

ABSTRACTS OF CURRENT LITERATURE

A Framework for Inference in Natural Language Front Ends to Databases

B.K. Boguraev, K. Sparck Jones

Computer Laboratory

University of Cambridge

Corn Exchange Street

Cambridge CB2 3QG, England

Technical Report TR 64, February 1985, 73pp.

£1.50 plus postage:

£5.00(1.50) air(surface)

This report discusses the inference requirements to be met by a natural language database front end of the kind developed in Cambridge, with detailed examples, and proposes a framework, exploiting a specific type of network for knowledge representation, for developing the front end to support inference.

An Experimental Interactive Natural Language Programming System

Kenji Sugiyama, Kouji Akiyama,

Masayuki Kameda, Akifumi Makinouchi

Fujitsu Laboratories, Ltd.

Kawasaki, Japan 211

Systems, Controls, Computers, 15(3): 28-37.

This paper discusses the problems encountered in the development of the interactive natural language programming system (KIPS) from three aspects: input sentence, target program, and communication between the user and the system. Based on the recognition of the problem, an interactive natural language programming system is proposed which is constructed according to a model of the task domain consisting of active objects in the object-oriented programming sense. The proposed system is composed of four modules: parser, specification acquirer, coder, and user interface. These modules realize the functions of information extraction from a Japanese sentence, assimilation of fragmentary information, automatic programming, and man-machine interface, respectively. Finally, future development of the system is discussed.

Plan Parsing for Intended Response Recognition in Discourse

Candace L. Sidner

BBN Laboratories, Inc.

Cambridge, MA 02238

Computational Intelligence, 1(1): 1-10.

In a discourse the hearer must recognize the response intended by the speaker. To perform this recognition, the hearer must ascertain what plans the speaker is undertaking and how the utterances in the discourse further that plan. To do so, the hearer can parse the initial intentions (recoverable from the utterance) and recognize the plans the speaker has in mind and intends the hearer to know about. This paper reports on a theory of parsing the intentions in discourse. It also discusses the role of another aspect of discourse, discourse markers, that are valuable to intended response recognition.

On the Adequacy of Predicate Circumscription for Closed-World Reasoning

David W. Etherington, Robert E.

Mercer, Raymond Reiter

Department of Computer Science

University of British Columbia

Vancouver, B.C., Canada V6T 1W5

Computational Intelligence, 1(1): 11-15.

We focus on McCarthy's method of predicate circumscription in order to establish various results about its ability to consistency, and about its conjecture new information. A basic result is that predicate circumscription cannot account for the standard kinds of default reasoning. Another is that predicate circumscription yields no new information about the equality predicate. This has important consequences for the unique names and domain closure assumptions.

The following technical reports are available from

Computer Science Department

College of Liberal Arts

Boston University

11 Cummington Street

Boston, MA 02215

A Computational Theory of Metaphor Comprehension and Analogical Reasoning

Bipin Indurkha

BUCS Tech Report 85-001, February 1985,

In this thesis we propose a formal theory of metaphors and analogies. We start from the assumption that a metaphor, or an analogy, is characterized by the description of one domain (target domain) in terms of another domain (source domain). We describe a formalism, called Schema-Language (SL), for representing domain knowledge which

195 pages.

is based on the First-Order Predicate Calculus. We then develop a theory of Constrained Semantic Transference (CST) which shows how the terms and structural relationships of the source domain can be coherently transferred to the target domain. The concept of a T-MAP, which is a partial coherent mapping from the terms of the source domain to the target domain, is central to CST. We show how to characterize metaphors and analogies by using T-MAPs which can explain many cognitive properties associated with them. A major limitation of CST is that the notion of coherency is not computational.

We propose a theory of Approximate Semantic Transference (AST), which is derived from CST by replacing the coherency requirement on T-MAPs by approximate coherency. The partial approximate-coherent mappings of AST, called AT-MAPs, are computational and can be used as a basis for developing models of cognitive processes involved in comprehending metaphors and analogies. We propose two alternative formulations of approximate coherency. Based on one of these versions, we present several algorithms, and principles that can be used in designing algorithms, for computing AT-MAPs from the knowledge of the source and target domains.

Automatic Constraint Generation for Semantic Query Optimization

Michael Siegel, Edward Sciore
BUCS Tech Report 85-006, April 1985, 11 pages.

Semantic query optimization uses expert knowledge in the form of integrity constraints to improve query execution. Due to the dependency on expert knowledge, this method of optimization is limited to data bases with useful constraint sets. However, the constraint sets as supplied by the expert are not dynamic with respect to changes in the state of the data base or changes in the database usage patterns. This paper describes a methodology that can be used to automatically select and generate constraints from the data base. This method provides a semantic query optimizer with a self-adapting set of relevant constraints. The method also includes various strategies for selecting the most promising constraints, minimizing the set of derived constraints and maintenance of the derived constraint set.

The following abstracts are from the Proceedings of the Second Conference of the European Chapter of the Association for Computational Linguistics, 27-29 March 1985 (in press). Copies will be available from

Donald E. Walker, ACL
Bell Communications Research
435 South Street MRE 2A-379
Morristown, NJ 07960-1961

Natural Languages and the Chomsky Hierarchy

Andres Kornai
Institute of Linguistics
Hungarian Academy of Sciences
Budapest, P.C.B. 19, H-1250 Hungary
Proc. EACL 1985, pp. 1-7

The central claim of the paper is that NL stringsets are regular. Three independent arguments are offered in favor of this position: one based on parsimony considerations, one employing the McCullough-Pitts (1942) model of neurons, and a purely linguistic one. It is possible to derive explicit upper bounds for the number of (live) states in NL acceptors: the results show that finite state NL parsers can be implemented on present-day computers. The position of NL stringsets within the regular family is also investigated: it is proved that NLs are counter-free, but not locally testable.

