

EDUCATION

Ph.D. in Computer Science **The Johns Hopkins University**
Baltimore, MD, USA May 2005

Thesis:

Language Model Adaptation for Automatic Speech Recognition and Statistical Machine Translation

Advisor: Prof. Sanjeev Khudanpur

M.S.E. in Computer Science **The Johns Hopkins University**
Baltimore, MD, USA May 2000

M.S. in Computer Science and Engineering **Pohang University of Science and Technology**
Pohang, Republic of Korea February 1992

Thesis:

A Design and Implementation of a Korean and Chinese Character Document Reader using Neural Networks

Advisor: Prof. Sung-Yang Bang

B.S. in Computer Science **Korea Advanced Institute of Science and Technology**
Taejon, Republic of Korea February 1990

ADDITIONAL EDUCATION

A Participant **The 4th ELSNET Summer school on Dialogue Systems**
Technical University of Budapest, Hungary July 1996

A Participant **The 1st ELSNET Summer school on Prosody**
University College London, U.K. July 1993

RECENT PROJECT WORK

Statistical Machine Translation (MURI: Muti University Research Initiative)

- Participated in NIST MT evaluations 2002-2004
- Building language Models for machine translation including adaptive language models

A Study on Statistical Language Identification

- Ph.D. Qualifying project #2, The Johns Hopkins University, Jan. 2001.
- Building text-based language identification systems

An Empirical Study of Smoothing Techniques in Language Modeling for Speech Recognition

- Ph.D. Qualifying project #1, The Johns Hopkins University, Aug. 2000.
- Implementation of Kneser-Ney smoothing methods in the CMU LM Toolkit

RESEARCH INTERESTS

Statistical Language Modeling, Language Model Adaptation, Automatic Speech Recognition, Statistical Machine Translation, (Cross-Lingual) Information Retrieval, Text Classification

COMPUTER SKILLS

Programming Languages: C, C++, Perl, LISP, ML, Awk, Sed, Shell scripts, Assembler

Toolkits:

- Language Modeling: CMU-Cambridge LM Toolkit, SRILM
- Speech Recognition: Hidden Markov Model Toolkit (HTK)
- Finite State Machine: AT&T FSM Toolkit
- Machine Translation: GIZA++ Statistical MT Toolkit
- Information Retrieval: LEMUR IR Toolkit
- Text Classification: Rainbow, SVMlight, AdaBoost (BoostTexter)

WORK EXPERIENCE

Senior Analyst

Cincinnati, OH.

Convergys Corporation

January 2005–Present

- Research on automatic speech transcriptions for call center dialogues
- Working on automatic call classifiers: e.g., support vector machines, AdaBoost

Research Assistant

Baltimore, MD.

The Johns Hopkins Univ.

September 2001–October 2004

- Conducted research on various smoothing methods for automatic speech recognition
- Language model adaptations for speech recognition and machine translation

Summer Intern

Boston, MA.

SpeechWorks International Inc.

June–August 2001

- Worked at the international product development team

Teaching Assistant

Baltimore, MD.

The Johns Hopkins Univ.

September 1998–May 2001

- For two courses: database systems (Falls) and information retrieval and web agents (Springs)

Researcher

Seoul, Republic of Korea

Korea Telecom

February 1992–August 1998

- Worked at spoken language processing team in R&D group
- Participated in the development of speech translation system between Korean and Japanese

Research Assistant

Pohang, Republic of Korea

Pohang Univ. of Science and Technology

March 1990–February 1992

- Development of a Korean document formatting system (Korean \LaTeX)
- Design of a neural network description language
- Study on the human recognition of hangul (Korean) characters

SELECTED PUBLICATIONS

Journal Articles

TALIP2004 W. Kim and S. Khudanpur

Lexical Triggers and Latent Semantic Analysis for Cross-Lingual Language Model Adaptation in *ACM Transactions on Asian Language Information Processing*, Vol. 3, No. 2, pp. 94–112, June 2004.

CSL2004 S. Khudanpur and W. Kim

Contemporaneous Text as Side-Information in Statistical Language Modeling in *Computer Speech and Language*, Vol. 18/2, pp. 143–162, April 2004.

- ICASSP2004** W. Kim and S. Khudanpur
Cross-Lingual Latent Semantic Analysis for Language Modeling
in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 257–260, May 2004, Montreal, Canada.
- MT-SUMMIT2003** W. Byrne, S. Khudanpur, W. Kim, S. Kumar, P. Pecina, P. Virga, P. Xu, and D. Yarowsky
The Johns Hopkins University 2003 Chinese-English Machine Translation System
in *Proc. of the Machine Translation Summit IX 2003*, Sep. 2003, New Orleans, Louisiana, USA.
- EUROSPEECH2003** W. Kim and S. Khudanpur
Language Model Adaptation Using Cross-Lingual Information
in *Proc. 8th European Conf. on Speech Communication and Technology*, pp. 3129–3132, Sep. 2003, Geneva, Switzerland.
- EMNLP2003** W. Kim and S. Khudanpur
Cross-Lingual Lexical Triggers in Statistical Language Modeling
in *Proc. Conf. on Empirical Methods in Natural Language Processing*, **Nominated for Best Paper Award**, pp. 17–24, July 2003, Sapporo, Japan.
- ICSLP2002** S. Khudanpur and W. Kim
Using Cross-Language Cues for Story-Specific Language Modeling
in *Proc. 7th International Conf. on Spoken Language Processing*, Vol. 1, pp. 513–516, Sep. 2002, Denver, CO USA.
- EUROSPEECH2001** W. Kim, S. Khudanpur and J. Wu
Smoothing Issues in the Structured Language Model
in *Proc. 7th European Conf. on Speech Communication and Technology*, Vol. 1, pp. 717–720, Sep. 2001, Aalborg, Denmark.
- EUROSPEECH97** W. Kim and M. W. Koo
A Korean Speech Corpus for Train Ticket Reservation Aid System Based On Speech Recognition
in *5th Proc. European Conf. on Speech Communication and Technology*, Vol. 4, pp. 1723–1726, Sep. 1997, Rhodes, Greece.
- ICSP97** W. Kim and M. W. Koo
Statistical Corpus Analysis for KT-TREASURE : Korea Telecom Train ticket REservation Aid System based Upon speech REcognition
in *Proc. 14th International Conf. on Speech Processing*, pp. 319–323, Aug. 1997, Seoul, Korea.
- KCOLC96** W. Kim, D. S. Chang, J. I. Kim and M. W. Koo
A Multi-lingual Speech Translation System for Hotel Reservation
in *Proc. '96 2nd KOREA-China Joint Symposium on Oriental Language Computing*, Oct. 1996, Seoul, Korea.
- EUROSPEECH95** M. W. Koo, I. H. Sohn, W. Kim and D. S. Chang
KT-STTS : A Speech Translation System for Hotel Reservation and a Continuous Speech Recognition System for Speech Translation
in *Proc. 4th European Conf. on Speech Communication and Technology*, Vol. 2, pp. 1227–1230, Sep. 1995, Madrid, Spain.
- IJCNN92** W. Kim and S. Y. Bang.
An Efficient Approach to Read Documents with Korean and Chinese Characters Mixed
in *Proc. International Joint Conf. on Neural Networks*, **Received Best Poster Award**, Nov. 1992, Beijing, China.