Abstracts of Current Literature

The following abstracts are from the papers selected for presentation at Coling84 – the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the Association for Computational Linguistics – to be held 2-6 July 1984 at Stanford University, California. *Proceedings of Coling84,* $30 a copy, is available from

Donald E. Walker, ACL
Bell Communications Research
445 South Street
Morristown, NJ 07960

Multilingual Text Processing in a Two-Byte Code

Lloyd B. Anderson
Ecological Linguistics
316 A Street, S.E.
Washington, DC 20003
Coling84, pp. 1-4

National and international standards committees are now discussing a two-byte code for multilingual information processing. This provides for 65,536 separate character and control codes, enough to make permanent code assignments for all the characters of all national alphabets of the world, and also to include Chinese/Japanese characters.

This paper discusses the kinds of flexibility required to handle both Roman and non-Roman alphabets. It is crucial to separate information units (codes) from graphic forms, to maximize processing power.

Comparing alphabets around the world, we find that the graphic devices (letters, digraphs, accent marks, punctuation, spacing, etc.) represent a very limited number of information units. It is possible to arrange alphabet codes to provide transliteration equivalence, the best of three solutions compared as a framework for code assignments.

Conveying Implicit Content in Narrative Summaries

Malcolm E. Cook, Wendy G. Lehnert,
David D. McDonald
Dept. of Computer and Information
Science
University of Massachusetts
Amherst, MA 01003
Coling84, pp. 5-7

One of the key characteristics of any summary is that it must be concise. To achieve this the content of the summary (1) must be focused on the key events, and (2) should leave out any information that the audience can infer on their own. We have recently begun a project on summarizing simple narrative stories. In our approach, we assume that the focus of the story has already been determined and is explicitly given in the story's long-term representation; we concentrate instead on how one can plan what inferences an audience will be able to make when they read a summary. Our conclusion is that one should think about inferences as following from the audience's recognition of the central concepts in the story's plot, and then plan the textual structure of the summary so as to reinforce that recognition.

Transforming English Interfaces to Other Natural Languages: An Experiment with Portuguese

Gabriel Pereira Lopes
Dept. de Matematica
Instituto Superior de Agronomia
Tapada da Ajuda
1399 Lisboa Codex, Portugal
Coling84, pp. 8-10

Nowadays it is common in the construction of English understanding systems (interfaces) that sooner or later one has to re-use them, adapting and converting them to other natural languages. This is not an easy task and in many cases the problems that arise are quite complex. In this paper an experiment that was accomplished for Portuguese language is reported, and some conclusions are explicitly stated. A knowledge information processing system with natural language comprehension capabilities, known as SSIPA, that interacts with users in Portuguese through a Portuguese interface, LUSO, was built. Logic was used as a mental aid and as a practical tool.

Un Outil Multidimensionnel de L'analyse du Discours

J. Chauche
·I.U.T. Le Havre
Place Robert Schuman
76610 Le Havre FRANCE
C.E.L.T.A.
23, Boulevard Albert 1er

Le traitement automatique du discourse suppose un traitement algorithmique et informatique. Plusieurs méthodes permettent d'appréhender cet aspect. L'utilisation d'un langage de programmation général (par exemple PL/I) ou plus orienté (par exemple LISP représente la première approche. A l'opposé, l'tuilisation d'un logiciel spécialisé permet d'éviter l'étude algorithmique nécessaire dans le premier cas et de concentrer cette étude sur les aspects réellement spécifiques de ce traitement. Les choix qui ont conduit à la définition du système sygmart sont exposés ici. L'aspect multi-

54000 Nancy FRANCE
Coling84, pp. 11-15

A Stochastic Approach to Sentence Parsing

Tetsunosuke Fujisaki
Science Institute, IBM Japan, Ltd.
No. 36 Kowa Building
5-19 Sanbancho, Chiyoda-ku
Tokyo 102, Japan
Coling84, pp. 16-19

Rounded Context Parsing and Easy Learnability

Robert C. Berwick
Room 820
MIT Artificial Intelligence Lab
Cambridge, MA 02139
Coling84, pp. 20-23

The Representation of Constituent Structures for Finite-State Parsing

D. Terence Langendoen,
Yedidyah Langsam
Depts. of English and Computer & Information Science
Brooklyn College of the City University of New York
Brooklyn, NY 11210
Coling84, pp. 24-27

Features and Values

Lauri Karttunen
University of Texas at Austin
Artificial Intelligence Center, SRI International
Center for the Study of Language and Information, Stanford University
Coling84, pp. 28-33

Applications of a Lexicographical Data Base for German

Wolfgang Teubert
Institut fur deutsche Sprache
Friedrich-Karl-Str. 12
6800 Mannheim 1, West Germany
Coling84, pp. 34-37

Denormalization and Cross Referencing in Theoretical Lexicography

Joseph E. Grimes
DMLL, Morrill Hall, Cornell University

dimensionnel est analysé du point de vue conceptuel et permet de situer cette réalisation par rapport aux différents systèmes existants.

A description will be given of a procedure to assign the most likely probabilities to each of the rules of a given context-free grammar. The grammar developed by S. Kuno at Harvard University was picked as the basis and was successfully augmented with rule probabilities. A brief exposition of the method with some preliminary results, when used as a device for disambiguating parsing English texts picked from natural corpus, will be given.

Natural languages are often assumed to be constrained so that they are either easily learnable or parsable, but few studies have investigated the connection between these two "functional" demands. Without a formal model of parsability or learnability, it is difficult to determine which is more "dominant" in fixing the properties of natural languages. In this paper we show that if we adopt one precise model of "easy" parsability, namely that of *bounded context parsability,* and a precise model of "easy" learnability, namely that of *degree 2 learnability,* then we can show that certain families of grammars that meet the bounded context parsability condition will also be degree 2 learnable. Some implications of this result for learning in other subsystems of linguistic knowledge are suggested.

A mixed prefix-postfix notation for representations of the constituent structures of the expressions of natural languages is proposed, which are of limited degree of center embedding if the original expressions are noncenter-embedding. The method of constructing these representations is applicable to expressions with center embedding, and results in representations which seem to reflect the ways in which people actually parse those expressions. Both the representations and their interpretations can be computed from the expressions from left to right by finite-state devices.

The paper discusses the linguistic aspects of a new general purpose facility for computing with features. The program was developed in connection with the course I taught at the University of Texas in the fall of 1983. It is a generalized and expanded version of a system that Stuart Shieber originally designed for the PATR-II project at SRI in the spring of 1983 with later modifications by Fernando Pereira and me. Like its predecessors, the new Texas version of the "DG (directed graph)" package is primarily intended for representing morphological and syntactic information, but it may turn out to be very useful for semantic representations too.

The Institut fur deutsche Sprache recently has begun setting up a LExicographical DAta Base for German (LEDA). This data base is designed to improve efficiency in the collection, analysis, ordering, and description of language material by facilitating access to textual samples within corpora and to word articles within machine readable dictionaries and by providing a frame to store results of lexicographical research for further processing. LEDA thus consists of the three components *Text Bank, Dictionary Bank,* and *Result Bank,* and serves as a tool to support monolingual German dictionary projects at the Institute and elsewhere.

A computational vehicle for lexicography was designed to keep to the constraints of meaning-text theory: sets of lexical correlates, limits on the form of definitions, and argument relations similar to lexical-functional grammar.

Ithaca, NY 14853
Summer Institute of Linguistics
7500 West Camp Wisdom Road
Dallas, TX 75236
Coling84, pp. 38-41

Relational data bases look like a natural framework for this. But linguists operate with a non-normalized view. Mappings between semantic actants and grammatical relations do not fit actant fields uniquely. Lexical correlates and examples are polyvalent, hence denormalized.

Cross referencing routines help the lexicographer work toward a closure state in which every term of a definition traces back to zero level terms defined extralinguistically or circularly. Dummy entries produced from defining terms ensure no trace is overlooked. Values of lexical correlates lead to other word senses. Cross references for glosses produce an index unilingual dictionary, the start of a fully bilingual one.

To assist field work a small structured editor for a systematically denormalized data base was implemented in PTP under RT-11; Mumps would now be easier to implement on small machines. It allowed fields to be repeated and non-atomic strings included, and produced cross reference entries. It served for a monograph on a language of Mexico and for student projects from Africa and Asia.

Lexicon Features for Japanese Syntactic Analysis in Mu-Project-JE

Yoshiyuki Sakamoto
Electrotechnical Laboratory
Sakura-mura., Niihari-gun.
Ibaraki, Japan
Masayuki Satoh
The Japan Information Center of Science and Technology
Nagata-cho. Chiyoda-ku
Tokyo, Japan
Tetsuya Ishikawa
University of Library and Information Science
Yatabe-machi., Tsukuba-gun.
Ibaraki, Japan
Coling84, pp. 42-47

In this paper, we focus on the features of a lexicon for Japanese syntactic analysis in Japanese-to-English translation. Japanese word order is almost unrestricted and *Kokujo-shi* (postpositional case particle) is an important device which acts as the case label (case marker) in Japanese sentences. Therefore case grammar is the most effective grammar for Japanese syntactic analysis.

The case frame governed by *Yougen* and having surface case *(Kokujo-shi)*, deep case (case label) and semantic markers for nouns is analyzed here to illustrate how we apply case grammar to Japanese syntactic analysis in our system.

The parts of speech are classified into 56 sub-categories.

We analyze semantic features for nouns and pronouns classified into sub-categories and we present a system for semantic markers. Lexicon formats for syntactic and semantic features are composed of different features classified by part of speech.

As this system uses LISP as the programming language, the lexicons are written as S-expressions in LISP, punched onto tapes, and stored as files in the computer.

Toward a Redefinition of Yes/No Questions

Julia Hirschberg
Dept. of Computer and Information Science
Moore School/D2
University of Pennsylvania
Philadelphia, PA 19104
Coling84, pp. 48-51

While both theoretical and empirical studies of question-answering have revealed the inadequacy of traditional definitions of *yes-no questions* (YNQs), little progress has been made toward a more satisfactory redefinition. This paper reviews the limitations of several proposed revisions. It proposes a new definition of YNQs based upon research on a type of *conversational implicature,* termed here *scalar implicature,* that helps define appropriate responses to YNQs. By representing YNQs as *scalar* queries, it is possible to support a wider variety of system and user responses in a principled way.