How Does Natural Language Quantify?

Michael Hess
University of Zurich
Seminar of General Linguistics
Plattenstrasse 54
CH-8032 Zurich, Switzerland
Proc. EACL 1985, pp. 8-15

It has traditionally been assumed that Natural Language uses explicit quantifier expressions (such as *all* and *most*, *the* and *a*) for the purpose of quantification. We argue that expressions of the first type are comparatively rare in real world Natural Language sentences, and that the latter (articles) cannot be considered straightforward quantifiers in the first place. However, practically all applications of Natural Language Processing require sentences to be quantified unambiguously. We list a few possible (syntactical, semantical, and "pragmatical") sources of "implicit" quantificational information in Natural Language; they combine in some-

times intricate ways to give a sentence a (more or less) unambiguous quantification.

With the intention of indicating some temporal/event-theoretic characteristics of distributive clauses, a generalisation is made over distributives and clauses marked for iterative aspect: two kinds of semantic phenomena which have normally been confined to separate theoretical domains. It is shown that in particular, both give rise to an 'inferential set construction' problem. An informal outline is given of what might constitute such a generalisation. The generalisation is proposed initially on grounds of *prima facie* plausibility, but its ultimate defensibility and explanatory value will depend on the validity of its consequence, that distributive clauses entail a multiplicity of temporal entities or events. This proposal is considered with respect to two types of discourse phenomena; anaphoric reference to event entities, and temporal binding. These provide further support for making the generalisation clarify its nature and indicate in what respect the entailment claim can be true of distributives. The set construction problem is of practical importance for computational models of natural language interaction, and since the concept of iterated action is central to planning, the generalisation across iteration and distributives, along with the observations about their nature, have interesting implications for work in this area.

Distributives, Quantifiers and a Multiplicity of Events

Lesley Stirling

School of Epistemics &
Department of Linguistics
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW, U.K.
Proc. EACL 1985, pp. 16-24

Montagovian Definite Clause Grammar

R. I. Bainbridge

Department of Computer Science
Teesside Polytechnic
Middlesbrough, Cleveland, England
Proc. EACL 1985, pp. 25-34

This paper reports a completed stage of ongoing research at the University of York. Landsbergen's advocacy of analytical inverses for compositional syntax rules encourages the application of Definite Clause Grammar techniques to the construction of a parser returning Montague analysis trees. A parser MDCG is presented which implements an augmented Friedman-Warren algorithm permitting post referencing, and interfaces with a language of intentional logic translator LILT so as to display the derivational history of corresponding reduced IL formulae. Some familiarity with Montague's PTG and the basic DCG mechanism is assumed.

The Specification of Time Meaning for Machine Translation

Frank van Eynde

Catholic University, Leuven
Blijde Inkomststraat, 21
3000 Leuven, Belgium

Louis des Tombe

Utrecht State University
Trans, 14, 3512 JK Utrecht, Holland

Fons Maes

Catholic University of Tilburg
Postbus 90153, 5000 LE Tilburg, Holland

Proc. EACL 1985, pp. 35-40

In this paper, we put forward some ideas on the representation of *time* in a machine translation system. In such a system, we usually have the following four representations:

- source text
- source representation
- target representation
- target text

In an interlingual system, there is no difference between source and target representation; in a transfer-based system, the step between the two is usually called transfer, and this step is meant to be as simple as possible.

The research described was originally done in the framework of the EUROTRA MT project, which is transfer-based. However, it can be used in other MT systems as well; in fact, it is very well suited for interlingual systems.

The problem with time meaning is that it is expressed in natural languages in a way that is non-universal and, moreover, not very perspicuous *prima facie*. As a consequence, it is difficult to find rules for the translation of the tense form of the verb.

In this paper we propose a conceptual calculus in which the meanings of language specific temporal expressions can be represented in an interlingual way, so that the translation of the latter can be achieved via the corresponding conceptual representations.

An ATN Treatment WH-Movement

Hans Haugeneder

An ATN-parser is represented with emphasis on the treatment of those phenomena which, in the framework of transformational grammar, are

Siemens AG
 ZT ZTI; Otto-Hahn-Ring 6
 8 Munchen 83, West Germany
Proc. EACL 1985, pp. 41-47

SAUMER: Sentence Analysis Using Metarules

Fred Popowich
 Natural Language Group
 Laboratory for Computer and
 Communications Research
 Dept. of Computer Science
 Simon Fraser University
 Burnaby, B.C., Canada V5A 1S6
Proc. EACL 1985, pp. 48-56

Effective Parsing with Generalised Phrase Structure Grammar

Allan Ramsay
 Cognitive Studies Program
 University of Sussex
 Brighton, BN1 9QN, England
Proc. EACL 1985, pp. 57-61

An Evaluation of METAL: the LRC the LRC Machine Translation System

Jonathan Slocum
 Microelectronics & Computer Technology
 Corp (MCC)
*Winfield S. Bennett, Lesley Whiffin,
 Edda Norcross*
 Siemens Communication Systems, Inc.
Proc. EACL 1985, pp. 62-69

A Two-Way Approach to Structural Transfer in MT

subsumed under the concept of WH-movement. The approach taken tries to embed these constructions into an ATN grammar in a general, linguistically motivated and, in terms of the ATN grammar formalism, descriptive way. To accomplish this goal, the approach described incorporates the basic principles governing such constructions as formulated in the framework of the trace theory proposed in the development of the Extended Standard Theory (EST). Thus a unified treatment for both relative clauses and wh-questions is achieved.

