COLING 94

The 15th International Conference on Computational Linguistics

PROCEEDINGS

Vol. II

August 5 – 9, 1994

Kyoto, Japan

COLING 94

The 15th International Conference on Computational Linguistics

PROCEEDINGS

Vol. II

August 5 – 9, 1994

Kyoto, Japan

PREFACE

COLING is held every other year under the auspices of ICCL.

This conference is to discuss theoretical and practical problems occurring when languages are handled by computer. The discussion covers a wide range of linguistic issues and their computer implementation. These include: (a) linguistic theories such as syntax, semantics, and discourse problems; (b) linguistic data such as dictionaries and text corpora; (c) algorithms for analysis and synthesis; (d) experimental systems for language understanding and dialogue; and (e) application systems such as those involving machine translation and man-machine interface. This conference provides the bases and technologies essential for the future information society.

The 1st International Conference was held in New York in 1965, and the 8th Conference was in Tokyo in 1980. The 15th Conference in Kyoto will be a good opportunity to learn what role languages will play in the information- and multimedia-oriented society. We expect this Conference to promote greater development in this field.

Professor Makoto Nagao
Conference Chairman

Program Chairman's Introduction

My first duty and pleasure is to thank all those involved in preparing the program for COLING94, as well as those who are to present papers and participate in panels. A very long list appears in this volume of those who were on the program committee or assisted its members directly by reviewing. The committee itself operated as a set of more or less autonomous members, differing in area of expertise, and making final recommendations to me based on the forms filled by the reviewers themselves. In only one case, I think, did I interfere in that recommendation: all other exercises of discretion by me were where the committee member remained uncertain and, more generally, where the final recommendations, as a whole, exceeded the slots available, given their type (long or short). We did change the names of the types between the first announcement and the final outcome, after it became clear how hard it would be to provide platforms for demonstrations in Kyoto with the resources available.

The acceptance/rejection ratio was almost exactly the same as in 1992, namely same as the last COLING, namely 1:2. This would be insufficiently macho for more abrasive conferences, but I feel sure the quality of this conference is good and will compare favorably with its predecessors. It is hard to gauge the changes in area of interest in the last two years since we have used a different taxonomy (without separating off Applications) from what was used at Nantes, but I suspect that the proportions are very similar, perhaps with syntax having a small resurgence, and empirical methods (broadly defined and not excluding syntax) still running very strong.

A very pleasing feature is the wide spread of countries represented here (28), which has always been an important matter in at COLINGs. They are, I believe, the most international meetings in our field, much more so that ACL or AAAI, which remain fundamentally American meetings, which makes the whole review process much simpler, a matter I shall return to below.

As always, the host country is well represented here, though if you add together the papers from Germany and the UK they total the same as those from Japan, and the summed population of the two countries is also roughly that of Japan. By that standard, Japan is not over represented here at all, and perhaps the US and other parts of Europe, as well as many other countries, are under represented, with respect both to their populations and the amount of relevant work being done in them.

US under-representation abroad is always a result of the no-repeat-papers rule imposed by some US conferences, and we have lost a number of accepted papers in that way this year, which is why we took twenty "reserve" papers into the printed proceedings. However, the US was well-represented on the program committee and the ratio of papers submitted to those accepted was pretty invariant across countries. So, this under-representation remains a pity, especially in view of the important role that US Government research programs, particularly the ARPA Speech and Language Program, have had on the field world-wide in recent years. That Program has done more than any other to reestablish the importance of evaluation to progress and the primacy of empiricism in CL/NLP; even though some of the growth points came from elsewhere, particularly example-based MT from Japan, and statistical corpus tagging from the UK. Hopefully, the International Projects Day, chaired by Antonio Zampolli, will give a platform for what has been done in the US under that program.

It must be said here that aspects of the program construction process have made some people unhappy, including me: especially (i) the fact that the date given for return of verdicts on papers was missed; (ii) that referees' comments on REJECTED papers were not returned to authors immediately, and (iii) there were mailing errors of various sorts.

No author has complained specifically about (iii), but at least one of the errors happened at Sheffield and may have contributed to problem (i). I apologize for all the above, since I am responsible though, as adults know, that is not the same as being to blame. As to (ii), I decided initially that, as this was not done

uniformly, and COLING has little in the way of a formal budget, I would not return the comments on rejected papers, even though I realised and appreciated that much work had gone into them. However, the volume of protests caused me to reconsider and these have now been sent out, but I very much take the view that this was done as a courtesy to colleagues and not because there is some mysterious right to them, and especially not because the Committee has to prove to any individual that its decisions are transparent and not open to question. The last is a very important point, and I will return to it below, but I am certain that the reviewing process world wide cannot continue to function if our community comes to believe that decisions of this kind are justiciable, and can be judged and questioned by legal criteria. That is a deep cultural error, I am certain. The reliability of the process as a whole rests on the collective reputations of the committee as individuals in our research community, and nowhere else.

