



American Journal of Computational Linguistics

Volume 9, Numbers 3-4

July-December 1983

SPECIAL ISSUE ON ILL-FORMED INPUT

Recovery Strategies for Parsing Exagrammatical Language	123
Jaime G. Carbonell and Philip J. Hayes	
Parse Fitting and Prose Fixing: Getting a Hold on Ill-Formedness	147
K. Jensen, G.E. Heidorn, L.A. Miller, and Y. Ravin	
Meta-rules as a Basis for Processing Ill-Formed Output	161
Ralph M. Weischedel and Norman K. Sondheimer	
Preference Semantics, Ill-Formedness, and Metaphor	178
Dan Fass and Yorick Wills	
The NOMAD System: Expectation-Based Detection and Correction of Errors during Understanding of Syntactically and Semantically Ill-Formed Text	188
Richard H. Granger	
<i>Letter to the Editor</i>	
Re Sparck Jones Re Ballard on the Need for Careful Description	197
Bruce Ballard	
<i>The FINITE STRING Newsletter</i>	
Site Report	199
Controlling Complex Systems of Linguistic Rules	
Announcements	202
Carl Engelman Memorial Fund; IMT Conference; CAI Society, Journal; Calls for Papers; ACL at NCC '84	
Abstracts of Current Literature	205
Thirty-six abstracts, including Proc. 21st ACL Conf.	
<i>Microfiche (AJCL Mf. 96)</i>	

AJCL

The American Journal of Computational Linguistics (ISSN 0362-613X) is published quarterly by the Association for Computational Linguistics.

James F. Allen **Editor**

Computer Science Department
University of Rochester
Rochester, NY 14627 USA
(716) 275-5288; ACL@ROCHESTER.ARPA

Donald E. Walker **Managing Editor**

Natural-Language and Knowledge-Resource Systems
SRI International
Menlo Park, CA 94025 USA
(415) 859-3071; WALKER@SRI-AI.ARPA

Michael C. McCord **Associate Editor**

IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598 USA
(914) 945-1308; MCCORD.YKTVMX.IBM@RAND-RELAY.ARPA

Madeleine Bates **Book Review Editor**

Bolt Beranek and Newman
10 Moulton Street
Cambridge, MA 02238 USA
(617) 497-3634; BATES@BBNG.ARPA

Ralph M. Weischedel **The FINITE STRING Editor**

Computer and Information Science
University of Delaware
Newark, DE 19716 USA
(302) 738-2712; WEISCHEDEL@UDEL-RELAY.ARPA

Cay Dietrich **Production Editor**

IBM Thomas J. Watson Research Center
(914) 945-3995; AJCL.YKTVMV.IBM@RAND-RELAY.ARPA

Editorial Board Members

1981-83:

Philip R. Cohen, *Fairchild Laboratory for AI Research*

Barbara J. Grosz, *SRI International*

Karen Sparck Jones, *University of Cambridge*

Jonathan Slocum, *Siemens Corporation*

Ralph M. Weischedel, *University of Delaware*

1982-84:

Jaime G. Carbonell, *Carnegie-Mellon University*

Daniel L. Chester, *University of Delaware*

Martha W. Evens, *Illinois Institute of Technology*

Ronald M. Kaplan, *Xerox Palo Alto Research Center*

S. Jerrold Kaplan, *Teknowledge*

1983-85:

Madeleine Bates, *Bolt Beranek and Newman*

Alan Biermann, *Duke University*

Beatrice T. Oshika, *System Development Corporation*

D. Terrence Langendoen, *CUNY Graduate Center*

Lenhart K. Schubert, *University of Alberta*

Copyright © 1984 by the Association for Computational Linguistics. The code and copyright notice appearing at the bottom of the first page of an article in this journal indicate the copyright owner's consent that a single copy of the article may be made for personal use. Any copier must pay for copying beyond that permitted by Sections 107 and 108 of the U.S. Copyright Law. The per-copy fee for each article appears after the dollar sign and is to be paid through the Copyright Clearance Center, Inc. This consent does not apply to other kinds of copying, such as copying for general distribution, for advertising and promotional purposes, for creating new collective works, or for resale. Such permission requests and other permission inquiries should be addressed to the *Managing Editor* at the address given above.

Subscriptions to the AJCL are available only through membership in the ACL. Each subscription includes both hard-copy and microfiche versions of the Journal.

Back issues of the AJCL may be ordered at annual rates of \$15 for orders from personal members, \$30 from institutional members. The first five volumes (1974-1978) are available only on microfiche; add \$3 for first class mailing in North America; \$6 for air printed mailing overseas. Volumes 6 - 8 (1980 - 1982) are available in both hard-copy and microfiche form; add \$6 for first class, \$12 for air printed mail. The AJCL was not published during 1979.

ACL

The Association for Computational Linguistics is an international scientific and professional society with membership open to any person or institution whose professional goals relate to computational linguistics.

