KEVIN KNIGHT

Address

Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292
(310) 448-8716
knight@isi.edu
www.isi.edu/~knight

Personal

- Birthdate: January 14, 1965

- Birthplace: Pasadena, California

Current

- Project Leader, Senior Research Scientist, ISI Fellow, USC/Information Sciences Institute.

- Research Associate Professor, Computer Science Department, University of Southern California.

- Research interests include machine translation, natural language processing, artificial intelligence, semantics, statistical modeling, heuristic search, decipherment.

Education

- **Carnegie Mellon University** (1986–1992). Ph.D. dissertation: "Integrating Knowledge Acquisition and Language Acquisition." Thesis committee: Masaru Tomita (chairman), Jaime Carbonell, Sergei Nirenburg, Elaine Rich.

- **Harvard University** (1982–1986). B.A. in computer science.

Professional

- Co-author of textbook *Artificial Intelligence*, McGraw-Hill, 1991.

- AAAI Council member, 2005-2007.

- **Editorial board.** *Computational Linguistics*, 1999-2001.

- **Editorial board.** *Journal of Artificial Intelligence Research*, 1999-2001.

- **Founding Editorial board.** *ACM Transactions on Speech and Language Processing*, 2004-.

- **General chair.** Conference of the Association for Computational Linguistics (ACL), 2005.

- **Program chair.** Conference of the North American Association for Computational Linguistics (NAACL), 2001.

- **Program committee.** Conference of the Association for Computational Linguistics (ACL), Conference of the North American Association for Computational Linguistics (NAACL), Conference of the Association for Machine Translation in the Americas (AMTA), International Joint Conference on Artificial Intelligence (IJCAI), Theoretical and Methodological Issues in Machine Translation (TMI), National Conference on Artificial Intelligence (AAAI).

- **Book review editor.** *AI Magazine*, 1994-1997.

- **Invited to lead** six-week intensive workshop on statistical machine translation (Language Engineering for Students and Professionals Integrating Research and Education, Johns Hopkins University, funded by NSF and DoD), summer 1999.

- **Invited talk.** "Statistical Machine Translation: Where it Went," Conference on Empirical Methods in Natural Language Processing (EMNLP), Madrid, Spain, 1997.

- **Invited talk.** "Every Time I Fire a Statistician, I Get a Warm Fuzzy Feeling," Conference of the Association for Machine Translation in the Americas (AMTA), Cuernavaca, Mexico, 2000.

- **Invited talk.** "Deeper Representations for Machine Translation: Ready or Not?," Conference on Empirical Methods in Natural Language Processing (EMNLP), Sapporo, Japan, 2003.

- **Invited talk.** "What's New in Statistical Machine Translation," Conference on Uncertainty in AI (UAI), Acapulco, Mexico, 2003.

- **Invited talk.** "What's New in Statistical Machine Translation," International Joint Conference on AI (IJCAI), Edinburgh, Scotland, 2005.

- **Invited talk.** "Tree Transducers for Natural Language Processing," European Workshop on Natural Language Generation, Aberdeen, Scotland, 2005.

- **Invited talk.** "Results on Decipherment Problems," ESSLLI, Heriot-Watt, Scotland, 2005.

- **Invited talk.** "I Shouldn't Have Slept Through That One Class When I Was a Grad Student at CMU," Carnegie Mellon Computer Science Anniversary, Pittsburgh, 2006.

- **Invited talk.** "Isn't Linguistic Structure Important, Asked the Engineer?" Conference of the European Association for Computational Linguistics, Trento, 2006.

- **Invited talk.** "Can a Machine Translate Without Knowing What a Verb Is?" Conference of the North American Association for Computational Linguistics, Workshop on Statistical Machine Translation, New York City, 2006.

- **Invited talk.** "No More Strings, Please," IEEE/ACL Spoken Language Technology Workshop, Aruba, 2006.

Teaching

- **University of Southern California**. *Natural Language Processing*, co-taught Spring 1993, Fall 1994, Spring 1996, Spring 1998. *Empirical Methods in Natural Language Processing*, Fall 2000, Spring 2002, Spring 2003, Fall 2004.

- **University of California, Los Angeles**. *Machine Translation*, short course, UCLA Extension (Spring 1996, Spring 1997).