The Syntax and Semantics of User-Defined Modifiers in a Transportable Natural Language Processor

Bruce W. Ballard
Dept. of Computer Science
Duke University
Durham, NC 27706
Coling84, pp. 52-56

The Layered Domain Class system (LDC) is an experimental natural language processor being developed at Duke University which reached the prototype stage in May of 1983. Its primary goals are (1) to provide English-language retrieval capabilities for structured but unnormalized data files created by the user; (2) to allow very complex semantics, in terms of the information directly available from the physical data file; and (3) to enable users to customize the system to operate with new types of data. In this paper we shall discuss (a) the types of modifiers LDC provides for; (b) how information about the syntax and semantics of modifiers is obtained from users; and (c) how this information is used to process English inputs.

Interaction of Knowledge Sources in a

This paper describes a general approach to the design of natural language

Portable Natural Language Interface

Carole D. Hafner
Computer Science Dept.
General Motors Research Laboratories
Warren, MI 48090
Coling84, pp. 57-60

interfaces that has evolved during the development of DATALOG, an English database query system based on Cascaded ATN grammar. By providing separate representation schemes for linguistic knowledge, general world knowledge, and application domain knowledge, DATALOG achieves a high degree of portability and extendibility.

Uses of C-Graphs in a Prototype for Automatic Translation

Marco A. Clemente-Salazar
Centro de Graduados e Investigacion
Instituto Technologico de Chihuahua
Av. Tecnologico No. 2909
31310 Chihuahua, Chih., MEXICO
Coling84, pp. 61-64

This paper presents a prototype, not completely operational, that is intended to use c-graphs in the translation of assemblers. Firstly, the formalization of the structure and its principal notions (substructures, classes of substructures, order, etc.) are presented. Next section describes the prototype which is based on a Transformational System as well as on a rewriting system of c-graphs which constitutes the nodes of the Transformational System. The following part discusses a set of operations on the structures. Finally, the implementation in its present state is shown.

Quasi-Indexical Reference in Propositional Semantic Networks

William J. Rapaport
Dept. of Philosophy, SUNY
Fredonia, NY 14063
Dept. of Computer Science, SUNY,
Buffalo, NY 14260
Stuart C. Shapiro
Dept. of Computer Science, SUNY
Buffalo, NY 14260
Coling84, pp. 65-70

We discuss how a deductive question-answering system can represent the beliefs or other cognitive states of users, of other interacting systems, and of itself. In particular, we examine the representation of first-person beliefs of others (e.g., the *system's* representation of a *user's* belief that he himself is rich). Such beliefs have as an essential component "quasi-indexical pronouns" (e.g., 'he himself'), and, hence, require for their analysis a method of representing these pronominal constructions and performing valid inferences with them. The theoretical justification for the approach to be discussed is the representation of nested *"de dicto"* beliefs (e.g., the system's belief that user-1 believes that system-2 believes that user-2 is rich). We discuss a computer implementation of these representations using the Semantic Network Processing System (SNePS) and an ATN parser-generator with a question-answering capability.

The Costs of Inheritance in Semantic Networks

Rob't F. Simmons
The University of Texas, Austin
Coling84, pp. 71-74

Questioning texts represented in semantic relations requires the recognition that synonyms, instances, and hyponyms may all satisfy a questioned term. A basic procedure for accomplishing such loose matching using inheritance from a taxonomic organization of the dictionary is defined in analogy with the unification algorithm used for theorem proving, and the costs of its application are analyzed. It is concluded that inheritance logic can profitably be included in the basic questioning procedure.

Functional Unification Grammar: A Formalism for Machine Translation

Martin Kay
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
and CSLI, Stanford
Coling84, pp. 75-78

Functional Unification Grammar provides an opportunity to encompass within one formalism and computational system the parts of machine translation systems that have usually been treated separately, notably analysis, transfer, and synthesis. Many of the advantages of this formalism come from the fact that it is monotonic allowing data structures to grow differently as different nondeterministic alternatives in a computation are system is that it is fundamental reversible, allowing a to translate as b only if b could translate as a.

Computer Simulation of Spontaneous Speech Production

Bengt Sigurd
Dept. of Linguistics and Phonetics
Helgonbacken 12, S-223 62 Lund
SWEDEN
Coling84, pp. 79-83

This paper pinpoints some of the problems faced when a computer text production model (COMMENTATOR) is to produce spontaneous speech, in particular the problem of chunking the utterances in order to get natural prosodic units. The paper proposes a buffer model which allows the accumulation and delay of phonetic material until a chunk of the desired size has been built up. Several phonetic studies have suggested a similar temporary storage in order to explain intonation slopes, rhythmical patterns, speech errors and speech disorders. Small-scale simulations of the whole verbalization process from perception and thought to sounds, hesitation behaviour, pausing, speech errors, sound changes and speech disorders are presented.

Limited Domain Systems for Language Teaching
S. G. Pulman
Lingustics, EAS
University of East Anglia
Norwich NR4 7TJ, UK
Coling84, pp. 84-87

This abstract describes a natural language system which deals usefully with ungrammatical input and describes some actual and potential applications of it in computer-aided second language learning. However, this is not the only area in which the principles of the system might be used, and the aim in building it was simply to demonstrate the workability of the general mechanism, and provide a framework for assessing developments of it.

GIT: A General Transducer for Teaching Computational Linguistics
P. Shann, J.L. Cochard
Dalle Molle Institute for Semantic and Cognitive Studies
University of Geneva
Switzerland
Coling84, pp. 88-91

The GIT-system is a tree-to-tree transducer developed for teaching purposes in machine translation. The transducer is a specialized production system giving the linguists the tools for expressing information in a syntax that is close to theoretical linguistics. Major emphasis was placed on developing a system that is user friendly, uniform and legible. This paper describes the linguistics data structure, the rule formalism and the control facilities that the linguist is provided with.

A Parsing Architecture Based on Distributed Memory Machines
Jon M. Slack
Dept. of Psychology
Open University
Milton Keynes MK7 6AA
ENGLAND
Coling84, pp. 92-95

The paper begins by defining a class of distributed memory machines which have useful properties as retrieval and filtering devices. These memory mechanisms store large numbers of associations on a single composite vector. They provide a natural format for encoding the syntactic and semantic constraints associated with linguistic elements. A computational architecture for parsing natural language is proposed which utilises the retrieval and associative features of these devices. The parsing mechanism is based on the principles of Lexical Functional Grammar and the paper demonstrates how these principles can be derived from the properties of the memory mechanisms.

Automated Determination of Sublanguage Syntactic Usage
Ralph Grishman, Ngo Thanh Nhan
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012
Elaine Marsh
Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory
Washington, DC 20375
Lynette Hirschman
Research and Development Division
System Development Corporation/A Burroughs Company
Paoli, PA 19301
Coling84, pp. 96-100

Sublanguages differ from each other, and from the "standard language", in their syntactic, semantic, and discourse properties. Understanding these differences is important if we are to improve our ability to process these sublanguages. We have developed a semi-automatic procedure for identifying sublanguage syntactic usage from a sample of text in the sublanguage. We describe the results of applying this procedure to three text samples: two sets of medical documents and a set of equipment failure messages.

Semantic Interpretation Using KL-ONE
Norman K. Sondheimer
USC/Information Sciences Institute
Marina del Rey, CA 90292
Ralph M. Weischedel
Dept. of Computer & Information Sciences
University of Delaware
Newark, DE 19716
Robert J. Bobrow
Bolt Beranek and Newman Inc.
Cambridge, MA 02238
Coling84, pp. 101-107

This paper presents extensions to the work of Bobrow and Webber on semantic interpretation using KL-ONE to represent knowledge. The approach is based on an extended case frame formalism applicable to all types of phrases, not just clauses. The frames are used to recognize semantically acceptable phrases, identify their structure, and relate them to their meaning representation through translation rules. Approaches are presented for generating KL-ONE structures as the meaning of a sentence, for capturing semantic generalizations through abstract case frames, and for handling pronouns and relative clauses.

Two Theories for Computing the Logical Form of Mass Expressions

Francis Jeffry Pelletier,
Lenhard K. Schubert
Dept. of Computing Science
University of Alberta
Edmonton, Alberta T6G 2E1 Canada
Coling84, pp. 108-111

There are various difficulties in accommodating the traditional mass/count distinction into a grammar for English which has as a goal the production of "logical form" semantic translations of the initial English sentences. The present paper surveys some of these difficulties. One puzzle is whether the distinction is a syntactic one or a semantic one, i.e., whether it is a well-formedness constraint or whether it is a description of the semantic translations produced. Another puzzle is whether it should be applied to simple words (as they occur in the lexicon) or whether it should apply only to longer units (such as entire NPs). Of the wide variety of possible theories, only two seem to produce the required results (having to do with plausible inferences and intuitively satisfying semantic representations). These two theories are developed and compared.

Syntactic and Semantic Parsability

Geoffrey K. Pullum
Syntax Research Center
Cowell College, UCSC
Santa Cruz, CA 95064
and
Center for the Study of Language and
Information
Stanford, CA 94305
Coling84, pp. 112-122

This paper surveys some issues that arise in the study of the syntax and semantics of natural languages (NLs) and have potential relevance to the automatic recognition, parsing, and translation of NLs. An attempt is made to take into account the fact that parsing is scarcely ever thought about with reference to syntax alone; semantic ulterior motives always underlay the assignment of a syntactic structure to a sentence. First I consider the state of the art with respect to arguments about the language-theoretic complexity of NLs: whether NLs are regular sets, deterministic CFLs, CFLs, or whatever. While English still appears to be a CFL as far as I can tell, new arguments (some not yet published) appear to show for the first time that some languages are not CFLs. Next I consider the question of how semantic filtering affects the power of grammars. Then I turn to a brief consideration of some syntactic proposals that employ more or less modes extensions of the power of context-free grammars.

The Semantics of Grammar Formalisms Seen as Computer Languages

Fernando C.N. Pereira,
Stuart M. Shieber
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and
Information
Stanford University
Coling84, pp. 123-129

The design, implementation, and use of grammar formalisms for natural language have constituted a major branch of computational linguistics throughout its development. By viewing grammar formalisms as just a special case of computer languages, we can take advantage of the machinery of denotational semantics to provide a precise specification of their meaning. Using Dana Scott's domain theory, we elucidate the nature of the feature systems used in augmented phrase-structure grammar formalisms, in particular those of recent versions of generalized phrase structure grammar, lexical functional grammar and PATR-II, and provide a denotational semantics for a simple grammar formalism. We find that the mathematical structures developed for this purpose contain an operation of feature generalization, not available in those grammar formalisms, that can be used to give a partial account of the effect of coordination on syntactic features.

