

Karine Megerdooian, Ph.D.

PROFILE

Scientist with over 12 years of experience in research and development in computational linguistics with focus in the areas of **social media analysis**, **machine translation**, **information extraction**, and **lexical semantics**. Expertise includes in-depth theoretical and hands-on knowledge of linguistic and computational approaches for **less commonly studied languages**, with a strong specialization in Persian. Experience in managing technical and linguistic aspects of Natural Language Processing projects in industry and academia. Linguistic specialization in the study of **complex predicates** and **Heritage Speaker Pedagogy** in Persian and Armenian.

EDUCATION

2002	PhD in Linguistics	University of Southern California, Los Angeles
1995	Master of Arts in Linguistics	University of Southern California, Los Angeles
1991–1993	Directed readings in Neural Computation	University of Southern California, Los Angeles
1991	Bachelor of Science in Physics	University of Southern California, Los Angeles

EMPLOYMENT HISTORY

Oct. 2006 to present	<i>Lead Artificial Intelligence Engineer; Group Lead</i> <i>Technical Leader in New Media</i> The MITRE Corporation, McLean, VA; <i>Human Language Technology Group</i> Developed standards and led projects in social media analysis, detecting topic and sentiment in Persian and Arabic language blogs, social computation in new media, automatic analysis of conversational language, and Tajiki machine translation. Contributed as 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.	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 SAP)	Inxight

Developed and maintained Persian, Arabic and Slovenian text mining systems as part of the the LinguistX platform, using Xerox finite-state technology. The projects included stemmer, tagger, shallow parsing and information extraction modules, and used a statistical disambiguator.

January 1998 to April 2001	Computational Linguist Computing Research Laboratory (CRL), New Mexico State University, Las Cruces	CRL
	<ul style="list-style-type: none"> • Led the development and acquisition of the linguistic aspects of the first large-scale Persian to English machine translation system. Responsible for building the linguistic components and for integration with the computational system (Shiraz project). • 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 to Aug. 1997	Research Assistant Information Sciences Institute (ISI), Marina del Rey	ISI
	Developed a grammar of Persian Machine Translation. PI: Prof. Kevin Knight	

TEACHING EXPERIENCE

10/2007	Invited Instructor, NATO Advanced Study Institute; Batumi, Georgia Advances in Language Engineering for Low- and Middle-Density Languages <i>Building Computational Resources and Processors for Persian and Armenian</i>
2006, 2004	Invited Instructor, Linguistics Department; University of Tehran, Iran A tutorial on 'Finite-State Morphology' with emphasis on issues of Persian (June 2006). One-day seminar on 'Introduction to Computational Linguistics' (May 2004).
1/2004-7/2005	Lecturer, Heritage Language Program; Linguistics Department, Univ. of California San Diego Responsible for all aspects of instruction, including curriculum development, for Persian and Armenian language analysis and linguistics for heritage speakers.
5/1999	Guest Instructor, Summer School in Language Engineering; New Mexico State University <i>Approaches to Computational Morphology: Applications for Persian</i>
9/1998-12/1998	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.
9/1994-5/1997	Assistant Lecturer, French and Italian Department; University of Southern California Full responsibility for all aspects of French language instruction.

COMPUTATIONAL AND LINGUISTIC SKILLS

- Expertise in several major areas of Natural Language Processing, especially pertaining to social media analysis, multilingual morphological analysis, information extraction, and machine translation. Specializing in low-density languages with emphasis on Persian (Farsi, Tajiki, Dari) 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, Python, and scripting languages. Basic knowledge of Java.
- Curriculum development for teaching languages and linguistics to heritage speakers.

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, Dari and Tajiki Persian.
Familiarity and experience with machine processing of: Arabic, Russian, Urdu.

