

Karine Megerdooian, Ph.D.
[personal information removed]
U.S. CITIZEN

PROFILE

Senior scientist with over 10 years of experience in research and development in computational linguistics with focus in the areas of **machine translation, information extraction, and lexical semantics**. Expertise includes in-depth theoretical and hands-on knowledge of linguistic and Natural Language Processing approaches for **less commonly studied languages**, with a strong specialization in Persian. Experience in managing technical and linguistic aspects of NLP projects in industry and academia. Linguistic specialization in the study of **complex predicates and language acquisition of Heritage Speakers** in Persian and Armenian.

EMPLOYMENT HISTORY

- | | | |
|-------------------------------|--|----------------|
| Oct. 2006
to present | <i>Lead Artificial Intelligence Engineer</i>
The MITRE Corporation, McLean, VA
<i>Human Language Technology Group</i>
Contributing as a computational linguist to projects in Persian and multilingual NLP: event analysis and temporal relations, multilingual information retrieval, cross-script name-matching, entity resolution, and evaluation of machine translation systems. Developed and led projects in computational analysis of Persian language weblogs, conversational Persian dialects, and Tajiki machine translation. | MITRE |
| Sept. 2005
to Oct. 2006 | <i>Assistant Research Scientist</i>
Center for Advanced Study of Language, University of Maryland
Developed test design and content for cross-linguistic investigation of language acquisition by heritage and L2 speakers. Project manager for the aspects involving linguistic and sociolinguistic analysis of Persian heritage speakers. | CASL |
| August 2004
to August 2005 | <i>Consultant</i>
Various projects: <ul style="list-style-type: none">• Inxight Software, Inc.; Sunnyvale, California; 8/04-7/05 and 7/06-10/06
Improved performance of the Inxight Persian text mining system. Developed an automatic transcription system of Persian named entities.• Multilingual Named Entity Detection and Transliteration Project; 2/05-8/05
University of Illinois, Urbana-Champaign
Developed Persian and Eastern Armenian entity extraction rules. PI: Prof. Richard Sproat• Diagnostic Testing Project (CAST); 2/05-8/05
Language Acquisition Resource Center; San Diego State University
Responsible for designing and developing online proficiency test items for Persian, based on the ACTFL proficiency classification. | |
| January 2002
to Aug. 2004 | <i>Computational Linguist, Technical Lead</i>
Inxight Software, Inc.; Sunnyvale, California (currently Business Objects)
Developed and maintained Persian, Arabic and Slovenian text mining systems as part of the Inxight LinguistX platform, using Xerox finite-state technology. The projects included stemmer, tagger, shallow parsing and information extraction modules, and used a statistical disambiguator. | Inxight |
| January 1998
to April 2001 | <i>Computational Linguist</i>
Computing Research Laboratory (CRL), New Mexico State University, Las Cruces <ul style="list-style-type: none">• Led the development and acquisition of the linguistic aspects of a large-scale Persian to English machine translation system. Responsible for building the linguistic components and for integration with the computational system (Shiraz project). | CRL |

- Acquired ontological knowledge for an English lexicon (MikroLite project).
- Developed the linguistic components for a Turkish to English machine translation system. Incorporated a pre-existing lexicon and morphological analyzer. Supervised additional lexical acquisition (Expedition project).
- Prepared the linguistic specification for the development of a structural transfer component, used in a machine translation system from low-density languages to English (Boas project).
- Supervised lexicon acquisition for five languages during an upgrade phase for a multilingual machine translation system (Corelli project).

Project managers: Prof. Sergei Nirenburg, Dr. Rémi Zajac, and Dr. Evelyne Viegas

May 1997 **Research Assistant** ISI
to Aug. 1997 Information Sciences Institute (ISI), Marina del Rey
Developed a grammar of Persian for a Machine Translation project. PI: Prof. Kevin Knight

EDUCATION

2002 **Ph.D. in Linguistics** University of Southern California, Los Angeles
Specialization: Syntax-Morphology interface; complex predicates; Armenian and Persian.
Dissertation title: *Beyond Words and Phrases: A Unified Theory of Predicate Composition*
Dissertation director: Prof. Jean-Roger Vergnaud
Committee members: Profs. Joseph Aoun, Maria Luisa Zubizarreta, Yen-Hui Audrey Li.

1995 **Master of Arts in Linguistics** University of Southern California, Los Angeles
M.A. thesis title: *Scrambling and Scopal Interactions in a Minimalist Perspective*
Advisor: Prof. Jean-Roger Vergnaud

1991–1993 **Directed readings in Neural Computation** University of Southern California, Los Angeles
Supervisor: Prof. William Wagner

1991 **Bachelor of Science in Physics** University of Southern California, Los Angeles
Coursework included Statistical Mechanics, Quantum Theory, Signal Processing, Integrated Circuits and Electrical Engineering.