I have a particular perspective on this because I was also the COLING84 Program Chair. Then as now, there was a slightly amateurish chaos about arrangements, if anything it may have been slightly worse that year because I was on sabbatical leave in Costa Rica at the time, and running things with a worse communication system than was generally available ten years ago. There were glitches as always, but the conference was as great success, as COLINGs normally are, and all was right on the night.

Costa Rica was, in that sense, a symbolic center from which to organize a COLING, in that the ICCL has always taken a more world-wide view of the field than have the professional societies which has meant, since global communications are not evenly distributed, that the most up-to-date methods have not been used: we have stuck to paper ard mail, perhaps wrongly, as the basis of reviews, and have had a wide-flung committee, all of whom must report before decisions can be taken. All that inevitably leads to a slower, bumpier review process and we shall perhaps be able to change it in the near future, given the rate at which the Internet is spreading.

However, in that decade, there have been other real changes in this process that I have noticed and would like to comment on, particularly in:
(a) the expectation level from authors and participants and, in particular, the belief of some of them that the client-provider model is an appropriate one for the author-program-official relationship and (b) the way the immediacy and density of email can produce damaging side effects in procedures like conference reviewing.

There have been, in recent times, four main modes of conference organization:

a. The old ICCL way of organizing COLING, where the ICCL itself reviewed abstracts and made quick decisions with no reports to authors. A magnificent version of that was Hans Karlgren's extraordinary performance as program chair of COLING 90 in Helsinki, where he read and commented on virtually every paper himself.

b. The fully funded style US meeting with a professional organization (AAAI) or secretariat (ACL) and a paid committee meeting where all issues MUST be settled to a schedule. This is a high cost method, but deadlines are kept because any late sub-reviewer will have their papers reviewed at the meeting by some part of the Committee. This requires strong geographical proximity in the Committee which COLING does not have, even if it had funds.

c. The conventional low-cost format that we have operated at this conference, with no committee meeting, and mailing paper backwards and forwards across the world. There can be fully electronic forms of that process, of course, but many COLING papers still come from email-poor areas and groups which would penalize exactly the people we have always favored, in some sense.

d. The streamlined conference arrangement that I see more frequently, particularly coming from Japan, and now being used elsewhere: rapid off-the-cuff comments and decisions from committee members by email and with little in the way of formal requirements, all done in full knowledge that a huge favor is beinq asked of busy reviewers, for no reward whatever, and with an assumption that it probably does not make make much difference to the outcome, which is pretty robust across different reviewing procedures.

As claims increase on the time of all of us, we will move, I am sure, towards form (d) which is also, paradoxically, close to the original COLING form (a). The ICCL might want to move towards the "full accountability model" (b) as used by most American professional bodies. It is high-cost, efficient, and can probably only work well in a tight-knit national community. Moreover, in many cases in can become bogged down in issues of political correctness that the international community usually finds less pressing that real political issues, on which COLING has had an excellent record: e.g. defending and supporting CL in Eastern Europe during very difficult periods.

COLING remains the most amateurish of such meetings, without budget or treasurer, and without fixed rules or a membership. It is not, in a strict sence, accountable to anyone and, however much that may be unacceptable to some, it does tend to produce nicer conferences. These reflections, and their relationship to accountablity, have been provoked by the changes over the decade I referred to above. While doing this job in 1994, I have received an enormous amount of email, demanding responses on everything from paper receipt to margin sizes to hotel reservations to serious complaints about my "defective consciousness of public duty" as one correspondent put it. He was one of two I know of who chose to go straight to the Internet with these complaints in very strong terms indeed, though without any detailed knowledge of what was happening in the review system and why.

I am sure all this criticism was richly deserved, but the danger I see, beyond the purely personal, is that it will become very hard to recruit colleagues to serve in any of these roles (in journals as well as conferences) for no reward, and certainly outside the fully-supported conference of type (b), if the density and unpleasantness of much of this email were to continue. The very immediacy of the email medium, just banging out a message world-wide or to individuals to relieve one's feelings and without thought or knowledge, encourages this situation.

The source of much of it, and it comes, shall we say, from the more litigious CL zones of the world in general, is a belief that an author's relationship to a program chair or committee member is one of client-to-provider, to one who is accountable, for a fee, and can be sued. Only that belief could explain the vehemence, the stridency, the demandingness for accountablity of decisions, that I have recently encountered. Such beliefs are, of course, very much culture specific, and represent a transfer of the attitudes associated with the fully-funded responsible-to-a-membership professional meetings in the US to the quite different situation of COLING.

