes around a University department in text and synthesised speech with appropriate intonation. She contributed to a project with Sydney's PowerHouse Museum that built an NLG system to compare thousands of museum objects in English, Spanish, French and Chinese.

1988 – 1997. Senior Professional, Natural Language Processing Group, British Telecom. Research Labs., U.K. Williams' research at BT Labs. (now BTextact Technologies) included spoken dialogue management, reference resolution, and text summarisation. She was leader of the Text Summarisation and Information Management project that developed BT's web-based Text

Summariser.

1985 – 1988. Undergraduate Student, University of Sussex, U.K.

1976 – 1985. Before becoming a mature student at University, Williams had 10 years of broad work experience including Seismic Survey Processing, Customer Relations, office, factory and farm work.

TEACHING EXPERIENCE

2009 - present Ph.D. Supervisor and M.Sc. Specialist Advisor, The Open University, U.K.

Williams supervises two Open University Ph.D. students and one visiting Ph.D. student and is Specialist Advisor to one Masters student.

2000 - 2005 Computer Laboratory Demonstrator, University of Aberdeen, U.K. Williams had five years laboratory demonstrating experience in a number of Computing Science courses, including Java Programming, NLP, HCI, and Data Structures and Algorithms.

1998 MSc module in Spoken Language Dialogue Systems, Macquarie University, Australia.

Williams designed and taught part of the above module.

1998 Student Supervisor, Macquarie University, Australia. Williams supervised an undergraduate student in his final year Honours project.

1996 Student Supervisor, BT Research Labs., U.K. Williams supervised an undergraduate student in his summer internship.

TRAINING COURSES

2011 - 2013 Research Supervisors Forum: Open University staff development series.

2010 PhD Supervision: Open University staff development course.

2009 Web Ontology Language Tutorial: University of Manchester course.

2009 Media Training: ESRC course.

2009 How to be a Mentor: Open University staff development course.

2008 Effective Research: Research Councils UK course.

2007 Widening Participation: Open University staff development course.

2006 Writing Course Materials: Open University staff development course.

2005 Bidding for Funding: Open University staff development course.

2002 Research Councils' Summer School: Personal and career development course.

1992 Presentation Skills: British Telecom. staff development course.

1995 Project Management: British Telecom. staff development course.

COMPUTER SKILLS

Programming: JAVA, PROLOG, LISP, C, C++ and PERL.

Other skills: SQL, SGML, HTML, DHTML and JavaScript, Tk/Tcl, UNIX, DOS, CGI, SPSS, Word, LaTeX, Excel, PowerPoint and Protégé ontology development tool.

MANAGEMENT EXPERIENCE

2008 - 2009 Principal Investigator, ESRC project: NumGen, Williams managed a one-year feasibility study. Continuing the research in 2010, she is supervising a visiting PhD student.

2005 Deliverable Manager, EU Semantic Mining Project, Williams managed and edited a literature survey with project partners at The Open University and the University of Goteburg.

1999 Project Manager, CTAD Ltd., Williams managed a language technology project.

1997 - 1998 Project Manager, Macquarie University, Williams managed a language technology project. She liaised with project partners at the University of New South Wales. She initiated a collaboration with Sydney's Powerhouse Museum.

1997 - 1999 Organiser, Macquarie University, Williams was local organiser for an international conference. She also organised a Seminar Series.

1991 - 1997 Project Leader, BT Research Labs, Williams led the Information Management project theme in the Natural Language Processing research group.

GRANTS AND AWARDS

2008 ESRC Small Grant, £70,000, Principal Investigator, NumGen: Generating intelligent descriptions of numerical quantities for people with different levels of numeracy.

2000 EPSRC PhD Studentship, Aberdeen University (Williams contributed to the proposal)

1999 DTI SMART Award, £40,000, feasibility study (jointly with C. Webb, CTAD Ltd.)

1998 Small Research Grant, AU\$7,000, Macquarie University, prosodic annotation.
1990 BT major training award, for MPhil at Cambridge University

PATENTS

Williams is named as an inventor in a patent for a French language tuition system developed at BT Labs.

CONFERENCE ORGANISATION

PITR 2013 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations, Sofia, Bulgaria.

PITR 2012 1st Workshop on Predicting and Improving Text Readability for Target Reader Populations, Montreal, Canada.

PROGRAMME COMMITTEE MEMBERSHIP AND JOURNAL REVIEWING

DO 2013 International Workshop on Definitions in Ontologies, Montreal, Canada.

ENLG 2013 14th European Workshop on Natural Language Generation, Sofia, Bulgaria.

RANLP 2013 Recent Advances in Natural Language Processing, Hissar, Bulgaria.