- **Conference of the Association for Computational Linguistics (ACL), 1996**. *Machine Translation*, tutorial. **North American ACL Conference (NAACL), 2000**. *Machine Translation*, tutorial. **Conference of the Association for Computational Linguistics (ACL), 2000**. *Statistical Machine Translation*, tutorial. **Conference of the Association for Machine Translation in the Americas (AMTA), 2000**. *Statistical Machine Translation*, tutorial. **MT Summit IX, 2003**. *Statistical Machine Translation*, tutorial. **Conference of the Association for Machine Translation in the Americas (AMTA), 2004**. *Statistical Machine Translation*, tutorial. **North American ACL Conference (NAACL), 2004**. *Machine Translation*, tutorial. **University of Aberdeen, 2005**, *Statistical Machine Translation*, tutorial. **Conference of the Association for Computational Linguistics (ACL), 2006**. *Synchronous Grammars and Tree Transducers*, tutorial.

- **Carnegie Mellon University**. Teaching Assistant.

- **Harvard University**. Teaching Assistant. Certificate of Distinction in Teaching (Committee on Undergraduate Education), recognizing a special contribution to the teaching of undergraduates.

- **Beijing, People's Republic of China**. English instructor at the Mining College in Beijing (Summer 1981).

Grants and Contracts Awarded

- "HARE: Improving Fundamental Accuracy of Speech-to-Speech Translation." 2006-2007. DARPA-funded at $370,000 (with Daniel Marcu; subcontract to SRI).

- "AGILE: Human Informed Statistical Machine Translation." 2005-. DARPA-funded at $2,000,000/year base (with Daniel Marcu; subcontract to BBN).

- "GRAMMAR: Advanced Language Modeling for Machine Translation." 2005-2006. Funded at $285,000 (with Daniel Marcu).

- "Rabbit: Iraqi-English Speech-to-Speech Deployment." 2005-2006. DARPA-funded at $700,000 (with Daniel Marcu; subcontract to SRI).

- "Treeworld: Probabilistic Tree Transducers for Machine Translation and Natural Language Processing." 2004-2009. NSF-funded at $1,250,000 (with Daniel Marcu and Daniel Gildea).

- "Translation Technology for Multilingual Translation Modeling." 2003-2007. NSF-funded at $1,260,000 (with Daniel Marcu; subcontract to University of Washington).

- "Machine Translation for a Handheld Speech Translator." 2002-2004. DARPA-funded at $670,000 (with Daniel Marcu).

- "Transmodel: US-Egypt Cooperative Project on Machine Translation." 2003-2004. NSF-funded at $25,000 (joint project with Prof. Ahmed Rafea).

- "Advanced Generation for Question Answering." 2002-2003. ARDA-funded at $450,000 (with Daniel Marcu).

- "ReWRITE: A Probabilistic Approach to Rewriting for Machine Translation, Language Generation, and Abstracting." 2000-2002. DARPA-funded at $4,200,000 (with Daniel Marcu).

- "Robust, Scalable Language Generation Using Symbolic and Statistical Techniques." 1999-2001. NSF-funded at $432,000.

- "Designing Robust Technology for Automatic Language Generation." 1998-1999. DARPA-funded at $100,000 (with Eduard Hovy).

- "Building a Standard Platform for Statistical Machine Translation." 1999. NSF-funded at $50,000.

- "Arabic-to-English Machine Translation." 1996-1999. NSA-funded at $2,000,000 (with Eduard Hovy).

- "FSTRANS: Performance- and User-Oriented Enhancements to Machine Translation." 1996. NSA-funded at $300,000 (with Eduard Hovy).

Ph.D. Students

- Dr. Ishwar Chander. Ph.D. thesis (1998), "Automated Postediting of Documents."

- Dr. Irene Langkilde. Ph.D. thesis (2002), "Automatic Sentence Generation Using a Hybrid Statistical Model of Lexical Collocations and Syntactic Relations." Now at Brigham Young University.

- Dr. Yaser Al-Onaizan. Ph.D. thesis (2002), 'Translating Named Entities Using Bilingual and Monolingual Resources." Now at IBM T. J. Watson Research Center.

- Dr. Kenji Yamada. Ph.D. thesis (2002), "A Syntax-Based Statistical Translation Model." Now at Language Weaver.

- Dr. Philipp Koehn. Ph.D. thesis (2003), "Noun Phrase Translation." Now at University of Edinburgh.