The ICCL will probably have to ask itself, at some stage, if it wants to follow model (b), though without a membership that would be hard to fund, or perhaps to state a firm position that COLING is not a conference of that sort and that older attitudes, of a sort that used to prevail among academic colleagues, where one was careful what one said because one's roles would be reversed next time, may continue to prevail. None of that would be incompatible with greater efficiency in the organization of the program, and perhaps with a fully automated form at the next COLING if it is thought that the Internet has now penetrated all parts of the known CL universe. Personally, I am sure COLINGs will continue to be a little different from other meetings, to the benefit of all of us and to the field we represent.

Yorick Wilks
Program Chairman

The ICCL are grateful to the following who participated in the review process

Appelt, Douglas
Armstrong, Susan
Ballim, Afzal
Barnett, Jim
Bennett, Paul
Bilange, Eric
Black, Bill
Boguraev, Bran
Broglio, John
Callan, Jamie
Calzolari, Nicolleta
Carberry, Sandra
Catizone, Roberta
Collier, Rob
Cowie, Jim
Dale, Judy
DiMarco, Chrysanne
Dohsaka, Koji
Dunning, Ted
Ehara, Terumasa
Estival, Dominique
Evens, Matha
Faewell, David
Finch, Steve
Furuse, Osamu
Gaizauskas, Rob
Guthrie, Louise
Hashida, Koichi
Helmreich, Steve
Hepple, Mark
Higuchi, Norio
Hirschberg, Julia
Hirschman, Lynette
Hirst, Graeme
Hovy, Eduard
Iida, Hitoshi
Ikehara, Satoru
Isahara, Hitoshi
Ishizaki, Syun
Jacob, Paul
Johnson, Ian
Kameyama, Megumi
Katoh, Naoto
Kawamori, Masahiro
Kikuchi, Jun'ichiro
King, Maggie
Klavans, Judith
Klavans, Judith
Kogure, Kiyoshi
Krovetz, Robert
Kyung-ho, Loken-Kim
Lascarides, Alex
Leech, Geoffrey
Lehnert, Wendy
Litman, Diane
Lochbaum, Karen
Luperfoy, Susann
Matsumoto, Yuji
McCarthy, Joe
McCoy, Kathy
McKevitt, Paul
McRoy, Susan
Mercer, Robert
Moore, Johanna
Morimoto, Tsuyoshi
Muraki, Kazunori
Naito, Shozo
Nakamura, Jun'ichi
Nakano, Mikio
Nomura, Hirosato
Oberlander, Jon
Ogden, Bill
Passoneau, Rebecca
Peterson, Jon
Popowich, Fred
Rensik, Philip
Ritchie, Graeme
Roesner, Dietmar
Ronser, Mike
Russell, Graham
Sabah, Gerard
Scott, Donia
Scott, Robert
Shimazu, Akira
Sidner, Candy
Siskind, Jeffrey
Smedt, de Koenraad
Somers, Harold
Stede, Manfred
Stevenson, Suzanne
Sugimura, Ryoichi
Sumita, Eiichiro
Takezawa, Toshiyuki
Tanaka, Hideki
Tanaka, Hozumi
Thompson, Henry
Tokunaga, Takenobu
Tombes, Louis des
Tsujii, Jun-ichi
Uratani, Noriyoshi
Wachtel, Tom
Wakao, Takahiro
Walker, Marilyn
Ward, Gregory
Warwick, Susan
Webber, Bonnie Lynn
Wilks, Yorick
Williams, Sheila
Yeun-Bae, Kim
Yusoff, Zaharin
Zaenen, Annie
Zampolli, Antonio
Zock, Michael

Conference Committee

Conference Chairman:

Prof. Makoto Nagao, Dept. of Electrical Engineering, Kyoto University , JAPAN

Program Chairman:

Prof. Yorick Wilks, University of Sheffield, ENGLAND

The International Committee on Computational Linguistics (ICCL):

C. Boitet (France), N. Calzolari (Italy), E. Hajicova (Chezco), B. Harris (Canada), D. Hays (Honorary, USA), K. Heggstadt (Norway), H. Karlgren (Sweden), M. Kay (President, USA), O. Kulagina (Russia), W. Lenders (Germany), M. Nagao (Japan), H. Schnelle (Germany), P. Sgall (Chezco), J. Tsujii (England), H. Wada (Honorary, Japan), Y. Wilks (England), A. Zampolli (Italy)

Cooperation:

The Institute of Electronics Information and Communication Engineers
Information Processing Society of Japan
Japan Society for Artificial Intelligence
Japan Society for Software Science and Technology
Japan Information Processing Development Center
Japan Electronic Industry Development Association
Asia-Pacific Association for Machine Translation

