

Personal Details:

Address: National Centre for Language Technology, School of Computing, DCU, Dublin 9
Nationality: Irish
Email Address: aoife.cahill@computing.dcu.ie
Web Page: <http://www.computing.dcu.ie/~acahill/>

Education:

Ph.D. [Dublin City University](#), Dublin 9
2001-04: Ph.D. under the supervision of [Prof. Josef van Genabith](#) and [Dr. Andy Way](#)
Thesis Title: Parsing with Automatically Acquired, Wide-Coverage, Robust, Probabilistic LFG Approximations
Brief Abstract: Traditionally, rich, constraint-based grammatical resources have been hand-coded. Scaling such resources beyond toy fragments to unrestricted, real text is knowledge-intensive, time-consuming and expensive. This thesis proposes a method for treebank-based, wide-coverage, deep, constraint-based grammar acquisition. The resulting probabilistic, LFG approximations parse the Penn-II treebank with wider coverage (measured in terms of complete spanning parse) and parsing results comparable to or better than those achieved by the best hand-crafted grammars, with, we believe, considerably less grammar development effort. We believe that our approach successfully addresses the knowledge-acquisition bottleneck (familiar from rule-based approaches to AI and NLP) in wide-coverage, constraint-based grammar development. Our approach can provide an attractive, wide-coverage, multilingual, deep, constraint-based grammar acquisition paradigm.

BSc. in Applied Computational Linguistics [Dublin City University](#), Dublin 9
1997 - 2001: BSc. (Hons) Applied Computational Linguistics - Grade 1.1
Final year project: A Probabilistic Context-Free Parser

Research Interests

Computational Linguistics: Statistical Parsing and Generation, Statistical methods in NLP, Unification Grammars (esp. LFG), Corpus linguistics, Empirical methods in NLP, Treebanks, Question Answering, Machine Translation

Research and Other Relevant Experience

Aug. 2004 – July 2007 Co-chair of the Program Committee for the International Conferences in Lexical Functional Grammar (LFG) in Bergen (2005), Konstanz (2006) and Stanford (2007).
March 2006-Sept. 2006 Visiting Researcher to the Natural Language Theory and Technology group at the Palo Alto Research Centre (PARC), California, USA
Oct. 2004 - present Post-doctoral Researcher on SFI Basic Research grant 04/BR/CS0370 entitled “Pro-Gen: Robust Generation with Treebank-based Unification Grammars”, NCLT, DCU
April – July 2003 Visiting Researcher at the Universität Stuttgart, Germany, where I worked with Dr. Helmut Schmid and Martin Forst on porting our previous work on English to German.
Oct. 2001 – Sept. 2004 Doctoral Student on Enterprise Ireland funded Basic Research Grant SC/2001/0186: “Automatic Annotation of the Penn-II Treebank with Feature-Structure Information”
Nov. 1999 – Feb. 2000 Worked as an Assistant Scientist in the University of Stuttgart, providing feedback to developers a machine translation project (Verbmobil) on the quality of translations.
Oct 1999 – Was awarded a highly competitive DAAD (German Academic Exchange Service)

- July 2000** Scholarship for the academic year 1999/2000 at IMS, Stuttgart University, Germany.
- June – Sept. 1998** Summer Student Internship in the Data Mining Department of IBM Stuttgart, Germany, working on the testing and development of DB2 products.