C. Raymond Perrault **President (1983)**

Artificial Intelligence Center
SRI International
Menlo Park, CA 94025 USA
(415) 859-6470; RPERRAULT@SRI-AI.ARPA

Martha W. Evens **Vice President (1983)**

Department of Computer Science
Illinois Institute of Technology
10 West 31st Street
Chicago, IL 60616 USA
(312) 567-5153; EVENS@SRI-AI.ARPA

Donald E. Walker **Secretary-Treasurer**

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025 USA
(415) 859-3071; WALKER@SRI-AI.ARPA

Executive Committee Members

William C. Mann, *USC/Information Sciences Institute* (1981-83)

Lauri Karttunen, *SRI International* (1982-84)

Ralph M. Weischedel, *University of Delaware* (1983-85)

Nominating Committee

Bonnie L. Webber, *University of Pennsylvania* (1981-83)

Norman K. Sondheimer, *USC/Information Sciences Institute* (1982-84)

Jane J. Robinson, *SRI International* (1983-85)

ACL EUROPEAN CHAPTER OFFICIALS

Eva Hajicova **Chair (1982-83)**

Charles University - Linguistics
Malostranske nam. 25
CS-118 00 Praha 1, Czechoslovakia
532 136

Harold L. Somers **Secretary (1982-83)**

Centre for Computational Linguistics
University of Manchester, PO Box 88
Manchester M60 1QD ENGLAND
(061) 236-3311, ext. 2296

Michael Rosner **Treasurer (1982-83)**

ISSCO
54 route des Acacias
CH-1227 Geneva, SWITZERLAND

Advisory Committee Members (1982-83)

Hubert Lehmann, *IBM Heidelberg*

Remko J. Scha, *Eindhoven, Netherlands*

Yorick A. Wilks, *University of Essex*

Nominating Committee (1982-83)

Giacomo Ferrari, *Istituto di Linguistica Computazionale*

Gerald Gazdar, *University of Sussex*

Peter Hellwig, *University of Heidelberg*

Bente Maegaard, *University of Copenhagen*

Annual dues are \$15 for personal and \$30 for institutional memberships. A joint membership for couples that provides one subscription is available for \$20. The dues payment includes bulk or surface mailing fees. First class mailing is available in North America for an additional \$6; air printed mailing is available overseas for an additional \$12.

Payments for dues or back issues should be made payable to the Association for Computational Linguistics (or ACL). Payments sent to the ACL Secretary-Treasurer should be payable in US dollars. Payments from people in Europe or with Swiss accounts should be sent to the European Chapter Treasurer; they can be paid in Swiss francs, by personal checks in the currency of that country, or by a banker's draft that credits account 141.880.LAV at the Union Bank of Switzerland, 8 rue de Rhone, CH-1211 Geneva 11, SWITZERLAND. European residents can get information on currency conversion values from the European Chapter Secretary.

Address changes should be sent to the ACL Secretary-Treasurer.

Notes from the Editor:

Change in journal name. With the first issue of Volume 10, which will be at the printers by the time you receive this issue, the journal name will be changed to COMPUTATIONAL LINGUISTICS to reflect better its international nature. In the next few years we look forward to the increased participation and representation of the global computational linguistics community in ACL affairs.

Special issues. This is the first in a series of special issues that will appear in the journal of the Association for Computational Linguistics in the upcoming years. Each special issue will deal with an important issue in depth, and will attempt to present a representative survey of all approaches to the problem. This issue was produced by actively soliciting papers from the researchers we knew to be working in the area, and by a call for papers published in various AI newsletters to notify those we inadvertently overlooked. All papers were reviewed through the regular journal reviewing process and the normal acceptance standards were applied. This will be the procedure for future issues as well.

In general, there will be a guest editor for each special issue. This issue, being the first, was edited by me in order that I might lay the groundwork for future issues. The next special issue, under the guest editorship of Ray Perrault at SRI, will appear in Volume 10 and will concern the complexity of grammatical formalisms. Any suggestions for further special issues would be appreciated and should be sent to me.

About this issue. One of the crucial problems facing natural language systems is the handling of ill-formed input. While significant progress has been made in the processing of correct text, a practical system must be able to handle many forms of ill-formedness gracefully. This includes lexical problems such as misspelled words, sentential problems such as missing words or phrases and bad word order, semantic problems such as anomalous or self-contradictory sentences, and contextual problems such as incoherent requests or continuations. Of course, often when a problem arises, it appears in many of the above classes simultaneously, since a problem at one level can easily cause a problem at another. This issue contains papers that address ill-formedness at a particular level, such as in syntactic structure, as well as papers that address ill-formedness across many levels.

continued overleaf

Two of these papers offer surveys of the approaches to ill-formedness, namely the Carbonell and Hayes paper and the Weischedel and Sondheimer paper. These give the reader an idea of the range of problems to be tackled and a rough breakdown of the various approaches. The Carbonell and Hayes paper offers the most comprehensive survey and describes approaches usable in semantically restricted domains.

The Jensen et al. and the Weischedel and Sondheimer papers deal mainly with ill-formedness at the syntactic level. Both suggest methods of relaxing the syntactic constraints imposed by the grammar in the face of ill-formedness. Jensen et al. propose a framework that produces an approximate parse using the grammatical rules for correct sentences, while Weischedel and Sondheimer propose a system of meta-rules that are used to diagnose problems and relax certain parsing constraints in order to produce an analysis.

The two remaining papers deal more with semantic and contextual ill-formedness. Fass and Wilks propose methods for choosing the "best" semantic interpretation in the presence of apparent semantic ill-formedness and relate that work to the problem of understanding metaphors. Granger proposes a range of techniques, emphasizing the use of semantic and contextual methods for dealing with ill-formed text in domains that have well defined contextual constraints.

Taken together, these five papers offer a comprehensive view of the state of the art in dealing with ill-formed input, a problem that is of increasing importance as practical systems are developed.

James F. Allen, Editor