INVITED TALKS AND TEACHING

September 2009 **Invited Speaker, Tools and Resources for Linguists: Special Emphasis on Af/Pak and African Languages**
Government Coordination Conference and Technical Forum, MITRE, Virginia
Unclassified Resources for Linguists

June 2009 **Invited Speaker, STARTALK 2009 Persian Teacher's Workshop**
University of Pennsylvania, National Middle East Language Resource Center
Teaching Heritage Persian Speakers

February 2009 **Invited Speaker, HLT Tools and Methods for OSINT (Open Source Intelligence)**
Army Research Laboratory, Maryland
The Language of Iranian Weblogs

May 2008 **Invited Speaker, USC Student Workshop**
University of Southern California, Los Angeles
Some bare nouns are more equal than others

January 2008 **Keynote Speaker, Teachers of Persian Workshop**
University of Maryland, College Park
Difficulties for Learning Persian as a Second Language

- October 2007 **Invited Instructor, NATO Advanced Study Institute**
 Advances in Language Engineering for Low- and Middle-Density Languages
 Batumi, Georgia
Building Computational Resources and Processors for Persian and Armenian
- June 2006 **Invited Instructor, Linguistics Department**
 University of Tehran, Iran
 Taught a course and tutorial on ‘Finite-State Morphology’ with emphasis on issues of Persian.
- June 2006 **Invited Speaker, 2nd Workshop on Persian Language and Computers**
 June 28-30; University of Tehran, Iran
Extending a Persian Morphological Analyzer to Blogs
- January 2004 to July 2005 **Lecturer, Heritage Language Program**
 Linguistics Department, University of California San Diego
 Responsible for all aspects of instruction, including curriculum development, for Persian and Armenian language analysis and linguistics for heritage speakers.
- May 2005 **Invited Speaker, Workshop on Armenian Syntax**
 May 22-25; Pithiviers, France
Reflexive Verbs and Voice Morphology in Armenian
- May 2004 **Invited Instructor, Linguistics Department**
 University of Tehran, Iran
 One-day seminar on ‘Introduction to Computational Linguistics’.
- May 2004 **Invited Speaker, 1st Workshop on Persian Language and Computers**
 May 25-26; University of Tehran, Iran
Developing a Persian Part-of-Speech Tagger
A Semantic Template for Light Verb Constructions
- April 2002 **Invited Speaker, USC-GSIL Student Workshop**
 University of Southern California, Los Angeles
Parallel Domains and the Projection of Arguments
- May 2001 **Invited Speaker, New Trends in Linguistics Workshop**
 May 27-29; Allameh Tabatabai University, Tehran, Iran
Theories of Verb Formation
- June 2001 **Invited Speaker, Workshop on Issues in Persian Linguistics**
 Islamic Azad University at Khorasgan, Isfahan, Iran
Complex Predicates in Persian
- May 1999 **Guest Instructor, Summer School in Language Engineering**
 New Mexico State University
Approaches to Computational Morphology: Applications for Persian
- Sept. 1998 to Dec. 1998 **Teacher’s Assistant, Linguistics Department**
 University of Southern California, Los Angeles
 Taught discussion sessions for the course ‘Language and Mind’ on acoustic and articulatory phonetics, phonology, speech perception and processing.
- Sept. 1994 to May 1997 **Assistant Lecturer, French and Italian Department**
 University of Southern California
 Full responsibility for all aspects of French language instruction.

LANGUAGES

Native speaker of Eastern Armenian and Persian (Farsi), and near-native speaker of English and French.
High working knowledge of Western Armenian, Italian, Spanish; basic knowledge of German, Russian and Arabic.

SKILLS

- Expertise in several major areas of Natural Language Processing, especially pertaining to multilingual morphological analysis, information extraction, and machine translation. Specializing in low-density languages with emphasis on Iranian Persian (Farsi), Tajiki Persian, and Armenian.
- Development expertise in unification-based systems, Xerox finite-state tools, morphological and syntactic analysis and generation, lexicography, ontological knowledge representation. Experience in corpus linguistics.
- Linguistic specialization in Distributed Morphology, Minimalist Program, Principles and Parameters Theory. Knowledge of HPSG.
- Familiar with Perl and scripting languages; basic knowledge of Java.
- Curriculum development for teaching languages and linguistics to heritage speakers.

