

LIBRARY OF CONGRESS



0 014 794 944 3

INTERNATIONAL  
CONFERENCE ON  
COMPUTATIONAL  
LINGUISTICS

COLING

94

P 98

.I55

1994

Vol 2

Copy 1

# COLING 94

The 15th International Conference on Computational Linguistics

## PROCEEDINGS

Vol. II



August 5 – 9, 1994

Kyoto, Japan



# **COLING 94**

**The 15th International Conference on Computational Linguistics**

## **PROCEEDINGS**

**Vol. II**

August 5 – 9, 1994

Kyoto, Japan

Pg8

I55

1994

vol. 2

Copy 2

2006438697

## PREFACE

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

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

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

Professor Makoto Nagao  
Conference Chairman

## Table of Contents (Vol. I)

<b>Machine Translation</b>	
<i>Improvement in Customizability Using Translation Templates</i> .....	25
S. Kinoshita, A. Kumano, H. Hirakawa (Toshiba Corp.)	
<i>Countability and Number in Japanese to English Machine Translation</i> .....	32
F. Bond, K. Ogura, S. Ikehara (NTT Communication Science Labs.)	
<i>A Method for Distinguishing Exceptional and General Examples in Example-based Transfer Systems</i> .....	39
H. Watanabe (IBM Japan, Ltd.)	
<i>Interpreting Compounds for Machine Translation</i> .....	45
B. Gawronska, A. Nordner, C. Johansson, C. Willners (University of Lund)	
<i>Towards Machine Translation Using Contextual Information</i> .....	51
T. Cornish, K. Fujita, R. Sugimura (Matsushita Electric Industrial Co., Ltd.)	
<i>Two Methods for Learning ALT-J/E Translation Rules from Examples and a Semantic Hierarchy</i> .....	57
H. Almuallim, Y. Akiba, T. Yamazaki, A. Yokoo, S. Kaneda (King Fahd Univ. of Petroleum and Minerals/NTT Communication Sci. Labs)	
<i>A Bidirectional, Transfer-driven Machine Translation System for Spoken Dialogues</i> .....	64
Y. Sobashima, O. Furuse, S. Akamine, J. Kawai, H. Iida (ATR Interpreting Telecommunication Research Laboratories)	
<i>Treating 'Free Word Order' in Machine Translation</i> .....	69
R. Steinberger (UMIST)	
<i>Building an MT Dictionary from Parallel Texts Based on Linguistic and Statistical Information</i> .....	76
A. Kumano, H. Hirakawa (Toshiba Corp.)	
<i>Dilemma - An Instant Lexicographer</i> .....	82
H. Karlgren, J. Karlgren, M. Nordström, P. Pettersson, B. Wahrolén (Swedish Institute of Computer Science)	
<i>Portable Knowledge Sources for Machine Translation</i> .....	85
K. Takeda (IBM Research)	
<i>Coping with Ambiguity in a Large-Scale Machine Translation System</i> .....	90
K. L. Baker, A. M. Franz, P. W. Jordan, T. Mitamura, E. H. Nyberg, 3rd (Carnegie Mellon University)	
<i>Evaluation Metrics for Knowledge-Based Machine Translation</i> .....	95
E. H. Nyberg, 3rd, T. Mitamura, J. G. Carbonell (Carnegie Mellon University)	
<i>A Matching Technique in Example-Based Machine Translation</i> .....	100
L. Cranias, H. Papageorgiou, S. Piperidis (Institute for Language & Speech Processing, Greece)	
<i>Constituent Boundary Parsing for Example-Based Machine Translation</i> .....	105
O. Furuse, H. Iida (ATR Interpreting Telecommunications Research Labs.)	
<i>The JaRAP Experimental System of Japanese-Russian Automatic Translation</i> .....	112
L. S. Modina, Z. M. Shalyapina (Russian Academy of Sciences)	
<i>Perspectives of DBMT For Monolingual Authors on the Basis of LIDIA-1, an Implemented Mock-up</i> .....	115
H. Blanchon (GETA-IMAG)	

<i>Modals as a Problem for MT</i> .....	120
B. Sigurd, B. Gawróńska (Lund University)	
<i>Two Types of Adaptive MT Environments</i> .....	125
S. Nirenburg, R. Frederking, D. Farwell, Y. Wilks (Carnegie Mellon Univ./New Mexico State Univ./Univ. of Sheffield)	
<i>An English-to-Korean Machine Translator:MATES/EK</i> .....	129
K.-S. Choi, S. Lee, H. Kim, D.-B. Kim, C. Kweon, G. Kim (Korea Advanced Institute of Science and Technology)	
<i>Interlanguage Signs and Lexical Transfer Errors</i> .....	134
Atle Ro (University of Bergen)	

### **Morphology & Tagging**

<i>Morphology with a Null-Interface</i> .....	141
H. Trost, J. Matiassek (Austrian Research Institute for Artificial Intelligence)	
<i>Automatic Model Refinement - with an Application to Tagging</i> .....	148
Y.-C. Lin, T.-H. Chiang, K.-Y. Su (National Tsing Hua University)	
<i>Disambiguation of Super Parts of Speech (or SUPERTAGS):Almost Parsing</i> .....	154
A. K. Joshi, B.Srinivas (University of Pennsylvania)	
<i>Probabilistic Tagging with Feature Structures</i> .....	161
A. Kempe (Univ. of Stuttgart)	
<i>A Part-of-Speech-Based Alignment Algorithm</i> .....	166
K.-H. Chen (National Taiwan University)	
<i>Part-of-Speech Tagging with Neural Networks</i> .....	172
H. Schmid (Univ. of Stuttgart)	
<i>The Rumors System of Russian Synthesis</i> .....	177
M. I. Kanovich, Z. M.Shalyapina (Russian Academy of Sciences)	
<i>Multi-Tape Two-Level Morphology: A Case Study in Semitic Non-linear Morphology</i> .....	180
G. A. Kiraz (University of Cambridge)	
<i>An Evaluation to Detect and Correct Erroneous Characters Wrongly Substituted, Deleted and Inserted in Japanese and English Sentences Using Markov Models</i> .....	187
T. Araki, S. Ikehara, N. Tsukahara, Y. Komatsu (Fukui Univ./NTT Communication Sci. Labs.)	
<i>An Efficient Treatment of Japanese Verb Inflection for Morphological Analysis</i> .....	194
T. Hisamitsu, Y. Nitta (ARL, Hitachi Ltd.)	
<i>A Stochastic Japanese Morphological Analyzer Using a Forward-DP Backward-A*N-Best Search Algorithm</i> .....	201
M. Nagata (NTT Network Information Systems Labs.)	
<i>Backtracking-Free Dictionary Access Method for Japanese Morphological Analysis</i> .....	208
H. Maruyama (IBM Research)	
<i>Modularity in a Connectionist Model of Morphology Acquisition</i> .....	214
M. Gasser (Indiana University)	
<i>Syllable-Based Model for the Korean Morphology</i> .....	221
S.-S. Kang, Y. T. Kim (Hansung Univ./Seoul National Univ.)	
<i>Segmenting a Sentence into Morphemes Using Statistic Information between Words</i> .....	227
S. Nobesawa, J. Tsutsumi, T. Nitta, K. Ono, S. D. Jiang, M. Nakanishi (Keio University)	
<i>'Derivational' Paradigms in Morphonology</i> .....	234
V. Pirrelli, S. Federici (ILC-CNR)	