The SAUMER system uses specifications of natural language grammars, which consist of rules and metarules, to provide a semantic interpretation of an input sentence. The SAUMER Specification Language (SSL) is a programming language which combines some of the features of *generalised* phrase structure grammars (Gazdar 1981), like the correspondence between syntactic and semantic rules, with **definite clause grammars** (DCGs) (Pereira and Warren 1980) to create an executable grammar specification. SSL rules are similar to DCG rules except that they contain a semantic component and may also be left recursive. Metarules are used to generate new rules from existing rules before any parsing is attempted. An implementation is tested which can provide semantic interpretations for sentences containing topicalisation, relative clauses, passivation, and questions.

Generalised phrase structure grammars (GPSGs) appear to offer a means by which the syntactic properties of natural languages may be very concisely described. The main reason for this is that the GPSG framework allows you to state a variety of meta-grammatical rules which generate new rules from old ones, so that you can specify rules with a wide variety of realisations via a very small number of explicit statements. Unfortunately, trying to analyse a piece of text in terms of such rules is a very awkward task, as even a small set of GPSG statements will generate a large number of underlying rules.

This paper discusses some of the difficulties of parsing with GPSGs, and presents a fairly straightforward bottom-up parser for them. This parser is, in itself, no more than adequate – all its components are implemented quite efficiently, but there is nothing tremendously clever about how it searches the space of possible rules to find an analysis of the text it is working on. Its power comes from the fact that it learns from experience: not new rules, but how to recognize realisations of complex combinations of its existing rules. The improvement in the system's performance after even a few trials is dramatic. This is brought about by a mechanism for recording the analysis of text fragments. Such recordings may be used very effectively to guide the subsequent analysis of similar pieces of text. Given such guidance it becomes possible to deal even with text containing unknown or ambiguous words with very little search.

The Linguistics Research Center (LRC) at the University of Texas at Austin is currently developing METAL, a fully-automatic high-quality machine translation system, for market introduction in 1985. This paper will describe the current status of METAL, emphasizing the results of the most recent post-editors' evaluation, and will briefly indicate some future directions for the system. A 6-page German original text and a raw (unedited but automatically reformatted) METAL translation of that text into English are included as appendices.

The METAL machine translation project incorporates two methods of structural transfer – direct transfer and transfer by grammar. In this paper

Rebecca Root

Linguistics Research Center
University of Texas
P.O. Box 7247
Austin, TX 78712
Proc. EACL 1985, pp. 70-72

I discuss the strengths and weakness of these two approaches in general and with respect to the METAL project, and argue that, for many applications, a combination of the two is preferable to either alone.

**Various Representations of Text
Proposed for EUROTRA**

Christian Boitet

Groupe d'Etudes pour La Traduction
Automatique, Universite Scientifique et
Medicale de Grenoble
BP 68 - 38402 Saint Martin d'Herès -
France

Nelson Verastegui, Daniel Bachuet

Institut de Formation et Conseil en
Informatique
27, rue Turenne - 38000 Grenoble - France
Proc. EACL 1985, pp. 73-78

We introduce several general notions concerning the texts and the particularities of text processing on a computer support, in relation to some problems which are specific to M(A)T. And we present the solution we have proposed for the duration of the EUROTRA project.

**Natural Language Processing and the
Automatic Acquisition of Knowledge:
A Simulative Approach**

Daniilo Fum

Laboratorio di Psicologia E.E.
Universita di Trieste
via Tigor 22, I - 34124 Trieste (Italy)
Proc. EACL 1985, pp. 79-88

The paper presents the general design and the first results of a research project whose long term goal is to develop and implement ALICE, an experimental system capable of augmenting its knowledge base by processing natural language texts. ALICE (an acronym for Automatic Learning and Inference Computerized Engine) is an attempt to model the cognitive processes that occur in humans when they learn a series of descriptive texts and reason about what they have learned. In the paper a general overview of the system is given with the description of its specifics, basic methodologies, and general architecture. How parsing is performed in ALICE is illustrated by following the analysis of a sample text.

**Right Attachment and Preference
Semantics**

Yorick Wilks

Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003
Proc. EACL 1985, pp. 89-92

The paper claims that the right attachment rules for phrases originally suggested by Frazier and Fodor are wrong, and that none of the subsequent patchings of the rules by syntactic methods have improved the situation. For each rule there are perfectly straightforward and indefinitely large clauses of simple counter-examples. We then examine suggestions by Ford et al., Schubert and Hirst which are quasi-semantic in nature and which we consider ingenious but unsatisfactory. We point towards a straightforward solution within the framework of preference semantics, set out in detail elsewhere, and argue that the principle issue is not the type and nature of information required to get appropriate phrase attachments, but the issue of where to store the information and with what processes to apply it.

How to Restrict Ambiguity of Discourse

Barbara Dunin-Keplicz

Institute of Informatics
University of Warsaw
P.O. Box 1210 - 00-901 Warszawa, Poland
Proc. EACL 1985, pp. 93-97

We single out a class of **prototypes**, i.e., a class of constructions forcing the obligatory coreference or obligatory noncoreference. An essential feature of prototypes is their undistinctiveness. In this sense they are the most natural and efficient means of communication in discourse.

The non-application of prototype should be well motivated. This leads to the *rule of restricted choice* stating that whenever it is possible the application of a prototype should be preferred.

The rule of the restricted choice suggests the general outline of interpreting ambiguous sentences, strictly speaking, the method of ordering admissible interpretations: those which can be equivalently expressed by means of a prototype are less probable. In other words, the rule of the restricted choice can be regarded as some kind of mechanism ordering the hypotheses for computation.