Donators:

Support Center for Advanced Telecommunications Technology Research
The Telecommunications Advancement Foundation
International Media Research Foundation
International Communication Foundation
International Information Science Foundation
Tateishi Science and Technology Foundation

Asia-Pacific Association for Machine Translation, Brother Industries, Ltd., Center for International Cooperation for Computarization, CSK Corporation, Fujitsu Ltd., Fuji Zerox Co., Ltd., Hitachi, Ltd., IBM Japan, Ltd., Japan Electronic Industry Development Association, Matsushita Electric Industrial Co., Ltd., Mitsubishi Electric Corp., NEC Corp., NHK, Nichigai Associates Inc., NOVA, Inc., NTT, NTT Data Communication Systems Corp., Oki Electric Industry Co., Ltd., Ricoh Co., Ltd., Sanyo Electric Co., Ltd., Sharp Corp., Sony Corp., The Toin Corp., Toshiba Corp.

Organizing Committee:

Nagao, Makoto (Chair, Kyoto University),Sato, Shigeru (Fujitsu Laboratories, Ltd.) Tanaka, Hozumi (Tokyo Institute of Technology), Habara, Kohei (Advanced Telecommunications Research Insutitute International), Yoshida, Sho (Kyushu Insutitute of Design), Aizawa, Teruaki (NHK Science Technology Research Laboratory), Asai, Shojiro (Hitachi, Ltd.), Iinuma, Kazumoto (NEC Corporation), Ikuno, Tatsuhiko (Sharp Corporation), Ishizaki, Shun (Keio University), Ishiwata, Toshio (Nihon Wemen's University), Otsu, Nobuyuki (Electrotechnical Laboratory), Okada, Naoyuki (Kyushu Insutitute of Technology), Kawaoka, Tsukasa (NTT Communication Science Laboratories), Kusanagi, Yutaka (Tsukuba University), Kushiki, Yoshiaki (Matsushita Electric Industrial Co.,Ltd.), Kurematsu, Akira (University of Electro-communications), Kuwano, Yukinori (Sanyo Electric Co.,Ltd.), Shirai, Katsuhiko (Waseda University), Suzuki, Norihisa (IBM Japan), Soga, Masakazu (Mitsubishi Electric Corporation), Nakano, Hiroshi (National Language Research), Narita, Hajime (Osaka University), Nishioka, Kazuhiko (CSK Corporation), Nose, Isamu (Oki Electric Industry Co., Ltd.), Nomura, Hiorosato (Kyushu Insutitute of Technology), Hirose, Keikichi (Tokyo University), Fuchi, Kazuhiro (The University of Tokyo), Makino, Takenori (Toho University), Amano, Shinya (Toshiba Corporation), Yokoi, Toshio (Japan Electronic Dictionary Research Institute, Ltd.), Yoneyama, Masahide (Richo Company., Ltd.)

Steering Committee

Tanaka, Hozumi (Chair, Tokyo Institute of Technology), Iida, Hitoshi (ATR Interpreting Telecommunications Research Laboratories), Ikehara, Satoru (NTT Network Information Systems Laboratories), Uchida, Yuji (Fujitsu Laboratories Ltd.) Ogino, Tsunao (University of Tsukuba), Kawada, Tsutomu (Toshiba Coraporation), Sakamoto, Masashi (Oki Electric Industry Co.,Ltd.), Shimazu, Akira (NTT Basic Research Laboratories), Sugimura, Ryoichi (Matsushita Electric Industrial Co.,Ltd.), Suzuki, Katsushi (Mitsubishi Electric Corp.), Suzuki, Hitoshi (Sharp Corporation), Tsutsumi, Taijiro (IBM Japan), Nishida, Yukiteru (Sanyo Electric Co., Ltd.), Nitta, Yoshihiko (Hitachi, Ltd.), Hashida, Koichi (Electrotechnical Laboratory), Matsumoto, Yuji (Nara Institute of Science and Technology), Muraki, Kazunori (NEC Corporation), Yamauchi, Satoshi (Richo Co., Ltd.), Yokoyama, Shoichi (Yamagata University), Nakamura, Junichi (Kyushu Institute of Technology), Tokunaga, Takenobu (Tokyo Institute of Technology)

Table of Contents (Vol. I)

Machine Translation

Morphology & Tagging

Lexicon

Generation

Parsing

Computational Linguistics

Corpus-based NLP

Table of Contents (Vol. II)

Semantics

Syntax

Speech & Phonology

Information Retrieval & Extraction

Discourse & Pragmatics

Reserve Paper

Author's Index

Author's Index

Author's Index

Author's Index

Author's Index