## Lexicon

<i>An Architecture for a Universal Lexicon. A Case Study on Shared Syntactic Information in Japanese, Hindi, Bengali, Greek, and English</i> .....	243
N. Nomura, D. A. Jones, R. C. Berwick (Massachusetts Institute of Technology)	
<i>Adjuncts and the Processing of Lexical Rules</i> .....	250
G. V. Noord, G. Bouma (BCN RUG)	
<i>Noun Phrasal Entries in the EDR English Word Dictionary</i> .....	257
A. Koizumi, M. Arioka, C. Harada, M. Sugimoto (Japan Electric Dictionary Research Institute)	
<i>KASSYS: A Definition Acquisition System in Natural Language</i> .....	263
P. Hernert (L.I.F.O., Research Lab. de Vinci)	
<i>Complex Syntax: Building a Computational Lexicon</i> .....	268
R. Grishman, C. Macleod, A. Meyers (New York University)	
<i>Word Knowledge Acquisition, Lexicon Construction and Dictionary Compilation</i> .....	273
A. Sanfilippo (SHARP Laboratories of Europe, Ltd.)	
<i>Interlingual Lexical Organisation for Multilingual Lexical Databases in NADIA</i> .....	278
G. Sérasset (GETA IMAC-campus)	
<i>Manipulating Human-Oriented Dictionaries with Very Simple Tools</i> .....	283
J. Gaschler, M. Lafourcade (GETA, IMAG-campus)	
<i>Towards Linguistic Knowledge Discovery Assistants: Application to Learning Lexical Properties of Chinese Characters</i> .....	287
G. Fafiotte, F. Tcheou (GETA, IMAG)	
<i>Logic Compression of Dictionaries for Multilingual Spelling Checkers</i> .....	292
B. Meddeb-Hamrouni (GETA, IMAG-campus)	
<i>Construction of a Bilingual Dictionary Intermediated by a Third Language</i> .....	297
K. Tanaka, K. Umemura (Univ. of Tokyo/NTT Basic Research Labs.)	
<i>Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries</i> .....	304
Y. Niwa, Y. Nitta (ARL, Hitachi Ltd.)	
<i>Analysis of Scene Identification Ability of Associative Memory with Pictorial Dictionary</i> .....	310
T. Tsunoda, H. Tanaka (University of Tokyo)	

## Generation

<i>Anticipating the Reader's Problems and the Automatic Generation of Paraphrases</i> .....	319
N. Lenke (Univ. of Duisburg)	
<i>TGE: Tlinks Generation Environment</i> .....	324
A. Ageno, F. Ribas, G. Rigau, H. Rodríguez, A. Samiotou (Univ. Politècnica de Catalunya)	
<i>Planning Argumentative Texts</i> .....	329
X. Huang (Universität des Saarlandes)	
<i>A Hybrid Approach to the Automatic Planning of Textual Structures</i> .....	334
G. Zhu, N. Shadbolt (Nottingham University)	
<i>Generating Multilingual Documents from a Knowledge Base: The TECHDOC Project</i> .....	339
D. Rosner, M. Stede (FAW Ulm)	
<i>Abstract Generation Based on Rhetorical Structure Extraction</i> .....	344
K. Ono, K. Sumita, S. Miike (Toshiba Corp.)	
<i>The Correct Place of Lexical Semantics in Interlingual MT</i> .....	349
L. Levin, S. Nirenburg (Carnegie Mellon University)	

<i>Default Handling in Incremental Generation</i> .....	356
K. Harbusch, G. Kikui, A. Kilger (German Research Center for Artificial Intelligence)	
<i>English Generation from Interlingua by Example-Based Method</i> .....	363
E. Komatsu, J. Cui, H. Yasuhara (Japan Electric Dictionary Res. Inst.)	
<i>On Lexically Biased Discourse Organization in Text Generation</i> .....	369
L. Wanner (University of Stuttgart)	

## Parsing

<i>Concurrent Lexicalized Dependency Parsing: The Parse Talk Model</i> .....	379
N. Bröker, U. Hahn, S. Schacht (Freiburg University)	
<i>Notes on LR Parser Design</i> .....	386
C. Samuelsson (Swedish Institute of Computer Science)	
<i>Parsing a Flexible Word Order Language</i> .....	391
V. Pericliev, A. Grigorov (Institute of Mathematics with Computing Center, Bulgaria)	
<i>Parsing as Tree Traversal</i> .....	396
D. Gerdemann (Universität Tübingen)	
<i>LR(k)-Parsing of Coupled-Context-Free Grammars</i> .....	401
G. Pitsch (Universität des Saarlandes)	
<i>Constructing Lexical Transducers</i> .....	406
L. Karttunen (Rank Xerox Research Centre, Grenoble)	
<i>An HPSG Parser Based on Description Logics</i> .....	412
J. Joachim Quantz (Technische Universität Berlin)	
<i>An Efficient Parser Generator for Natural Language</i> .....	417
M. Ishii, K. Ohta, H. Saito (Fujitsu Inc./Apple Technology, Inc./Keio Univ.)	
<i>Exploring the Role of Punctuation in Parsing Natural Text</i> .....	421
B. E.M. Jones (University of Edinburgh)	
<i>The "Whiteboard" Architecture: A Way to Integrate Heterogeneous Components of NLP Systems</i> .....	426
C. Boitet, M. Seligman (GETA, IMAG/ATR Interpreting Telecommunications Res. Labs)	
<i>Two Parsing Algorithms by Means of Finite State Transducers</i> .....	431
E. Roche (Mitsubishi Electric Research Labs.)	
<i>DISCO—An HPSG-Based NLP System and its Application for Appointment Scheduling</i> ---	436
H. Uszkoreit, R. Backofen, S. Busemann, A. K. Diagne, E. A. Hinkelman, W. Kasper, B. Kiefer, H.-U. Krieger, K. Netter, G. Neumann, S. Oepen, S. P. Spackman (DFKI)	
<i>A Corpus-Based Learning Technique for Building a Self-Extensible Parser</i> .....	441
R.-L. Liu, V.-W. Soo (National Tsing-Hua University)	
<i>A "Not-so-shallow" Parser for Collocational Analysis</i> .....	447
R. Basili, M. T. Pazienza, P. Velardi (Università di Roma)	
<i>A Modular Architecture for Constraint-based Parsing</i> .....	454
F. Barthélemy, F. Rouaix (Universidade Nova de Lisboa/INRIA Rocquencourt)	
<i>Minimal Change and Bounded Incremental Parsing</i> .....	461
M. Wirén (Universität des Saarlandes)	
<i>Emergent Parsing and Generation with Generalized Chart</i> .....	468
K. Hashida (Electrotechnical Laboratory, Japan)	
<i>Syntactic-Head-Driven Generation</i> .....	475
E. König (Institute for Computational Linguistics, Germany)	