Language-Based Environment for Natural Language Parsing

A. Lehtola, H. Jappinen, E. Nelimarkka
Sitra Foundation
POBox 329, SF-00121 Helsinki, Finland
and Helsinki University of Technology
Proc. EACL 1985, pp. 98-106

This paper introduces a special programming environment for the definition of grammars and for the implementation of corresponding parsers. In natural language processing systems it is advantageous to have linguistic knowledge and processing mechanisms separated. Our environment accepts grammars consisting of binary dependency relations and grammatical functions. Well-formed expressions of functions and relations provide constituent surroundings for syntactic categories in the form of two-way automata. These relations, functions, and automata are described in a special definition language.

In focusing on high level descriptions a linguist may ignore computational details of the parsing process. He writes the grammar into a DPL-description and a compiler translates it into efficient LISP-code. The environment has also a tracing facility for the parsing process, grammar-sensitive lexical maintenance programs, and routines for the interactive graphic display of parse trees and grammar definitions. Translator routines are also available for the transport of compiled code between various LISP-dialects. The environment itself exists currently in INTERLISP and FRANZLISP. This paper focuses on knowledge engineering issues and does not enter linguistic argumentation.

Parametrized Abstract Objects for Linguistic Information Processing

Helene Bestougeff, Gerard Ligozat
CNRS-Universite Paris VII
2, Place Jussieu, 75005 PARIS, France
Proc. EACL 1985, pp. 107-115

Programming languages which have adequate primitives for linguistic information processing and a clear semantics at the formal computational level are now slowly emerging as a convergent effort from computer science, linguistics, and artificial intelligence. Our work on the processing of a special kind of linguistic information, namely temporal information, has led us to advocate the use of a language with the following characteristic features:

- high level of abstraction;
- capacity for inference;
- modularity.

A high level of abstraction is needed to deal with complex linguistic notations which are not easily reducible to elementary data structures.

A capacity for inference is required, as most criteria or tests in linguistics make use of particular kinds of deductions, at different levels of the linguistic analysis.

As for modularity, a typical situation in linguistics has to do with a hierarchy of concepts or units, and the relations between these units at different levels.

This paper discusses the relevance of the choice of parameterized abstract objects as tools for linguistic information processing and exemplifies the use of such objects for temporal information.

On the Representation of Query Term Relations by Soft Boolean Operators

Gerard Salton
Department of Computer Science
Cornell University
Ithaca, NY 14853
Proc. EACL 1985, pp. 116-122

The language analysis component in most text retrieval systems is confined to a recognition of noun phrases of the type normally included in back-of-the-book indexes, and an identification of related terms included in a preconstructed thesaurus of quasi-synonyms. Even such a restricted language analysis is fraught with difficulties because of the well-known problems in the analysis of compound nominals, and the hazards and cost of constructing word synonym classes valid for large text samples.

In this study an extended (soft) Boolean logic is used for the formulation of information retrieval queries which is capable of representing both the use of compound noun phrases as well as the inclusion of synonym constructions in the query statements. The operations of the extended Boolean logic are described, and evaluation output is included to demonstrate the effectiveness of the extended logic compared with that of ordinary text retrieval systems.

The Resolution of Local Syntactic

The resolution of local syntactic ambiguity by the Human Sentence Proc-

**Ambiguity by the Human Sentence
Processing Mechanism**

Gerry Altmann

Department of Linguistics

University of Edinburgh

George Square, Edinburgh EH8 9LL (GB)

Proc. EACL 1985, pp. 123-127

essing Mechanism is a topic which has provoked considerable interest in recent years. At issue is whether such ambiguities are resolved on the basis of syntactic information alone (cf. Minimal Attachment – Frazier 1979), or whether they are resolved on some other basis. Crain & Steedman (1982) suggest that the resolution process is governed not by Minimal Attachment but instead by whether or not a referring expression provides sufficient information with which to identify a unique referent. Such an approach relies on the provision of adequate contextual information, something which has been lacking in experiments which have been claimed to support Minimal Attachment. In this paper I shall consider a number of such experiments, and the different patterns of results which emerge once contextual information is provided. Although the importance of contextual information will be stressed, I shall briefly consider reasons why parsing preferences arise in the absence of any explicit prior context. The conclusion is that computational models of syntactic ambiguity resolution which are based on evidence which has ignored contextual considerations are models of something other than *natural* language processing.

A Parser That Doesn't

S.G. Pulman

University of Cambridge

Computer Laboratory

Corn Exchange Street

Cambridge CB2 3QG, UK

Proc. EACL 1985, pp. 128-135

This paper describes an implemented parser-interpreter which is intended as an abstract formal model of part of the process of sentence comprehension. It is illustrated here for Phrase Structure Grammars with a translation into a familiar type of logical form, although the general principles are intended to apply to any grammatical theory sharing certain basic assumptions, which are discussed in the paper. The procedure allows for incremental semantic interpretation as a sentence is parsed, and provides a principled explanation for some familiar observations concerning properties of deeply recursive constructions.

**Parsing Difficulties and Phonological
Processing in Italian**

Rodolfo Delmonte

Istituto di Linguistica e Didattica

Ca' Garzoni-Moro - S.Marco 3417

Universita degli Studi di Venezia(I)

Proc. EACL 1985, pp. 136-145

A recognition grammar to supply information to a text-to-speech system for the synthesis of Italian must rely heavily upon lexical information, in order to instantiate the appropriate grammatical relations.

Italian is an almost free word order language which nonetheless adopts fairly analysable strategies to move major constituents: some of these can strongly affect the functioning of the phonological component. Two basic claims will be made: (i) Difficulties in associating grammatical functions to constituent structure can be overcome only if Lexical Theory is adopted as a general theoretical framework, and translated into adequate computational formalisms like ATN or CHART; (ii) Decisions made at a previous point affect focus structure construal rules, which are higher level phonological rules which individuate intonation centre, build up adequate Intonational Groups and assign pauses to adequate sites, all being very sensitive to syntactic and semantic information.