The Resolution of Quantificational Ambiguity in the Tendum System

Harry Bunt
Computational Linguistics Research
Unit
Dept. of Language and Literature
Tilburg University
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
Coling84, pp. 130-133

A method is described for handling the ambiguity and vagueness that is often found in quantifications – the semantically complex relations between nominal and verbal constituents. In natural language certain aspects of quantification are often left open; it is argued that the analysis of quantification in a model-theoretic framework should use semantic representations in which this may also be done. This paper shows a form for such a representation and how "ambiguous" representations are used in an elegant and efficient procedure for semantic analysis, incorporated in the TENDUM dialogue system.

Preventing False Inferences

Aravind Joshi, Bonnie Webber
Dept. of Computer and Information
Science

In cooperative man-machine interaction, it is taken as *necessary* that a system truthfully and informatively respond to a user's question. It is not, however, *sufficient.* In particular, if the system has reason to believe that its planned response might lead the user to draw an inference that it

Moore School/D2
University of Pennsylvania
Philadelphia PA 19104
Ralph M. Weischedel
Dept. of Computer & Information Sciences
University of Delaware
Newark, DE 19716
Coling84, pp. 134-138

knows to be false, then it must block it by modifying or adding to its response. The problem is that a system neither can nor should explore all conclusions a user might possibly draw: its reasoning must be constrained in some systematic and well-motivated way.

Such cooperative behavior has been investigated previously, and a modification of Grice's *Maxim of Quality* is proposed:

Grice's *Maxim of Quality*

Do not say what you believe to be false or for which you lack adequate evidence.

Joshi's *Revised Maxim of Quality*

If you, the speaker, plan to say anything which may imply for the hearer something that you believe to be false, then provide further information to block it.

This behavior was studied in the context of interpreting certain definite noun phrases. In this paper, we investigate this revised principle as applied to question answering. In particular the goals of the research described here are to:

1. characterize tractable cases in which the system as respondent (R) can anticipate the possibility of the user/questioner (Q) drawing false conclusions from its response and can hence alter or expand its response so as to prevent it happening;
2. develop a formal method for computing the projected inferences that Q may draw from a particular response, identifying those factors whose presence or absence catalyzes the inferences;
3. enable the system to generate modifications of its response that can defuse possible false inferences and that may provide additional useful information as well.

Problem Localization Strategies for Pragmatics Processing in Natural-Language Front Ends

Lance A. Ramshaw,
Ralph M. Weischedel
Dept. of Computer and Information
Science
University of Delaware
Newark, DE 19716
Coling84, pp. 139-143

Problem localization is the identification of the most significant failures in the AND-OR tree resulting from an unsuccessful attempt to achieve a goal, for instance, in planning, backward-chaining inference, or top-down parsing. We examine heuristics and strategies for problem localization in the context of using a planner to check for pragmatic failures in natural language input to computer systems, such as a cooperative natural language interface to Unix[1]. Our heuristics call for selecting the most hopeful branch at ORs, but the most problematic one at ANDs. Surprise scores and special-purpose rules are the main strategies suggested to determine this.

[1]Unix is a trademark of Bell Laboratories.

A Connectionist Model of Some Aspects of Anaphor Resolution

Ronan G. Reilly
Education Research Centre
St Patrick's College, Drumcondra
Dublin 9, Ireland
Coling84, pp. 144-149

This paper describes some recent developments in language processing involving computational modes which more closely resemble the brain in both structure and function. These models employ a large number of interconnected parallel computational units which communicate via weighted levels of excitation and inhibition. A specific model is described which uses this approach to process some fragments of connected discourse.

Concurrent Parsing in Programmable Logic Array (PLA-) Nets: Problems and Proposals

Helmut Schnelle
RUHR-Universitat Bochum

This contribution attempts a conceptual and practical introduction into the principles of wiring or constructing special machines for language processing tasks instead of programming a universal machine. Construction would in principle provide higher descriptive adequacy in computationally based linguistics. After all, our heads do not apply programs on stored symbol

Sprachwissenschaftliches Institut
D-4630 Bochum 1, West Germany
Coling84, pp. 150-153

arrays but are appropriately wired for understanding or producing language.

A Case Analysis Method Cooperating with ATNG and Its Application to Machine Translation

Hitoshi Iida, Kentaro Ogura,
Hirosato Nomura
Musashino Electrical Communication
Laboratory, N.T.T.
Musashino-shi, Tokyo, 180, Japan
Coling84, pp. 154-158

This paper presents a new method for parsing English sentences. The LUTE-EJ parser is combined with case analysis and ATNG-based analysis. The LUTE-EJ parser has two interesting mechanical characteristics. One is providing a structured buffer, Structure Constituent Buffer, so as to hold previous fillers for a case structure, instead of case registers before a verb appears in a sentence. The other is extended HOLD mechanisms (in ATN), in whose use an embedded clause, especially a "be-deleted" clause, is recursively analyzed by case analysis. This parser's features are (1) extracting a case filler, basically as a noun phrase, by ATNG-based analysis, including recursive case analysis, and (2) mixing syntactic and semantic analysis by using case frames in case analysis.

A Proper Treatment of Syntax and Semantics in Machine Translation

Yoshihiko Nitta, Atsushi Okajima,
Hiroyuki Kaji, Youichi Hidano,
Koichiro Ishihara
Systems Development Laboratory
Hitachi, Ltd.
1099 Ohzenji Asao-Ku, Kawasaki-shi,
215 JAPAN
Coling84, pp. 159-166

A proper treatment of syntax and semantics in machine translation is introduced and discussed from the empirical viewpoint. For English-Japanese machine translation, the syntax directed approach is effective where the Heuristic Parsing Model (HPM) and the Syntactic Rule System play important roles. For Japanese-English translation, the semantics directed approach is powerful where the Conceptual Dependency Diagram (CDD) and the Augmented Case Marker System (which is a kind of Semantic Role System) play essential roles. Some examples of the difference between Japanese sentence structure and English sentence structure, which is vital to machine translation, are also discussed together with various interesting ambiguities.

A Consideration on the Concepts Structure and Language in Relation to Selections of Translation Equivalents of Verbs in Machine Translation Systems

Sho Yoshida
Dept. of Electronics
Kyushu University 36
Fukuoka 812, Japan
Coling84, pp. 167-169

To give appropriate translation equivalents for target words is one of the most fundamental problems in machine translation systems, especially when the MT systems handle languages that have completely different structures like Japanese and European languages as source and target languages. In this report, we discuss the data structure that enables appropriate selections of translation equivalents for verbs in the target language. This structure is based on the concepts structure with associated information relating source and target languages. Discussions have been from the standpoint of realizability of the structure (e.g., from the standpoint of ease of data collection and arrangement, ease of realization and compactness of the size of storage space).

Detecting Patterns in a Lexical Data Base

Nicoletta Calzolari
Dipt. di Linguistics, Universita di Pisa
Istituto di Linguistica Computazionale del
CNR
Via della Faggiola 32
56100 Pisa - Italy
Coling84, pp. 170-173

In a well-structured Lexical Data Base, a number of relations among lexical entries can be interactively evidenced. The present article examines hyponymy, as an example of paradigmatic relation, and "restriction" relation, as a syntagmatic relation. The theoretical results of their implementation are illustrated.

Linguistic Problems in Multilingual Morphological Decomposition

G. Thurmair
Siemens AG, ZT ZTI
Otto-Hahn-Ring 6
Munich 83, West Germany
Coling84, pp. 174-177

An algorithm for the morphological decomposition of words into morphemes is presented. The application area is information retrieval, and the purpose is to find morphologically related terms to a given search term. First, the parsing framework is presented, then several linguistic decisions are discussed: morpheme selection and segmentation, morpheme classes, morpheme grammar, allomorph handling, etc. Since the system works in several languages, language-specific phenomena are mentioned.

A General Computational Model for Word-Form Recognition and Production

Kimmo Koskenniemi
Dept. of General Linguistics
University of Helsinki
Hallituskatu 11-13
Helsinki 10, Finland
Coling84, pp. 178-181

A language independent model for recognition and production of word forms is presented. This "two-level model" is based on a new way of describing morphological alternations. All rules describing the morpho-phonological variations are parallel and relatively independent of each other. Individual rules are implemented as finite state automata, as in an earlier model due to Martin Kay and Ron Kaplan. The two-level model has been implemented as an operational computer program in several places. A number of operational two-level descriptions have been written or are in progress (Finnish, English, Japanese, Rumanian, French, Swedish, Old Church Slavonic, Greek, Lappish, Arabic, Icelandic). The model is bidi-rectional and it is capable of both analyzing and synthesizing word-forms.

Panel: Natural Language and Data-bases, Again

Karen Sparck Jones
Computer Laboratory
University of Cambridge
Corn Exchange Street
Cambridge CB2 3QG, England
Coling84, pp. 182-183

Natural Language and Databases has been a common panel topic for some years, partly because it has been an active area of work, but more impor-tantly, because it has been widely assumed that database access is a good test environment for language research. I thought the time had come to look again at this assumption, and that it would be useful, for COLING 84, to do this. I therefore invited the members of the Panel to

speak to the proposition (developed below) that database query is no longer a good, let alone the best, test environment for language pro-cessing research, because it is insufficiently demanding in its linguistic aspects and too idiosyncratically demanding in its non-linguistic ones;

and to

propose better task environments for language understanding research, without the disadvantages of database query, but with its crucial advantage of an independent evaluation test.

There Still Is Gold in the Database Mine

Madeleine Bates
BBN Laboratories
10 Moulton Street
Cambridge, MA 02238
Coling84, pp. 184-185

Let me state clearly at the outset that I disagree with the premise that the problem of interfacing to database systems has outlived its usefulness as a productive environment for NL research. But I can take this stand strongly only by being very liberal in defining both "natural language interface" and "database systems".

Is There Natural Language after Data Bases?

Jaime G. Carbonell
Computer Science Dept.
Carnegie Mellon University
Pittsburgh, PA 15213
Coling84, pp. 186-187

The undisputed favorite application for natural language interfaces has been data base query. Why? The reasons range from the relative simplici-ty of the task, including shallow semantic processing, to the potential real-world utility of the resultant system. Because of such reasons, the data base query task was an excellent paradigmatic problem for computational linguistics, and for the very same reasons it is now time for the field to abandon its protective cocoon and progress beyond this rather limiting task. But, one may ask, what task shall then become the new paradigmatic problem? Alas, such a question presupposes that a single, universally acceptable, syntactically and semantically challenging task exists. I will argue that better progress can be made by diversification and focusing on different theoretically meaningful problems, with some research groups opting to investigate issues arising from the development of integrated multi-purpose systems.