<i>PRINCIPAR-An Efficient, Broad-coverage, Principle-based Parser</i> .....	482
D. Lin (University of Manitoba)	
<i>Concurrent Lexicalized Dependency Parsing: A Behavioral View on Parse Talk Events</i> ---	489
S. Schacht, U. Hahn, N. Broker (Freiburg University)	
<i>Parsing Turkish Using the Lexical Functional Grammar Formalism</i> .....	494
Z. Güngördü, K. Oflazer (University of Edinburgh/Bilkent Univ.)	
<i>LHIP: Extended DCGs for Configurable Robust Parsing</i> .....	501
A. Ballim, G. Russell (ISSCO, University of Geneva)	
<i>Categorial Grammar and Discourse Representation Theory</i> .....	508
R. Muskens (Institute for Language Technology and Artificial Intelligence)	
<i>Towards Automatic Extraction of Monolingual and Bilingual Terminology</i> .....	515
B. Daille, E. Gaussier, J.-M. Langé (Talana Univ. Paris 7/Centre Scientifique IBM France)	

### Computational Linguistics

<i>FAX: An Alternative to SGML</i> .....	525
K. W. Church, W. A. Gale, J. I. Helfman, D. D. Lewis (AT&T Bell Labs.)	
<i>Referring to World Objects with Text and Pictures</i> .....	530
E. André, T. Rist (DFKI)	
<i>A Two-level Morphological Analysis of Korean</i> .....	535
D.-B. Kim, S.-J. Lee, K.-S. Choi, G.-C. Kim (Korea Advanced Institute of Science and Technology)	
<i>Character-based Collocation for Mandarin Chinese</i> .....	540
C.-R. Huang, K.-J. Chen, Y.-Y. Yang (Academia Sinica/National Taiwan Univ.)	
<i>Lexical Knowledge Representation in an Intelligent Dictionary Help System</i> .....	544
E. Agirree, X. Arregi, X. Artola, A. Díaz de Iñaraza, K. Sarasola (Univ. of Basque Country)	
<i>Reversible Resolution with an Application to Paraphrasing</i> .....	551
M. Hurst (The University of Sheffield)	
<i>Classifier Assignment by Corpus-Based Approach</i> .....	556
V. Sornlertlamvanich, W. Pantachat, S. Meknabin (Linguistics and Knowledge Science Lab., Ministry of Science Technology and Environment, Thailand)	

### Corpus-based NLP

<i>Annotating 200 Million Words: The Bank of English Project</i> .....	565
T. Järvinen (University of Helsinki)	
<i>Restructuring Tagged Corpora with Morpheme Adjustment Rules</i> .....	569
T. Tashiro, N. Uratani, T. Morimoto (ATR Interpreting Telecommunications Research Labs.)	
<i>Encoding Standards for Large Text Resources: The Text Encoding Initiative</i> .....	574
N. Ide (Vassar College)	
<i>INTEX: A Corpus Processing System</i> .....	579
M. D. Silberztein (Université Paris 7)	
<i>An IBM-PC Environment for Chinese Corpus Analysis</i> .....	584
R. W. P. Luk (City Polytechnic of Hong Kong)	
<i>Multext: Multilingual Text Tools and Corpora</i> .....	588
N. Ide, J. Véronis (CNRS & Université de Provence)	

<i>A Parser Coping with Self-Repaired Japanese Utterance and Large Corpus-Based Evaluation</i> .....	593
Y. Sagawa, N. Ohnishi, M. Sugie (Nagoya University)	
<i>A Tool for Collecting Domain Dependent Sortal Constraints from Corpora</i> .....	598
F. Andry, M. Gawron, J. Dowding, R. Moore (SRI International/CAP GEMINI Innovation)	
<i>Building a Lexical Domain Map from Text Corpora</i> .....	604
T. Strzalkowski (New York University)	
<i>A New Method of N-gram Statistics for Large Number of n and Automatic Extraction of Words and Phrases from Large Text Data of Japanese</i> .....	611
M. Nagao, S. Mori (Kyoto University)	
<i>Non-directionality and Self-Assessment in an Example-based System Using Genetic Algorithms</i> .....	616
Y. Lepage (Universiti Sains Malaysia)	
<i>CLAWS4: The Tagging of the British National Corpus</i> .....	622
G. Leech, R. Garside, M. Bryant (Lancaster University)	
<i>Syntactic Analysis of Natural Language Using Linguistic Rules and Corpus-Based Patterns</i> ----	629
P. Tapanainen, T. Järvinen (Rank Xerox Research Centre/Univ. of Helsinki)	