We will concentrate on Subject/object function association to c-structure in Italian, and its relation to ATN formalism, in particular HOLD mechanism and FLAGging. Then we will show how syntactic decisions interact with an intonation grammar. We shall also introduce two functional notions: STRUCTURE REVERSIBILITY vs. FUNCTIONAL REVERSIBILITY in Italian.

**Design and Implementation of a Lexical
Data Base**

Eric Wehrli

Department of Linguistics

U.C.L.A.

405 Hilgard Avenue

Los Angeles, CA 90024

Proc. EACL 1985, pp. 146-153

This paper is concerned with the specifications and the implementations of a particular concept of word-based lexicon to be used for large natural language processing systems such as machine translation systems, and compares it with the morpheme-based conception of the lexicon traditionally assumed in computational linguistics.

It will be argued that, although less concise, a relational word-based lexicon is superior to a morpheme-based lexicon from a theoretical, computational and also practical viewpoint.

Lexifanis: A Lexical Analyzer of Modern Greek*Yannis Kotsanis, Yanis Maistros*

Computer Science Department
National Tech. University
Herion Polytechniou 9
GR - 157 73 - Athens, Greece
Proc. EACL 1985, pp. 154-158

Lexifanis is a Software Tool designed and implemented by the authors to analyze Modern Greek Language. This system assigns grammatical classes (parts of speech) to 95-98% of the words of a text which is read and normalized by the computer.

By providing the system with the appropriate grammatical knowledge (i.e.: dictionaries of non-inflected words, affixation morphology and limited surface syntax rules) any "variant" of Modern Greek Language (dialect or idiom) can be processed.

In designing the system, special consideration is given to the Greek Language morphological characteristics, primarily to the inflection and the accentuation.

In Linguistics, lexifanis can assist the generation of indexes or lemmata; on the other hand readability or style analysis can be performed using this software as a basic component. In Word Processing this software may serve as a background to build dictionaries for a spelling checking and error detection package.

Through this study our research group has set the basis in designing an "expert system" which is intended to "understand" and process Modern Greek texts. Lexifanis is the first working tool for Modern Greek Language.

A Probabilistic Approach to Grammatical Analysis of Written English by Computer*Andrew David Beale*

Unit for Computer Research on the English Language
University of Lancaster, Bowland College
Bailrigg, Lancaster, England LA1 4YT
Proc. EACL 1985, pp. 159-165

Work at the Unit for Computer Research on the English Language at the University of Lancaster has been directed towards producing a grammatically annotated version of the Lancaster-Oslo/Bergen (LOB) Corpus of written British English texts as the preliminary stage in developing computer programs and data files for providing a grammatical analysis of unrestricted English text.

From 1981-83, a suite of PASCAL programs was devised to automatically produce a single level of grammatical description with one word tag representing the word class or part of speech of each word token in the corpus. Error analysis and subsequent modification to the system resulted in over 96 per cent of word tags being correctly assigned automatically. The remaining 3 to 4 per cent were corrected by human post-editors.

Work is now in progress to devise a suite of programs to provide a constituent analysis of the sentences in the corpus. So far, sample sentences have been automatically assigned phrase and clause tags using a probabilistic system similar to word tagging. It is hoped that the entire corpus will eventually be parsed.

A Probabilistic Parser*Roger Garside, Fanny Leech*

Unit for Computer Research on the English Language
University of Lancaster, Bowland College
Bailrigg, Lancaster, England LA1 4YT
Proc. EACL 1985, pp. 166-170

The UCREL team at the University of Lancaster is engaged in the development of a robust parsing mechanism, which will assign the appropriate grammatical structure to sentences in unconstrained English text. The techniques used involve the calculation of probabilities for competing structures, and are based on the techniques successfully used in tagging (i.e., assigning grammatical word classes) to the LOB (Lancaster-Oslo/Bergen) corpus.

The first step in the parsing process involves dictionary lookup of successive pairs of grammatically tagged words, to give a number of possible continuations to the current parse. Since this lookup will often not be able unambiguously to distinguish the point at which a grammatical constituent should be closed, the second step of the parsing process will have to insert closures and distinguish between alternative parses. It will generate trees representing these possible alternatives, insert closure points for the constituents, and compute a probability for each parse tree from the probability of each constituent within the tree. It will then be able to select a preferred parse or parses for output.

The probability of a grammatical constituent is derived from a bank of manually parsed sentences.

Towards a Dictionary Support Environment for Real Time Parsing

*Hiyun Alshawi, Bran Boguraev,
Ted Briscoe*

Computer Laboratory
Cambridge University
Corn Exchange Street
Cambridge CB2 3QG, UK
Proc. EACL 1985, pp. 171-178

In this article we describe research on the development of large dictionaries for natural language processing. We detail the development of a dictionary support environment linking a restructured version of the *Longman Dictionary of Contemporary English* to natural language processing systems. We describe the process of restructuring the information in the dictionary and our use of the Longman grammar code system to construct dictionary entries for the PATR-II parsing system and our use of the Longman word definitions for automated word sense classification.