PANEL: Natural Language and Data bases

Daniel P. Flickinger
Computer Research Center
Hewlett-Packard Company
1501 Page Mill Road
Palo Alto, CA 94304
Coling84, pp. 188-189

While I disagree with the proposition that database query has outlived its usefulness as a test environment for natural language processing (for reasons that I give below), I believe there are other reasonable tasks which can also spur new research in NL processing. In particular, I will suggest that the task of providing a natural language interface to a rich program-ming environment offers a convenient yet challenging extension of work already being done with database query.

Natural Language for Expert Systems: Comparisons with Database Systems
Kathleen R. McKeown
Dept. of Computer Science
Columbia University
New York, NY 10027
Coling84, pp. 190-193

Do natural language database systems still provide a valuable environment for further work on natural language processing? Are there other systems which provide the same hard environment for testing, but allow us to explore more interesting natural language questions? In order to answer *no* to the first question and *yes* to the second (the position taken by our panel's chair), there must be an interesting language problem which is more naturally studied in some other system than in the database system.

Representing Knowledge about Knowledge and Mutual Knowledge
Said Soulhi
Equipe de Comprehension du Raisonnement Naturel
LSI - UPS
118 route de Narbonne
31062 Toulouse - FRANCE
Coling84, pp. 194-199

In order to represent speech acts, in a multi-agent context, we choose a knowledge representation based on the modal logic of knowledge KT4 which is defined by Sato. Such a formalism allows us to reason about knowledge and represent knowledge about knowledge, the notions of truth value and of definite reference.

Understanding Pragmatically Ill-Formed Input
M. Sandra Carberry
Dept. of Computer Science
University of Delaware
Newark, DE 19716
Coling84, pp. 200-206

An utterance may be syntactically and semantically well-formed yet violate the pragmatic rules of the world model. This paper presents a context-based strategy for constructing a cooperative but limited response to pragmatically ill-formed queries. Suggestion heuristics use a context model of the speaker's task inferred from the preceding dialogue to propose revisions to the speaker's ill-formed query. Selection heuristics then evaluate these suggestions based upon semantic and relevance criteria.

Referring as Requesting
Philip R. Cohen
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and Information
Stanford University
Coling84, pp. 207-211

Searle has argued forcefully that referring is a speech act; that people refer, not just expressions. This paper considers what kind of speech act referring might be. I propose a generalization of Searle's "propositional" act of referring that treats it as an illocutionary act, a request, and argue that the propositional act of referring is unnecessary.

The essence of the argument is as follows: First, I consider Searle's definition of the propositional act of referring (which I term the PAA, for Propositional Act Account). This definition is found inadequate to deal with various utterances in discourse used for the sole purpose of referring. Although the relevance of such utterances to the propositional act has been defined away by Searle, it is clear that any comprehensive account of referring should treat them. I develop an account of their use in terms of a speaker's *requesting* the act of referent identification, which is to be understood in a perceptual sense. This *illocutionary act analysis* (IAA) is shown to satisfy Searle's conditions for referring yet captures utterances that the PAA cannot. The converse position is then examined: Can the IAA capture the same uses of referring expressions as the PAA? If one extends the perceptually-based notion of referent identification to include Searle's concept of identification, then by associating a complex propositional attitude to one use of the definite determiner, a request can be derived. The IAA thus handles the referring use of definite noun phrases with independently motivated rules. Referring becomes a kind of requesting. Hence, the propositional act of referring is unnecessary.

Entity-Oriented Parsing
Philip J. Hayes
Computer Science Dept.
Carnegie-Mellon University
Pittsburgh, PA 15213
Coling84, pp. 212-217

An entity-oriented approach to restricted domain parsing is proposed. In this approach the definitions of the structure and surface representation of domain entities are grouped together. Like semantic grammar, this allows easy exploitation of limited domain semantics. In addition, it facilitates fragmentary recognition and the uses of multiple parsing strategies, and so is particularly useful for robust recognition of extragrammatical input. Several advantages from the point of view of language definition are also noted. Representative samples from an entity-oriented language definition

are presented, along with a control structure for an entity-oriented parser, some parsing strategies that use the control structure, and worked examples of parses. A parser incorporating the control structure and the parsing strategies is currently under implementation.

Combining Functionality and Object-Orientedness for Natural Language Processing

Toyoaki Nishida, Shuji Doshita
Dept. of Information Science
Kyoto University
Sakyo-ku, Kyoto 606, Japan
Coling84, pp. 218-221

This paper proposes a method for organizing linguistic knowledge in both a systematic and a flexible fashion. We introduce a purely applicative language (PAL) as an intermediate representation and an object-oriented computation mechanism for its interpretation. PAL enables the establishment of a principled and well-constrained method of interaction among lexicon-oriented linguistic modules. The object-oriented computation mechanism provides a flexible means of abstracting modules and sharing common knowledge.

Use of Heuristic Knowledge in Chinese Language Analysis

*Yiming Yang, Yoyaki Nishida,
Shuji Doshita*
Dept. of Information Science
Kyoto University
Sakyo-ku, Kyoto, 606, Japan
Coling84, pp. 222-225

This paper describes an analysis method which uses heuristic knowledge to find local syntactic structures of Chinese sentences. We call it a preprocessing, because we use it before we do global syntactic structure analysis of the input sentence. Our purpose is to guide the global analysis through the search space, to avoid unnecessary computation.

To realize this, we use a set of special words that appear in commonly used patterns in Chinese. We call them "characteristic words". They enable us to pick out fragments that might figure in the syntactic structure of the sentence. Knowledge concerning the use of characteristic words enables us to rate alternative fragments, according to pattern statistics, fragment length, distance between characteristic words, and so on. The preprocessing system proposes to the global analysis level a most "likely" partial structure. In case this choice is rejected, backtracking looks for a second choice, and so on.

For our system, we use 200 characteristic words. Their rules are written by 101 automata. We tested them against 120 sentences taken from a Chinese physics text book. For this limited set, correct partial structures were proposed as first choice for 94% of the sentences. Allowing a second choice, the score is 98%; with a third choice, the score is 100%.

The Design of the Kernel Architecture for the Eurotra Software

R.L. Johnson
U.M.I.S.T.
P.O. Box 88
Manchester M60 1QD, U.K.
S. Krauwer
Rijksuniversiteit, Trans 14
3512 JK Utrecht, Holland
M.A. Rosner
ISSCO
University of Geneva
1211 Geneve 4, Switzerland
G.B. Varile
Commission of the European
Communities
P.O. Box 1907
Luxembourg
Coling84, pp. 226-235

Starting from the assumption that machine translation (MT) should be based on theoretically sound grounds, we argue that, given the state of the art, the only viable solution for the designer of software tools for MT is to provide the linguists building the MT system with a *generator* of highly specialized, problem oriented systems. We propose that such theory sensitive systems be generated automatically by supplying a set of definitions to a kernel software, of which we give an informal description in the paper. We give a formal functional definition of its architecture and briefly explain how a prototype system was built.

Machine Translation: What Type of Post-Editing on What Types of Documents for What Type of Users

Ann-Marie Laurian
Centre National de la Recherche

Various typologies of technical and scientific texts have already been proposed by authors involved in multilingual transfer problems. They were usually aimed at a better knowledge of the criteria for deciding if a document has to be or can be machine translated. Such a typology could also lead to a better knowledge of the typical errors occurring, and so lead to

Scientifique
Universite de la Sorbonne Nouvelle -
Paris III
19 rue des Bernardins
75005 Paris (France)
Coling84, pp. 236-238

more appropriate post-editing, as well as to improvements in the system.

Raw translations being usable, as they are quite often for rapid information needs, it is important to draw the limits between a style adequate for rapid information, and an elegant, high quality style such as required for information for large dissemination. Style could be given a new definition through a linguistic analysis based on machine translation, on communication situations, and on the users' requirements and satisfaction.

Simplifying Deterministic Parsing
Alan W. Carter
Dept. of Computer Science
University of British Columbia
Vancouver, B.C. V8T 1W5
Michael J. Freiling
Dept. of Computer Science
Oregon State University
Corvallis, OR 97331
Coling84, pp. 239-242

This paper presents a model for deterministic parsing which was designed to simplify the task of writing and understanding a deterministic grammar. While retaining structures and operations similar to those of Marcus's PARSIFAL parser, the grammar language incorporates the following changes. (1) The user of productions operating in parallel has essentially been eliminated and instead the productions are organized into sequences. Not only does this improve the understandability of the grammar, it is felt that this organization corresponds more closely to the task of performing the sequence of buffer transformations and attachments required to parse the most common constituent types. (2) A general method for interfacing between the parser and a semantic representation system is introduced. The interface is independent of the particular semantic representation used and hides all details of the semantic processing from the grammar writer. (3) The interface also provides a general method for dealing with syntactic ambiguities which arise from the attachment of optional modifiers such as propositional phrases. This frees the grammar writer from determining each point at which such ambiguities can occur.

Dealing with Conjunctions in a Machine Translation Environment
Xiuming Huang
Institute of Linguistics
Chinese Academy of Social Sciences
Beijing, China
Coling84, pp. 243-246

The paper presents an algorithm, written in PROLOG, for processing English sentences which contain either Gapping, Right Node Raising, or Reduced Conjunction. The DCG (Definite Clause Grammar) formalism (Pereiera & Warren 1980) is adopted. The algorithm is highly efficient and capable of processing a full range of coordinate constructions containing any number of coordinate conjunctions ('and', 'or', and 'but'). The algorithm is part of an English-Chinese machine translation system which is in the course of construction.

On Parsing References
Lenhart K. Schubert
Dept. of Computing Science
University of Alberta, Edmonton
Coling84, pp. 247-250

It is argued that syntactic preference principles such as *Right Association* and *Minimal Attachment* are unsatisfactory as usually formulated. Among the difficulties are: (1) dependence on ill-specified or implausible principles of parser operation; (2) dependence on questionable assumptions about syntax; (3) lack of provision, even in principle, for integration with semantic and pragmatic preference principles; and (4) apparent counterexamples, even when discounting (1)-(3). A possible approach to a solution is sketched.

A Computational Theory of the Function of Clue Words in Argument Understanding
Robin Cohen
Dept. of Computer Science
University of Toronto
Toronto, CANADA M5S 1A4
Coling84, pp. 251-258

This paper examines the use of clue words in argument dialogues. These are special words and phrases directly indicating the structure of the argument to the hearer. Two main conclusions are drawn: 1) clue words can occur in conjunction with coherent transmissions, to reduce processing of the hearer; 2) clue words must occur with more complex forms of transmission, to facilitate recognition of the argument structure. Interpretation rules to process clues are proposed. In addition, a relationship between use of clues and complexity of processing is suggested for the case of exceptional transmission strategies.