## Table of Contents (Vol. II)

### Semantics

<i>Word Sense Acquisition for Multilingual Text Interpretation</i> .....	665
P. S. Jacobs (GE Research and Development Center)	
<i>A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English</i> .....	672
H. Nakaiwa, A. Yokoo, S. Ikehara (NTT Communication Science Labs.)	
<i>Semantics of Complex Sentences in Japanese</i> .....	679
H. Nakagawa, S. Nishizawa (Yokohama National University)	
<i>Content Characterization Using Word Shape Tokens</i> .....	686
P. Sibun, D. S. Farrar (Fuji Xerox Palo Alto Lab.)	
<i>The Nature of Near-synonymic Relations</i> .....	691
C. DiMarco (University of Waterloo)	
<i>Virtual Polysemy</i> .....	696
A. Sanfilippo, K. Benkerimi, D. Dwehus (Sharp Labs. of Europe)	
<i>Building a Windows-based Bilingual Functional Semantic Processor</i> .....	701
J. J. Webster (City Polytechnic of Hong Kong)	
<i>On the Proper Role of Coercion in Semantic Typing</i> .....	706
J. Pustejovsky, P. Bouillon (Brandeis Univ./Univ. of Geneva)	
<i>Word Sense Ambiguation: Clustering Related Senses</i> .....	712
W. B. Dolan (Microsoft Corp.)	
<i>A Best-match Algorithm for Broad-coverage Example-based Disambiguation</i> .....	717
N. Uramoto (IBM Japan Ltd.)	
<i>Drawing Pictures with Natural Language and Direct Manipulation</i> .....	722
M. Hiyoshi, H. Shimazu (NEC Corp.)	
<i>Verbal Case Frame Acquisition from a Bilingual Corpus: Gradual Knowledge Acquisition</i> ---	727
H. Tanaka (NHK Science & Technical Research Labs.)	
<i>An Empirical Study on the Generation of Zero Anaphors in Chinese</i> .....	732
C.-L. Yeh, C. Mellish (University of Edinburgh)	
<i>Representing Information Need with Semantic Relations</i> .....	737
A. S. Chakravarthy (MIT Media Lab.)	
<i>Generalizing Automatically Generated Selectional Patterns</i> .....	742
R. Grishman, J. Sterling (New York University)	
<i>Incremental Interpretation: Applications, Theory, and Relationship to Dynamics Semantics</i>	748
D. Milward, R. Cooper (University of Edinburgh)	
<i>Word Sense Disambiguation and Text Segmentation Based on Lexical Cohesion</i> .....	755
M. Okumura, T. Honda (Japan Advanced Institute of Science and Technology)	
<i>Automatic Recognition of Verbal Polysemy</i> .....	762
F. Fukumoto, J. Tsujii (UMIST)	
<i>An Experiment on Learning Appropriate Selectional Restrictions from a Parsed Corpus</i> ---	769
F. R. Framis (Universitat Politècnica de Catalunya)	

<i>A Discrete Model of Degree Concept in Natural Language</i> .....	775
S. Kamei, K. Muraki (NEC Corp.)	
<i>Algorithm for Automatic Interpretation of Noun Sequences</i> .....	782
L. Vanderwende (Microsoft Corp.)	
<i>A Treatment of Functional Definite Descriptions</i> .....	789
H. Wada (Intelligent Text Processing, Inc.)	
<i>Bottom-up Earley Deduction</i> .....	796
G. Erbach (University of Saarland)	
<i>The Merged Upper Model: A Linguistic Ontology for German and English</i> .....	803
R. Henschel, J. Bateman (GMD/IIPI)	

### Syntax

<i>Modeling Dialogue by Functional Subcategorization</i> .....	813
J.R. Zubizarreta Aizpuru, C. Jones (U.P.V.-E.H.U./Univ. of Aberdeen)	
<i>Dutch Cross Serial Dependencies in HPSG</i> .....	818
G. Rentier (Tilburg University)	
<i>HPSG Lexicon without Lexical Rules</i> .....	823
K. Oliva (University of Saarland)	
<i>A Lexicon of Distributed Noun Representations Constructed by Taxonomic Traversal</i> .....	827
R. Sutcliffe, D. O'Sullivan, F. Meharg (University of Limerick)	
<i>Multi-modal Definite Clause Grammar</i> .....	832
H. Shimazu, S. Arita, Y. Takashima (NEC Corp.)	
<i>Hypothesis Selection in Grammar Acquisition</i> .....	837
M. Kiyono, J. Tsujii (UMIST)	
<i>Achieving Flexibility in Unification Formalisms</i> .....	842
L. Strömbäck (Linköping University)	
<i>TWP: How to Assist English Production on Japanese Word Processor</i> .....	847
K. Muraki, S. Akamine, K. Satoh, S. Ando (NEC Corp.)	
<i>Long-distance Dependencies and Applicative Universal Grammar</i> .....	853
S. Shaumyan, F. Segond (Yale University/Rank Xerox Res. Centre)	
<i>A Reestimation Algorithm for Probabilistic Recursive Transition Network</i> .....	859
Y. S. Han, K.-S. Choi (Korea Advanced Institute of Science and Technology)	
<i>Analysis of Japanese Compound Nouns Using Collocational Information</i> .....	865
Y. Kobayashi, T. Tokunaga, H. Tanaka (Tokyo Institute of Technology)	
<i>Free-ordered CUG on Chemical Abstract Machine</i> .....	870
S. Tojo (Mitsubishi Research Institute, Inc.)	
<i>Computing FIRST and FOLLOW Functions for Feature-theoretic Grammars</i> .....	875
A. Trujillo (University of Cambridge)	
<i>Focus on "Only" and "Not"</i> .....	881
A. Ramsay (University College Dublin)	
<i>Structure Sharing Problem and its Solution in Graph Unification</i> .....	886
K. Kogure (NTT Basic Research Labs.)	
<i>TDL-A Type Description Language for Constraint-based Grammars</i> .....	893
H.-U. Krieger, U. Schäfer (German Research Center for Artificial Intelligence)	
<i>On the Portability of Complex Constraint-based Grammars</i> .....	900
C.J. Rupp, R. Johnson (IDSIA)	

<i>A Grammar Based Approach to a Grammar Checking of Free Word Order Languages</i> ----	906
V. Kubon, M. Plátek (Charles University)	
<i>Table-driven Neural Syntactic Analysis of Spoken Korea</i> -----	911
W. Lee, G. Lee, J.-H. Lee (Phonag Institute of Science and Technology)	
<i>Universal Guides and Finiteness and Symmetry of Grammar Processing Algorithms</i> -----	916
M. Martinovic (New York University)	
<i>XTAG System-A Wide Coverage Grammar for English</i> -----	922
C. Doran, D. Egedi, B. A. Hockey, B. Srinivas, M. Zaidel (University of Pennsylvania)	
<i>Weakly Restricted Stochastic Grammars</i> -----	929
R. O. D. Akker, H. ter Doest (University of Twente)	
<i>Non-constituent Coordination: Theory and Practice</i> -----	935
D. Milward (University of Edinburgh)	
<i>Hypothesis Scoring over Theta Grids Information in Parsing Chinese Sentences with Serial Verb Constructions</i> -----	942
K. H.C. Lin, V.-W. Soo (National Tsing-Hua University)	
<i>An Efficient Syntactic Tagging Tool for Corpora</i> -----	949
M. Zhou, C. Huang (Tsinghua University)	
<i>The Correct and Efficient Implementation of Appropriateness Specifications for Typed Feature Structures</i> -----	956
D. Gerdemann, P. J. King (Universität Tübingen)	
<i>A Classification Method for Japanese Signs Using Manual Motion Descriptions</i> -----	961
H. Adachi, K. Kamata (Utsunomiya University)	