Towards a New Type of Morphemic Analysis

Eva Koktova

9. kevtna 1576
39001 Tabor, Czechoslovakia
Proc. EACL 1985, pp. 179-186

The present paper provides a report on a new system of an automated morphemic analysis of technical texts in Czech as a highly inflectional language, which is being prepared by the linguistic team of the Faculty of Mathematics and Physics in Prague, within the project of man-machine communication without a pre-arranged data base (TIBAQ). The kind of morphemic analysis presented here is based on a retrograde (right-to-left) analysis of words by means of morphemically unambiguous or irresolvably ambiguous word-ends, which do not coincide with the etymological word-endings but correspond to the structure of the accidental cases of morphemic ambiguity in an inflectional language (word-endings being accountable for in a certain way by word-ends). The algorithm of analysis can thus dispense with any dictionary (of morphemic irregularities and exceptions), economically accounting especially for productive word-endings. The word-ends of the analysis are assigned several kinds of morphemic information, concerning morphemic categories and lemmatization. The analysis is based on the absolute frequency of word-ends in technical texts and is able to interact with the semantic analysis.

A Computational Theory of Prose Style for Natural Language Generation

*David D. McDonald,
James D. Pustejovsky*

Department of Computer and
Information Science
University of Massachusetts at Amherst
Proc. EACL 1985, pp. 187-193

In this paper we report on initial research we have conducted on a computational theory of prose style. Our theory speaks to the following major points:

1. Where in the generation process style is taken into account.
2. How a particular prose style is represented; what "stylistic rules" look like;
3. What modifications to a generation algorithm are needed; what the decision is that evaluates stylistic alternatives;
4. What elaborations to the normal description of surface structures are necessary to make it usable as a plan for the text and a reference for these decisions;
5. What kinds of information decisions about style have access to.

Our theory emerged out of design experiments we have made over the past year with our natural language generation system, the Zetalisp program MUMBLE. In the process we have extended MUMBLE through the addition of an additional process that now mediates between content planning and linguistic realization. This new process, which we call "attachment", provides the further significant benefit that text structure is no longer dictated by the structure of the message; the sequential order and dominance relationships of concepts in the message no longer force one form onto the words and phrases in the text. Instead, rhetorical and intentional directives can be interpreted flexibly in the context of the ongoing discourse and stylistic preferences. The text is built up through composition under the direction of linguistic organizing principles, rather than having to follow conceptual principles in lockstep.

We will begin by describing what we mean by prose style and then introducing the generation task that lead us to this theory, the reproduction of short encyclopedia articles on African tribes. We will then use that task to outline the parts of our theory and the operations of the attachment process. Finally, we will compare our techniques to the related work of

An English Generator for a Case-Labelled Dependency Representation

John Irving Tait

Fulborn Road, Cherry Hinton

Cambridge CB1 4JN, UK

Proc. EACL 1985, pp. 194-197

Davey, McKeown and Derr, and Gabriel, and consider some of the possible psycholinguistic hypotheses that it may lead to.

The paper describes a program which has been constructed to produce English strings from a case-labelled dependency representation. The program uses an especially simple and uniform control structure with a well-defined separation of the different knowledge sources used during generation. Furthermore, the majority of the system's knowledge is expressed in a declarative form, so in principle the generator's knowledge bases could be used for purposes other than generation. The generator uses a two-pass control structure, the first translating from the semantically oriented, case-labelled dependency structures into surface syntactic trees and the second translating from these trees into English strings.

The generator is very flexible: it can be run in such a way as to produce all the possible syntactically legitimate variations on a given utterance, and has built in facilities to do some synonym substitution. It has been used in a number of application domains: notably as a part of a free text retrieval system and as part of a natural language front end to a relational database system.

Augmented Dependency Grammar: A Simple Interface between the Grammar Rule and the Knowledge

Kazunori Muraki, Shunji Ichiyama

C & C Systems Research Laboratories

NEC Corporation

Kawasaki-city, 213 JAPAN

Yasutomo Fukumochi

Software Development Division

NSIS Corporation

Kawasaki-city, 213 JAPAN

Proc. EACL 1985, pp. 198-204

This paper describes some operational aspects of a language comprehension model which unifies the linguistic theory and the semantic theory in respect to operations. The computational model, called Augmented Dependency Grammar (ADG), formulates not only the linguistic dependency structure of sentences but also the semantic dependency structure using the extended deep case grammar and field-oriented fact-knowledge based inferences. Fact knowledge base and ADG model clarify the qualitative difference between what we call semantics and logical meaning. From a practical viewpoint, it provides clear image of syntactic/semantic computation for language processing in analysis and synthesis. It also explains the gap in semantics and logical meaning, and gives a clear computational image of what we call conceptual analysis.

This grammar is used for analysis of Japanese and synthesis of English, in the Japanese-to-English machine translation system called VENUS (Vehicle for Natural Language Understanding and Synthesis) currently being developed by NEC.

A Natural Language Interface Using a World Model

Yoshio Izumida, Hiroshi Ishikawa,

Toshiaki Yoshino, Tadashi Hoshiai,

Akifumi Makinouchi

Software Laboratory

Fujitsu Laboratories Ltd.

1015 kamikodar.aka, Nakahara-ku

Kawasaki, 211 JAPAN

Proc. EACL 1985, pp. 205-212

Databases are nowadays used by varied and diverse users, many of whom are unfamiliar with the workings of a computer, but who, nevertheless, want to use those databases more easily. Rising to meet this demand, the authors are developing a Japanese language interface, called KID, as a database front-end system. KID incorporates a world model representing application and database knowledge to help make databases easier to use. KID has the following features: (1) parser extendibility and robustness, (2) independence from the application domain, (3) ease of knowledge editing, (4) independence from the database. This paper focuses on the first three features. KID has already been applied to the fields of housing, sales, and drug testing, thus confirming its transportability and practicality.

Interpreting Singular Definite Descriptors in Database Queries

Genevieve Berry-Rogghe

Department of Computer and Information Science

Temple University

Philadelphia, PA 19122

Proc. EACL 1985, pp. 213-217

The paper examines some of the characteristic features of natural language interaction with a database system and its implications for the processing of singular definite descriptions. Some proposals are made for assessing the uniqueness claim of the singular definite article in the context of retrieval from a relational database. Other standard assumptions such as the extensional evaluation and referent evaluation exclusively in the database – rather than within the discourse model – are critically examined.