NAACL-HLT 2012 Student Research Workshop and Doctoral Consortium, Montreal, Canada.

QG 2012 Fifth International Workshop on Question Generation, Cancelled after reviewing.

COLING 2012 Natural Language Generation area, December, Mumbai, India.

Patient Education and Counseling Journal 2012

ENLG 2011 13th European Workshop on Natural Language Generation, Nancy, France.

RANLP 2011 Recent Advances in Natural Language Processing, Hissar, Bulgaria.

Dialogue and Discourse (open access journal), Special Issue on Question Generation 2011.

ACL-HLT 2011 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies Portland, Oregon, USA.

ICL 2010 International Conference on Interactive Computer Aided Learning, Hasselt, Belgium.

INLG 2010 6th International Natural Language Generation Conference, Dublin.

MOG 2010 3rd Workshop on Multimodal Output Generation, Dublin.

EACL 2009 12th European Chapter of the Association for Computational Linguistics, Athens.

ICL 2009 Interactive Computer Aided Learning, Special Track on Computer-based Knowledge & Skill Assessment and Feedback, Villach, Austria.

Transactions on Information Systems The official journal of the ACM.

INLG 2008 5th International Natural Language Generation Conference, Ohio.

RANLP 2007 International Conference on Recent Advances in NLP, Borovets.

JoLLi Journal of Logic, Language and Information, Special Issue on Coherence in Dialogue and Generation, 16:4, 2007.

Research on Language & Computation Special Issue on Formal Issues in Language Generation, 4:1, 2006.

ESSLLI 2006 18th European Summer School in Logic, Language and Information, Malaga.

COLING 2002 19th International Conference on Computational Linguistics, Taipei.

ACL/EACL 1997 Association for Computational Linguistics, Madrid.

INVITED TALKS

2012 Seminar at University of Aberdeen, Computing Department, Scotland.

2009 Seminar at Macquarie University, Computing Department, Sydney, Australia.

2008 Seminar at Aberdeen University, Computing Department, Scotland.

2007 Seminar at National Informatics Institute, Tokyo, Japan.

2005 Seminar at Aberdeen University, Centre for Linguistics Research, Scotland.

2002 Seminar at Monash University, Computer Science Department, Melbourne, Australia.

2002 Seminar at Stirling University, English Studies Department, Scotland.

PUBLICATIONS

Book Chapters

Sandra Williams [1996] Anaphoric reference and ellipsis resolution in a telephone-based spoken language system for accessing email. In Simon Botley and Anthony McEnery (eds.) *Corpus-based and Computational Approaches to Discourse Anaphora*, John Benjamins Publishing Company, ISBN 902722272X, 2000.

Ehud Reiter and Sandra Williams [2009] *Generating Texts in Different Styles*. In S Argamon, K Burns, S Dubnov (eds.) *The Structure of Style: Algorithmic Approaches to Manner and Meaning*. Springer-

Verlag.

Journal Articles

- Williams, Sandra and Richard Power [2013] Hedging and Rounding in Numerical Expressions. Accepted for publication in the Journal of Pragmatics and Cognition, John Benjamins Publishing Co.
- Power, R. and S. Williams. 2012. Generating numerical approximations. Computational Linguistics, Volume 38, No. 1.
- Robert Stevens, James Malone, Sandra Williams, Richard Power, Alan Third (2011) Automating generation of textual class definitions from OWL to English. Journal of Biomedical Semantics, 2011 May 17, Vol. 2 Suppl 2:S5.
- Sandra Williams and Ehud Reiter [2008]. Generating basic skills reports for low-skilled readers. Journal of Natural Language Engineering, 14:4, 495-525, doi: 10.1017/S1351324908004725.
- Peter Wyard, Alison Simons, Steve Appleby, Edward Kaneen, Sandra Williams and Keith Preston [1996] Spoken Language Systems. BT Technology Journal, January 1996.
- Keith Preston and Sandra Williams [1994] Managing the Information Overload. Physics in Business, Institute of Physics, June 1994.