PROFESSIONAL ACTIVITIES AND MEMBERSHIPS

- Editorial Board, *The Linguistics Journal*.
- Serving as reviewer for IEEE, Linguistic Inquiry, Lingua, Natural Language and Linguistic Theory, NSF Linguistics Program, Language Sciences, Word Structure, Computers and the Humanities, Journal of Linguistics, Linguistica Atlantica, Natural Sciences and Engineering Research Council of Canada, and SALT conference.
- Organizer for the **International Conference on Complex Predicates in Iranian Languages**. Paris, France. July 5-6, 2008. *URL*: www.iranianlinguistics.org/complexpredicates.
- Organizer for **Computational Approaches to Arabic Script-based Languages (CAASL)**. Ottawa, Canada; August 23, 2009; Stanford, California; July 21-22, 2007 and Geneva, Switzerland; August 28, 2004. *URL*: arabicscript.org
- Founder of the **Society for Iranian Linguistics**. *URL*: iranianlinguistics.org.
- Professional memberships: Linguistic Society of America (LSA), Association for Computational Linguistics (ACL), International Society for Iranian Studies (ISIS), Society for Armenian Studies (SAS).

PROFESSIONAL HONORS AND AWARDS

- MITRE Mission Oriented Investigation and Experimentation (MOIE) grant for *Exploring Soft Power in Weblogistan*, 2009.
- MITRE Innovation Grant for *FvER: Framework for Evaluation of Event Resolution*, 2009.
- MITRE Spot Award, 2008.
- MITRE Innovation Grant for *Rapid Generation of Translation Capabilities for Low-Resourced Languages*, 2008.
- MITRE Tech Initiative in *Social Media*, Information Discovery and Understanding dept., 2008.
- Member of the Technical Program Committee for the Conference on Language and Technology (CLT07), held in University of Peshawar, Pakistan. August 7-11, 2007.
- Member of the editorial board of *The Linguistics Journal*.
- Recipient of Ahmanson Foundation Dissertation Fellowship; University of Southern California, LA, 2000-2001.

PUBLICATIONS

Books

2009 *Beyond Words and Phrases: A Unified Theory of Predicate Composition*. VDM Verlag, Germany.

Book Contributions

- 2009 “Low-Density Language Strategies for Persian and Armenian”. In *Advances in Language Engineering for Lesser-Studied Languages*; edited by Sergei Nirenburg. NATO Science for Peace and Security Series – D: Information and Communication Security; Volume 21. IOS Press of Amsterdam.
- 2008 “Parallel Nominal and Verbal Projections”. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*; edited by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta. MIT Press.
- 2003 “Text Mining, Corpus Building and Testing”. In *A Handbook for Language Engineers*; edited by Ali Farghaly. CSLI Publications: Stanford, CA.

Book and Proceedings Editing

- To Appear *Foundational Issues in the Computational Processing of Persian*. Coeditor: Nick Pendar.
- 2009 *Proceedings of the 3rd Workshop on Computational Approaches to Arabic Script-Based Languages*. MT Summit XII. Coeditor: Ali Farghaly and Hassan Sawaf.
- 2004 *Proceedings of the 2nd Workshop on Computational Approaches to Arabic Script-Based Languages*. Coeditor: Ali Farghaly. Available at www.arabicscript.org.
- 2004 *Proceedings of the 1st Workshop on Computational Approaches to Arabic Script-Based Languages*. COLING 2004. Coeditor: Ali Farghaly.
- 2001 *Proceedings of the 20th West Coast Conference on Formal Linguistics*. Cascadilla Press. Coeditor: Leora Ann Bar-el.

Journal Publications

- To Appear “The Status of the Nominal in Persian Complex Predicates”. *Natural Language and Linguistic Theory*, Springer Netherlands.
- 2005 “Transitivity Alternations and Causative Constructions in Eastern Armenian”. In *Annual of Armenian Linguistics* 21.
- 2001 “Event Structure and Complex Predicates in Persian”. In *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 46(1/2): 97-125.

Peer-Reviewed Conference Proceedings

- 2008 “Low-Density Language Bootstrapping: The Case of Tajiki Persian”. In *Proceedings of Language Resources and Evaluation Conference (LREC 2008)*. Marrakech, Morocco. Coauthor: Dan Parvaz.
- 2004 “Finite-State Morphological Analysis of Persian”. In *Proceedings of the Workshop on Computational Approaches to Arabic Script-Based Languages, COLING 2004*. University of Geneva, Switzerland.
- 2000 “Rapid Development of Translation Tools: Application to Persian and Turkish”. In *Proceedings of COLING 2000*, Saarbrücken, Germany. Coauthors: Jan W. Amtrup and Rémi Zajac.
- 2000 “Unification-Based Persian Morphology”. In *Proceedings of CICLing 2000*. Alexander Gelbukh (ed.). Centro de Investigacion en Computacion-IPN, Mexico.
- 1999 “Rapid Development of Translation Tools”. In *Proceedings of Machine Translation Summit VII*, Singapore, pp. 385-389, September 1999. Coauthors: Jan W. Amtrup and Rémi Zajac.