### Speech & Phonology

<i>Machine-readable Dictionaries in Text-to-speech Systems</i> -----	971
J. L. Klavans, E. Tzoukermann (Columbia University)	
<i>Issues in Text-to-speech for French</i> -----	976
E. Tzoukermann (AT&T Bell Labs.)	
<i>CHATR: A Generic Speech Synthesis System</i> -----	983
A. W. Black, P. Taylor (ATR Interpreting Telecommunications Labs.)	
<i>Pause as a Phrase Demarcator for Speech and Language Processing</i> -----	987
J. Hosaka, M. Seligman, H. Singer (ATR Interpreting Telephony Research Labs.)	
<i>The Parsody System: Automatic Prediction of Prosodic Boundaries for Text-to-speech</i> ---	992
S. Minnis (BT Labs.)	
<i>Anytime Algorithms for Speech Parsing?</i> -----	997
G. Görz, M. Kessler (University of Erlangen-Nürnberg)	
<i>Towards a Proper Linguistic and Computational Treatment of Scrambling: An Analysis of Japanese</i> -----	1002
S. Fong (NEC Research Institute, Inc.)	
<i>Phonological Derivation in Optimality Theory</i> -----	1007
M. T. Ellison (University of Edinburgh)	
<i>A Grammar and a Parser for Spontaneous Speech</i> -----	1014
M. Nakano, A. Shimazu, K. Kogure (NTT Basic Research Labs.)	
<i>Catching the Cheshire Cat</i> -----	1021
C. Johansson (Lund University)	

## Information Retrieval & Extraction

<i>A Dutch to SQL Database Interface Using Generalized Quantifier Theory</i> .....	1029
D. Speelman, G. Adriaens (University of Leuven)	
<i>A Methodology for Automatic Term Recognition</i> .....	1034
S. Ananiadou (Manchester Metropolitan University)	
<i>Knowledge Extraction from Texts: A Method for Extracting Predicate-argument Structures from Texts</i> .....	1039
F. Pugeault, P. Saint-Dizier, M.-G. Monteil (Université Paul Sabatier/EDF, D.E.R.)	
<i>Thesaurus-based Efficient Example Retrieval by Generating Retrieval Queries from Similarities</i> .....	1044
T. Utsuro, K. Uchimoto, M. Matsumoto, M. Nagao (Nara Institute of Science and Technology/Kyoto University)	
<i>Tools for Extracting and Structuring Knowledge from Texts</i> .....	1049
A. Ogonowski, M. L. Herviou, E. Dauphin (GSI-ERLI/AEROSPATIALE CCR/EDF)	
<i>N-Gram Cluster Identification During Empirical Knowledge Representation Generation</i> ..	1054
R. Collier (University of Sheffield)	
<i>Document Classification by Machine: Theory and Practice</i> .....	1059
L. Guthrie, E. Walker (New Mexico State University)	
<i>Pattern Matching in the Textract Information Extraction System</i> .....	1064
T. Kitani, Y. Eriguchi, M. Hara (Carnegie Mellon University)	
<i>Recognizing Text Genres with Simple Metrics Using Discriminant Analysis</i> .....	1071
J. Karlgren, D. Cutting (Swedish Institute of Computer Science/Apple Computer)	
<i>Bilingual Text Matching Using Bilingual Dictionary and Statistics</i> .....	1076
T. Utsuro, H. Ikeda, M. Yamane, Y. Matsumoto, M. Nagao (Nara Institute of Science and Technology/Kyoto University)	
<i>Towards a More User-friendly Correction</i> .....	1083
D. Genthial, J. Courtin, J. Ménézo (LGI-IMAG)	
<i>Reverse Queries in DATR</i> .....	1089
H. Langer (University of Osnabrück)	
<i>K-vec: A New Approach for Aligning Parallel Texts</i> .....	1096
P. Fung, K. W. Church (Columbia University/AT&T Bell Labs.)	

## Discourse & Pragmatics

<i>A Formal Representation of the Thematic-rhematic Structure of Sentences Based of a Typed A-calculus</i> .....	1105
Y. Uetake (Tokyo University of Mercantile Marine)	
<i>Customizing and Evaluating a Multilingual Discourse Module</i> .....	1109
C. Aone (Systems Research and Applications Corp. [SRA])	
<i>Consequence Relations in DRT</i> .....	1114
S. Akama, Y. Nakayama (Teikyo Univ. of Technology/Nihon Unisys Ltd.)	
<i>Collaboration on Reference to Objects that are not Mutually Known</i> .....	1118
P. G. Edmonds (University of Toronto)	
<i>Automatic Detection of Discourse Structure by Checking Surface Information in Sentences</i>	1123
S. Kurohashi, M. Nagao (Kyoto University)	

<i>Extending DRT with a Focusing Mechanism for Pronominal Anaphora and Ellipsis Resolution</i> .....	1128
J. Abraços, J. G. Lopes (CRIA/UNINOVA Faculdade de Ciências e Tecnologia)	
<i>Reference Resolution Using Semantic Patterns in Japanese Newspaper Articles</i> .....	1133
T. Wakao (University of Sheffield)	
<i>Exploiting Reference Interaction in Resolving Temporal Reference</i> .....	1138
K. Dohsaka (NTT Basic Research Labs.)	
<i>A Grammatico-statistical Approach to Discourse Partitioning</i> .....	1145
T. Nomoto, Y. Nitta (Hitachi Ltd., ARL)	
<i>Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-pronouns</i> .....	1151
S. Takada, N. Doi (Keio University)	
<i>Robust Method of Pronoun Resolution Using Full-text Information</i> .....	1157
T. Nasukawa (IBM Research)	
<i>Towards a Dynamic Theory of Belief-sharing in Cooperative Dialogues</i> .....	1164
H. Komatsu, N. Ogata, A. Ishikawa (Toin Corp./Univ. of Tsukuba/Sophia Univ.)	
<i>An Integrated Model for Anaphora Resolution</i> .....	1170
R. Mitkov (Institute of Mathematics)	
<i>Breaking Down Rhetorical Relations for the Purpose of Analysing Discourse Structures</i> --	1177
J. Fukumoto, J. Tsujii (UMIST)	
<i>Presupposition &amp; VP-Ellipsis</i> .....	1184
J. Bos (Universität des Saarlandes)	
<i>Communicating with Multiple Agents</i> .....	1191
E. A. Hinkelman, S. P. Spackman (DFKI)	
<i>A Rule-based Approach to Prepositional Phrase Attachment Disambiguation</i> .....	1198
E. Brill, P. Resnik (Massachusetts Institute of Technology/Sun Microsystems Labs.)	
<i>Discourse and Deliberation: Testing a Collaborative Strategy</i> .....	1205
M. A. Walker (Mitsubishi Electric Research Labs.)	
<i>A Bayesian Approach for User Modeling in Dialogue Systems</i> .....	1212
T. Akiba, H. Tanaka (Tokyo Institute of Technology)	

