

Personal Details:

Nationality: Irish
Email Address: aoife.cahill@gmail.com
Web Page: <http://www.ims.uni-stuttgart.de/~cahillae>

Education:

Ph.D. Dublin City University, Dublin 9

2001-2004: 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 at [Dublin City University](#), Dublin 9: **1997 – 2001**

Research Interests

My main interest is in Natural Language Processing, in particular integrating statistical techniques with traditional rule-based approaches. To date, I have predominantly worked on parsing and generation (within the LFG framework) and have contributed to substantial advances in the field. Some other areas of NLP that I find particularly interesting are: Language Modelling, Corpus linguistics, Question Answering, Treebanks, Machine Translation, Machine Learning and Information Extraction.

Work Experience

Nov. 2006-present Post-doctoral Researcher on the D2 project of the German Science Foundation funded SFB-732 “Incremental Specification in Context” at the University of Stuttgart.

The project I am working on deals mostly with natural language generation, specifically ranking of candidate strings. Using an existing hand-crafted grammar of German, we have developed a state-of-the-art maximum entropy ranking model, that outperforms a simple language model baseline. The focus of my research has been to design and develop linguistically-motivated features for the ranking. We have tested our models on the TIGER corpus (around 50K sentences of German newspaper text). I was also heavily involved in the writing of a proposal to extend and continue this sub-project. Our proposal was successful and the DFG (German Research Council) awarded funding for the larger project for a further 4 years.

Nov. 2007 Visiting Researcher at the University of Bergen, Norway

Here I carried out experiments on parsing Norwegian, with the specific focus of making the parsing process faster. We carried out experiments on English, German and Norwegian and achieved a significant improvement in parse-times, with almost no loss in parse accuracy.

March 2006- Visiting Researcher to the Natural Language Theory and Technology group at the Palo

- Sept. 2006** Alto Research Centre (PARC), California, USA
Here I worked on integrating insights from my Ph.D. thesis into the XLE parsing framework, developed at PARC. We carried out preliminary experiments on increasing the efficiency of the parser, which eventually led to a fully integrated module in the XLE allowing for (optional) pruning of the parse chart to speed up the process.
- Oct. 2004 – Sept. 2006** Post-doctoral Researcher on SFI Basic Research grant 04/BR/CS0370 entitled “Pro-Gen: Robust Generation with Treebank-based Unification Grammars”, NCLT, DCU
I was a post-doctoral researcher and co-manager on a large Science Foundation Ireland project (2 post-doctoral researchers and 6 Ph.D. students). The project I worked on was on natural language generation, specifically automatically learning a generation grammar from annotated corpora of English. The work built on what I had achieved during my Ph.D on the automatic acquisition of "deep" grammars for parsing. We used simple maximum likelihood models to rank alternative solutions and achieved encouraging results. During this time, I was also the joint administrator of the National Center for Language Technology in DCU, together with Prof. Josef van Genabith.
- April – July 2003** Visiting Researcher at the University of Stuttgart, Germany
Here 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”
- Oct 1999 – July 2000** Was awarded a highly competitive DAAD (German Academic Exchange Service) 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**
- Cahill, A.** and Forst, M. Human Evaluation of a German Surface Realisation Ranker. In Emiel Kraemer and Mariet Theune (eds.), *To appear in Empirical Methods in Natural Language Generation*. Springer, 2010.
 - Cahill, A.**, McCarthy, M., Burke, M., van Genabith, J. and Way, A (2007) *Deriving Quasi-Logical Forms from F-Structures for the Penn Treebank* in (eds.) Harry Bunt et. al. *Computing Meaning*, Vol-3, Kluwer Academic Publishers, pages 33-54
- Journal Articles**
- Cahill, A.**, Burke, M., O'Donovan, R., Riezler, S., van Genabith, J. and Way, A. (2008) *Wide-Coverage Deep Statistical Parsing using Automatic Dependency Structure Annotation* in *Computational Linguistics*, Vol. 34, No. 1, pages 81-124
 - Cahill, A.** (2008) *Treebank-Based Probabilistic Phrase Structure Parsing* in *Language and Linguistics Compass* 2 (1), 18-40.
 - 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*; Vol 3, pages 247 – 279.
 - 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.
 - 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
- Peer-reviewed Conference**
- Zarriß S., **Cahill A.**, Kuhn J. and Rohrer C. (2010) *Cross-Lingual Induction for Deep Broad-Coverage Syntax: A Case Study on German Participles* In *Proceedings of COLING 2010*, Beijing, China

