Diarmuid Ó Séaghdha

Address:	Computer Laboratory University of Cambridge 15 JJ Thomson Avenue Cambridge CB3 0FD United Kingdom	Telephone (work): Fax:	01223 763582 01223 334678
Nationality:	Irish	Email:	do242@cam.ac.uk

Education

2004 – 2008 **Computer Laboratory, University of Cambridge, UK**
PhD in Computer Science
Supervisor: Dr. Ann Copestake
Thesis Title: *Learning Compound Noun Semantics*
Defended September 2008, Conferred July 2009

2003 – 2004 **Department of Oriental Studies, University of Cambridge, UK**
MPhil in Sanskrit
Supervisor: Dr. Eivind Kahrs
Thesis Title: *Object Language and Metalanguage in Sanskrit Grammatical Texts*

1999 – 2003 **Trinity College Dublin, Ireland**
BA(Mod) in Computer Science, Linguistics and German
Grade: 1st Class Honours

Post-Education Employment

Aug 09 – **Computer Laboratory, University of Cambridge, UK**
Postdoctoral **Research Associate** on EPSRC project *Lexical Acquisition for the Biomedical Domain* (PI Anna Korhonen).

Aug 08 – Jul 09 **Computer Laboratory, University of Cambridge, UK**
Postdoctoral **Research Assistant** on EPSRC-funded Knowledge Transfer project, in conjunction with Linguamatics Ltd.

Publications

Journal articles

Tom Lippincott, Diarmuid Ó Séaghdha and Anna Korhonen. 2011. Exploring subdomain variation in biomedical language. *BMC Bioinformatics* 12:212.

Peer-reviewed international conference proceedings

Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the 5th ACM Conference on Web Search and Data Mining (WSDM 2012)*. Seattle, WA. [21% acceptance rate]

Diarmuid Ó Séaghdha and Anna Korhonen. 2011. Probabilistic models of similarity in syntactic context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*. Edinburgh, UK. [23% acceptance rate]

Tom Lippincott, Diarmuid Ó Séaghdha, Lin Sun and Anna Korhonen. 2010. Exploring variation across biomedical subdomains. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*. Beijing, China. [19% acceptance for oral presentation]

Diarmuid Ó Séaghdha. 2010. Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*. Uppsala, Sweden. [25% acceptance rate]

Diarmuid Ó Séaghdha. 2009. Semantic classification with WordNet kernels. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies Conference (NAACL-HLT 2009)*. Boulder, CO. (Short paper)

Diarmuid Ó Séaghdha and Ann Copestake. 2009. Using lexical and relational similarity to classify semantic relations. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)*. Athens, Greece. [28% acceptance rate]

Diarmuid Ó Séaghdha and Ann Copestake. 2008. Semantic classification with distributional kernels. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*. Manchester, UK. [24% acceptance rate]

Diarmuid Ó Séaghdha. 2007. Designing and evaluating a semantic annotation scheme for compound nouns. In *Proceedings of Corpus Linguistics 2007*.

Peer-reviewed international workshop proceedings

Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano and Stan Szpakowicz. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. In *Proceedings of the SemEval-2 Workshop*. Uppsala, Sweden.

Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz and Tony Veale. 2010. SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *Proceedings of the SemEval-2 Workshop*. Uppsala, Sweden.

Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano and Stan Szpakowicz. 2009. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. *In Proceedings of the NAACL-HLT-09 Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-09)*. Boulder, CO.

Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz and Tony Veale. 2009. SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. *In Proceedings of the NAACL-HLT-09 Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-09)*. Boulder, CO.

Andreas Vlachos, Paula Buttery, Diarmuid Ó Séaghdha and Ted Briscoe. 2009. Biomedical Event Extraction without Training Data. In *Proceedings of BioNLP 2009*. Boulder, CO.

Diarmuid Ó Séaghdha. 2007. Annotating and learning compound noun semantics. In *Proceedings of the ACL-07 Student Research Workshop*. Prague, Czech Republic. [31% acceptance rate]

Diarmuid Ó Séaghdha and Ann Copestake. 2007. Co-occurrence contexts for noun compound interpretation. In *Proceedings of the ACL-07 Workshop A Broader Perspective on Multiword Expressions*. Prague, Czech Republic.

Selected Academic Awards and Scholarships

Undergraduate Awards

2001 – 2003 Trinity College Dublin Foundation Scholarship
2003 TCD Gold Medal for Academic Excellence

Postgraduate Scholarships
2003 – 2006 Robert Gardiner Memorial Scholarship
2003 – 2007 Corpus Christi College Taylor Scholarship
2004 – 2007 EPSRC Doctoral Studentship

Skills and Research Interests

Languages (natural): English (native), Irish (fluent), German (fluent), French(intermediate), Sanskrit (reading).

Languages (computer): Java, Perl, C, C++, Matlab, PHP, SQL, Prolog, Visual Basic, HTML.

Natural Language Processing: Lexical semantics, identification of semantic relations, multiword expressions, lexical disambiguation, semantic structured prediction

Machine Learning: Supervised and semi-supervised classification, kernel methods for vectorial and structured data, unsupervised Bayesian methods, graph-based learning

Professional Activities

Occasional reviewer for *Computational Linguistics*, *Journal of Semantics*, *Journal of Linguistics, Computer Speech and Language, Journal of Artificial Intelligence Research, Journal of Natural Language Engineering*

Reviewer for journal special issues on *Computational Approaches to the Semantics of Noun Compounds* (Journal of Natural Language Engineering) and *Multiword Expressions* (ACM Transactions on Speech and Language Processing).

Reviewer for EACL 2009-, ACL 2010-, COLING 2010, IJCNLP 2011, NAACL-HLT 2012, AAAI 2012 and numerous workshops.

Co-organiser of ACL-11 Workshop on Relational Models of Semantics (RELMS-11)

Co-organiser of Computational Linguistics UK 2007 postgraduate student conference.

Co-organiser of SemEval-2010 Task 8 (*Multi-Way Classification of Semantic Relations Between Pairs of Nominals*) and Task 9 (*The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions*).

Co-organiser of SemEval-2013 Task 24 (*Free Paraphrases of Noun Compounds*).

Teaching **(at Computer Laboratory, University of Cambridge, UK unless otherwise stated)**

2011
Created and lectured summer school course on *Distributional Methods for Semantic Analysis* at HIT-MRSA Summer School on Language Technology, Harbin Institute of Technology, China.

2009 – present
Occasional lecturing as stand-in on undergraduate courses**.**

2009 – present
Supervised MPhil research projects:
Clemens Heppner (2010) *A probabilistic framework for noun compound paraphrasing*
Yuan Zhang (2011) *Auralist: Introducing serendipity into music recommendation systems*

2008
Lectured MPhil course on Semantics

2003 – 2008
Undergraduate supervisions for *Databases, Artificial Intelligence I, Natural Language Processing* and *Information Retrieval*.

Other Interests and Activities

Athletics: Cambridge University cross-country half-blue, marathon full blue (PB 2:28:36), Member of Cambridge and Coleridge and Donore Harriers athletic clubs.

Photography, Cookery, Travel, Film, Music.

References available on request