- Served on PhD dissertation committee of: Feng Pan (USC), Chris Callison-Burch (Edinburgh), Dragos Stefan Munteanu (USC), Yuan Ding (Penn), Deepak Ravichandran (USC), Noemie Elhadad (Columbia), David Evans (Columbia), Hal Daumé III (USC), Soomin Kim (USC), Radu Soricut (USC), Chad Langley (CMU), Liang Zhou (USC), Fei Huang (CMU), Shou-de Lin (USC), Greg Kondrak (Toronto), Sheila Tejada (USC), Michael Harm (USC).

- Currently serving on PhD dissertation committee of: Colin Cherry (Alberta), Yuk Wah Wong (UTexas), Dagen Wang (USC), Alexander Fraser (USC), Qing Luo (USC), Hao Zhang (Rochester), Ashish Venugopal (CMU).

Awards

- Best Paper Award, "Fast Decoding and Optimal Decoding for Machine Translation" (U. Germann, M. Jahr, K. Knight, D. Marcu, and K. Yamada), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2001.

- Outstanding Paper Award, "Statistics-Based Summarization—Step One: Sentence Compression," (K. Knight and D. Marcu), National Conference on Artificial Intelligence (AAAI), 2000.

- Meritorious Service Award, for outstanding work in building an automated research-software distribution web site, securing approval from USC lawyers.

Industry

- Applied artificial intelligence research at the **Microelectronics and Computer Technology Corporation** (MCC), Austin, Texas (1986–1991).

- Cartoonist for *IEEE Intelligent Systems* (1996–1999).

- Technical illustrations and typesetting for the textbook *Artificial Intelligence, Second Edition* (E. Rich and K. Knight, see Publications). Programs for generating technical illustrations used in *Functional Syntax* (S. Kuno, 1987).

- Co-founder, Language Weaver, Inc. (2002).

Languages

- English. Native language.

- Spanish. Working knowledge.

- Japanese. Intensive equivalent of 2 years study at Harvard, both written and spoken.

- Chinese. Two years study of written and spoken Chinese, prior to six weeks living in Beijing (1981).

Publications

- "Unsupervised Analysis for Decipherment Problems" (K. Knight, A. Nair, N. Rathod, and K. Yamada), Proc. ACL-COLING (poster), 2006.

- "Scalable Inference and Training of Context-Rich Syntactic Models" (M. Galley, J. Graehl, K. Knight, D. Marcu, S. DeNeefe, W. Wang, and I. Thayer), Proc. ACL-COLING, 2006.

- "Discovering the Linear Writing Order of a Two-Dimensional Ancient Hieroglyphic Script" (Shou-de Lin and Kevin Knight), *Artificial Intelligence*, volume 170(4-5), 2006.

- "A Better N-Best List: Practical Determinization of Weighted Finite Tree Automata", (J. May and K. Knight), Proc. NAACL-HLT, 2006.

- "Capitalizing Machine Translation", (W. Wang, K. Knight, D. Marcu), Proc. NAACL-HLT, 2006.

- "Synchronous Binarization for Machine Translation", (H. Zhang, L. Huang, D. Gildea, K. Knight), Proc. NAACL-HLT, 2006.

- "Relabeling Syntax Trees to Improve Syntax-Based Machine Translation Quality", (B. Huang and K. Knight), Proc. NAACL-HLT, 2006.

- "Transonics: A Practical Speech-to-Speech Translator for English-Farsi Medical Dialogs," (Robert Belvin, Emil Ettelaie, Sudeep Gandhe, Panayiotis Georgiou, Kevin Knight, Daniel Marcu, Scott Millward, Shrikanth Narayanan, Howard Neely, David Traum), ACL-2005 poster/demo.

- "Interactively Exploring a Machine Translation Model," (Steve DeNeefe, Kevin Knight, and Hayward H. Chan), ACL-2005 poster/demo.

- "Machine Translation in the Year 2004," (K. Knight, D. Marcu), Proc. ICASSP, 2005.

- "An Overview of Probabilistic Tree Transducers for Natural Language Processing," (K. Knight, J. Graehl), Proc. of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Lecture Notes in Computer Science, Springer Verlag, 2005.