Other Conference Proceedings

- 2006 “Extending a Persian Morphological Analyzer to Blogs”. In *Proceedings of the Second Workshop on Persian Language and Computers*. Tehran University, Iran. June 28-30.

- 2004 “Developing a Persian Part-of-Speech Tagger”. In *Proceedings of the First Workshop on Persian Language and Computers*. Tehran University, Iran. May 25-26.
- 2004 “A Semantic Template for Light Verb Constructions”. In *Proceedings of the First Workshop on Persian Language and Computers*. Tehran University, Iran. May 25-26.
- 2001 “Against Optional *Wh*-Movement”. In *Proceedings of WECOL 2000*, Volume 12. California State University, Fresno. Coauthor: Shadi Ganjavi.
- 2001 “Aspect and Partitive Objects in Finnish”. In *Proceedings of WCCFL 2000*. Cascadilla Press.
- 2000 “Projection of Direct Objects”. In *Proceedings of WECOL 1999*, Volume 11. California State University, Fresno.

Technical Reports & Unpublished Manuscripts

- 2008 “Analysis of Farsi Weblogs”. MITRE Technical Report (MP080206), McLean, VA.
- 2008 “The Structure of Persian Names”. MITRE Technical Report (MP080034), McLean, VA.
- 2004 “Developing a Persian Text Mining System”. Ms., Inxight Software, Sunnyvale, California.
- 2004 “Dichotomy in the Structure of the Lexicon: Transitivity Splits in Armenian”. Ms., University of California, San Diego.
- 2000 “Persian-English Machine Translation: An Overview of the Shiraz Project”. NMSU, CRL, Memoranda in Computer and Cognitive Science (MCCS-00-319). Coauthors: Jan W. Amtrup, Hamid Mansouri Rad, and Rémi Zajac.
- 2000 “Persian Computational Morphology: A Unification-Based Approach”. NMSU, CRL, Memoranda in Computer and Cognitive Science (MCCS-00-320).
- 2000 “A Computational Analysis of the Persian Noun Phrase”. NMSU, CRL, Memoranda in Computer and Cognitive Science (MCCS-00-321).
- 2000 “Processing Persian Text: Tokenization in the Shiraz Project”. NMSU, CRL, Memoranda in Computer and Cognitive Science (MCCS-00-322). Coauthor: Rémi Zajac.

RECENT CONFERENCE PRESENTATIONS (UNPUBLISHED)

- 2009 D-Linked *Wh*-Phrases and Focus Fronting in Persian (with Shadi Ganjavi). Talk presented at the *Third International Conference on Iranian Languages (ICIL3)*, Paris, France. September 2009.
- 2009 Linguistic Patterns in Heritage Persian Instruction. Talk presented at the *National Council of Less Commonly Taught Languages (NCOLTCL)* conference, Madison, Wisconsin. April 2009.
- 2008 Second Position, Focus and Multiple Spell-Out: Toward an Analysis of the Armenian Auxiliary (with Arsalan Kahnemuyipour). Talk presented at the *Meeting Clitics* workshop, Barcelona, Spain. August 2008.
- 2008 Preverbal Nominals and Telicity in Persian Complex Predicates. Talk presented at the *International Conference on Complex Predicates in Iranian Languages*, Paris, France. July 2008.
- 2008 Second Position Clitics in the *vP* Phase: The Case of the Armenian Auxiliary (with Arsalan Kahnemuyipour). Talk presented at the *LSA 2008 Meeting*, Chicago. January 2008.
- 2007 The Auxiliary Clitic and Sentential Stress in Eastern Armenian. Talk presented at the *GURT 2007: Little words – their history, phonology, syntax, semantics, pragmatics, and acquisition*, Georgetown University, Washington, DC. March 2007.
- 2007 Testing the Linguistic Competence of Persian Heritage Speakers (with Ilhan Cagri and Scott Jackson). Talk presented at the *Second Language Research Forum: Colloquium on understanding the interlanguage systems of heritage language learners*, Urbana-Champaign, Illinois. October 2007.
- 2007 On Incorporation in Persian Verbal Predicates. Talk presented at the *Second International Conference on Aspects of Iranian Linguistics*, Hamburg, Germany. June 2007.
- 2007 Machine Translation of Persian Complex Predicates (with Jan W. Amtrup). Talk presented at the *Second International Conference on Aspects of Iranian Linguistics*, Hamburg, Germany. June 2007.
- 2005 A Typology of Persian Complex Predicates. Talk presented at the *First International Conference on Aspects of Iranian Linguistics*, Leipzig, Germany. June 2005.
- 2002 Asymmetries in Form and Meaning: Surface Realization and Interface Conditions. Talk presented at *Approaching Asymmetry at the Interfaces* workshop, held in Université de Québec à Montréal, Canada. Oct. 24-25, 2003.