

Associate Professor,  
Dept. of Information Communications Engineering, Chungnam National University.

**Address:**

College of Engineering III-421  
Dept. of Information Communications Engineering  
Chungnam National University  
220 Gung-dong, Yuseong-gu, Daejeon  
305-764, Korea  
TEL: 82-42-821-5662  
FAX: 82-42-823-5586  
email: [kjoolee@cnu.ac.kr](mailto:kjoolee@cnu.ac.kr) [kjoolee@gmail.com](mailto:kjoolee@gmail.com)  
homepage: <http://marble.cnu.ac.kr>

**Education:**

- **Korea Advanced Institute of Science and Technology (KAIST)**. Daejeon, Korea  
Ph.D in Computer Science, 1998  
Thesis title: Probabilistic Parsing of Korean based on Language-Specific Properties  
Advisor: Professor Gil Chang Kim
- **Korea Advanced Institute of Science and Technology (KAIST)**. Daejeon, Korea  
M.S in Computer Science, 1994  
Thesis title: A Study on a Robust Parser for Extragrammatical English Sentences  
Advisor: Professor Gil Chang Kim
- **Sogang University**. Seoul, Korea  
B.A in Computer Science, 1992

**Professional Experience:**

- 2005.4 ~ current: Associate Professor of **Chungnam National University**, Daejeon, Korea.
- 2004.3 ~ 2005.3: Lecturer of **KyungIn Woman's College**, Incheon, Korea.
- 2003.3 ~ 2004.2: Lecturer of **Ewha Women's University**, Seoul, Korea.
- 1998.2 ~ 2003.2: Researcher of **Microsoft Korea**, Seoul, Korea.
- 1994.3 ~ 1998.2: Teaching and Research Assistant of **Korea Advanced Institute for Science and Technology (KAIST)**, Daejeon, Korea.

## Research Interest/Areas:

Statistical Syntactic Analysis

Statistical Machine Translation

Intelligent Computer-Assisted Language Learning

Learning methods in Natural Language Processing, Language Modeling, Corpus Representation

Sentiment Analysis, Opinion Mining

Information Extraction

## Research Grants:

1. **"Developing a bidirectional Japanese-Korean Machine Translation System based on MOSES"**, funded by NHN, 2009.09 ~ 2010.08, ₩100,000,000
2. **"A study on extracting sentimental patterns from a Korean Corpus"**, funded by Electronic Telecommunication Research Institute (ETRI), 2009.05 ~ 2010.1, ₩35,000,000
3. **"Developing an Automated Scoring System for English Writings III"**, funded by Korea Institute of Curriculum & Evaluation (KICE), 2008.03 ~ 2008.12, ₩35,000,000  
<http://marble.cnu.ac.kr/~kice/StarSun/kjoollee.php>
4. **"Improving Performance of a Korean word breaker"**, funded by Microsoft Korea, 2008.05 ~ 2008.06., ₩30,000,000
5. **"Developing a Sentiment Dictionary for Opinion Mining"**, funded by Moran Soft, 2008.6 ~ 2008.8, ₩15,000,000
6. **"Developing an Automated Scoring System for English Writings II"**, funded by Korea Institute of Curriculum & Evaluation (KICE), 2007.03 ~ 2007.12, ₩75,000,000
7. **"Developing an Automated Scoring System for English Writings I"**, funded by Korea Institute of Curriculum & Evaluation (KICE), 2006.03 ~ 2006.12, ₩90,000,000
8. **"Developing a Korean word breaker"**, funded by Microsoft Korea, 2006.11 ~ 2007.05., ₩60,000,000

## Projects Participated:

1. **"Developing a Korean Opinion Mining System"**, Co-researcher, Yahoo Korea, 2005-2006.
2. **"Developing a Korean Grammar Checker for Microsoft Office XP Word"**, Program Manager, MS Korea, 1999-2000.
3. **"Developing a Korean Stemmer for Microsoft Office XP,"** Program Manager, MS Korea, 1998-1999.
4. **"Developing a Korean Spell Checker for Microsoft Office XP"**, Program Manager, MS Korea, 1998-1999.
5. **"Developing a Korean Smart Tag for Microsoft Office 2003"**, Program Manager, MS Korea, 2001.
6. **"Developing a Korean Word Breaker for Microsoft SharePoint Portal Server"**, Program Manager, MS Korea, 1998-2001.

