A L I C I A    S A G A E, Ph.D.

alicia@sagae.org

https://asagae.github.io

Professional Summary

*Dr. Alicia Sagae is a researcher in natural language processing with experience in semantic analysis, machine translation, dialogue systems and computer-assisted language learning. She holds a Ph.D. in Language and Information Technologies from Carnegie Mellon University. Alicia has experience with both symbolic and statistical methods, and has programmed and run experiments in every phase of development for multi-lingual NLP systems. She has served as manager and Principal Investigator for multi-year projects. She has also published under the name Alicia Tribble.*

Education

### Ph.D. in Language Technologies                                          2010
Language Technologies Institute at Carnegie Mellon University
Pittsburgh, PA
Thesis Title: *Textual Inference for Retrieving Labeled Object Descriptions*
Thesis Committee: **Scott E. Fahlman**, chair (CMU); Eric Nyberg (CMU); Carolyn Rosè (CMU);
Bruce W. Porter (UT Austin); Vibhu O. Mittal (Edmodo)

### M.Sci. in Language Technologies                                         2002
Language Technologies Institute at Carnegie Mellon University
Pittsburgh, PA
Advisors: **Alex Waibel**; Stephan Vogel; Alon Lavie; Lori Levin

### Bachelor of Science in Mathematical Sciences, Computer Science Concentration    2000
### Bachelor of Arts in Slavic Linguistics (Dual Degree)
University of North Carolina, Chapel Hill
Chapel Hill, NC

Employment

### Consulting Scientist in Natural Language Processing                    2014-present
Davis, CA
Design NLP solutions for large and small companies in the technology sector. Implement components including text processing pipelines, classifiers, and knowledge bases using a variety of programming languages suited to the task and computing environment. Evaluate current NLP tools; provide enhancements; validate improvements. Communicate results to a variety of audiences as oral presentations, technical reports, and peer-reviewed publications.

### Senior Research Scientist                                              2014
E Touch Systems Corp. Fremont, CA
Consultant on assignment to Google. Designed, implemented, and evaluated a system to classify tickets automatically and reduce manual effort to resolve them. Created a Python experimentation framework to train, test, and evaluate classifiers on this task, leveraging existing Machine Learning (ML) tools in the Google code base and creating new ones. Evaluated, reported, and presented results, with successful outcomes supporting desired use cases.

### Senior Research Scientist                                              2010-2013
Alelo, Inc. Los Angeles, CA
Principal Investigator (PI) for funded projects in the area of dialogue systems for language and culture training software; supervised interns and collaborated with interdisciplinary team (Anthropology, Linguistics, Software Engineering); Implemented learner models using Python, Perl, Java, and Scala; Constructed company research agenda with Chief Scientist.

**Research Scientist**　　　　　　　　　　　　　　　　　　　　　　　　**2008-2010**

Alelo, Inc. Los Angeles, CA

Programmed and maintained machine-generated bilingual dictionaries to support language instruction software. Designed and implemented improvements to a dialogue system where language learners practice conversational skills. Implemented a learner performance tracking system including pruning of lesson content based on learner mastery. Managed corpus construction for new automatic speech recognizer for Malaysian language.

**Visiting Researcher**　　　　　　　　　　　　　　　　　　　　　　　　**2006-2008**

Tsujii Laboratory, University of Tokyo, Japan

Assisted in evaluation and update of the GENIA ontology for biotext mining.

**Doctoral Research Assistant**　　　　　　　　　　　　　　　　　　　　　**2003-2006**

Language Technologies Institute at Carnegie Mellon University

Delivered software components for funded research on semantic interpretation, including: English language semantic interpreter; noun compound resolver; semantic feature spotter for email classification; ontology editing tools (http://www.cs.cmu.edu/~atribble/SconeEdit/).

**Visiting Researcher**　　　　　　　　　　　　　　　　　　　　　　　　**2002-2003**

Karlsruhe University, Karlsruhe, Germany

Responsible for maintenance and updates to a Statistical Machine Translation system; trained system in new languages and domains; implemented phrase-translation improvements and naive morphological analysis.

**Master's Research Assistant**　　　　　　　　　　　　　　　　　　　　　**2000-2002**

Language Technologies Institute at Carnegie Mellon University

Responsible for maintenance and updates to a Statistical Machine Translation system.

