Karine Megerdoomian, Ph.D.
www.zoorna.org

PROFILE

Computational Linguist with specialization in social media analytics, and associated linguistic and socio-political issues, such as sentiment analysis, detection of ideology and radicalization, and frame and narrative analysis. Expertise includes in-depth theoretical and hands-on knowledge of computational approaches for less commonly studied languages, with a strong specialization in Persian (Farsi, Dari, Tajik), and linguistically-informed Natural Language Processing approaches. Years of experience in developing and managing research programs. Linguistic specialization in the study of complex predicates and Heritage Speaker pedagogy in Persian and Armenian.

EDUCATION

2016 expected	**Graduate Certificate in Computational Social Science**	George Mason University, Fairfax, VA
2002	**PhD in Linguistics**	University of Southern California, Los Angeles
1995	**Master of Arts in Linguistics**	University of Southern California, Los Angeles
1991	**Bachelor of Science in Physics**	University of Southern California, Los Angeles

EMPLOYMENT HISTORY

2006 -present　　*Principal Computational Linguist*　　　　　　　　　　　　　　　**MITRE**
The MITRE Corporation, McLean, VA
- Project lead for research project on automatic cause identification in aviation incident narratives.
- Contributed to and influenced the company research portfolio by developing and managing projects in social media analysis in languages of the Middle East. Research focus on the relationship between language in online media and associated socio-political issues -- with emphasis on online community detection, language characteristics of social groups with emphasis on radicalization, automatic framing and narrative analysis.
- Developed standards and led tasks in support of sponsor needs for evaluation of language technology tools, transcription guidelines, automatic analysis of conversational and dialectal 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.

2005-2006　　*Assistant Research Scientist*　　　　　　　　　　　　　　　　　**CASL**
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.

2004-2005　　*Consultant*
- 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.

2002-2004 ***Computational Linguist, Technical Lead***   **Inxight**
      Inxight Software, Inc.; Sunnyvale, California (currently SAP)
      Developed and maintained Persian, Arabic and Slovenian text mining systems as part of the LinguistX platform, using Xerox finite-state technology. The projects included morphological analyzer, tagger, shallow parsing and information extraction modules.

1998-2001 ***Computational Linguist***   **CRL**
      Computing Research Laboratory (CRL), New Mexico State University, Las Cruces
- 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).
- Contributed to various projects in ontology development and machine translation (developed and integrated linguistic components for a Turkish to English machine translation system, and supervised lexicon acquisition for a multilingual machine translation system).

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

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

TEACHING EXPERIENCE

Adjunct Professor, **Georgetown University;** Washington, DC (August 2012-Present)
    Communication, Culture and Technology Department
    Social Media Analytics: Culture and Ideology in the Middle East
    Narrative Networks: Analysis of Framing and Narratives in Online Media
    Sentiment Analysis and Opinion Mining in Social Media
Instructor, **NATO Advanced Study Institute;** Batumi, Georgia (October 2007)
    Advances in Language Engineering for Low- and Middle-Density Languages
    Building Computational Resources and Processors for Persian and Armenian
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).
Lecturer, **Heritage Language Program;** Linguistics Department, Univ of Calif. San Diego (Jan 2004-Jul 2005)
    Responsible for all aspects of instruction, including curriculum development, for Persian and Armenian language analysis and linguistics for heritage speakers.

COMPUTATIONAL AND LINGUISTIC SKILLS

- Expertise in several major areas of Human Language Technology, especially pertaining to social media analysis, sentiment 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.
- Knowledge of Python and scripting languages. Experience with NLTK, social network analysis (Gephi, NetworkX), NetLogo, WEKA.
- 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.

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.
- 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), Computational Social Science Society of the Americas (CSSSA), Association of American Teachers of Persian (AATP).

PROFESSIONAL HONORS AND AWARDS

- Member of Program Committee:
    - Language Resources and Evaluation Conference (LREC 2014, 2016)
    - Arabic Natural Language Processing Workshop (ANLP 2014)
    - Arabic/Amazigh/Farsi/Urdu Language Processing track of the ACS/IEEE International Conference on Computer systems and Applications (AICCSA 2013)
    - Coling 2012, Area: Named Entity Recognition
    - Conference on Language and Technology (CLT07)
- STEM Women of Color in Technology nominee in New Media/IT Leadership, MITRE, 2011.
- Invited to serve as NSF panel judge for interdisciplinary scientific competition, 2010.
- MITRE research grants awarded: *Cause Identification from Aviation Safety Reports* (2014-2015), *Predicting Revolutionary Triggers in Social Media* (2012), *Mining Socio-Cultural Fault Lines* (2012-2013), *Exploring Soft Power in Weblogistan* (2010-2011).
- MITRE Innovation Grants awarded: *Domain-Focused Evaluation of Linguisitc Inquiry and Word Count (LIWC)* (2014), *FEvER: Framework for Evaluation of Event Resolution* (2009), *Rapid Generation of Translation Capabilities for Low-Resourced Languages* (2008), Tech initiative in *social media* (2008).
- MITRE Spot Awards, 2008, 2009, 2013.
- Member of the editorial board of *The Linguistics Journal*, 2007-2009.
- Recipient of Ahmanson Foundation Dissertation Fellowship; Univ. of Southern California, LA, 2000-2001.

INVITED TALKS

April 2012	**Academedia conference; Georgetown University**
	Panelist on Social Media Analysis
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*

PUBLICATIONS

Books

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

Book Contributions

ToAppear Persian Computer-Mediated Communication: The Language of Twitter and Blogs. In *Essays in Persian and Iranian Linguistics*. Mouton de Gruyter.

ToAppear Persian Computational Linguistics. In *Handbook of Persian Linguistics*. Oxford University Press.

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

In Press On the Positional Distribution of an Armenian Auxiliary: Second Position Clisis, Focus and Phases. *Syntax*.

2015 Human Language Reveals a Universal Positivity Bias. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 8: 2389-2394. Co-authors: P. S. Dodds, E. M. Clark, S. Desu, M.

R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, M. T. McMahon, B. F. Tivnan, and C. M. Danforth.

2012 The Status of the Nominal in Persian Complex Predicates. *Natural Language and Linguistic Theory* 30 (1): 179-216.

2011 Second Position Clitics in the *v*P 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

2015 Modeling Community Resilience for a Post-Epidemic Society. In *Proceedings of CSSSA 2015.* Coauthor: Shaun Michel.

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 20*00, Saarbrucken, 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.

Technical Reports

2014 State of the Research in Human Language Technology: A Study of ACL and NAACL Publications from 2007 through 2014. MITRE Technical Report (MTR140208), McLean, VA.

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., Inxight 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.