- "Text Simplification for Information Seeking Applications," (B. Beigman Klebanov, K. Knight, and D. Marcu), On the Move to Meaningful Internet Systems, (R. Meersman, Z. Tari, eds), Lecture Notes in Computer Science (3290), Springer-Verlag, 2004.

- "What's in a Translation Rule?" (M. Galley, M. Hopkins, K. Knight, and D. Marcu), Proc. NAACL-HLT, 2004.

- "Training Tree Transducers," (J. Graehl and K. Knight), Proc. NAACL-HLT, 2004.

- "The Transonics Spoken Dialogue Translator: An Aid for English-Persian Doctor-Patient Interviews," (S. Narayanan, S. Ananthakrishnan, R. Belvin, E. Ettaile, S. Gandhe, S. Ganjavi, P. G. Georgiou, C. M. Hein, S. Kadambe, K. Knight, D. Marcu, H. E. Neely, N. Srinivasamurthy, D. Traum, and D. Wang), AAAI Fall Symposium, 2004.

- "Syntax-based Language Models for Machine Translation," (E. Charniak, K. Knight, and K. Yamada), Proc. MT Summit IX, 2003.

- "Syntax-based Alignment of Multiple Translations: Extracting Paraphrases and Generating New Sentences," (B. Pang, K. Knight, and D. Marcu, Proc. NAACL-HLT, 2003.

- "Transonics: A Speech to Speech System for English-Persian Interactions," (S. Narayanan et al), Proc. IEEE ASRU, 2003.

- "Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays" (J. Burstein, D. Marcu, and K. Knight), *IEEE Intelligent Systems*, Jan/Feb, 2003.

- "Translation with Scarce Bilingual Resources" (Y. Al-Onaizan, U. Germann, U. Hermjakob, K. Knight, P. Koehn, D. Marcu, K. Yamada), *Machine Translation*, 2003.

- "Fast Decoding and Optimal Decoding for Machine Translation" (U. Germann, M. Jahr, K. Knight, D. Marcu, and K. Yamada), *Artificial Intelligence*, 2003.

- "Using a Large Monolingual Corpus to Improve Translation Accuracy" (R. Soricut, K. Knight, and D. Marcu), Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), 2002.

- "Learning a Translation Lexicon from Monolingual Corpora" (P. Koehn and K. Knight), Proceedings of the ACL Workshop on Unsupervised Lexical Acquisition, 2002.

- "Summarization Beyond Sentence Extraction: A Probabilistic Approach to Sentence Compression" (K. Knight and D. Marcu), *Artificial Intelligence*, 139(1), 2002.

- "Named Entity Translation: Extended Abstract" (Y. Al-Onaizan and K. Knight), Proceedings of the Human Language Technology Conference (HLT), 2002.

- "Translating Named Entities Using Monolingual and Bilingual Resources" (Y. Al-Onaizan and K. Knight), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2002.

- "Machine Transliteration of Names in Arabic Text" (Y. Al-Onaizan and K. Knight), Proceedings of ACL Workshop on Computational Approaches to Semitic Languages, 2002.

- "A Decoder for Syntax-Based Statistical MT" (K. Yamada and K. Knight), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2002.

- "A Syntax-Based Statistical Translation Model" (Yamada, K. and Knight, K.), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2001.

- "Fast Decoding and Optimal Decoding for Machine Translation" (U. Germann, M. Jahr, K. Knight, D. Marcu, and K. Yamada), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2001.

- "Knowledge Sources for Word-Level Translation Models" (P. Koehn and K. Knight), Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2001.

- "Statistics-Based Summarization—Step One: Sentence Compression," (K. Knight and D. Marcu), National Conference on Artificial Intelligence (AAAI), 2000.

- "Translating with Scarce Resources" (Y. Al-Onaizan, U. Germann, U. Hermjakob, K. Knight, P. Koehn, D. Marcu, K. Yamada), National Conference on Artificial Intelligence (AAAI), 2000.

- "Preserving Ambiguities in Generation via Automata Intersection" (K. Knight and I. Langkilde), National Conference on Artificial Intelligence (AAAI), 2000.

- "Estimating Word Translation Probabilities from Unrelated Monolingual Corpora Using the EM Algorithm" (P. Koehn and K. Knight), National Conference on Artificial Intelligence (AAAI), 2000.

- "Decoding Complexity in Word-Replacement Translation Models," *Computational Linguistics*, 25(4), 1999.

