COLING 2004를 통해 본 COLING과 문법공학
- COLING 2004를 다녀와서 -

류병례
중앙대학교 언어학과
http://www.cnu.ac.kr/~ryu/
ryu@cnu.ac.kr

Contents

1 COLING & COLING 2004 ................................................................. 1
  1.1 COLING ...................................................................................... 1
  1.2 COLING 2004 ............................................................................. 3
  1.3 COLING 2004 Tutorials × 3 ..................................................... 3
  1.4 COLING 2004 Workshops - Day 1 ........................................... 4
  1.5 COLING 2004 Workshops - Day 2 ........................................... 4

2 The Xerox Linguistics Environment (XLE) ..................................... 5

3 Postscript ....................................................................................... 6

1 COLING & COLING 2004

1.1 COLING

• International Committee on Computational Linguistics (ICCL) organises the International Conference on Computational Linguistics (COLING).

• ICCL:

  – The International Committee on Computational Linguistics was set up by David Hays in the mid-Sixties as a permanent body to run international computational linguistics conferences but in an original way: with no permanent secretariat, subscriptions or funds.

  – COLINGs have striven for inclusiveness, both geographical, when the world was more harshly divided than it is now, and theoretical, in that COLING has been less prone to mood swings of theory than societies that are run in different, and more conventional, ways.
– In recent years, The Association for Computational Linguistics (ACL) has given great assistance and cooperation in keeping COLING proceedings available and distributed, a relationship due to Don Walker’s participation in both bodies.
– Members of the ICCL who have died include Don, Dave Hays, Bernard Vauquois, Hans Karlgren, Andre Ljudskanov, Guy Rondeau and Antonio Zampolli.


- ACL & EACL
  – The Association for Computational Linguistics (ACL):
  – The European Chapter of the ACL (EACL):

  – 2004
    • ACL-04 was held in Barcelona, Spain in the second half of July.
    • COLING-04 was held in Geneva, Switzerland from August 23 to August 27.
  – 2003
    • EACL-03 was held in Budapest, Hungary from April 12 to April 17, 17, 2003.
    • ACL-03 was held in Sapporo, Japan from July 7 to July 12, 2003.
  – 2002
    • ACL-02 (40th Anniversary!!) was held at the University of Pennsylvania, Philadelphia, PA, USA, July 7-12, 2002.
    • COLING 2002 will be on the campus of Academia Sinica in Taipei, Taiwan from August 26 to August 30, 2002.
  – 2001
    • Second meeting of the NAACL (NAACL’01), Pittsburgh, PA (June 2-7, 2001)
    • 39th Annual Meeting of the ACL (ACL’01) - joint with EACL’01, Toulouse, France (July 6-11, 2001)
  – 2000
    • Language Technology Joint Conference (ANLP-NAACL’00), Seattle, WA (April 29 - May 3, 2000)
    • The 18th International Conference on Computational Linguistics (Coling’00), Luxembourg (Tutorials: July 29, 2000), Saarbrucken, Germany (Conference: July 31 - August 4, 2000), and Nancy, France (Workshops: August 5-6, 2000)
1.2 COLING 2004

- The 20th International Conference on Computational Linguistics COLING 2004 took place at the University of Geneva, Switzerland, on August 23rd-27th, 2004 (main conference dates), together with satellite events before and after the main conference.

- Further information:

- COLING 2006 will take place in Sydney, Australia.

1.3 COLING 2004 Tutorials × 3

- Sunday 22nd August 2004

- T1 Multilingual and Cross-Lingual Information Access Gareth J.F. Jones, Dublin City University, Ireland 10:00 - 13:00 Room R160

  - This tutorial will introduce information searching and access for multilingual environments.

  - The tutorial will begin with a brief overview of current approaches to rank-based information retrieval and then explain how these have been adapted successfully to a range of different language types.

  - The presentation will then go on to describe the challenges of cross-language and multilingual information retrieval and techniques that are being explored to overcome them.

- T2 The XLE Grammar Development Platform and Parser/Generator Miriam Butt, Universität Konstanz, Germany and Tracy Holloway King, Palo Alto Research Center, USA 10:00 - 13:00 Room R170

  - This tutorial provides an introduction to writing large-scale, robust, deep grammars with XLE.

  - XLE is a parsergenerator with special tools built in to aid grammar development.

  - The tutorial will include sections on writing grammars, incorporating robustness techniques, generation, and using the grammars in applications.

- T3 - CANCELLED Natural Language Information Assurance and Security Victor Raskin, Purdue University, USA

- T4 Modeling Information Structure for Discourse and Dialog Processing Ivana Kruijff-Korbayova, Universität des Saarlandes, Germany 14:00 - 17:00 Room R160

  - Information Structure (IS) concerns structural and semantic properties of utterances reflecting communicative intentions and discourse context.
Among the means to realize IS are word order, intonation and marked syntactic constructions. Various dichotomies are used to describe IS, e.g. Theme-Rheme, Topic-Comment, Background-Focus, Given-New and Contextually Bound-Nonbound. The proliferating terminologies are one reason why it is difficult to orient in the existing formal and computational work on IS.