INVITED TALKS

- June 2011 **Department of Computer and Information Science, Linköpings Universitet; Sweden**
Computing Linguistic Creativity in the Persian Blogosphere
- July 2010 **STARTALK 2010 Persian Teacher's Workshop**
University of Pennsylvania, National Middle East Language Resource Center
Computational Tools for Learners of Persian
- September 2009 **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 **STARTALK 2009 Persian Teacher's Workshop**
University of Pennsylvania, National Middle East Language Resource Center
Teaching Heritage Persian Speakers
- February 2009 **HLT Tools and Methods for OSINT (Open Source Intelligence); Army Research Lab, MD**
The Language of Iranian Weblogs
- May 2008 **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
- June 2006 **Keynote Speaker, 2nd Workshop on Persian Language and Computers; Univ. of Tehran, Iran**
Extending a Persian Morphological Analyzer to Blogs
- May 2005 **Workshop on Armenian Syntax; Pithiviers, France**
Reflexive Verbs and Voice Morphology in Armenian
- May 2004 **Keynote Speaker, 1st Workshop on Persian Language and Computers; Univ. of Tehran, Iran**
- *Developing a Persian Part-of-Speech Tagger*
- *A Semantic Template for Light Verb Constructions*

PROFESSIONAL ACTIVITIES AND MEMBERSHIPS

- Serving as reviewer for IEEE, Linguistic Inquiry, Lingua, Natural Language and Linguistic Theory, National Science Foundation 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.
- Founder of the **Society for Iranian Linguistics**. URL: iranianlinguistics.org.
- Conference organizations:
 - **International Conference on Complex Predicates in Iranian Languages**. Paris, France. July 2008.
 - **Computational Approaches to Arabic Script-based Languages (CAASL)**. Ottawa, Canada (August 2009); Stanford, California (July 2007) and Geneva, Switzerland (August 2004).
- Professional memberships: Linguistic Society of America (LSA), Association for Computational Linguistics (ACL), International Society for Iranian Studies (ISIS), Association of American Teachers of Persian (AATP), Society for Armenian Studies (SAS).

PROFESSIONAL HONORS AND AWARDS

- STEM Women of Color in Technology nominee in New Media/IT Leadership, MITRE, 2011.
- Member of the Technical Program Committee for the Conference on Language and Technology (CLT07), University of Peshawar, Pakistan (August 7-11, 2007).

- Invited to serve as NSF panel judge for interdisciplinary scientific competition, 2010.
- MITRE research grants awarded: *Predicting Revolutionary Triggers in Social Media* (2012), *Mining Socio-Cultural Fault Lines* (2012), *Exploring Soft Power in Weblogistan* (2010-2011).
- MITRE Innovation Grants awarded: *FEvER: Framework for Evaluation of Event Resolution* (2009), *Rapid Generation of Translation Capabilities for Low-Resourced Languages* (2008).
- MITRE Tech Initiative awarded in *Social Media*, Information Discovery and Understanding dept., 2008.
- MITRE Spot Awards, 2008 and 2009.
- Member of the editorial board of *The Linguistics Journal*, 2007-2009.
- Recipient of Ahmanson Foundation Dissertation Fellowship; Univ. of Southern California, LA, 2000-2001.

PUBLICATIONS

Books

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

Book Contributions

- 2010 Extending a Persian Morphological Analyzer to Blogs. In *Zabane Farsi va Rayane [Persian Language and Computers]*. SAMT Publishers, Organization for Research and Editing of University Publications in the Humanities, Tehran, Iran.
- 2010 Developing a Persian Part-of-Speech Tagger. In *Zabane Farsi va Rayane [Persian Language and Computers]*. SAMT Publishers, Organization for Research and Editing of University Publications in the Humanities, Tehran, Iran.
- 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

- 2009 *Proceedings of the 3rd Workshop on Computational Approaches to Arabic Script-Based Languages*. MT Summit XII. Coeditor: Ali Farghaly and Hassan Sawaf.
- 2007 *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.