- "Mining Online Text," *Communications of the ACM*, 42(11), 1999.

- "A Computational Approach to Deciphering Unknown Scripts," (K. Knight and K. Yamada), Proceedings of the ACL Workshop on Unsupervised Learning in Natural Language Processing, 1999.

- "Machine Transliteration," (K. Knight and J. Graehl), *Computational Linguistics* 24(4), 1998. (Earlier version appeared in Proceedings of the Conference of the Association for Computational Linguistics (ACL), 1997).

- "Translation with Finite-State Devices," (K. Knight and Y. Al-Onaizan), Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), 1998.

- "Translating Names and Technical Terms in Arabic Text," (B. Stalls and K. Knight), Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, 1998.

- "The Practical Value of N-Grams in Generation," (I. Langkilde and K. Knight), Proceedings of the International Natural Language Generation Workshop, 1998.

- "Generation that Exploits Corpus-based Statistical Knowledge," (I. Langkilde and K. Knight), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 1998.

- "Automating Knowledge Acquisition for Machine Translation," *AI Magazine* 18(4), 1997.

- "Learning Word Meanings by Instruction," Proceedings of the National Conference on Artificial Intelligence (AAAI), 1996.

- "Two-Level, Many-Paths Generation," (K. Knight and V. Hatzivassiloglou), Proceedings of the Conference of the Association for Computational Linguistics (ACL), 1995.

- "Filling Knowledge Gaps in a Broad-Coverage MT System," (K. Knight, I. Chander, M. Haines, V. Hatzivassiloglou, E. Hovy, M. Iida, S. K. Luk, R. Whitney, and K. Yamada), Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1995.

- "Unification-Based Glossing," (V. Hatzivassiloglou and K. Knight), Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1995.

- "Integrating Knowledge Bases and Statistics in MT," (K. Knight, I. Chander, M. Haines, V. Hatzivassiloglou, E. Hovy, M. Iida, S. Luk, A. Okumura, R. Whitney, K. Yamada), Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA), 1994.

- "Building a Large-Scale Knowledge Base for Machine Translation," (K. Knight and S. Luk), Proceedings of the National Conference on Artificial Intelligence (AAAI), 1994.

- "Automated Postediting of Documents," (K. Knight and I. Chander), Proceedings of the National Conference on Artificial Intelligence (AAAI), 1994.

- "Are Many Reactive Agents Better Than a Few Deliberative Ones?" Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1993.

- "Building a Large Ontology for Machine Translation," ARPA Workshop on Human Language Technology, Princeton, NJ, March 1993.

- "Unification," *The Encyclopedia of Artificial Intelligence, Second Edition*, John Wiley and Sons, 1992.

- "Integrating Knowledge Acquisition and Language Acquisition," *Applied Intelligence*, 1(1), 1992.

- "Integrating Knowledge Acquisition and Language Acquisition," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, 1991. (Available as technical report CMU-CS-91-209, December 1991).

- "Automated Rationalization," (K. Knight and Y. Gil), Proceedings of the Workshop on Multistrategy Learning, Harpers Ferry, WV, November 1991.

- "Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT," (J. Barnett, I. Mani, E. Rich, C. Aone, K. Knight, and J. C. Martinez), Proceedings of Machine Translation Summit III, Washington, DC, June 1991.

- *Artificial Intelligence, Second Edition* (E. Rich and K. Knight), McGraw-Hill Book Company, 1991.

- "Connectionist Ideas and Algorithms," *Communications of the ACM*, 33(11), November, 1990.

- "Knowledge and Natural Language Processing" (J. Barnett, K. Knight, I. Mani, and E. Rich), *Communications of the ACM*, 33(8), August, 1990.

- "Integrating Language Acquisition and Knowledge Acquisition" (K. Knight, E. Rich, D. Wroblewski), Proceedings of the First International Workshop on Automated Lexical Acquisition, International Joint Conference on Artificial Intelligence (IJCAI), Detroit, MI, 1989.

- "Unification: A Multidisciplinary Survey," *ACM Computing Surveys*, 21(1), March, 1989.

- "Full Unification and Pseudo Unification" (M. Tomita and K. Knight), technical report, Center for Machine Translation, Carnegie Mellon University, 1988.

- "Treeprint: A Program That Draws Parse Trees," technical memo, Center for Machine Translation, Carnegie Mellon University, 1987.