Non Standard Uses of If*D.S. Bree, R.A. Smith*

Rotterdam School of Management

Erasmus University

P.O. Box 1738

3000 DR Rotterdam, The Netherlands

Proc. EACL 1985, pp. 218-225

The present study examines the semantic problems involved in computing the meaning of the non standard uses of *if*. The central question is whether or not it is necessary to introduce different meanings of *if*.

Austin proposed two non standard meaning for *if*. We show that these can be accounted for by the standard meaning together with shifts in the the position of the speech set within the sentence. These uses of *if* are among the 9 different non standard uses which we found in a sample of *if* sentences taken from the Brown University corpus:

1. Counterfactual:
If E had stuck to his plan he'd still be famous.
2. Factual:
If R was a liar, he was also a canny gentleman.
3. Conditional speech act:
You may come back to Strasbourg, now, if you wish.
4. Performative speech act:
He vowed vengeance on L, if ever the chance came his way.
5. Noun clause:
He wondered if the audience would let him finish.
6. Doubtful presupposition:
Perfect entities, if they move at all, don't move to . . .
7. Restrictive:
Social relations impose courtesy, if not sympathy, . . .
8. Concessive:
9. Protasis only: "If you want to see -" "Never mind", she said sternly.

Each use was examined to see whether it could be accounted for by the standard meaning of *if*, together with other features of the sentence. Similar differences in usage should then be found with other SCs. This was the case for the first four uses. In three uses (6, 7, 8) *if* may/must occur in a phrase rather than in a full clause. The hypothesis that these uses can be derived from the standard meaning of *if* in an equivalent clause was explored and rejected. Two of these uses (6, 7) require a material implication interpretation of *if*, also necessary for a few of the standard conditional sentences.

Two uses (5, 9) require only that the truth value of the following clause/phrase is unspecified. This is a property that all the uses have in common (with the exception of the factual use where the truth of the protasis is used to emphasize the truth of the apodosis) and is thus the feature that relates the different meanings of *if*. The standard use and the non standard uses using the standard meaning (1, 2, 3, 4) require, in addition, that there is an inference relation from the protasis (the *if* sub clause) to the apodosis (the main clause in which the *if* clause is embedded).

So we propose that three different meanings of *if* are required, inference (including the standard use), material implication (uses 6, 7) and just doubting the truth value of the following proposition (uses 5, 9). Each of these three uses may be expected to be translated by different words in other languages, e.g., in Dutch by *ais*, *zo*, and *of* (except for use 8) respectively.

Using a Text Model for Analysis and Generation*E. Fimbel, H. Groscolt, J.M. Lancel,**N. Simonin*

CAP SOGETI INNOVATION

129, rue de l'Universite

75007 Paris, France

Proc. EACL 1985, pp. 226-231

The following paper concerns a general scheme for multilingual text generation, as opposed to just translation. Our system processes the text as a whole, from which it extracts a representation of the meaning of the text. From this representation, a new text is generated, using a text model and action rules.

This process is done in six steps: word analysis, sentence analysis using a Functional Grammar, reference solving and inference, construction of the text pattern, sentence generation, and word generation. Different kinds of information are used at each step of the process: text organization, syntax, semantic, etc.

The Simulation of Stress Patterns in Synthetic Speech -- A Two-Level Program

Timothy J. Gillett

Department of Artificial Intelligence
University of Edinburgh
Hope Park Square
Edinburgh EH9 2NH Scotland
Proc. EACL 1985, pp. 232-238

Automated Speech Recognition: A Framework for Research

Anne Johnstone

Department of Artificial Intelligence
Edinburgh University
Hope Park Square, Meadow Lane
Edinburgh EH8 9LL (GB)
Gerry Altmann
Department of Linguistics
Edinburgh University
George Square, Edinburgh EH8 9LL (GB)
Proc. EACL 1985, pp. 239-243

A Rule-Based Approach to Evaluating Importance in Descriptive Texts

Daniilo Fum, Giovanni Guida, Carla Tasso

Istituto di Matematica, Informatica e Sistemistica
Universita di Udine
Udine, Italy
Proc. EACL 1985, pp. 244-250

A Problem Solving Approach to Generating Text from Systemic Grammars

Terry Patten

All the knowledge, as well as the text, is given in a declarative manner. It is expressed in a single formalism named Functional Descriptions. It consists of lexical data, a Functional Grammar, a knowledge network, action rules for reference solving and sentence generation, models of text, rules of structuration, and sentence schema.

Text representation, included in the semantic network, is composed of different kinds of objects (not necessarily distinct): text organization, syntactical information, objects introduced by the discourse, affirmations on these objects, and links between these affirmations.

This paper is part of an MSc. report on a program called GENIE (Generator of Inflected English), written in CProlog, that acts as a front end to an existing speech synthesis program. It allows the user to type a sentence in English text, and then processes it so that the synthesizer will output it with natural-sounding inflection; that is, as well as transcribing text to a phonemic form that can be read by the system, it assigns this text an f0 contour. The assigning of this stress is described in this paper, and it is asserted that the problem can be solved with reference to two main levels, the sentential and the syllabic.

This paper reflects the view that the decoding of speech, either by computer systems or people, must to a large extent be determined by the ways in which the speaker has encoded the information necessary for its comprehension. We therefore place great emphasis on the use of psycholinguistics as a tool for the construction of models essential to the characterisation of the speech understanding task.