# Reserve Paper

<i>World Class Discovery for Postprocessing Chinese Handwriting Recognition</i> -----	1221
C.-H. Chang (Industrial Technology Research Institute)	
<i>A Simple Transformation for Offline-Parsable Grammars and Its Termination Properties</i> --	1226
M. Dymetman (Rank Xerox Research Center)	
<i>The Evaluation of Machine-Tractable Dictionaries</i> -----	1231
C.-M. Guo, C.-N. Huang, J.-P. Gong, J. Li (Tsinghua University)	
<i>Discontinuity and the Lambek Calculus</i> -----	1235
M. Hepple (University of Sheffield)	
<i>Lexical Functions and Machine Translation</i> -----	1240
D. Heylen, K.G. Maxwell, M. Verhagen (OTS)	
<i>Chinese Segmentation Disambiguation</i> -----	1245
W. Jin (New Mexico State University)	
<i>Typed Feature Structures as Descriptions</i> -----	1250
P.J. King (Eberhard-Karls-Universität)	
<i>Portuguese Analysis with Tree Adjoining Grammars</i> -----	1255
K.C. Kipper, V.L.S. de Lima (PUCRS)	
<i>Incremental Construction of a Lexical Transducer for Korean</i> -----	1262
H.-C. Kwon, L. Karttunen (Pusan National University)	
<i>Dynamic Logic with Possible World</i> -----	1267
R. Lu (Shanghai Jiao Tong University)	
<i>Humor-based Applications</i> -----	1270
G. Prószéky, M. Pál, L. Tihanyi (Morphologic/OPKM Comp. Centre/Inst. for Linguistics)	
<i>Blending Segmentation with Tagging in Chinese Language Corpus Processing</i> -----	1274
Z. Qiang, Y. Shiwen (Peking University)	
<i>Syllable-based Phonetic Transcription by Maximum Likelihood Methods</i> -----	1279
R.A. Sharman (IBM UK Labs., Ltd.)	
<i>A Knowledge Acquisition and Management System for Morphological Dictionaries</i> -----	1284
P. Ten Hacken, S. Bopp, M. Domening, D. Holz, A. Hsiung, S. Pedrazzini (Universitat Basel/Vrije Univ. Amsterdam/IDSIA)	
<i>NL Understanding with a Grammar of Constructions</i> -----	1289
W. Zadrozny, M. Szummer, S. Jarecki, D.E. Johnson, L. Morgenstern (IBM)	

# NL Understanding with a Grammar of Constructions

Wlodek Zadrozny and Marcin Szummer and Stanislaw Jarecki  
and David E. Johnson and Leora Morgenstern \*

IBM Research Division  
T.J.Watson Research Lab  
Yorktown Heights, NY 10598 USA

## Abstract

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system. The architecture has two key aspects: (a) the use of constructions, integrating descriptions of form, meaning and context into one whole; and (b) the separation of *domain knowledge* (about calendars) from *application knowledge* (about the particular on-line calendar).

## 1 Introduction: an overview of the system

We present an approach to natural language understanding based on a *computable grammar of constructions*. A *construction* consists of a set of features of form and a description of meaning in a context. A grammar is a set of constructions. This kind of grammar is the key element of MINCAL, an implemented natural language speech-enabled interface to an on-line calendar system.

The system consists of a NL grammar, a parser, an on-line calendar, a domain knowledge base (about dates, times and meetings), an application knowledge base (about the calendar), a speech recognizer, a speech generator.

In this paper we describe two key aspects of the system architecture: (a) the use of constructions, where instead of separating NL processing into the phases of syntax, semantics and pragmatics, we integrate descriptions of form, meaning and context into one whole, and use a parser that takes into account all this information (see [10] for details); (b) the separation of the *domain knowledge* (about calendars) and the *application knowledge* (about the particular on-line calendar).

\*M. Szummer and S. Jarecki are also from MIT.

## The dialogs

The system allows users to engage in dialogs like:

- *Schedule a meeting with Bob!*
- *At what time and date?*
- *On August 30th.*
- *At what time?*
- *At 8.*
- *Morning or afternoon?*
- *In the evening.*

The parser recognizes *Schedule a meeting with Bob* as an instance of *sent(imp)*, the imperative construction consisting of a verb and an NP, here *np(event)*. The context is used to prevent another reading in which *with Bob* modifies *schedule*, as in *Dance a tango with Bob!*. That is, a contextual rule is used which says that for calendar applications, people do not modify actions or places. Context also plays an important role in understanding answers, e.g. *At 8*. This is understood as a time expression (and not place or rate or something else) only because of the context.

The parameters of a meeting can be given in many ways, e.g. synonyms or different constructions can be used, users can include as many parameters in a sentence as they wish, and the parameters can be given in any order. As a result there are about 10,000 ways of scheduling meetings (with a given set of parameters).

## How are the dialogs understood

With respect to parsing, grammars of constructions can be parsed like "standard" grammars, except that the set of features is richer. Given a string (representing a sentence, a fragment of a discourse or a paragraph), the parser assigns it a construction. From this viewpoint, the situation is similar to "regular" parsing, and the possible algorithms are similar. We have implemented a prototype chart parser for construction grammars, discussed further in Section 3.

But, clearly, having understood the sentence as a linguistic entity in isolation is not the ultimate goal. Here the message of an utterance must be understood in the context of an intended action. This is done in two steps. First, the system determines the intended

action and its parameters, using domain knowledge (meetings+time+places). Second, once all the parameters have been extracted from the dialog, the system executes the action. To do this, the program uses application-specific knowledge to translate the action and its parameters into a form that can be executed by the application (Xdiary).

## 2 Constructions as data structures

A construction is given by the matrix:

$$\begin{bmatrix} \mathbf{N} : \text{name\_of\_construction} \\ \begin{bmatrix} \mathbf{C} : \text{context} \\ \mathbf{V} : \text{structure} \\ \mathbf{M} : \text{message} \end{bmatrix} \end{bmatrix}$$

The *vehicle* **V** consists of formulas describing presence (or perhaps absence) of certain *taxemes*, or features of form, within the structure of the construction. Such a structure is given by a list of subconstructions and the way they have been put together (in all our examples this is concatenation, but there are other possibilities, e.g. wrapping). The *context*, **C**, consists of a set of semantic and pragmatic constraints limiting the application of the construction. It can be viewed as a set of *preconditions* that must be satisfied in order for a construction to be used in parsing. The *message*, **M**, describes the meaning of the construction, via a set of syntactic, semantic and pragmatic constraints.

