

Smaranda Muresan
- CURRICULUM VITAE -

Address University of Maryland Institute for Advanced Computer Studies
 A.V. Williams Building
 University of Maryland
 College Park, MD 20742

Office +1 301-405-1139

Email smara@umiacs.umd.edu

WWW http://umiacs.umd.edu/~smara

RESEARCH INTERESTS

Natural Language Processing (language learning, machine translation, constraint-based grammar formalisms, computational semantics), **Machine Learning** (relational learning, text and data mining), **Artificial Intelligence** (knowledge acquisition from text, ontologies)

EDUCATION

1999 - 2006 Ph.D. in Computer Science
 Columbia University, New York, NY
 Dissertation: Learning Constraint-based Grammars from Representative Examples: Theory and Applications
 Advisors: Dr. Owen Rambow and Dr. Judith L. Klavans

 M.Sc. in Computer Science, *October 2002*
 Columbia University, New York, NY

1994-1999 B.Sc. in Computer Science and Engineering, *Cum Laude*
 Thesis: "Automatic Logic Program Generation using Mode-Direct Inverse Entailment"
 Technical University of Cluj-Napoca, Romania

AWARDS

2004 AAAI/SIGART Doctoral Consortium Scholarship

Spring 2001 Extraordinary Teaching Assistant Award
 School of Engineering and Applied Sciences, Columbia University, New York, NY

1998-1999 Honor Fellowship
 Faculty of Automation and Computer Science, Technical Univ. of Cluj-Napoca, Romania

1996 Mention (4th place) at the Undergraduate National Mathematics Olympiad, Romania

1994-1998 Merit Fellowship
 Department of Computer Science, Technical University of Cluj-Napoca, Romania

RESEARCH EXPERIENCE

- Fall 2006 - **Postdoctoral Research Associate, University of Maryland Institute for Advanced Computer Studies, working with Prof. Philip Resnik**
Leading the development, experimentation, and evaluation effort for UMD's participation in the DARPA GALE program for the years 2007-2008, as well as the upcoming NIST 2008 Machine Translation Evaluation, together with Chris Dyer. Developing new models for statistical MT based on grammar learning from representative data.
- 1999-2006 **Graduate Research Assistant, NLP Group led by Professor Kathleen Mckeown, Columbia University**
Member of PERSIVAL project (NSF Digital Library Phase 2) under the supervision of Dr. Owen Rambow and Dr. Judith Klavans. Defined a new type of constraint-based grammars, Lexicalized Well-Founded Grammars, which capture syntax and semantics, and are learnable. Designed, implemented and evaluated a relational learning framework for LWFGs from representative examples. Applied this framework to the acquisition of medical terminological knowledge from natural language definitions. To automatically extract an input corpus of definitions from consumer-oriented medical text, I designed and implemented the DEFINDER system. I conducted several user-centered evaluations of DEFINDER, using both lay users and medical specialists (nurses and medical residents).
- Summer 2000 **Research Intern, Bell Labs, Lucent Technologies, Murray Hill, NJ**
Worked on Automatic Email Summarization. Employed machine learning methods using linguistically motivated features. The system was integrated in a TTS system for voice mail.
- 1997-1999 **Undergraduate Research Assistant, Technical University of Cluj-Napoca, Romania**
Member of the Multiparadigm Logic Programming Group. Designed and implemented an efficient system for automatic generation of logic programs using a small set of ordered representative examples. The system learned recursive, deterministic and non-deterministic logic programs.

TEACHING EXPERIENCE

- Spring 2007 **Guest lecture** on "Graph-based methods in IR and NLP", part of the graduate course *Computational Linguistics II* (Instructor: Professor Philip Resnik). This course is cross-listed between Linguistics and Computer Science.
- Fall 2006 **Guest lectures** on "Context-Free Grammars" and "Features and Unification", part of the graduate course *Computational Linguistics I* (Instructor: Professor Philip Resnik). This course is cross-listed between Linguistics and Computer Science.
- 2002-2003 **Supervised a high school student** from Bergen County Academies for his Senior Experience Program. Topic: Creation of a terminological database for the medical domain.
- Fall 2002 **Advised an undergraduate student** for a research project on medical ontologies (coverage, compatibility between ontologies).
- Spring 2001 **Head Teaching Assistant**, Columbia University, New York
Course: Programming Languages and Translators
Responsible for more than 50 students. Graded homework assignments and part of the

midterm and final exam. Held weekly office hours. **Guest lecturer** for the “Logic Programming” class.

PATENTS

2003 *System and Method of Generating Dictionary Entries*. Judith Klavans and Smaranda Muresan, Columbia University, New York, NY. Serial No 398,535.

INVITED TALKS

02/26/2008 *Learning Constraint-Based Grammars from Representative Data*. The Center for Language and Speech Processing, Johns Hopkins University.

03/01/2006 *Inducing Constraint-based Grammars using a Domain Ontology*. Computational Linguistics and Information Processing Laboratory (CLIP) at the University of Maryland Institute for Advanced Computer Studies (UMIACS).

10/22/2004 *Learning Constraint-based Grammars using a Small Semantic Treebank*. Computational Linguistics Group, Department of Computer Science, University of Toronto.

10/30/2000 *DEFINDER: Extracting Definitions of Medical Terms from On-line Articles*
IBM Open House, Watson Research Center, Hawthorne, NY.

PROFESSIONAL ACTIVITIES

Reviewer for ACM TOIS Journal, Computational Intelligence Journal, ACL 2008, AAAI-06 “Member Posters” Track, COLING 2004.

Program Committee member for FLAIRS 2008 Special Track on Applied Natural Language Processing, FLAIRS 2006 Trends in Natural Language Processing, and The 2nd International Workshop on Data Engineering in E-Commerce and Services (DEECS 2006) .