Refereed Conference Proceedings

- Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power and Sandra Williams [2013] A System for the Simplification of Numerical Expressions at Different Levels of Understandability. Proceedings of the workshop on Natural Language Processing for Improving Textual Accessibility (NLP4ITA), Atlanta, USA, pp.10-19.
- Nguyen, Tu Anh T., Power, Richard, Piwek, Paul and Williams, Sandra [2013] Predicting the understandability of OWL inferences. In: Extended Semantic Web Conference 2013 (ESWC 2013) - Research Track, 26th May to 30th May, 2013, Montpellier, France.
- Sandra Williams [2013] An Information Extraction System for English Ontology Identifier Names. Technical Report TR2013/03, Department of Computing, The Open University.
- Sandra Williams [2013] An Analysis of POS Tag Patterns in Ontology Identifiers and Labels. Technical Report TR2013/02, Department of Computing, The Open University.
- Nguyen, Tu Ahn, Power, Richard, Piwek, Paul and Williams, Sandra [2012] Measuring the understandability of deduction rules for OWL. In: First International Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM12), 8th October 2012, Galway, Ireland.
- Nguyen, Tu Anh T, Power, Richard, Piwek, Paul and Williams, Sandra [2012] Planning accessible explanations for entailments in OWL ontologies. In: 7th International Natural Language Generation Conference (INLG 2012), 30 May-1 June 2012, Utica, IL, USA.
- Sandra Williams, Allan Third and Richard Power [2011] Levels of organisation in ontology verbalisation. Proceedings of the 13th European Workshop on Natural Language Generation, September 2011, Nancy, France, pp. 158-163.
- Sandra Williams [2011] Generating Mathematical Word Problems. Proceedings of the AAAI Fall Symposium on Question Generation, November 2011, Arlington, VA, USA
- Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, Sandra Williams [2011] Experimental Identification of the Use of Hedges in the Simplification of Numerical Expressions. The Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), 30 July 2011, Edinburgh, Scotland.
- Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, Sandra Williams [2011] How to Make Numerical Information Accessible: Experimental Identification of Simplification Strategies. In Part 1 Proceedings of 13th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT 2011), P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque and M. Winckler (Eds.), Lecture Notes in Computer Science (LNCS) Volume 6946, pp. 57-64, September 5-9, DOI: 10.1007/978-3-642-23774-4
- Allan Third, Sandra Williams and Richard Power [2011] OWL to English: a tool for generating organised easily-navigated hypertexts from ontologies. Proceedings of the 10th International Semantic Web Conference (ISWC2011).
- Sandra Williams and Richard Power [2010] Grouping axioms for more coherent ontology descriptions.

- 6th International Natural Language Generation Conference (INLG 2010), Dublin.
- Robert Stevens, James Malone, Sandra Williams and Richard Power [2010] Automating Class Definitions From OWL to English. *Bio-Ontologies 2010: Semantic Applications in Life Sciences SIG at 18th Annual International conference on Intelligent Systems for Molecular Biology (ISMB 2010)*, Boston.
- Sandra Williams and Richard Power [2010] A fact-aligned corpus of numerical expressions. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*, Malta, pp. 2396-2400.
- Sandra Williams and Richard Power [2009] Precision and mathematical form in first and subsequent mentions of numerical facts and their relation to document structure. *Proceedings of 12th European Workshop on Natural Language Generation*, Athens.
- Sandra Williams and Ehud Reiter [2008] SkillSum: basic skills screening with personalised, computer-generated feedback. *Interactive Computer Aided Learning (ICL 2008)*, Special Track: Computer-based Knowledge & Skill Assessment and Feedback in Learning Settings.
- Sandra Williams, Richard Power and Paul Piwek [2008] Simulating emotional reactions in medical dramas. *Proceedings of the Symposium on Affective Language in Human and Machine, Volume 2, The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) 2008 Convention: Communication, Interaction and Social Intelligence*, Aberdeen, April 2008, pp. 25-32.
- Ehud Reiter and Sandra Williams [2008] Three Approaches to Generating Texts in Different Styles. *Proceedings of the Symposium on Style in text: creative generation and identification of authorship, Volume 7, The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) 2008 Convention: Communication, Interaction and Social Intelligence*, Aberdeen, April 2008, pp. 26-33.
- Sandra Williams and Richard Power [2008] Deriving rhetorical complexity data from the RST-DT Corpus. *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, Marrakech, Morocco, 28-30 May, 2008.
- Sandra Williams, Paul Piwek and Richard Power [2007] Generating monologue and dialogue to present personalised medical information to patients. *Proceedings of the 11th European Workshop on Natural Language Generation*, pp. 167-170.
- Ehud Reiter, Sandra Williams and Lesley Crichton [2005] Generating Feedback Reports for Adults Taking Basic Skills Tests. *Proceeding of the The Twenty-fifth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, UK. In A Macintosh, R Ellis, and T Allen (ed) *Applications and Innovations in Intelligent Systems XIII (Proceedings of ES-05)*, pp. 50-63.
- Sandra Williams and Ehud Reiter [2005] Generating readable texts for readers with low basic skills. *Proceeding of the 10th European Workshop on Natural Language Generation*, Aberdeen.
- Sandra Williams and Ehud Reiter [2005] Appropriate Microplanning Choices for Low-Skilled Readers. *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh.
- Sandra Williams and Ehud Reiter [2005] Deriving content selection rules from a corpus of non-naturally occurring documents for a novel NLG application. *Workshop on Using Corpora for Natural Language Generation*, Information Technology Research Institute (ITRI), Report ITRI-05-03, University of Brighton.
- Sandra Williams and Ehud Reiter [2004] Reading errors made by skilled and unskilled readers: evaluating a system that generates reports for people with poor literacy. *Fourteenth Annual Meeting of the Society for Text and Discourse*, Chicago.
- Sandra Williams [2003] Language choice models for microplanning and readability. *Student Workshop of the Human Language Technology and North American Chapter of the Association for Computational Linguistics Conference (HLT-NAACL03 Student Workshop)*, pp. 13-18.
- Sandra Williams, Ehud Reiter and Liesl Osman [2003] Experiments with discourse-level choices and readability. *Proceedings of the 9th European Workshop on Natural Language Generation*, Budapest, pp. 127-134, April 2003.
- Ehud Reiter, Somayajulu Sripada and Sandra Williams [2003] Acquiring and Using Limited User Models in NLG. *Proceedings of the 9th European Workshop on Natural Language Generation*, Budapest, pp. 87-94, April 2003.