SPECIALIZED SKILLS

**Machine Translation:** Training, maintenance, algorithmic updates to language modeling and alignment modeling tools for Statistical MT. Language modeling with the CMU–Cambridge SLM Toolkit and others. Grammar writing with unification–based tools for interlingual MT systems.

**Speech Recognition & Synthesis**: Phoneme recognition and language modeling using the JANUS toolkit for LVCSR. Installation and custom programming with the Festival software for speech synthesis. Building new diphone and unit-selection voices for Festival.

**Programming Languages & Operating Systems:** Perl, Python, Lisp, Java, HTML, Scala, C++, CLIPS; Linux, UNIX, Windows

**Natural Language:**

Intermediate-Advanced speaker and reader of Brazilian Portuguese (Portuguese spoken at home)

Intermediate speaker and reader of Spanish

Beginning/Intermediate speaker of Russian (resided in Russia, 5 months)

Beginning/intermediate speaker and reader of French and Polish

Beginning/intermediate speaker and reader of German (resided in Germany, 12 months)

Limited training in Japanese (Hiragana/Katakana, resided in Tokyo, 24 months)

TEACHING & SUPERVISION

### Intern Project Supervisor                                          2008-2013
Alelo, Inc. Los Angeles, CA
Supervised four graduate interns and one undergraduate in work contributing to funded research projects at the company. Four out of five interns produced publications for peer-reviewed venues based on this work.

### Teaching Assistant                                                    Fall 2005
Carnegie Mellon University, Pittsburgh, PA
Supported the professors of *Grammars and Lexicons*, an introductory course in Computational Linguistics. Prepared and delivered four recitation sessions and one 90-minute class lecture.

### Lab Assistant                                                        Spring 2000
UNC Computer Science Dept. Chapel Hill, NC
Supported professor of *Comp 14*, Introduction to Algorithms and Programming. Created assignments and solutions in Java. Coordinated efforts of 5 other Lab assistants and 3 Teaching Assistants. Taught biweekly help session.

FUNDING

### MICS                                                                      2012
Metaphor Interpretation for Cultural Schemas
Role: co-PI; Budget: 460K; Duration: 1 base + 4 option years; Sponsor: IARPA

### C-Core                                                                    2010
Cultural Content, Ontology, and Resource Engineering
Role: PI; Budget: 100K; Duration: 6 months; Sponsor: DARPA

### CultureCom                                                                2009
Commonsense Socio-cultural Models for Culture Training in Serious Games
Role: PI; Budget: 809K; Duration: 3 years; Sponsor: Office of Naval Research

PUBLICATIONS

Alicia Sagae and Scott E. Fahlman. Image Retrieval with Textual Label Similarity Features. in *INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT*, Volume 22, Issue 1, January/March 2015, Pages: 101-113.

Alicia Sagae and Scott E. Fahlman. "Knowledge Resource Development for Identifying Matching Image Descriptions." in *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development (KEOD 2013)*. September 2013. *acceptance rate 13%*

Alicia Sagae, Jerry R. Hobbs, Suzanne Wertheim, Michael Agar, Emily Ho, & W. Lewis Johnson. "Efficient Cultural Models of Verbal Behavior for Communicative Agents" in *Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA 2012)*. September 2012.

Alicia Sagae, W. Lewis Johnson, & Andre Valente. "Conversational Agents in Language and Culture Training" in D. Perez-Marin & I. Pascual-Nieto (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 358-377). Madrid: IGI Global. 2011.

Jerry R. Hobbs & Alicia Sagae. "Toward a Commonsense Theory of Microsociology: Interpersonal Relationships" in *Proceedings of the AAAI Spring Symposium*. March, 2011.

Alicia Sagae, W. Lewis Johnson, & Stephen Bodnar. "Validation of a Dialog System for Language Learners." in *Proceedings of the The 11th annual SIGdial Meeting on Discourse and Dialogue*. Tokyo, Japan. September, 2010.

W. Lewis Johnson and Alicia Sagae. "Adaptive Agent-Based Learning of Intercultural Communication Skills" in *Proceedings of the User Modeling and Personalization Conference, Demonstration Sessions (UMAP 2010)*. June 2010.