Refereed Journal Publications

- InReview Shift Happens: Language Change and Maintenance in the Persian Diaspora. *Journal of Iranian Studies*.
- 2011 The Status of the Nominal in Persian Complex Predicates. *Natural Language and Linguistic Theory*, Online First, 16 September 2011.
- 2011 Second Position Clitics in the vP Phase: the Case of the Armenian Auxiliary. *Linguistic Inquiry* 42(1): 152-162. Coauthor: Arsalan Kahnemuyipour.
- 2009 Telicity in Persian Complex Predicates. *Snippets*, Issue 19, July 2009.
- 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

- 2010 Mining and Classification of Neologisms in Persian Blogs. In *Proceedings of CALC-10, NAACL-HLT workshop on Computational Approaches to Linguistic Creativity*. Coauthor: Ali Hadjarian.

- 2009 Automated Metrics for Speech Translation. In *PerMIS '09 Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*. New York. Coauthors: Sherri Condon, Mark Arehart, Christy Doran, Dan Parvaz, John Aberdeen, Beatrice Oshika, Greg Sanders.
- 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 Investigación en Computación-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

- 2010 Linguistic Patterns in Heritage Persian Instruction. In *STARTALK Persian Teacher's Workshop for Professional Curriculum and Materials Development*. University of Pennsylvania, Philadelphia. July 2010.
- 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*. Cascadia Press.
- 2000 Projection of Direct Objects. In *Proceedings of WECOL 1999*, Volume 11. California State Univ., Fresno.

Technical Reports

- 2011 The Format of Middle Eastern Names in Official Documents. MITRE Technical Report (MTR110136), McLean, VA.
- 2011 Issues in Dari Transcription. MITRE Product (MP110054), McLean, VA.
- 2009 The Structure of Afghan Names. MITRE Product (MP090315), McLean, VA.
- 2008 Analysis of Farsi Weblogs. MITRE Product (MP080206), McLean, VA.
- 2008 The Structure of Persian Names. MITRE Product (MP080034), McLean, VA.
- 2008 Argus Search and Stemming. MITRE Product (MP080256). McLean, VA.
- 2008 Persian Supplement to the TIDES Standard for the Annotation of temporal Expressions. MITRE Technical Report. McLean, VA.
- 2007 The Structure of Arabic Names. MITRE Technical Report. McLean, VA.
- 2004 Developing a Persian Text Mining System. Ms., Inlight Software, Sunnyvale, California.
- 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)

- 2011 Language and gender correlations in Persian blogs. Talk presented at the *Georgetown University Round Table on Languages and Linguistics (GURT) – Discourse 2.0: Language and New Media*. Washington, DC. March 2011.

- 2011 Focus and the auxiliary in Eastern Armenian. Talk presented at the 37th *Annual Meeting of the Berkeley Linguistics Society (BLS)*, Special session on Languages of the Caucasus. Berkeley, CA. February 2011.
- 2010 On the domain of second position clitics (with Arsalan Kahnemuyipour). Talk presented at the 28th *West Coast Conference for Formal Linguistics (WCCFL)*, Special session on Prosody in Grammar, Acquisition, and Processing. Los Angeles, CA. February 2010.
- 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 MITRE Infrastructure and Evaluation for Identity Resolution Technology (with Keith Miller, Elizabeth Schroeder, Sarah McLeod, Azar Ulrich, James Finley, Gail Hamilton, Andre Milota, Ken Samuel, Sherri Condon and Mark Arehart), presented at the 77th *MORS Symposium*. Fort Leavenworth, Kansas. June 2009.
- 2009 Method for Comparing Identity Resolution Technology (with Keith Miller, Elizabeth Schroeder, Sarah McLeod, Azar Ulrich, James Finley, Gail Hamilton, Andre Milota, Ken Samuel, Sherri Condon and Mark Arehart), Talk presented at the 77th *MORS Symposium*. Fort Leavenworth, Kansas. June 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.
- 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.