

---

---

**IJCNLP-05**

---

---

**Sixth International Workshop  
on  
Linguistically Interpreted Corpora  
(LINC-2005)**

Proceedings of the Workshop

**15 October 2005  
Jeju Island, Korea**



## Editorial

Large linguistically interpreted corpora continue to play an important role for machine learning, evaluation, psycholinguistics, and (increasingly so) theoretical linguistics. Research groups world-wide are engaged in the creation of corpus resources annotated with morphological, syntactic, semantic, and discourse information for an ever-increasing variety of languages. Since 1999, the LINC workshop series has brought together these activities in order to identify and disseminate best practice in the development and utilization of linguistically interpreted corpora. LINC workshops, thus, have traditionally served a dual purpose, viz. to (i) present an up-to-date survey of ongoing work on linguistically interpreted corpora and (ii) facilitate the exchange, propagation, and further development of research results with respect to the annotation, conversion, and exploitation of such corpora. We hope to continue this tradition at LINC-05, as a post-conference workshop at IJCNLP on beautiful Jeju Island.

The aim of the workshop is to exchange and propagate research results with respect to the annotation, conversion and exploitation of corpora taking into account different applications and theoretical investigations in the field of language technology and research. The call for papers encouraged submissions of papers constituting substantial, original, and unpublished work on all aspects of linguistically interpreted corpora, including, but not limited to: (a) creation of practical annotation schemes, (b) efficient annotation techniques, (c) automation of corpus annotation, (d) tools supporting corpus conversions, (e) validation including consistency checking of corpora, (f) browsing corpora and searching for instances of linguistic phenomena, (g) interpretation of quantitative results, and (h) automatic induction of linguistic competence through machine learning techniques. Looking at the final program for the workshop, we believe all of the above are represented at LINC-05.

We would like to thank all the authors who submitted papers. There were 14 submissions, of which 11 appear in the proceedings. We would like to particularly thank all the members of the program committee for their time and effort in ensuring that all submissions were fairly assessed.

Finally, we wish to thank the organizers of the main conference, in particular the conference workshop co-chairs, Yuji Matsumoto and Laurent Romary,

Keihanna & Oslo, August 2005

Kyonghee Paik,  
Francis Bond, and  
Stephan Oepen

## **LINC-05 Program Committee**

- Anne Abeillé, Paris
- Olga Babko-Malaya, Pittsburgh
- Colin Baker, Berkely
- Pierrette Bouillon, Geneva
- Thorsten Brants, Palo Alto
- John Carroll, Sussex
- Huang Chu-Ren, Taipei
- Montserrat Civit Torruella, Barcelona
- Tomaz Erjavec, Ljubljana
- Jan Hajic, Prague
- Chung-hye Han, British Columbia
- Silvia Hansen, Saarbrücken
- Erhard Hinrichs, Tübingen
- Beom-mo Kang, Seoul
- Sadao Kurohashi, Tokyo
- Frank Keller, Edinburgh
- Brigitte Krenn, Vienna
- Joakim Nivre, Vaxjö
- Stephan Oepen (co-chair), Oslo and Stanford
- Kyonghee Paik (co-chair), Keihanna
- Laurent Romary, Nancy
- Kiril Simov, Sofia
- Kiyotaka Uchimoto, Keihanna
- Hans Uszkoreit, Saarbrücken
- Atro Voutilainen, Helsinki
- Nianwen Xue, Pittsburgh

# Table of Contents

EDITORIAL .....	i
LINC-05 PROGRAM COMMITTEE .....	ii
<i>The TIGER 700 RMRS Bank: RMRS Construction from Dependencies</i> Kathrin Spreyer and Anette Frank.....	1
<i>Obtaining Japanese Lexical Units for Semantic Frames from Berkeley FrameNet using a Bilingual Corpus</i> Toshiyuki Kanamaru, Masaki Murata, Kow Kuroda and Hitoshi Isahara .....	11
<i>A Study of Applying BTM Model on the Chinese Chunk Bracketing</i> Jia-Lin Tsai.....	21
<i>Integration of a Lexical Type Database with a Linguistically Interpreted Corpus</i> Chikara Hashimoto, Francis Bond, Takaaki Tanaka and Melanie Siegel .....	31
<i>Building Dialogue Corpora for Nursing Activity Analysis</i> Hiromi itoh Ozaku, Akinori Abe, Noriaki Kuwahara, Futoshi Naya, Kiyoshi Kogure and Kaoru Sagara.....	41
<i>The Syntactically Annotated ICE Corpus and the Automatic Induction of a Formal Grammar</i> Alex Chengyu Fang .....	49
<i>Syntactic Identification of Attribution in the RST Treebank</i> Peter Rossen Skadhauge and Daniel Hardt.....	57
<i>Linguistically Enriched Corpora for Establishing Variation in Support Verb Constructions</i> Begoña Villada Moirón .....	63
<i>Error Annotation for Corpus of Japanese Learner English</i> Emi Izumi, Kiyotaka Uchimoto and Hitoshi Isahara .....	71
<i>Some remarks on the Annotation of Quantifying Noun Groups in Treebanks</i> Kristina Spranger .....	81
<i>Annotating Honorifics Denoting Social Ranking of Referents</i> Shigeiko Nariyama, Hiromi Nakaiwa and Melanie Siegel.....	91
AUTHOR INDEX.....	101