Control Structures and Theories of Interaction in Speech Understanding Systems
E.J. Briscoe, B.K. Boguraev

In this paper, we approach the problem of organisation and control in automatic speech understanding systems, firstly, by presenting a theory of the non-serial interactions necessary between two processors in the systems, namely, the morphosyntactic and the prosodic, and secondly, by

Computer Laboratory
University of Cambridge
Corn Exchange Street
Cambridge CB2 3QG, England
Coling84, pp. 259-266

showing how, when generalised, this theory allows one to specify a highly efficient architecture for a speech understanding system with a simple control structure and genuinely independent components. The theory of non-serial interactions we present predicts that speech is temporally organised in a very specific way; that is, the system would not function effectively if the temporal distribution of various types of information in speech were different. The architecture we propose is developed from a study of the task of speech understanding and, furthermore, is specific to this task. Consequently, the paper argues that general problem solving methods are unnecessary for speech understanding.

Analysis Grammar of Japanese in the Mu-Product — A Procedural Approach to Analytic Grammar

*Jun-ichi Tsujii, Jun-ichi Nakamura,
Makoto Nagao*
Dept. of Electrical Engineering
Kyoto University
Kyoto, Japan
Coling84, pp. 267-274

Analysis grammar of Japanese in the Mu-project is presented. It is emphasized that rules expressing constraints on single linguistic structures and rules for selecting the most preferable readings are completely different in nature, and that rules for selecting preferral readings should be utilized in analysis grammars of practical MT systems. It is also claimed that procedural control is essential in integrating such rules into a unified grammar. Some sample rules are given to make the points of discussion clear and concrete.

Lexicon-Grammar and the Syntactic Analysis of French

Maurice Gross
Laboratoire d'Automatique Documentaire
et Linguistique
University of Paris 7
2 place Jussieu
75251 Paris CEDEX 05 France
Coling84, pp. 275-282

A lexicon-grammar is constituted of the elementary sentences of a language. Instead of considering words as basic syntactic units to which grammatical information is attached, we use simple sentences (subject-verb-objects) as dictionary entries. Hence, a full dictionary item is a simple sentence with a description of the corresponding distributional and transformational properties.

The systematic study of French has led to an organization of its lexicon-grammar based on three main components:

- the lexicon-grammar of free sentences, that is, of sentences whose verb imposes selectional restrictions on its subject and complements (e.g., *to fall, to eat, to watch*);
- the lexicon-grammar of frozen or idiomatic expressions (e.g., *N takes N into account, N raises a question*);
- the lexicon-grammar of support verbs. These verbs do not have the common selectional restrictions, but more complex dependencies between subject and complement (e.g., *to have, to make* in *N has an impact on N, N makes a certain impression on N*).

These three components interact in specific ways. We present the structure of the lexicon-grammar built for French and we discuss its algorithmic implications for parsing.

Building a Large Knowledge Base for a Natural Language System

Jerry R. Hobbs
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and
Information
Stanford University
Coling84, pp. 283-286

A sophisticated natural language system requires a large knowledge base. A methodology is described for constructing one in a principled way. Facts are selected for the knowledge base by determining what facts are linguistically presupposed by a text in the domain of interest. The facts are sorted into clusters, and within each cluster they are organized according to their logical dependencies. Finally, the facts are encoded as predicate calculus axioms.

Linguistically Motivated Descriptive Term Selection

K. Sparck Jones and J.I. Tait
Computer Laboratory
University of Cambridge

A linguistically motivated approach to indexing, that is the provision of descriptive terms for texts of any kind, is presented and illustrated. The approach is designed to achieve good, i.e. accurate and flexible, indexing by identifying index term sources in the meaning representations built by a powerful general purpose analyser, and providing a range of text

Corn Exchange Street
Cambridge CB2 3QG, U.K.
Coling84, pp. 287-290

expressions constituting semantic and syntactic variants for each term concept. Indexing is seen as a legitimate form of shallow text processing, but one requiring serious semantically based language processing, particularly to obtain well-founded complex terms, which is the main objective of the project described. The type of indexing strategy described is further seen as having utility in a range of applications environments.

Inferencing on Linguistically Based Semantic Structures

Eva Hajicova, Milena Hnatkova
Dept. of Applied Mathematics
Faculty of Mathematics and Physics
Charles University
Malostranske n. 25
118 00 Praha 1, Czechoslovakia
Coling84, pp. 291-297

The paper characterizes natural language inferencing in the TIBAQ method of question-answering, focussing on three aspects: (i) specification of the structures on which the inference rules operate, (ii) classification of the rules that have been formulated and implemented up to now, according to the kind of modification of the input structure the rules invoke, and (iii) discussion of some points in which a properly designed inference procedure may help the search of the answer, and vice versa.

Semantic Relevance and Aspect Dependency in a Given Subject Domain

Burghard B. Rieger
Arbeitsgruppe fur mathematisch-
empirische Systemforschung (MESY)
German Dept.
Technical University of Aachen
Aachen, West Germany
Coling84, pp. 298-301

Cognitive principles underlying the (re-)construction of word meaning and/or world knowledge structures are poorly understood yet. In a rather sharp departure from more orthodox lines of introspective acquisition of structural data on meaning and knowledge representation in cognitive science, an empirical approach is explored that analyses natural language data statistically, represents its numerical findings fuzzy-set theoretically, and interprets its intermediate constructs (stereotype meaning points) topologically as elements of semantic space. As connotative meaning representations, these elements allow an aspect-controlled, contents-driven algorithm to operate which reorganizes them dynamically in dispositional dependency structures (DDS-trees) which constitute a procedurally defined meaning representation format.

A Plan Recognition Model for Clarification Subdialogues

Diane J. Litman, James F. Allen
Dept. of Computer Science
University of Rochester
Rochester, NY 14627
Coling84, pp. 302-311

One of the promising approaches to analyzing task-oriented dialogues has involved modeling the plans of the speakers in the task domain. In general, these models work well as long as the topic follows the task structure closely, but they have difficulty in accounting for clarification subdialogues and topic change. We have developed a model based on a hierarchy of plans and metaplans that accounts for the clarification subdialogues while maintaining the advantages of the plan-based approach.

A Computational Theory of Dispositions

Lotfi Z. Zadeh
Computer Science Division
University of California
Berkeley, CA 94720
Coling84, pp. 312-318

Informally, a *disposition* is a proposition which is preponderantly, but not necessarily always, true. For example, *birds can fly* is a disposition, as are the propositions *Swedes are blond* and *Spaniards are dark*.

An idea which underlies the theory described in this paper is that a disposition may be viewed as a proposition with implicit fuzzy quantifiers which are approximations to $\forall$ and *always*, e.g., *almost all*, *almost always*, *most, frequently*, etc. For example, *birds can fly* may be interpreted as the result of suppressing the fuzzy quantifier *most* in the proprosition *most birds can fly*. Similarly, *young men like young women* may be read as *most young men like mostly young women*. The process of transforming a disposition into a proposition is referred to as *explication* or *restoration*.

Explication sets the stage for representing the meaning of a proposition through the use of test-score semantics (Zadeh 1978, 1982). In this approach to semantics, the meaning of a proposition, *p*, is represented as a procedure which tests, scores and aggregates the elastic constraints which are induced by *p*.

The paper closes with a description of an approach to reasoning with dispositions which is based on the concept of a fuzzy syllogism. Syllogistic reasoning with dispositions has an important bearing on commonsense reasoning as well as on the management of uncertainty in expert systems.

As a simple application of the techniques described in this paper, we formulate a definition of *typicality* – a concept which plays an important role in human cognition and is of relevance to default reasoning.

Using Focus to Generate Complex and Simple Sentences

Marcia A. Derr
AT&T Bell Laboratories
Murray Hill, NJ 07974
and
Dept. of Computer Science
Columbia University
Kathleen R. McKeown
Dept. of Computer Science
Columbia University
New York, NY 10027
Coling84, pp. 319-326

One problem for the generation of natural language text is determining when to use a sequence of simple sentences and when a single complex one is more appropriate. In this paper, we show how focus of attention is one fact that influences this decision and describe its implementation in a system that generates explanations for a student advisor expert system. The implementation uses tests on functional information such as focus of attention within the Prolog definite clause grammar formalism to determine when to use complex sentences, resulting in an efficient generator that has the same benefits as a functional grammar system.

A Rational Reconstruction of the Proteus Sentence Planner

Graeme Ritchie
Dept. of Artificial Intelligence
University of Edinburgh
Hope Park Square
Edinburgh 3H8 9NW
Coling84, pp. 327-329

A revised and more structured version of Davey's discourse generation program has been implemented, which constructs the underlying forms for sentences and clauses by using rules which annotate and segment the initial sequence of events in various ways.

Software Tools for the Environment of a Computer Aided Translation System

Daniel Bachut
IFCI
INPG, 46, av. Felix-Viallet
38031 Grenoble Cedex, FRANCE
Nelson Verastegui
GETA
Universite de Grenoble
38402 Saint-Martin-d'Heres, France
Coling84, pp. 330-333

In this paper we will present three systems, ATLAS, THAM and VISULEX, which have been designed and implemented at GETA (Study Group for Machine Translation) in collaboration with IFCI (Institut de Formation et de Conseil en Informatique) as tools operating around the ARIANE-78 system. We will describe in turn the basic characteristics of each system, their possibilities, actual use, and performance.

Design of a Machine Translation System for a Sublanguage

Beat Buchmann, Susan Warwick,
Patrick Shann
Dalle Molle Institute for Semantic and
Cognitive Studies
University of Geneva
Switzerland
Coling84, pp. 334-337

This paper describes the design of a prototype machine translation system for a sublanguage of job advertisements. The design is based on the hypothesis that specialized linguistic subsystems may require special computational treatment and that therefore a relatively shallow analysis of the text may be sufficient for automatic translation of the sublanguage. This hypothesis and the desire to minimize computation in the transfer phase has led to the adoption of a flat tree representation of the linguistic data.

Grammar Writing System (GRADE) of Mu-Machine Translation Project and its Characteristics

Jun-ichi Nakamura, Jun-ichi Tsujii,
Makoto Nagao
Dept. of Electrical Engineering
Kyoto University
Sakyo, Kyoto, Japan
Coling84, pp. 338-343

A powerful grammar writing system has been developed. This grammar writing system is called GRADE (GRAmmar DEscriber). GRADE allows a grammar writer to write grammars including analysis, transfer, and generation using the same expression. GRADE has a powerful grammar writing facility. GRADE allows a grammar writer to control the process of a machine translation. GRADE also has a function to use grammatical rules written in a word dictionary. GRADE has been used for more than a year as the software of the machine translation project from Japanese into English, which is supported by the Japanese Government and called Mu-project.