Book reviewer for Linguist List “Dynamical Grammar: Minimalism, Acquisition and Change” by P. Culicover and A. Nowak.

<http://linguistlist.org/issues/15/15-2358.html#1>.

M.Sc. Admission Committee at Columbia University The committee includes several faculty and 3 Ph.D. students.

SOCIETY MEMBERSHIPS

AAAI, ACL, AIS, and ACL special interest groups: SIGSEM (computational semantics), SIGNLL (natural language learning), MOL (mathematics of language).

DISSERTATION

Smaranda Muresan (2006). “Learning Constraint-based Grammars from Representative Examples: Theory and Applications”. *PhD Dissertation, Columbia University, New York, 2006* (http://www1.cs.columbia.edu/~smara/muresan_thesis.pdf).

My thesis work focused on a new class of grammars that effectively integrates syntax and semantics, and on the problem of inducing these grammars using a new relational learning paradigm. The new grammar formalism, Lexicalized Well-Founded Grammar, captures syntax and semantics, has ontology constraints at the grammar rule level, and is learnable from a small set of annotated examples. The semantic representation is an ontology-based representation, which is expressive enough to capture various phenomena of natural language, yet restrictive enough to facilitate learning. The new grammar learning model, Grammar Approximation by Representative Sublanguage, is based on the concept of representative examples, defining the importance to the model linguistically, and not simply by frequency, as in most previous work. I have proven that the search space for grammar induction is a complete grammar lattice, which guarantees the uniqueness of the solution. I have embedded all these theoretical concepts in a practical, implemented system for grammar learning, and showed the linguistic relevance of this framework and its usefulness for knowledge acquisition from text in the context of a medical digital library project.

REFEREED PUBLICATIONS

URL of Selected Publications

(<http://www.umiacs.umd.edu/~smara/publications/selected.html>)

URL Full List of Publications

(<http://www.umiacs.umd.edu/~smara/publications/papers.html>)

- [1] Christopher Dyer, Smaranda Muresan and Philip Resnik. “Generalizing Word Lattice Translation”. *Proceedings of the Association for Computational Linguistics, ACL 2008*.
- [2] Smaranda Muresan and Owen Rambow (2007). “Grammar Approximation by Representative Sublanguage: A New Model for Language Learning”. *Proceedings of the Association for Computational Linguistics, ACL 2007, Prague, Czech Republic*.
- [3] Smaranda Muresan (2004). “Inducing Constraint-based Grammars using a Domain Ontology”. *Proceedings of the Ninth AAAI/SIGART Doctoral Consortium*.
- [4] Smaranda Muresan, Tudor Muresan, and Judith Klavans (2004). “Inducing Constraint-based Grammars from a Small Semantic Treebank”. *AAAI Spring Symposium on Language Learning: An Interdisciplinary Perspective*, Stanford University.
- [5] Smaranda Muresan, Samuel D. Popper, Peter T. Davis, and Judith L. Klavans (2003). “Building a Terminological Database from Heterogeneous Definitional Sources”. *Proceedings of the National Conference on Digital Government Research*, Boston, Massachusetts.
- [6] Smaranda Muresan and Judith L. Klavans (2002). “A Method for Automatically Building and Evaluating Dictionary Resources”. *Proceedings of the Language Resources and Evaluation Conference (LREC 2002)*.
- [7] Smaranda Muresan, Tudor Muresan, and Rodica Potolea (2002). “Data Flow Coherence Constraints for Pruning the Search Space in ILP Tools”. *International Journal of Artificial Intelligence Tools*, volume 11(2), 2002.
- [8] Judith Klavans and Smaranda Muresan (2001). “Evaluation of DEFINDER: A System to Mine Definitions from Consumer-oriented Medical Text”. *Proceedings of The ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001)*.
- [9] Smaranda Muresan, Tudor Muresan, and Rodica Potolea (2001). “Data flow criteria in ILP

- tools”. *Proceedings of IEEE 13th International Conference on Tools with Artificial Intelligence (ICTAI 2001)*.
- [10] Judith Klavans and Smaranda Muresan (2001). “Evaluation of the DEFINDER System for Fully Automatic Glossary Construction”. *Proceedings of the American Medical Informatics Association Symposium (AMIA 2001)*.
- [11] Smaranda Muresan, Evelyne Tzoukermann, and Judith Klavans (2001). “Combining Linguistic and Machine Learning Techniques for Email Summarization”. *Proceedings of the Conference on Natural Language Learning (CoNLL-2001), held as a Workshop at the 39th Annual meeting of the Association for Computational Linguistics (ACL-2001)*.
- [12] Judith Klavans and Smaranda Muresan (2000). “DEFINDER: Rule-based Methods for the Extraction of Medical Terminology and their Associated Definitions from On-line Text”. *Proceedings of the American Medical Informatics Association Symposium (AMIA 2000)*.
- [13] Smaranda Muresan, Tudor Muresan, and Rodica Potolea (1999). “An Automatic Logic Program Generation Kernel”. *Scientific Journal of Automation, Computers and Applied Mathematics (ACAM), Technical University of Cluj-Napoca*, 8(1).
- [14] Tudor Muresan, Rodica Potolea, and Smaranda Muresan (1998). “Amalgamating CCP with Prolog”. *Scientific Journal of Polytechnic University, Timisoara*, 43(4).

TECHNICAL REPORTS

- [15] Smaranda Muresan, Tudor Muresan, and Judith Klavans (2005). “Lexicalized Well-Founded Grammars: Learnability and Merging”. *Technical Report CUCS-027-05, Columbia University, New York, NY*.
- [16] Smaranda Muresan (2005). “Parsing Preserving Techniques in Grammar Induction”. *Technical Report CUCS-032-05, Columbia University, New York, NY*.