To make this concrete, let us consider a few examples. We begin with a simple "command construction" consisting of an action verb followed by its argument.

$$\begin{bmatrix} \mathbf{N} : \text{sent}(\text{cmnd}, v.\text{np}) \\ \mathbf{C} : [\langle \text{hr attends} \rangle = \text{sr}] \\ \mathbf{V} : \begin{bmatrix} \text{struc} = (\text{V.NP}) \\ \langle \text{V cons\_n} \rangle = \text{verb} \\ \langle \text{V M } v.\text{type} \rangle = \text{action\_verb} \\ \langle \text{NP cons\_n} \rangle = \text{np} \\ \text{sem\_cat} = \text{command} \end{bmatrix} \\ \mathbf{M} : \begin{bmatrix} a.\text{type} = \langle \text{V M } \text{sem\_type} \rangle \\ a.\text{obj} = \langle \text{NP M } \text{sem\_type} \rangle \\ \text{agent} = \text{hr} \end{bmatrix} \end{bmatrix}$$

The context of the construction describes all situations in which the hearer *hr* (human or machine) is paying attention to the speaker *sr* (a "ready" state). The feature *struc* is a list of variables and/or words/tokens; it is used to describe the structure of a construction, and its role is similar to a rule in a generative grammar. (We will write names of variables in capital letters, e.g. *NP*, inside matrices of constructions). The attribute *cons\_n* gives the name of a construction that could be assigned to a string. We use it here to say that the form of the construction can be described as a concatenation of two strings,

of which one is a verb (construction) and the other an np (construction). Furthermore, the verb type  $\langle \text{V M } v.\text{type} \rangle$  is "action\_verb". (The expression  $\langle \text{V M } v.\text{type} \rangle$  should be read "the *v.type* of the message of *V*").

The message **M** describes the meaning of the construction as that of a command in which the type of action is described by the meaning of the verb, and the object of the action is given by the meaning of the noun phrase. The attribute *sem.type* stands for the "semantic type" and we identify it currently with the word sense. Thus "erase the file" is understood as a command to *delete* the file, if  $\langle \text{erase M } \text{sem.type} \rangle = \text{delete}$ , but "erase the picture" might refer to the type of action associated with *rub\_out*. In both cases the hearer *hr* is supposed to be the agent of the action.

### Constructions: from words to discourse

Words, phrases, and fragments of discourse can be analyzed as constructions. We view languages as collections of constructions which range from words to discourse. We claim that the same representation scheme can be used for all constructions.

The examples we are going to present have been developed with a specific purpose in mind, namely for scheduling calendar events. In other papers ([10] and [6]), we have presented examples showing that we can give a good descriptions of non-standard constructions. However, in either case descriptions of meanings and contexts are general, and hence applicable to other tasks.

We now turn our attention to words. The verb "cancel" can be represented as follows:

$$\begin{bmatrix} \mathbf{N} : \text{verb}(\text{cancel}) \\ \mathbf{C} : \begin{bmatrix} \text{lang\_code} = \text{english} \\ \text{lang\_channel} = \text{text} \end{bmatrix} \\ \mathbf{V} : \text{struc} = (\text{cancel}) \\ \mathbf{M} : \begin{bmatrix} \text{cat} = \text{verb} \\ \text{sem\_type} = \text{delete} \\ \text{v\_type} = \text{action\_verb} \end{bmatrix} \end{bmatrix}$$

Notice that even simple words require context to be (properly) interpreted. In **C** we say that English text is expected (but in other cases it could also be French text, or French speech, etc.). Some aspects of context do not have to be explicitly specified and can be replaced by defaults.

Although the vehicle and the message are both very simple in this example, the simplicity of the message is a result of deliberate simplification. We have restricted it to the specification of the semantic type, identified with one sense of the word, and to describing the verb type of "cancel" as a verb of action. Notice that the other sense of "cancel" – "offset, balance out" – would appear in another entry.

Of course, in reality, the lexical meaning of any word is a much more complicated matter [1]. For instance, in our lexicon the messages of words may contain many of the attributes that appear in the

explanatory combinatorial dictionary of Melcuk [7].

**Discourse constructions:** To illustrate discourse constructions, we consider the following dialog:

*Have you arranged the room yet?*

*No, but I'll do it right away.*

We view the pattern of the answer *no.but.S* as a discourse construction. It can be represented by the following array of features:

$$\left[ \begin{array}{l} \mathbf{N} : \text{sent}(\text{assrt}, \text{no.but.S}) \\ \mathbf{C} : \left[ \begin{array}{l} < p\_utter \text{ cons\_n} > = \text{sent}(\text{ques}, *) \\ \text{struc} = (\text{no.but.S}) \end{array} \right] \\ \mathbf{V} : \left[ \begin{array}{l} < S \text{ cons\_n} > = \text{sent}(\text{assrt}, *) \\ < p\_sent \text{ truth\_value} > = 0 \end{array} \right] \\ \mathbf{M} : \left[ \begin{array}{l} < S \mathbf{M} > \end{array} \right] \end{array} \right]$$

As we can see, the construction applies only in the context of a previously asked question, and its message says that the answer to the question is negative, after which it elaborates the answer with a sentence *S*.

### 3 System Architecture

#### The parts

MINCAL consists of a NL grammar, a parser, a domain knowledge base (about dates, times and meetings), an on-line calendar (Xdiary), an application knowledge base (about Xdiary), a continuous speech recognizer (IBM, ICSS), a speech generator (Speech Plus, Text to Speech Converter), and the interfaces.

At present, the grammar consists of a few hundred lexical constructions, and about 120 "productions", i.e. constructions describing combinations of other constructions.<sup>1</sup> It covers the basic forms of assertive sentences, but it emphasizes commands. Thus a command can, for example, be given either by *v.np* (also with "please", or "kindly"), or by an assertive sentence recognized as an indirect speech act ("I'd like to ...", "Leora wants you to ...", etc.). The next large group of constructions covers PPs, with particular emphasis on time and places. Finally, it covers a few discourse constructions, since it is important to deal with sentence fragments in dialogs, e.g. understanding "evening" as "in the evening", when it is an answer to the question "when?".