This tutorial will explain basic notions, provide an overview of IS approaches, and survey work employing IS in NLP systems, sketching future challenges.

1.4 COLING 2004 Workshops - Day 1

- Saturday, August 28th
  - W1 International Joint workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP) 2004 Nigel Collier, Patrick Ruch and Adeline Nazarenko - 2 day workshop 9:15 - 17:45 Room R070
  - W2 Psycho-Computational Models of Human Language Acquisition William Gregory Sakas 8:40 - 17:40 Room R040
  - W4 Recent Advances in Dependency Grammar Geert-Jan M. Kruijff and Denys Duchier 9:30 - 17:45 Room R170
  - W5 Computational Approaches to Arabic Script-based Languages Ali Farghaly and Karine Megerdoomian 8:30 - 18:00 Room 1150
  - W6 eLearning for Computational Linguistics and Computational Linguistics for eLearning Erhard Hinrichs, Lothar Lemnitzer, Detmar Meurers 8:30 - 18:30 Room 1160
  - W11 Multilingual Linguistic Resources Gilles Serasset, Susan Armstrong, Christian Boitet, Andrei Popescu-Belis, Dan Tufis 8:30 - 18:00 Room R160
  - SE1 DUMAS final workshop: Robust and adaptive information processing for mobile speech interfaces Bjorn Gamback, Kristiina Jokinen - 2 day workshop 9:15 - 18:00 Room R060

1.5 COLING 2004 Workshops - Day 2

- Sunday, August 29th
  - W1 International Joint workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP) 2004 Nigel Collier, Patrick Ruch and Adeline Nazarenko - 2 day workshop 9:30 - 13:00 Room R070

4
• W7 CompuTerm 2004: 3rd International Workshop on Computational Terminology
  Sophia Ananiadou and Pierre Zweigenbaum 8:55 - 17:30 Room R170

• W8 5th International Workshop on Linguistically Interpreted Corpora Silvia Hansen,
  Stephan Oepen and Hans Uszkoreit 9:00 - 18:00 Room R040

• W9 Romand 2004: 3rd workshop on RObust Methods in Analysis of Natural Language
  Data Vincenzo Pallotta, Amalia Todirascu 8:30 - 18:30 Room R150

• W10 Enhancing and using electronic dictionaries Michael Zock and Patrick St. Dizier
  8:45 - 18:00 Room R160

• SE1 DUMAS final workshop: Robust and adaptive information processing for mobile
  speech interfaces Bjorn Gamback, Kristiina Jokinen - 2 day workshop

2 The Xerox Linguistics Environment (XLE)

• The XLE Grammar Development Platform and Parser/Generator Miriam Butt (Uni.
  Konstanz) and Tracy Holloway King (PARC) This tutorial provides an introduction to
  writing large-scale, robust grammars with XLE. XLE is a parser/generator with special
  tools built in to aid grammar development. The grammars use the Lexical-Functional
  Grammar (LFG) formalism.

• Organization: First, we discuss the process of building a grammar: the formulation of
  rules, the integration of lexicons and morphological analyzers, and the use of templates
  to capture generalizations. We then discuss the use of several robustness techniques:
  FSM guessers, the fragment grammar, and optimality theory (OT) marks. During
  the second half of the tutorial, we discuss reversing the grammar for generation and
  the integration of stochastic methods for disambiguation. Finally, we briefly outline
  how such grammars can be used in applications like machine translation, sentence
  condensation, and computer assisted language learning.

• Writing a Grammar: The XLE grammars produce both a canonical phrase-structure
  tree and an attribute-value matrix which contains predicate-argument structure as well
  as information about tense, statement type, etc. XLE grammars consist minimally of
  annotated phrase-structure rules and a lexicon. Large-scale grammars also employ
  templates to capture generalizations and finite-state morphologies to bootstrap the
  lexicon.

• Robustness: Deep grammars are often avoided because of their brittleness. Three
  robustness techniques used by the XLE system are discussed: 1) Language-specific
  finite-state guessers to deal with novel vocabulary; 2) XLE incorporates a version of
  chunk parsing, by which strings which do not receive a spanning parse are broken into
smaller well-formed chunks with both phrase structure and functional structure parses; 3) A version of optimality theory in which lexical entries and rules can be annotated with OT marks to relax constraints on the grammar.

- **Generation**: The XLE system provides a generator as well as a parser. Generation is done from (underspecified) functional structures. The generation grammar is the reverse of the parsing grammar. Exactly reversing the parsing grammar can result in ambiguous output. OT marks can be used to constrain the output.

- **Applications**: The XLE grammars produce packed output for ambiguous input sentences. The XLE system integrates a stochastic component which can disambiguate parses for applications by picking the most probable parse. Large-scale XLE grammars are being incorporated into applications, such as machine translation, sentence condensation, and computer assisted language learning.

- **Pre-Requisites**: no prerequisites assumed, but a basic knowledge of syntax will be helpful.

3 Postscript