A Discovery Procedure for Certain Phonological Rules
Mark Johnson
Linguistics
University of California
San Diego, CA
Coling84, pp. 344-347

Acquisition of phonological systems can be insightfully studied in terms of discovery procedures. This paper describes a discovery procedure, implemented in Lisp, capable of determining a set of ordered phonological rules, which may be in opaque contexts, from a set of surface forms arranged in paradigms.

What Not to Say
Jan Fornell
Dept. of Linguistics & Phonetics
Lund University
Helgonabacken 12, Lund, Sweden
Coling84, pp. 348-351

A problem with most text production and language generation systems is that they tend to become rather verbose. This may be due to neglection of the pragmatic factors involved in communication. In this paper, a text production system, COMMENTATOR, is described and taken as a starting point for a more general discussion of some problems in Computational Pragmatics. A new line of research is suggested, based on the concept of unification.

When Is the Next ALPAC Report Due?
Margaret King
Dalle Molle Institute for Semantic and Cognitive Studies
University of Geneva
Switzerland
Coling84, pp. 352-353

Machine translation has a somewhat checqured history. There were already proposals for automatic translation systems in the '30s, but it was not until after the second world war that real enthusiasm led to heavy funding and unrealistic expectations. Traditionally, the start of intensive work on machine translation is taken as being a memorandum of Warren Weaver, then Director of the Natural Sciences Division of the Rockefeller Foundation, in 1949. . . .

Weaver's memorandum led to a great deal of activity in research on machine translation, and eventually to the first conference on the topic, organised by Bar-Hillel in 1952. . . .

In 1964, the US National Academy of Sciences set up an investigatory committee, the Automatic Language Processing Advisory Committee (ALPAC), with the task of investigating the results of machine translation. . . . The committee came to a strong negative conclusion: " . . . we do not have useful machine translation. Further, there is no immediate or predictable prospect of useful machine translation."

The ALPAC report effectively killed machine translation research in the States, although some European projects survived. . . .

There are people who see strong parallels between the present situation and that immediately before the publication of the ALPAC report, foreseeing a second 'failure' for machine translation as a discipline. Others believe that advances in linguistics and in computer science, together with the results of the last twenty years, justify a cautious optimism, especially when the more realistic expectations of today's research workers (and of their funding authorities) are taken into account.

The panel discussion will aim at clarifying similarities and differences in the two states of the world, weighing both scientific considerations and other relevant factors.

LR Parsers for Natural Languages
Masaru Tomita
Computer Science Dept.
Carnegie-Mellon University
Pittsburgh, PA 15213
Coling84, pp. 354-357

MLR, an extended LR parser, is introduced, and its application to natural language parsing is discussed. An LR parser is a shift-reduce parser which is deterministically guided by a parsing table. A parsing table can be obtained automatically from a context-free phrase structure grammar. LR parsers cannot manager ambiguous grammars such as natural language grammars, because their parsing tables would have multiply-defined entries, which precludes deterministic parsing. MLR, however, can handle multiply-defined entries, using a dynamic programming method. When an input sentence is ambiguous, the MLR parser produces all possible parse trees without parsing any part of the input sentence more than once in the same way, despite the fact that the parser does not maintain a chart as in chart parsing. Our method also provides an elegant solution to the problem of multi-part-of-speech words such as "that". The MLR parser and its

parsing table generator have been implemented at Carnegie-Mellon University.

LFG System in Prolog
Hideki Yasukawa
The Second Laboratory
Institute for New Generation Technology
(ICOT)
Tokyo, 108, Japan
Coling84, pp. 358-361

In order to design and maintain a large scale grammar, the formal system for representing syntactic knowledge should be provided. Lexical Function Grammar (LGF) (Kaplan, Bresnan 1982) is a powerful formalism for that purpose. In this paper, the Prolog implementation of and LFG system is described. Prolog provides a good tool for the implementation of LFG. LFG can be translated into DCG (Pereira, Warren 1980) and functional structures (f-structures) are generated during the parsing process.

The Design of a Computer Language for Linguistic Information
Stuart M. Shieber
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and Information
Stanford University
Coling84, pp. 362-366

A considerable body of accumulated knowledge about the design of languages for communicating information to computers has been derived from the subfields of programming language design and semantics. It has been the goal of the PATR group at SRI to utilize a relevant portion of this knowledge in implementing tools to facilitate communication of linguistic information to computers. The PATR-II formalism is our current computer language for encoding linguistic information. This paper, a brief overview of that formalism, attempts to explicate our design decisions in terms of a set of properties that effective computer languages should incorporate.

Discourse Structures for Text Generation
William C. Mann
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Coling84, pp. 367-375

Text generation programs need to be designed around a theory of text organization. This paper introduces *Rhetorical Structure Theory*, a theory of text structure in which each region of text has a central *nuclear* part and a number of *satellites* related to it. A natural text is analyzed as an example, the mechanisms of the theory are identified, and their formalization is discussed. In a comparison, Rhetorical Structure Theory is found to be more comprehensive and more informative about text function than the text organization parts of previous text generation systems.

Semantic Rule Based Text Generation
Michael L. Mauldin
Dept. of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213
Coling84, pp. 376-380

This paper presents a semantically oriented, rule based method for single sentence text generation and discusses its implementation in the Kafka generator. This generator is part of the XCALIBUR natural language interface developed at CMU to provide natural language facilities for a wide range of expert systems and data bases. Kafka takes as input the knowledge representation used in the XCALIBUR system and incrementally transforms it first into conceptual dependency graphs and then into English.

Controlling Lexical Substitution in Computer Text Generation
Robert Granville
MIT Laboratory for Computer Science
545 Technology Square
Cambridge, MA 02130
Coling84, pp. 381-384

This report describes *Paul*, a computer text generation system designed to create cohesive text through the use of lexical substitutions. Specifically, this system is designed to deterministically choose between pronominalization, superordinate substitution, and definite noun phrase reiteration. The system identifies a *strength of antecedence recovery* for each of the lexical substitutions, and matches them against the *strength of potential antecedence* of each element in the text to select the proper substitutions for these elements.

Understanding of Japanese in an Interactive Programming System
Kenji Sugiyama, Masayuki Kameda,
Kouji Akiyama, Akifumi Makinouchi
Software Laboratory
Fujitsu Laboratories, Ltd.
1015 Kamikodanaka, Nakahara-ku
Kawasaki 211, Japan
Coling84, pp. 385-388

KIPS is an automatic programming system which generates standardized business application programs through interactive natural language dialogue. KIPS models the program under discussion and the content of the user's statements as organizations of dynamic objects in the object-oriented programming sense. This paper describes the statement-model and the program-model, their use in understanding Japanese program specifications, and how they are shaped by the linguistic singularities of Japanese input sentences.

Two-Way Finite Automata and Dependency Grammar: A Parsing Method for Inflectional Free Word Order Languages
Esa Nelimarkka, Harri Jappinen, Aarno Lehtola
Helsinki University of Technology
Helsinki, Finland
Coling84, pp. 389-392

This paper presents a parser of an inflectional free word order language, namely Finnish. Two-way finite automata are used to specify a functional dependency grammar and to actually parse Finnish sentences. Each automaton gives a functional description of a dependency structure within a constituent. Dynamic local control of the parser is realized by augmenting the automata with simple operations to make the automata, associated with the words of an input sentence, activate one another.

Interruptable Transition Networks
Sergei Nirenburg
Colgate University
Chagit Attiya
Hebrew University of Jerusalem
Coling84, pp. 393-397

A specialized transition network mechanism, the interruptible transition network (ITN), is used to perform the last of three stages in a multiprocessor syntactic parser. This approach can be seen as an exercise in implementing a parsing procedure of the active chart parser family.

Automatic Construction of Discourse Representation Structures
Franz Guenthner
Universitat Tubingen
Wilhelmstr. 50
D-7400 Tubingen, FRG
Hubert Lehman
IBM Deutschland GmbH
Heidelberg Scientific Center
Tiergartenstr. 15
D-6900 Heidelberg, FRG
Coling84, pp. 398-401

Kamp's Discourse Representation Theory is a major breakthrough regarding the systematic translation of natural language discourse into logical form. We have therefore chosen to marry the User Specialty Languages System, which was originally designed as a natural language front end to a relational database system, with this new theory. In the paper we try to show, taking – for the sake of simplicity – Kamp's fragment of English, how this is achieved. The research reported is going on in the context of the project *Linguistics and Logic Based Legal Expert System* undertaken jointly by the IBM Heidelberg Scientific Center and the Universität Tübingen.

Textual Expertise in Word Experts: An Approach to Text Parsing Based on Topic/Comment Monitoring
Udo Hahn
Universitaet Konstanz
Informationswissenschaft:
Projekt TOPIC
Postfach 5560
D-7750 Konstanz 1, West Germany
Coling84, pp. 402-408

In this paper prototype versions of two word experts for text analysis are dealt with which demonstrate that word experts are a feasible tool for parsing texts on the level of text cohesion as well as text coherence. The analysis is based on two major knowledge sources: context information is modelled in terms of a frame knowledge base, while the co-text keeps record of the linear sequencing of text analysis. The result of text parsing consists of a text graph reflecting the thematic organization of topics in a text.

Some Linguistic Aspects for Automatic Text Understanding
Yutaka Kusanagi
Institute of Literature and Linguistics
University of Taukuba
Sakura-mura, Ibaraki 305 JAPAN
Coling84, pp. 409-412

This paper proposes a system of mapping classes of syntactic structures as instruments for automatic text understanding. The system illustrated in Japanese consists of a set of verb classes and information on mapping them together with noun phrases, tense and aspect. The system, having information on direction of possible inferences between the verb classes with information on tense and aspect, is supposed to be utilized for reasoning in automatic text understanding.

A Syntactic Approach to Discourse Semantics
Livia Polanyi, Remko Scha
English Department
University of Amsterdam
Amsterdam, The Netherlands
Coling84, pp. 413-419

A correct structural analysis of a discourse is a prerequisite for understanding it. This paper sketches the outline of a discourse grammar which acknowledges several different levels of structure. This grammar, the "Dynamic Discourse Model", uses an Augmented Transition Network parsing mechanism to build a representation of the semantics of a discourse in a stepwise fashion, from left to right, on the basis of the semantic representations of the individual clauses which constitute the discourse. The intermediate states of the parser model the intermediate states of the social situation which generates the discourse.