7. "**Integrated Korean Information Base: Korean syntactic tree-tagging system**", Researcher, financed by Ministry of Science and Technology of Korea, 1995.3-1997.8
8. "**Development of English-Korean machine translation system of military intelligence**", Researcher financed by the Republic of Korea Air Force, 1996.4-1997.12
9. "**Developing disambiguation algorithms for multiple interpretations arising in the analysis of Korean sentences**", Researcher, financed by Korea Science and Engineering Foundation (KOSEF), 1994.8-1997.8

### Awards/Honors:

- Alpha Sigma Nu, 1991.
- Magna Cum Laude, Sogang Univ. 1992

### Teaching:

- Undergraduate course:
  - Operating Systems
  - Database
  - Data Structure
  - Discrete Mathematics
- Graduate course:
  - Natural Language processing
  - Information Retrieval
  - Pattern Recognition

### Publications:

#### International Journal:

1. **Kong Joo Lee** and Jee Eun Kim, "Improving Automatic English Writing Assessment Using Regression Trees and Error Weighting," submitted to IEICE transactions of Information and Systems, 2009.
2. **Kong Joo Lee**, Yong-Seok Choi and Jee Eun Kim, "Building an Automated English Composition Evaluation System for Students Learning English as a Second Language," submitted to Computer Speech and Language, 2009.
3. Young-Hwan CHO and **Kong Joo Lee**, "Automatic Affect Recognition Using Natural Language Processing Techniques and Manually built Affect Lexicon," IEICE transactions of Information and Systems, Vol. E89D, NO.12, Dec. 2006.
4. Sanghwa Yuh, **Kong Joo Lee** and Jungyun Seo, "Multilingual Closed Caption Translation System for Digital Television," IEICE transactions of Information and Systems, Vol. E89D, NO.6, Jun. 2006.
5. **Kong Joo Lee** and Hyokyung Bahn, "A Compressed Display Technique for 240 x 320 Resolution Personal Digital Assistants (PDAs)," IEEE Transactions on Consumer Electronics, Vol.

51, No. 4, Nov. 2005.

6. Chul Su Lim, **Kong Joo Lee**, and Gil Chang Kim, "Multiple sets of features for automatic genre classification of web documents," *Information Processing & Management*, Volume 41, Issue 5, Sept. 2005.
7. Myung-Seok Choi, **Kong Joo Lee**, Key-Sun Choi, Gil Chang Kim, "Extracting partial parsing rules from Tree-annotated Corpus: Toward Deterministic Global Parsing," *IEICE transactions of Information and Systems*, Vol. E88D, NO.6, Jun. 2005.
8. **Kong Joo Lee**, Jae-Hoon Kim, "Sentence Compression Learned by News Headline for Displaying in Small Device, *Lecture Notes in Computer Science*, Vol.3411, Feb. 2005.
9. Chul Su Lim, **Kong Joo Lee**, and Gil Chang Kim, "Automatic Genre Detection of Web Documents, " *Lecture Notes in Computer Science*, Vol. 3248/2005. Jan. 2005.
10. **Kong Joo Lee**, "Document Genre Classification for User Interface of Web Search Engine," *IEICE transactions on Information and Systems*, VOL.E87-D NO. 7, Jul. 2004.
11. Seongyong KIM, **Kong Joo Lee**, Key-Sun CHOI. "Normalizing Syntactic Structures using POS Tags and Binary Rules," *IEICE transactions on Information and Systems*, VOL.E86-D, NO. 10, Oct. 2003.
12. Seongyong KIM, **Kong Joo Lee**, Key-Sun CHOI, "Automatic Generation of Composite Labels Using Part-of-Speech Tags for Parsing Korean," *International Journal of COMPUTER PROCESSING OF ORIENTAL LANGUAGES*, VOL 16. NO. 3, Sept. 2003.
13. **Kong Joo Lee**, Jae-Hoon Kim and Gil Chang Kim, "An Efficient Parsing of Korean Sentences Using Restricted Phrase Structure Grammar," *International Journal of COMPUTER PROCESSING OF ORIENTAL LANGUAGES*, VOL 11. NO. 1, 1997.
14. **Kong Joo Lee**, Jae-Hoon Kim Young S. Han and Gil Chang Kim, "Restricted Representation of Phrase Structure Grammar for Building a Tree Annotated Corpus of Korean," *International Journal of Natural Language Engineering*, VOL 3. ISSUE 2. 1997.

#### **Korean Journal:**

1. Hyun Ah Lee, Jee Eun Kim, **Kong Joo Lee**, "Accuracy Improvement of an Automated Scoring System through Removing Duplicately Reported Errors," *The Transaction of the Korean Information Processing Society*, Part B, Vol. 16-B, NO.2, April 2009. (in Korean)
2. Jee Eun Kim, **Kong Joo Lee**, Kyung-Ae Jin, "Building an Automated Scoring System for a Single English Sentences," *The Transaction of the Korean Information Processing Society*, Part B, Vol. 14-B, NO.3, June 2007. (in Korean)
3. Jee Eun Kim, **Kong Joo Lee**, "Implementing Automated English Error Detecting and Scoring System for Junior High School Student," *Journal of the Korea Contents Association*, Vol.7 NO.5, May, 2007. (in Korean)
4. **Kong Joo Lee**, Bo Hyun Yun, "Automatic Extraction of Paraphrases from a Parallel Bible Corpus," *Korean Journal of Cognitive Science*, Vol. 17, NO. 4, Dec, 2006. (in Korean)
5. **Kong Joo Lee**, Chul-Soo Lim, and Jae-Hoon Kim, "Automatic classification of web documents