We are primarily concerned with the interactions *between* the various levels at which a fragment of speech can be described (e.g., acoustic-phonetic, lexical, syntactic, etc.), and the ways in which the knowledge bases associated with each of the "levels" contribute towards a final interpretation of an utterance. We propose to use the Chart Parser as a general computational framework for simulating such interactions, since its flexibility allows various models to be implemented and evaluated.

Within this general framework we discuss problems of information flow and search strategy in combining evidence across levels of description and across time, during the extension of an hypothesis. We stress the importance of both psychological and computational theory in developing a particular control strategy which could be implemented within the framework.

Importance evaluation is one of the most challenging problems in the field of text processing. In the paper we focus on the notion of importance from a computational standpoint, and we propose a procedural, rule-based approach to importance evaluation. This novel approach is supported by a prototype experimental system, called importance evaluator, that can deal with descriptive texts taken from computer science literature on operating systems. The evaluator relies on a set of importance rules that are used to assign importance values to the different parts of a text and to resolve or explain conflicting evaluations. The system utilizes world knowledge on the subject domain contained in an encyclopedia and takes into account a goal assigned by the user for specifying the pragmatic aspects of the understanding activity. The paper describes the role of the evaluator in the frame of a larger system for text summarization (SUSY); it illustrates its overall mode of operation, and discusses some meaningful examples.

Systemic grammar has been used for AI text generation work in the past, but the implementations have tended to be ad hoc or inefficient. This paper presents an approach to systemic text generation where AI problem

Department of Artificial Intelligence
Edinburgh University
Hope Park Square, Meadow Lane
Edinburgh EH8 9NW Scotland
Proc. EACL 1985, pp. 251-257

GEMS: A Model of Sentence Production

Domenico Parisi, Alessandra Giorgi
Istituto di Psicologia del C.N.R.
Reparto Processi Cognitivi e
Intelligenza Artificiale
Via dei Monti Tiburtini, 509
00157 Roma, Italy
Proc. EACL 1985, pp. 258-262

Towards an Automatic Identification of Topic and Focus

Eva Hajicova, Petr Sgall
Faculty of Mathematics and Physics
Charles University
Malcstranske n. 25
118 00 Praha 1, Czechoslovakia
Proc. EACL 1985, pp. 263-267

User Modelling, Dialog Structure, and Dialog Strategy in HAM-ANS

Katharina Morik
Technische Universitaet Berlin
Project group KIT, Sekr, FR 5-8
Franklinstr. 28/29
D-1000 Berlin 10 (Fed. Repl. Germany)
Proc. EACL 1985, pp. 268-274

The Structure of Communicative Context of Dialogue Interaction

A.S. Narin'yan, O.P. Simonova
AI Laboratory, Computer Center
Siberian Division of the USSR Ac.Sci.
Novosibirsk 630090, USSR

solving techniques are applied directly to an unadulterated systemic grammar. This approach is made possible by a special relationship between systemic grammar and problem solving; both are organized primarily as choosing from alternatives. The result is simple, efficient text generation, firmly based in a linguistic theory.

The paper describes GEMS, a system for Generating and Expressing the Meaning of Sentences, focussing on the generation task, i.e., how GEMS extracts a set of propositional units from a knowledge store that can be expressed with a well-formed sentence in a target language. GEMS is lexically distributed. After a central processor has selected the first unit(s) from the knowledge store and activated the corresponding lexical entry, the further construction of the sentence meaning is entrusted to the entries in the vocabulary. Examples of how GEMS constructs the meaning of a number of English sentence types are briefly described.

The purpose of this paper is (i) to substantiate the claim that the output of an automatic analysis should represent, among other things, the hierarchy of topic-focus articulation, and (ii) to present a general procedure for determining the topic-focus articulation in Czech and English.

(i) The following requirements on the output of an automatic analysis are significant:

- (a) in the output of the analysis it should be marked which elements of the the analyzed sentence belong to its topic and which to the focus;
- (b) the scale of communicative dynamism (CD) should also be identified for every representation of a meaning of the analyzed sentence, since the degrees of CD correspond to the unmarked distribution of quantifier scopes in the semantic interpretation of the sentence;
- (c) the analysis should also distinguish topicless sentences from those having a topic, which is relevant for the scope of negation.

(ii) For an automatic recognition of topic, focus and the degrees of CD, two points are crucial:

- (a) either the input language has (a considerable degree of) the so-called free word order (as in Czech, Russian), or its word order is determined mainly by the grammatical relations (as in English, French);
- (b) either the input is spoken discourse (and the recognition procedure includes an acoustic analysis), or written (printed) texts are analyzed.

In accordance with these points, a general procedure for determining topic, focus and the degree of CD is formulated for Czech and English, with some hints how the preceding context can be taken into account.

AI dialog systems are now developing from question-answering systems toward advising systems. This includes

- structuring dialog
- understanding and generating a wider range of speech acts than simply information request and answer
- user modelling

User modelling in HAM-ANS is closely connected to dialog structure and dialog strategy. In advising the user, the system generates and verbalizes speech acts. The choice of the speech act is guided by the user profile and the dialog strategy of the system.

We propose a draft scheme of the model formalizing the structure of communicative context in dialogue interaction. The relationships between the interacting partners are considered as a system of three automata representing the partners of the dialogue and environment.

The communicative competence of the partners is defined by

- the set M of all propositions reflecting the possible states of the three

Proc. EACL 1985, pp. 274-276

- automata within the model;
- the set K of "contracts" representing all kinds of human-to-human relationships (social, interpersonal, professional, etc.) which include fixation of particular roles for the partners;
 - the set T of possible topics related to given "contract".
- The authors believe the system of the notions presented may be used as a basis for forming the communicative component in the dialogue system.