Selected Peer-Reviewed Publications

- Book Chapters** **1. Cahill, A.,** McCarthy, M., Burke, M., van Genabith, J. and Way, A (to appear) *Deriving Quasi-Logical Forms from F-Structures for the Penn Treebank* in (eds.) Harry Bunt et. al. Computing Meaning, Vol-3, Kluwer Academic Publishers
- Journal Articles** **1. Cahill, A.,** Burke, M., O'Donovan, R., Riezler, S., van Genabith, J. and Way, A. (under review) *Shallow and Deep Parser Comparison with Automatically Generated Dependency Relations*. Under review in Computational Linguistics
- 2. Cahill, A.,** Forst, M., Burke, M., McCarthy, M., O'Donovan, R., Rohrer, C., van Genabith, J., and Way, A. (2005) *Treebank-Based Acquisition of Multilingual Unification Grammar Resources*. In Journal of Research on Language and Computation; Volume 3, Number 2, pages 247 – 279.
- 3. O'Donovan, R.,** Burke, M., **Cahill, A.,** van Genabith J., and Way, A. (2005) *Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks*. In Computational Linguistics, Vol. 31, Issue 3, pages 329 – 365.
- 4. Burke, M., Cahill, A.,** McCarthy, M., O'Donovan, R., van Genabith J., and Way, A. (2004). *Evaluating Automatic F-Structure Annotation for the Penn-II Treebank*. In Journal of Language and Computation, pages 523-547
- Conference Proceedings:** **1. O'Donovan, R., Cahill, A.,** van Genabith J., and Way, A. (to appear) *Automatic Acquisition of Spanish LFG Resources from the CAST3LB Treebank*. In Proceedings of the Tenth International Conference on LFG, Bergen, Norway
- 2. Judge, J.,** Burke, M., **Cahill, A.,** O'Donovan, R., van Genabith J., and Way, A. (to appear) *Strong Domain Variation and Treebank-Induced LFG Resources*. In Proceedings of the Tenth International Conference on LFG, Bergen, Norway)
- 3. Burke, M., Cahill, A.,** van Genabith J., and Way, A. (to appear) *Evaluating Automatically Acquired F-Structures against Propbank*. In Proceedings of the Tenth International Conference on LFG, Bergen, Norway
- 4. Cahill, A.,** Burke, M., O'Donovan, R., van Genabith, J., and Way, A. (2004) *Long-Distance Dependency Resolution in Automatically Acquired Wide-Coverage PCFG-Based LFG Approximations*. In Proceedings of ACL '04, pages 320–327, Barcelona.
- 5. O'Donovan, R.,** Burke, M., **Cahill, A.,** van Genabith, J., and Way, A. (2004). *Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II Treebank*. In Proceedings of ACL '04, pages 368–375, Barcelona, Spain.
- 6. Burke, M., Cahill, A.,** O'Donovan, R., van Genabith, J., and Way, A. (2004). *The Evaluation of an Automatic Annotation Algorithm against the PARC 700 Dependency Bank*. In Proceedings of LFG '04, pages 101-121, Christchurch, NZ.
- 7. Cahill, A.,** McCarthy, M., van Genabith, J., and Way, A. (2003). *Quasi-Logical Forms for the Penn Treebank*. In Proceedings of the Fifth International Workshop on Computational Semantics, pages 55–71, Tilburg, The Netherlands.
- 8. Cahill, A.,** Forst, M., McCarthy, M., O'Donovan, R., Rohrer, C., van Genabith, J., and Way, A. (2003a). *Treebank-Based Multilingual Unification-Grammar Development*. In Proceedings of the Workshop on Ideas and Strategies for Multilingual Grammar Development, at ESSLLI 2003, pages 17–24, Vienna, Austria.
- 9. Cahill, A.** and van Genabith, J. (2002). *TTS – A Treebank Tool Suite*. In Proceedings of LREC 2002, volume V, pages 1712–1717, Las Palmas, Spain.
- 10. Cahill, A.,** McCarthy, M., van Genabith, J., and Way, A. (2002a). *Automatic Annotation of the Penn Treebank with LFG F-Structure Information*. In Proceedings

of the LREC Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data, pages 8–15, Las Palmas.

11. Cahill, A., McCarthy, M., van Genabith, J., and Way, A. (2002). *Parsing with PCFGs and Automatic F-Structure Annotation*. In Proceedings of LFG '02, pages 76–95, Athens, Greece.

Selected Presentations

- 2005** PARC (Palo Alto Research Center, USA), FujiXerox (Japan), Tokyo University (Japan), International Lexical Functional Grammar Conference (Norway)
- 2004** Annual Meeting of the Association of Computational Linguistics International Conference (Spain), PARC (USA)
- 2003** European Summer School in Logic, Language and Information (Austria), Stuttgart University (Germany), PARC (USA), Dublin Computational Linguistics Seminar Series (Ireland)
- 2002** PARC (USA), International LFG Conference (Greece), International Conference on Linguistic Resources and Evaluation (Spain), FujiXerox (Japan)

Training and Skills

Leadership Training I have completed two modules of a Leadership Training course provided by DCU's Training and Development team.

Administration I am joint administrator of the National Centre for Language Technology (NCLT)

Project Supervision I am currently jointly supervising one Ph.D. student. His thesis focuses on exploiting and improving Natural Language resources in a Question-Answering scenario. The expected submission date of the thesis is July 2006. I am also jointly supervising a final year Applied Computational Linguistics B.Sc. student. His project concentrates on probabilistic parsing of text into Propbank-style predicate-argument structures

Organisation I am co-managing the SFI Principal Investigator Grant GramLab project (2 post-doctoral researchers and 6 Ph.D. students) with Prof. van Genabith. I organise weekly meetings and administer a cluster of hi-spec computers used by 30 NCLT members.

Awards and Achievements

2003 Won an ESSLLI student scholarship to attend the 15th European Summer School in Logic, Language and Information, Vienna, Austria in 2003

2001 Won a prize for the best 4th year project in Applied Computational Linguistics sponsored by TRADOS Ireland

1999/2000 Won a highly competitive DAAD (German Academic Exchange Service) Scholarship for the academic year 1999/2000 Stuttgart University, Germany

1998/1999 Committee Member (Helpdesk) of DCU's Networking society (Redbrick)

1996 Completed Senior Certificate Piano, in the Royal Irish Academy of Music

1995 Member of the Irish Team that competed in the 7th International Olympiad in Informatics in Eindhoven, the Netherlands

References

Prof. Josef van Genabith (josef@computing.dcu.ie)
School of Computing, DCU, Dublin 9, Ireland
Tel: +353 1 700 5074

Dr. Andy Way (away@computing.dcu.ie)
School of Computing, DCU, Dublin 9, Ireland
Tel: +353 1 700 5644