- Sandra Williams and Ehud Reiter [2003] A corpus analysis of discourse relations for Natural Language Generation. Proceedings of Corpus Linguistics 2003, pp. 899-908, Lancaster University, March 2003.
- Sandra Williams [2002] Natural language generation of discourse connectives for different reading levels. The UK special interest Group for computational linguistics, 5th Annual CLUK. Research Colloquium.
- Sandra Williams and Catherine I. Watson [1999] A Profile of the Discourse and Intonational Structures of Route Descriptions. Proceedings of the 6th European Conference on Speech Communication and Technology, Eurospeech'99, September 5-9, 1999, Budapest, Hungary, Volume 4, pp. 1659-1662.
- Sandra Williams [1998] Generating Pitch Accents in a Concept-To-Speech System Using a Knowledge Base. Proceedings of the 5th International Conference on Spoken Language Processing, (ICSLP'98), Volume 4, pp. 1159-1162.
- Robert Dale, Stephen Green, Maria Milosavljevic, Cécile Paris, Cornelia Verspoor and Sandra Williams [1998] Dynamic Document Delivery: Generating Natural Language Texts on Demand. 9th International Conference and Workshop on Database and Expert Systems Applications. August 24-28, Vienna, Austria.
- Robert Dale, Stephen Green, Maria Milosavljevic, Cécile Paris, Cornelia Verspoor and Sandra Williams [1998] Using Natural Language Generation Techniques to Produce Virtual Documents. Proceedings of the Third Australian Document Computing Symposium, August 21st, Sydney, Australia.
- Robert Dale, Stephen Green, Maria Milosavljevic, Cécile Paris, Cornelia Verspoor and Sandra Williams [1998] The Realities of Generating Natural Language from Databases. 11th Australian Joint Conference on Artificial Intelligence, 12-17 July, Brisbane, Australia.
- Cornelia Verspoor, Robert Dale, Stephen Green, Maria Milosavljevic, Cécile Paris, and Sandra Williams [1998] Intelligent Agents for Information Presentation: Dynamic Description of Knowledge Base Objects. In Proceedings of the International Workshop on Intelligent Agent on the Internet and Web, Mexico City, Mexico, 16-20 March 1998, pp. 75-86.
- Sandra Williams, Mark Harvey and Keith Preston [1996] Rule-based reference resolution for unrestricted text using part-of-speech tagging and noun phrase parsing. Discourse Anaphora and Anaphor Resolution Colloquium (DAARC), Lancaster, U.K., July 1996
- Sandra Williams [1996] Anaphoric reference and ellipsis resolution in a telephone-based spoken language system for accessing email. Discourse Anaphora and Anaphor Resolution Colloquium (DAARC), Lancaster, U.K., July 1996.
- Sandra Williams [1996] Dialogue management in a mixed-initiative, cooperative, spoken language system. 11th Twente Workshop on Language Technology (TWLT11) Dialogue Management in Natural Language Systems, Enschede, Netherlands, June 1996
- Peter Wyard, Steven Appleby, Edward Kaneen, Sandra Williams and Keith Preston [1995] A Combined Speech and Visual Interface to the BT Business Catalogue. ESCA Workshop on Spoken Dialogue Systems, 30th May - 2nd June 1995.

PhD Thesis

- Sandra Williams [2004] Natural Language Generation (NLG) of discourse relations for different reading levels. PhD Thesis, University of Aberdeen.