This paper attempts to demonstrate that a discourse may indeed be

viewed as constructed by means of sequencing and recursive nesting of discourse constituents. It gives rather detailed examples of discourse structures at various levels, and shows how these structures are described in the framework proposed here.

Dealing with Incompleteness of Linguistic Knowledge in Machine Translation – Transfer and Generation Stage of MU Machine Translation Project

Makoto Nagao, Toyoaki Nishida,
Jun-ichi Tsujii
Dept. of Electrical Engineering
Kyoto University
Sakyo-ku, Kyoto 606, JAPAN
Coling84, pp. 420-427

Linguistics knowledge usable for machine translation is always imperfect. We cannot be free from the uncertainty of knowledge we have for machine translation. Especially at the transfer stage of machine translation, the selection of target language expression is rather subjective and optional.

Therefore the linguistic contents of a machine translation system always fluctuate, and make gradual progress. The system should be designed to allow such constant change and improvements. This paper explains the details of the transfer and generation stages of the Japanese-to-English system of the machine translation project by the Japanese Government, with the emphasis on the ideas to deal with the incompleteness of linguistic knowledge for machine translation.

Lexical Semantics in Human-Computer Communication

Jarrett Rosenberg
Xerox Office Systems Division
3333 Coyote Hill Road
Palo Alto, CA 94304
Coling84, pp. 428-431

Most linguistic studies of human-computer communication have focused on the issues of syntax and discourse structure. However, another interesting and important area is the lexical semantics of command languages. The names that users and system designers give the objects and actions of a computer system can greatly affect its usability, and the lexical issues involved are as complicated as those in natural languages. This paper presents an overview of the various studies of naming in computer systems, examining such issues as suggestiveness, memorability, descriptions of categories, and the use of non-words as names. A simple featural framework for the analysis of these phenomena is presented.

A Response to the Need for Summary Responses

J.K. Kalita, M.J. Colbourn,
G.I. McCalla
Dept. of Computational Science
University of Saskatchewan
Saskatoon, Saskatchewan
S7N OWO CANADA
Coling84, pp. 432-436

In this paper we argue that natural language interfaces to databases should be able to produce summary responses as well as listing actual data. We describe a system (incorporating a number of heuristics and a knowledge base built on top of the database) that has been developed to generate such summary responses. It is largely domain-independent, has been tested on many examples, and handles a wide variety of situations where summary responses would be useful.

Coping with Extrammaticality

Jaime G. Carbonell, Philip J. Hayes
Computer Science Dept.
Carnegie-Mellon University
Pittsburgh, PA 15213
Coling84, pp. 437-443

Practical natural language interfaces must exhibit robust behaviour in the presence of extragrammatical user input. This paper classifies different types of grammatical deviations and related phenomena at the lexical and sentential levels, discussing recovery strategies tailored to specific phenomena in the classification. Such strategies constitute a tool chest of computationally tractable methods for coping with extragrammaticality in restricted domain natural language. Some of the strategies have been tested and proven viable in existing parsers.

Correcting Object-Related Misconceptions: How Should the System Respond?

Kathleen F. McCoy
Dept. of Computer & Information
Science
University of Pennsylvania
Philadelphia, PA 19104
Coling84, pp. 444-447

This paper describes a computational method for correcting users' misconceptions concerning the objects modelled by a computer system. The method involves classifying object-related misconceptions according to the knowledge-base feature involved in the incorrect information. For each resulting class sub-types are identified, according to the structure of the knowledge base, which indicate what information may be supporting the misconception and therefore what information to include in the response. Such a characterization, along with a model of what the user knows, enables the system to reason in a domain-independent way about how best to correct the user.

An Algorithm for Identifying Cognates between Related Languages

Jacques B.M. Guy
Linguistics Dept. (RSPacS)
Australian National University
GPO Box 4
Canberra 2601 AUSTRALIA
Coling84, pp. 448-451

The algorithm takes as only input a list of words, preferably but not necessarily in phonemic transcription, in any two putatively related languages, and sorts it into decreasing order of probable cognition. The processing of a 250-item bilingual list takes about five seconds of CPU time on a DEC KL1091 and requires 56 pages of core memory. The algorithm is given no information whatsoever about the phonemic transcription used, and even though cognate identification is carried out on the basis of a context-free one-for-one matching of individual characters, its cognation decisions are bettered by a trained linguist using more information only in cases of wordlists sharing less than 40% cognates and involving complex, multiple sound correspondences.

From HOPE en l'ESPERANCE: On the Role of Computational Neurolinguistics in Cross-Language Studies

Helen M. Gigley
Dept. of Computer Science
University of New Hampshire
Durham, NH 03824
Coling84, pp. 452-456

Computational neurolinguistics (CN) is an approach to computational linguistics which includes neurally-motivated constraints in the design of models of natural language processing. Furthermore, the knowledge representations included in such models *must be supported* with documentated behavioral evidence, normal and pathological.

This paper will discuss the contribution of CN models to the understanding of linguistic "competence" within recent research efforts to adapt HOPE, an implemented CN model for "understanding" English, to l'ESPERANCE, one which "understands" French.

PANEL: Machine-Readable Dictionaries

Donald E. Walker
Natural-Language and Knowledge-Resource Systems
SRI International
Menlo Park, CA 94025
and
Artificial Intelligence and Information Science Research
Bell Communications Research
445 South Street
Morristown, NJ 07960
Coling84, p. 457

The papers in this panel consider machine-readable dictionaries from several perspectives: research in computational linguistics and computational lexicology, the development of tools for improving accessibility, the design of lexical reference systems for educational purposes, and applications of machine-readable dictionaries in information science contexts. As background and by way of introduction, a description is provided of a workshop on machine-readable dictionaries that was held at SRI International in April 1983.

Lexical Knowledge Bases

Robert A. Amsler
Natural-Language and Knowledge-Resource Systems
SRI International
Menlo Park, CA 94025
Coling84, pp. 458-459

A lexical knowledge base is a repository of computational information about concepts intended to be generally useful in many application areas including computational linguistics, artificial intelligence, and information science. It contains information derived from machine-readable dictionaries, the full text of reference books, the results of statistical analyses of text usages, and data manually obtained from human world knowledge.

Machine-Readable Dictionaries, Lexical Data Bases and the Lexical System

Nicoletta Calsolari
Dipt. di Linguistica
Universita di Pisa, ITALY
and
Istituto di Linguistica Computasionale
del CNR
Pisa, ITALY.
Coling84, p. 460

I should like to raise some issues concerning the conversion from a traditional Machine-Readable Dictionary (MRD) on tape to a Lexical Data Base (LDB), in order to highlight some important consequences for computational linguistics which can follow from this transition. The enormous potentialities of the information implicitly stored in a standard printed dictionary or a MRD can only be evidenced and made explicit when the same data are given a new logical structure in a data base model, and exploited by appropriate software.

The Dictionary Server

Martin Kay
Intelligent Systems Laboratory

The term "machine-readable dictionary" can clearly be taken two ways. In its strong and better established interpretation, it presumably refers to dictionaries intended for machine consumption and use as in a language

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
Coling84, p. 461

processing system of some sort. In a somewhat weaker sense, it has to do with dictionaries intended for human consumption, but through the intermediary of a machine. Ideally, of course, the two enterprises would be conflated, material from a single basic store of lexical information being furnished to different clients in different forms. Such a conflation would, if it could be accomplished, be beneficial to all parties. Certainly human users could surely benefit from some of the processes that the machine-oriented information in a machine-readable dictionary usually makes available. They can profit even more from other processes specifically oriented to the human user but which have not yet received much attention. For these reasons, I believe that machine-readable dictionaries should, and probably soon will, come to replace traditional book-form dictionaries. I do not have in mind machine-readable dictionaries that single users load into their personal machines so much as centralized services to which individual clients subscribe.

How to Misread a Dictionary

George A. Miller
Dept. of Psychology
Princeton University
Princeton, NJ 08544
Coling84, p. 462

A dictionary is an extremely valuable reference book, but its familiarity tends to blind adults to the high level of intelligence required to read it. This aspect becomes apparent, however, when school children are observed learning dictionary skills.

Machine-Readable Components in a Variety of Information-System Applications

Howard R. Webber
Reference Publishing Division
Houghton-Mifflin Company
2 Park Street
Boston, MA 02108
Coling84, p. 463

Components of the machine-readable dictionary can be applied in a number of information systems. The most direct applications of the kind are in word processing or in " writing-support" systems built on a word processing base. However, because a central function of any dictionary is in fact data verification, there are other proposed applications in communications and data storage and retrieval systems. Moreover, the complete interrelational electronic dictionary is in some sense the model of the language, and there are, accordingly, additional implications for language-based information search and retrieval.

Transfer in a Multilingual MT System

Steven Krauwer, Louis des Tombe
Institute for General Linguistics
Utrecht State University
Trans 14, 3512 JK Utrecht
The Netherlands
Coling84, pp. 464-467

In the context of transfer-based MT systems, the nature of the intermediate representations, and particularly their 'depth', is an important question. This paper explores the notions of "independence of languages' and 'simple transfer', and provides some principles that may enable linguists to study this problem in a systematic way.

Expert Systems and Other New Techniques in MT Systems

Christian Boitet, Rene Gerber
Groupe d'Etudes pour la Traduction
Automatique
BP no. 68, Universite de Grenoble
38402 Saint-Martin d'Heres, FRANCE
Coling84, pp. 468-471

Our MT systems integrate many advanced concepts from the fields of computer science, linguistics, and AI: specialized languages for linguistic programming based on production systems, complete linguistic programming environment, multilevel representations, organization of the lexicons around "lexical units", units of translation of the size of several paragraphs, possibility of using text-driven heuristic strategies.

We are now beginning to integrate new techniques: unified design of an "integrated" lexical data base containing the lexicon in "natural" and language, addition of expert systems equipped with "extralinguistic" or "metalinguistic" knowledge, and design of a kind of structural metaeditor (driven by a static grammar) allowing the interactive construction of a document in the same way as syntactic editors are used for developing programs. We end the paper by mentioning some projects for long-term research.

Robust Processing in Machine Translation

Doug Arnold

In this paper we provide an abstract characterisation of different kinds of robust processing in Machine Translation and Natural Language Processing systems in terms of the kinds of problem they are supposed to solve.

Centre for Cognitive Studies
University of Essex
Colchester, CO4 3SQ, U.K.
Rod Johnson
Centre for Computational Linguistics
UMIST, Manchester
M60 8QD, U.K.
Coling84, pp. 472-475

We focus on one problem which is typically exacerbated by robust processing, and for which we know of no existing solutions. We discuss two possible approaches to this, emphasising the need to correct or repair processing malfunctions.