- according to their styles," The Transaction of the Korean Information Processing Society, vol. 11-B, no. 5, pp. 555-562, August 2004 (in Korean).
6. **Kong Joo Lee** and Jae-Hoon Kim, "Construction of linearly aligned corpus using unsupervised learning," The Transaction of the Korean Information Processing Society, vol. 11-B, no. 3, pp. 387-394, June 2004 (in Korean).
  7. Seongyong Kim, **Kong Joo Lee**, Key-Sun Choi, "Korean Syntactic Rules using Composite Labels," Journal of the Korea Information Science Society, Vol.31, NO.2, Feb. 2004. (in Korean).
  8. **Kong Joo Lee**, Sun Young Hwang, and Jee Eun Kim, "A Korean Grammar Checker based on the Trees Resulted from a Full Parser," Journal of the Korea Information Science Society, Vol.30, NO.9~10, Sept. 2003. (in Korean)
  9. **Kong Joo Lee**, Jae-Hoon Kim, "Implementing Korean Partial Parser based on Rules," The Transaction of the Korean Information Processing Society Part B. Vol 10-B, NO.4, Aug. 2003. (in Korean)
  10. Jae-Hoon Kim and **Kong Joo Lee**, "Segmenting and Classifying Korean Words based on Syllables Using Instance-Based Learning," The Transaction of the Korean Information Processing Society Part B. Vol 10-B, NO.1, Feb. 2003. (in Korean)
  11. **Kong Joo Lee** and Jae-Hoon Kim, "Korean probabilistic syntactic model using head cooccurrence," The Transaction of the Korean Information Processing Society, vol. 9-B, no. 6, pp.809-816, Dec. 2002 (in Korean).
  12. Sun Young Hwang and **Kong Joo Lee**, "An Analysis of Noun-modifying Adverbs for Structural Disambiguation," Korean Journal of Cognitive Science, Vol. 13, NO.4, Dec, 2002. (in Korean)
  13. **Kong Joo Lee**, Jae-Hoon Kim, and Gil Chang Kim, "Syntactic analysis of Korean sentences based on restricted phrase structure grammar," Journal of the Korea Information Science Society (B): Software and Applications, vol. 25. no. 4, pp. 722-732, April 1998 (in Korean).
  14. **Kong Joo Lee**, Jae-Hoon Kim, Kye-Sun Choi, and Gil Chang Kim, "Korean syntactic tagset for building a tree annotated corpus," Korean Journal of Cognitive Science, vol. 7, no. 4, pp. 7-24, 1996 (in Korean).
  15. **Kong Joo Lee**, Cheol Jung Kwon, and Gil Chang Kim, "A Robust Parser for Extragrammatical Sentences Based on Syntactic Information," Journal of the Korea Information Science Society (B), Vol.23, No.4, April 1996. (in Korean).

### **Dissertation Abstract:**

#### ***Probabilistic Parsing of Korean based on Language-Specific Properties***

*Dissertation Committee:* Gil Chang Kim (Chair), Jungyun Seo, Sung-Deok Cha, Jae-Heon Yang, Chang Sup Sung,

Natural language parsing is a central component to many natural language processing tasks. Since a natural language has inherently structural ambiguities, one of the difficulties of parsing is resolving the structural ambiguities. The ambiguities arise whenever a sentence can be interpreted in more than one way. Recently, a probabilistic approach to tackle this disambiguation problem

has received considerable attention because it has some attractions such as automatic learning, wide-coverage, and robustness. Many probabilistic parsing models have been developed for the past few years, and they are mainly for English not Korean. In this thesis, we focus on Korean syntax and a probabilistic parsing model for Korean.

We investigate two problems: representing Korean syntax, and building a language model for Korean syntax. The representation of a language is to describe the structure of the language, and it is the direct reflection of the features of the language. A language model is a probability distribution  $P(S)$  over string  $S$  that attempts to reflect how frequently a string  $S$  occurs as a sentence. The claim of this thesis is that syntax of Korean in fact can be represented more efficiently using a grammar representation scheme that exploits the characteristics of Korean. Also, we argue that the consideration of the language specific features can produce a more accurate natural language parser than one without such consideration. These claims are justified by constructing a parser for Korean based on the specific properties of the Korean structures, and comparing its performance to a state-of-the-art parser for English on a common task.

The representation scheme of the syntactic structures is a specification about the kind of the knowledge to be used and the form of the knowledge to be annotated. Korean has different structural properties from English, for example, the word composition as well as the word order in a sentence. In addition, English is more or less fixed word order language, while Korean is partially free word order. It can be easily said that a representation scheme of the syntactic structures for English will not be apt to Korean. We propose a new representation scheme for Korean, which emphasizes the structural properties of Korean.

A probabilistic language model has direct relation with the syntactic representation which the model handles. Furthermore, the language model has to encode the language specific properties to improve the performance of the parser based on it. Hence, a probabilistic parsing model cannot be completely independent of the representation of the syntactic structure nor the properties of the language that the model is concerned on. We build the probabilistic Korean parser which reflects the language specific properties, such as the strong tendency of topicalization, and the freedom of the word order. Experiments show that the accuracy of the parser based on the model proposed in this thesis is higher than that of state-of-the-art parser proposed for English on the Korean texts.