- Proceedings:**
2. **Cahill, A.** and Riester, A. (2009) *Incorporating Information Status into Generation Ranking* In Proceedings of ACL-IJCNLP 2009, pages 817-825, Singapore
 3. **Cahill, A.** (2009) *Correlating Human and Automatic Evaluation of a German Surface Realiser* In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers (ACL-IJCNLP 2009), pages 97--100, Singapore
 4. **Cahill, A.**, Heid, U., Rohrer, C. and Weller, M. (2009) *Using tri-lexical dependencies in LFG parse disambiguation* In the Proceedings LFG'09. Cambridge, England, pages 208-221.
 5. **Cahill, A.** and Forst, M. (2009) *Human Evaluation of a German Surface Realisation Ranker* In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 112 - 120, Athens, Greece, March. Association for Computational Linguistics.
 6. **Cahill, A.**, Maxwell, J. Meurer, P, Rohrer, C. and Rosén, V. (2008) *Speeding up LFG Parsing using C-Structure Pruning* , in Coling 2008: Proceedings of the workshop on Grammar Engineering Across Frameworks, pages 33 - 40, Manchester, England
 7. **Cahill, A.**, King, T. and Maxwell, J. (2007) *Pruning the Search Space of a Hand-Crafted Parsing System with a Probabilistic Parser* In Proceedings of the ACL Workshop on Deep Linguistic Processing, pages 65-72, Prague, Czech Republic
 8. **Cahill, A.**, Forst, M. and Rohrer, C. (2007) *Stochastic Realisation Ranking for a Free Word Order Language* In Proceedings of the 11th European Workshop on Natural Language Generation, Schloss Dagstuhl, Germany, pages 17-24
 9. **Cahill, A.** and van Genabith, J. (2006) *Robust PCFG-Based Generation using Automatically Acquired LFG Approximations*, In Proceedings of the joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics 2006, Sydney, Australia, pages 1033-1040
 10. Judge, J. **Cahill, A.** and van Genabith, J. (2006) *QuestionBank: Creating a Corpus of Parse-Annotated Questions*, In Proceedings of the joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics 2006, Sydney, Australia, pages 497-504
 11. **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.
 12. 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.
 13. **Cahill, A.**, McCarthy, M., van Genabith, J., and Way, A. (2002). *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.

Professional Activities

- 2009-2012** Member of the editorial board for the Journal of Computational Linguistics
- 2009** External expert for the hiring committee of Uppsala University Post-doctoral Fellow in Computational Linguistics UFV-PA 2009/126
- 2008** Member of the doctoral committee of Erik Velldal, who successfully defended his thesis in May 2008.
- 2005-2007** Co-chair of the programme committee of the International Lexical Functional Grammar Conference
- 2006-present** Reviewer for various international conferences and workshops including: ACL, EACL, NAACL, EMNLP, LFG, GEAF

Training and Skills

Relevant Technical Skills	I have worked with several machine learning algorithms, using off-the shelf implementations. I have a specific interest in feature design and determining useful training data. I have a very good working knowledge of the following programming and scripting languages: Java, Perl, Prolog. I have a working knowledge of the following: shell scripting, python, C++.
Organisation	I co-managed the SFI Principal Investigator Grant GramLab project (2 post-doctoral researchers and 6 Ph.D. students) with Prof. van Genabith. I organised weekly meetings and administered a cluster of hi-spec computers used by 30 NCLT members.
Thesis Supervision	I co-supervised John Judge who successful defended his Ph.D thesis in October 2006. His thesis focused on exploiting and improving Natural Language resources in a Question-Answering scenario. I jointly supervised Marion Weller's Studienarbeit (German undergraduate thesis) on the topic of improving the output of machine translation systems. This project was awarded the prestigious IVS prize at the University of Stuttgart in 2010. I jointly supervised Conor Cafferky's undergraduate thesis in 2006. His project concentrated on probabilistic parsing of text into Propbank-style predicate-argument structures and received a grade of 90%.
Administration	I was joint administrator of the National Centre for Language Technology (NCLT) from 2004-2006
Leadership Training	I have completed two modules of a Leadership Training course provided by DCU's Training and Development team.

Awards and Achievements

2010	Awarded a "Best Reviewer" award for EMNLP 2010
2003	Won a student scholarship to attend the 15th ESSLI Summer School in Vienna
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
1995	Member of the Irish Team that competed in the 7th International Olympiad in Informatics in Eindhoven, the Netherlands

References Available on Request