Disambiguating Grammatically Ambiguous Sentences by Asking

Masaru Tomita
Computer Science Dept.
Carnegie-Mellon University
Pittsburgh, PA 15213
Coling84, pp. 476-480

The problem addressed in this paper is to disambiguate grammatically ambiguous input sentences by asking the user, who need not be a computer specialist or a linguist, without showing any parse trees or phrase structure rules. Explanation List Comparison (ELC) is the technique that implements this process. It is applicable to all parsers which are based on phrase structure grammar, regardless of the parser implementation. An experimental system has been implemented at Carnegie-Mellon University, and it has been applied to English-Japanese machine translation at Kyoto University.

Ambiguity Resolution in the Human Syntactic Parser: An Experimental Study

Howard S. Kurtzman
Dept. of Psychology
Massachusetts Institute of Technology
Cambridge, MA 02139
Coling84, pp. 481-485

Models of the human syntactic parsing mechanism can be classified according to the ways in which they operate upon ambiguous input. Each mode of operation carries particular requirements concerning such basic computational characteristics of the parser as its storage capacities and the scheduling of its processes, and so specifying which mode is actually embodied in human parsing is a useful approach to determining the functional organization of the human parser. In Section 1, a preliminary taxonomy of parsing models is presented, based upon a consideration of modes of handling ambiguities; and then, in Section 2, psycholinguistic evidence is presented which indicates what type of model best describes the human parser.

Conceptual Analysis of Garden-Path Sentences

Michael J. Pazzani
The MITRE Corporation
Bedford, MA 01730
Coling84, pp. 486-490

By integrating syntactic and semantic processing, our parser (LAZY) is able to deterministically parse sentences which syntactically appear to be garden-path sentences although native speakers do not need conscious reanalysis to understand them. LAZY comprises an extension to conceptual analysis which yields an explicit representation of syntactic information and a flexible interaction between semantic and syntactic knowledge.

Language Generation from Conceptual Structure: Synthesis of German in a Japanese/German MT Project

*L. Laubsch, D. Roesner, K. Hanakata,
A. Lesniewski*
Project SEMSYN
Institut fuer Informatik
Universitaet Stuttgart
Herdweg 51, D-7000 Stuttgart 1
West Germany
Coling84, pp. 491-494

This paper describes the current state of the SEMSYN project, whose goal is to develop a module for generation of German from a semantic representation. The first application of this module is within the framework of a Japanese/German machine translation project. The generation process is organized into three stages that use distinct knowledge sources. The first stage is conceptually oriented and language independent, and exploits case and concept schemata. The second stage employs realization schemata which specify choices to map from meaning structures into German linguistic constructs. The last stage constructs the surface string using knowledge about syntax, morphology, and style. This paper describes the first two stages.

Natural Language Driven Image Generation (NALIG)

*Giovanni Adorni, Mauro Di Manzo,
Fausto Giunchiglia*
Dept. of Communication, Computer and
System Sciences
University of Genoa
Via Opera Pia 11 A - 16145 Genoa
ITALY

In this paper the experience made through the development of a NAtural Language drive Image Generation (NALIG) is discussed. This system is able to imagine a static scene described by means of a sequence of simple phrases. In particular, a theory for equilibrium and support will be outlined together with the problem of object positioning.

Coling84, pp. 495-500

Conceptual and Linguistic Decisions in Generation
Laurence Danlos
LADL (CNRS)
Universite de Paris 7
2, Place Jussieu
75005 Paris, France
Coling84, pp. 501-504

Generation of texts in natural language requires making conceptual and linguistic decisions. This paper shows first that these decisions involve the use of a discourse grammar, secondly that they are all dependent on one another but that there is a priori no reason to give priority to one decision rather than another. As a consequence, a generation algorithm must not be modularized in components that make these decisions in a fixed order.

A Computational Analysis of Complex Noun Phrases in Navy Messages
Elaine Marsh
Navy Center for Applied Research in
Artificial Intelligence
Naval Research Laboratory - Code 7510
Washington, DC 20375
Coling84, pp. 505-508

Methods of text compression in Navy messages are not limited to sentence fragments and the omissions of function words such as the copula *be*. Text compression is also exhibited within "grammatical" sentences and is identified within noun phrases in Navy messages. Mechanisms of text compression include increased frequency of complex noun sequences and also increased usage of nominalizations. Semantic relationships among elements of a complex noun sequence can be used to derive a correct bracketing of syntactic constructions.

Another Look at Nominal Compounds
Pierre Isabelle
Dept. de Linguistique
Universite de Montreal
C.P. 6128, Succ. A, Montreal, Quebec
CANADA H3C 3J7
Coling84, pp. 509-516

We present a progress report on our research on nominal compounds (NCs). Recent approaches to this problem in linguistics and natural language processing (NLP) are reviewed and criticized. We argue that the notion of "role nominal", which is at the interface of linguistic and extra-linguistic knowledge, is crucial for characterizing NCs as well as other linguistic phenomena. We examine a number of constraints on the semantic interpretation rules for NCs. Proposals are made that should improve the capability of NLP systems to deal with NCs.

Semantic Parsing as Graph Language Transformation - A Multidimensional Approach to Parsing Highly Inflectional Languages
Eeero Hyvonen
Helsinki University of Technology
Digital Systems Laboratory
Otakaari 54
02150 Espoo 15 FINLAND
Coling84, pp. 517-520

The sturcture of many languages with "free" word order and rich morphology like Finnish is configurational rather than linear. Although non-linear structures can be represented by linear formalisms, it is often more natural to study multidimensional arrangement of symbols. Graph grammars are a multidimensional generalization of linear string gram mars. In graph grammars, string rewrite rules are generalized into graph rewrite rules. This paper presents a graph grammar formalism and parsing scheme for parsing languages with inherent configurational flavor. A small experimental Finnish parsing system has been implemented.

Handling Syntatical Ambiguity in Machine Translation
Vladimir Pericliev
Institute of Industrial Cybernetics and
Robotics
Acad. G. Bontchev Str., bl. 12
1113 Sofia, Bulgaria
Coling84, pp. 521-524

The difficulties to be met with the resolution of syntactical ambiguity in MT can be at least partially overcome by means of preserving the syntactical ambiguity of the source language in the target language. An extensive study of the correspondence between the syntactically ambiguous structures in English and Bulgarian has provided a solid empirical basis in favor of such an approach. Similar results could be expected for other sufficiently related languages as well. The paper concentrates on the linguistic grounds for adopting the approach proposed.

Argumentation in Representation Semantics
Pierre-Yves Raccah
ERA 430 - C.N.R.S.
Conseil d'Etat, Palais Royal
75100 Paris RP
Coling84, pp. 525-529

It seems rather natural to admit that language use is governed by rules that relate signs, forms and meanings to possible intentions or possible interpretations, in function of utterance solutions. Not less natural should seem the idea that the meaning of a natural language expression conveys enough **material** to the input of these rules, so that, given the situation of utterance, they determine the appropriate interpretation. If this is correct, the semantic description of a natural language expression should output not only the 'informative content' of that expression, but also all sorts of indications concerning the way this expression may be used or interpreted. In particular, the argumentative power of utterances is due to **argumentative**

indications conveyed by the sentences uttered, indications that are not part of their information content. This paper emphasizes the role of argumentation in language and shows how it can be accounted for in a formal Representation semantics framework. An example of an analysis is provided in order to show the "system at work".

Voice Simulation: Factors Affecting Quality and Naturalness

B. Yegnanarayana
Dept. of Computer Science and
Engineering
Indian Institute of Technology
Madras-600 036, India
J.M. Naik, D.G. Childers
Dept. of Electrical Engineering
University of Florida
Gainesville, FL 32611
Coling84, pp. 530-533

In this paper we describe a flexible analysis-synthesis system which can be used for a number of studies in speech research. The main objective is to have a synthesis system whose characteristics can be controlled through a set of parameters to realize any desired voice characteristics. The basic synthesis scheme consists of two steps: Generation of an excitation signal from pitch and gain contours and excitation of the linear system model described by linear prediction coefficients. We show that a number of basic studies such as time expansion/compression, pitch modifications and spectral expansion/compression can be made to study the effect of these parameters on the quality of synthetic speech. A systematic study is made to determine factors responsible for unnaturalness in synthetic speech. It is found that the shape of the glottal pulse determines the quality to a large extent. We have also made some studies to determine factors responsible for loss of intelligibility in some segments of speech. A signal dependent analysis-synthesis scheme is proposed to improve the intelligibility of dynamic sounds such as stops. A simple implementation of the signal dependent analysis is proposed.

Interpreting Syntactically Ill-Formed Sentences

Leonardo Lesmo, Piettro Torasso
Dipt. di Informatica
Universita di Torino
Corso Massimo D'Azeglio 42
10125 Torino - ITALY
Coling84, pp. 534-539

The paper discusses three different kinds of syntactic ill-formedness: ellipsis, conjunctions, and actual syntactic errors. It is shown how a new grammatical formalism, based on a two-level representation of the syntactic knowledge is used to cope with ill-formed sentences. The basic control structure of the parser is briefly sketched; the paper shows that it can be applied without any substantial change both to correct and to ill-formed sentences. This is achieved by introducing a mechanism for the hypothesization of syntactic structures, which is largely independent of the rules defining the well-formedness. On the contrary, the second level of syntactic knowledge embodies those rules and is used to validate the hypotheses emitted by the first level. Alternative hypotheses are obtained, when needed, by means of local reorganizations of the parse tree. Sentence fragments are handled by the same mechanism, but in this case the second level rules are used to detect the absence of one (or more) constituents.

An International Delphi Poll on Future Trends in "Information Linguistics"

Rainer Kuhlen
Universitaet Konstanz
Informationswissenschaft, Box 6650
D-7750 Konstanz 1, West Germany
Coling84, pp. 540-545

The results of an international Delphi poll on information linguistics, which was carried out between 1982 and 1983, are presented.

Machine Translation: Its History, Current Status and Future Prospects

Jonathan Slocum
Siemens Communications Systems, Inc.
Linguistics Research Center
University of Texas
Austin, TX
Coling84, pp. 546-561

Elements of the history, state of the art, and probable future of Machine Translation (MT) are discussed. The treatment is largely tutorial, based on the assumption that this audience is, for the most part, ignorant of matters pertaining to translation in general, and MT in particular. The paper covers some of the major MT R&D groups, the general techniques they employ(ed), and the roles they play(ed) in the development of the field. The conclusions concern the seeming permanence of the translation problem, and potential re-integration of MT with mainstream Computational Linguistics.