W. Lewis Johnson, Naveen Ashish, Stephen Bodnar, Alicia Sagae. "Expecting the Unexpected: Warehousing and Analyzing Data from ITS Field Use". in *Proceedings of ITS 2010*, Part II, LNCS 6095. June 2010.

Alicia Sagae. *Textual Inference for Retrieving Labeled Object Descriptions.* Ph.D. Thesis. Carnegie Mellon University, Pittsburgh, PA. May 2010.

Rohit Kumar, Alicia Sagae, and W. Lewis Johnson. "Evaluating an Authoring Tool for Mini-Dialogs" in *Proceedings of AIED 2009 (Poster)*, Brighton, UK, July 2009.

Alicia Sagae, Rohit Kumar, W. Lewis Johnson. "Scaling up in Speaking Proficiency by Supporting Robust Learning Behaviors" in *Proceedings of AIED 2009 Workshop: Scalability Issues in AIED*. Brighton, UK, July 2009.

Alicia Sagae, Baylor Wetzel, Andre Valente, W. Lewis Johnson. "Culture-Driven Response Strategies for Virtual Human Behavior in Training Systems". in *Proceedings of SLaTE-2009.* Warwickshire, England, 2009.

Alicia Tribble, Jin-Dong Kim, Tomoko Ohta, Jun'ichi Tsujii. "A Comparison of Knowledge Resource Designs: Supporting Term-Level Text Annotation" in *Proceedings of LREC-2008 Workshop: Building and evaluating resources for biotext mining.* pp 3–10, 2008.

Alicia Tribble and Scott E. Fahlman. "CMU-AT: Semantic Distance and Background Knowledge for Identifying Semantic Relations" in *Proceedings of SemEval-2007 Workshop of ACL 2007.* Prague, Czech Republic, 2007.

Alicia Tribble and Scott E. Fahlman. "Resolving Noun Compounds with Multi-Use Domain Knowledge" in *Proceedings of FLAIRS-2006.* Melbourne, Florida, May 2006.

Alicia Tribble, Benjamin Lambert, and Scott E. Fahlman. "SconeEdit: A Text-Guided Domain Knowledge Editor." in *Demonstration Sessions of HLT/NAACL-2006.* New York, June 2006.

Alicia Tribble and Scott E. Fahlman. "Resolving Noun Compounds with Multi-Use Domain Knowledge" in *Proceedings of FLAIRS-2006.* Melbourne, Florida, May 2006.

Alicia Tribble and Carolyn Rosé. "Usable Browsers for Ontological Knowledge Acquisition." in *Poster Presentations of CHI-2006.* Montréal, Quebec, April 2006.

Alicia Tribble, Stephan Vogel, Alex Waibel. "Overlapping Phrase-level Translation Rules in an SMT Engine." in *Proceedings of NLPKE-2003 Natural Language Processing and Knowledge Engineering.* Beijing, China, October 2003.

Stephan Vogel, Alicia Tribble. "Improving Statistical Machine Translation for a Speech-to-Speech Translation Task" in *Proceedings of ICSLP-2002 Workshop on Speech-to-Speech Translation.* Boulder, CO, September 2002.

Stephan Vogel, Alicia Tribble. "Direct Translation Approaches: Statistical Machine Translation" presentation for *ESSLLI-2002 European Summer School in Logic Language and Information, Workshop on Speech-to-Speech Translation.* Trento, Italy, August 2002.

Alicia Tribble, Alon Lavie, Lori Levin. "Rapid Adaptive Development of Semantic Analysis Grammars" in *Proceedings of TMI-2002 9th International Conference of Theoretical and Methodological Issues in Machine Translation.* Keihanna, Japan, March 2002.

Tanja Schultz, Qin Jin, Kornel Laskowski, Alicia Tribble, Alex Waibel. "Speaker, Accent, and Language Identification Using Multilingual Phone Strings" in *Proceedings of HLT-2002 Human Language Technology Conference.* San Diego, California, March 2002.

A. Lavie, L. Levin, T. Schultz, C. Langley, B Han, A. Tribble, D. Gates, D. Wallace, K. Peterson. "Domain Portability in Speech–to–Speech Translation" in *Proceedings of HLT-2001 Human Language Technology Conference.* San Diego, California, March 2001.