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: Prof. Ann Copestake
Thesis Title: *Learning Compound Noun Semantics*

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 (with gold medal for excellence)

POST-EDUCATION EMPLOYMENT

Sep 14 –   **Vocal IQ Ltd, Cambridge, UK**
Senior NLP Researcher

May 14 – Sep 14   **Computer Laboratory, University of Cambridge, UK**
Senior Research Associate

Aug 12 – May 14   **Computer Laboratory, University of Cambridge, UK**
Postdoctoral **Research Associate** on IARPA-funded project *Foresight and Understanding from Scientific Exposition (FUSE)* (PI Simone Teufel).

Aug 09 – Jul 12   **Computer Laboratory, University of Cambridge, UK**
Postdoctoral **Research Associate** on EPSRC-funded 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.

PRIMARY RESEARCH INTERESTS

- **Dialogue:** understanding and generation in spoken dialogue systems
- **Lexical semantics:** distributional semantics, disambiguation, multiword expressions
- **Information extraction:** identifying semantic relations, structured prediction for IE
- **Discourse:** power/group dynamics in language, structure and rhetoric in scientific discourse
- **Social media:** language use in social media, social data mining, recommender systems
- **Machine learning:** Bayesian modelling, unsupervised learning, graph-based NLP
- **Applications:** NLP for digital humanities, biomedical NLP

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

2011, 2013
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:

Meng Zhang (2014)	*Tracking word meaning through time*
Adrian Scoică (2013)	*Multilingual generative models for selectional preference learning*
Simo Tchokni (2013)	*The language of influence in social media*
Zhe Wang (2012)	*Linguistic analysis for computational social science*
Yuan Zhang (2011)	*Auralist: Introducing serendipity into music recommendation systems*
Clemens Heppner (2010)	*A probabilistic framework for noun compound paraphrasing*

2008
Lectured MPhil course on Semantics

2006 – present
Supervised final-year undergraduate projects:

Russ Williams (2006)	*Machine translation of noun-noun compounds*
Iulia-Raluca Turc (2013)	*Mapping sentiments in tweets onto the world map*

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

PROFESSIONAL ACTIVITIES

Visiting Industrial Fellow at University of Cambridge Computer Laboratory, 2014 –

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/PC member for EACL, ACL, COLING, IJCNLP, NAACL-HLT, AAAI, EMNLP, *SEM, IJCAI, IWCS, RANLP, CSCW 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*).

Invited keynote speaker at COLING-14 *Workshop on Computational Approaches to Compound Analysis* (ComAComA-14).

PUBLICATIONS

Books

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha and Stan Szpakowicz. 2013. *Semantic Relations Between Nominals*. Morgan and Claypool.

Journal articles

Kathleen McKeown, Hal Daumé III, Snigdha Chaturvedi, John Paparrizos, Kapil Thadani, Pablo Barrio, Or Biran, Suvarna Bothe, Michael Collins, Kenneth R. Fleischmann, Luis Gravano, Rahul Jha, Ben King, Kevin McInerney, Taesun Moon, Arvind Neelakantan, Diarmuid Ó Séaghdha, Dragomir Radev, Clay Templeton and Simone Teufel. Predicting the Impact of Scientific Concepts using Full Text Features. To appear in *Journal of the Association for Information Science and Technology.*

Diarmuid Ó Séaghdha and Anna Korhonen. 2014. Probabilistic distributional semantics with latent variable models. To appear in *Computational Linguistics*.

Diarmuid Ó Séaghdha and Ann Copestake. 2013. Interpreting compound nouns with kernel methods. *Journal of Natural Language Engineering* 19(3): 331-356.

Anna Korhonen, Diarmuid Ó Séaghdha, Ilona Silins, Lin Sun, Johan Hogberg and Ulla Stenius. 2012. Text mining for literature review and knowledge discovery in cancer risk assessment and research. *PLoS ONE* 7(4):e33427.

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

Peer-reviewed international conference proceedings

Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gašić, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen and Steve Young. 2015. Multi-domain Dialog State Tracking using Recurrent Neural Networks. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2015)*. Beijing, China.

Diarmuid Ó Séaghdha and Simone Teufel. 2014. Unsupervised learning of rhetorical structure with un-topic models. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*. Dublin, Ireland.   [20% acceptance for oral presentation]

Yufan Guo, Diarmuid Ó Séaghdha, Ilona Silins, Lin Sun, Johan Högberg, Ulla Stenius and Anna Korhonen. 2014. CRAB: A text mining tool for supporting literature review in chemical cancer risk assessment. In *Proceedings of the COLING-14 Demonstration Session*. Dublin, Ireland.

Simo Tchokni, Diarmuid Ó Séaghdha and Daniele Quercia. 2014. Emoticons and Phrases: Status Symbols in Social Media. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM 2014)*. Ann Arbor, MI.   [23% acceptance rate]

Zhe Wang, Daniele Quercia and Diarmuid Ó Séaghdha. 2013. Reading Tweeting Minds: Real-time Analysis of Short Text for Computational Social Science. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HYPERTEXT 2013)*. Paris, France.

Sandro Bauer, Anastasios Noulas, Diarmuid Ó Séaghdha, Stephen Clark and Cecilia Mascolo. Talking Places: Modelling and analysing linguistic content in Foursquare. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing (SocialCom 2012)*. Amsterdam, The Netherlands.   [23% acceptance rate]

Diarmuid Ó Séaghdha and Anna Korhonen. 2012. Modelling selectional preferences in a lexical hierarchy. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics (*SEM 2012)*. Montreal, QC.   [28% acceptance rate]

Tom Lippincott, Anna Korhonen and Diarmuid Ó Séaghdha. 2012. Learning syntactic verb frames using graphical models. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*. Jeju, Korea.   [19% acceptance rate]

Daniele Quercia, Diarmuid Ó Séaghdha and John Crowcroft. 2012. Talk of the City: Our Tweets, Our Community Happiness. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM 2012)*. Dublin, Ireland.   [26% poster acceptance]

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, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz and Tony Veale. 2013. SemEval-2013 Task 4: Free Paraphrases of Noun Compounds. In *Proceedings of the SemEval-2013 Workshop*. Atlanta, GA.

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.

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.

OTHER SKILLS/ACTIVITIES

Languages (natural):   English (native), Irish (fluent), German (fluent), French (advanced)

Languages (computer):   Java, Perl, Scala, Matlab, SQL, Prolog, C++.

Recreational activities: Marathon running (PB 2:28:36), cycle touring, photography