#### The interaction of the modules

**The calendar and the application knowledge base:** Xdiary is an on-line calendar for which we have not written a complete interface, but have focused on the three most important functions: appointment, moving, and canceling appointments. Other functions, such as "to do" lists, window management, listing somebody's appointments, etc., can

be dealt with in a similar fashion, and we plan to extend the interface to deal with them. At this point the application knowledge base is very simple. It consists of rules that say how to interpret the data given by the semantic interpreter, for instance the rules for formatting parameters and renaming slots (e.g. *event.duration* → *duration*). Such rules are necessary; if the distinction between application and domain knowledge is to be maintained.

**The domain knowledge base:** This has two kinds of facts: (1) background ontology, i.e., is, basic facts about time and places, and (2) linguistic knowledge associated with the domain. The former includes such obvious facts as the number of days in a month, which month follows the other, that offices are places etc. The latter includes facts about how the language is used. For example, the *filters* saying that places do not modify people, so that *I want to meet my manager in the cafeteria* can be unambiguously parsed, with "cafeteria" being a meeting place, and not an attribute of the manager.

**The organization of knowledge:** The issue of the organization of knowledge has been discussed at length in [8] and [9] and the formal model developed there is applicable in the present context. At this point, however, this formal model has only been implemented very crudely. Still the model is worth briefly discussing, because the conceptual distinctions made guide our work and have important practical consequences. The most important thing about it is that we discard the model of background knowledge as a logical theory, and replace it by a model consisting of collection of theories and mechanisms for putting them together depending on circumstances. Thus, the usual, two-part logical structures, consisting of a *metalevel* and an *object level*, are augmented by a third level — a *referential level*. The referential level is a partially ordered collection of theories; it encodes background knowledge in a way resembling a dictionary or an encyclopedia.<sup>2</sup>

#### Parser, construction grammar and linguistic knowledge

**Parser:** The parser does not produce (syntactic) structural descriptions of sentences. Instead, it computes meaning representations. For example, it converts adjuncts directly into attributes of place, time, participant etc., once they can be computed, and thus the message of the sentence does not contain any information about how these attributes were expressed or about the attachment of PPs that appear in it. For example, the sentence *I want you to arrange a conference in my office at 5* is analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, a verb, an object and a complement.

<sup>1</sup>These are constructions we used in MINCAL. In addition in various experiments we have used a few dozen other constructions, e.g. those covering "open idioms" (see Section 4).

<sup>2</sup>As usual, current situations are described on the object level, and the *metalevel* is a place for rules that can eliminate some of the models permitted by the object level and the referential level.

The latter and the message of the imperative that is passed to *sent(assert, svoc)* does not contain any structural information about the attachment of the PPs. This message is combined with the messages of the verb and the noun, yielding

```
[ den want(other_agent)]
[ agent hearer]
[ mental_agent
  [ [ type person]
    [ den speaker]
  [ [ action
    [ [ den arrange]
    [ action_object
      [ type event]
      [ den conference]
      [ number 1]
      [ mods
        [ [ det a]
          [ pp_mag
            [ [ prep at]
              [ type time(hour)]
              [ den
                [ [ hour
                  [ 5 am_or_pm]
                [ minute 0]
              [ [ prep in]
                [ type place ]
                [ den office]
                [ mods
                  [ [ det my]
```

This result of parsing is then interpreted by the domain interpreter to produce:

```
***Slots:
[ [ action_name schedule]
  [ event_name
    [ a conference]
  [ event_time
    [ [ minute 0]
      [ hour
        [ 5 am_or_pm]
  [ event_place
    [ my office]
```

Application-specific defaults then produce yet another interpretation where, in addition to filling the slots of Xdiary, [ hour [ 5 am\_or\_pm] ] is interpreted as [ hour [ 17 ] ].

The parser is a chart parser, working left to right, with no lookahead. The grammar is L-attributed, i.e., has both synthesized and inherited attributes, but each inherited attribute depends only on inherited attributes of the parent or attributes of the sisters to the left. Hence, although the parser does not have a lookahead step at present, such a step can be added following [2].

## 4 Comparisons with related work

Linguistic arguments for constructions-based grammars has been worked out chiefly by Ch. Fillmore and his colleagues (cf. [3]). Their motivation for advocating such an approach comes from the fact that typical generative theories of grammar cannot deal properly with open idioms illustrated by constructions such as:

*The more carefully you work, the easier it will get.*

*Why not fix it yourself?*

*Much as I like Ronnie, I don't approve of anything he does.*

*It's time you brushed your teeth.*

*Him be a doctor?*

The same is true about even so-called robust parsers of English. The reason for this failure can be attributed to the fact that expressions like these "exhibit properties that are not fully predictable from independently known properties of its lexical make-up and its grammatical structure" – [3], p.511. However we do not need a list of "strange" constructions to conclude that thoroughly integrating syntax with semantics and pragmatics could provide us with a better handle on natural language understanding. On a closer examination "normal" constructions exhibit enough complexity to warrant the new approach (see [10] for details).

Jurafsky [4] has independently come up with a proposal for a computable grammar of constructions. We compare our work with his in [10]. Here, we limit ourselves to a few remarks. What is common in both approaches is the centrality of the concept of grammatical construction as a data structure that represents lexical, semantic and syntactic knowledge. However, there are important differences between the two formalisms. First, the actual data structures used to represent constructions are different. The most important difference has to do with the presence of the *context* field in our version of the construction grammar. This allows us to account for the importance of pragmatics in representing many constructions, and to deal with discourse constructions.

Secondly, while Jurafsky acknowledges the need for abstract constructions (pp.43-51), his abstract constructions (*weak constructions*) are not first class citizens — they are defined only extensionally, by specifying the set of constructions they abstract over, and their abstract meaning (e.g. **entity** for NOUN). They are used to simplify descriptions of constituents of other constructions. However, because they do not have a separate *vehicle* part, they cannot be used to assign default meanings. For instance, since *verb* is defined as a collection of all verbs *is + read + cancel + know + look-up + ...*, it cannot be assigned a feature *action.verb* without introducing a contra-

diction – its semantics is therefore given as RELATION/PROCESS. For us the important feature of "abstract" constructions is not that they simplify descriptions of other constructions, but that they have default meanings. (A similar critique of [5] can be found in [10]).

## 5 Summary of results

Our approach to NLU is based both on linguistic arguments and on our dissatisfaction with the state of the art. State of the art systems typically are too "syntax-driven", failing to take context into account in determining the intended meaning of sentences. A related further weakness is that such systems are typically "sentence oriented", rather than "conversation/discourse oriented". In our view, this makes even the most robust systems "brittle" and ultimately impractical.

To test whether a construction-based approach is feasible built a "complete" working system that would include a representation for constructions. To do this, we focused on the "calendar domain", a domain with limited complexity and simple but not uninteresting semantics. We have chosen to deal with simple actions, and not e.g. with question answering, where deeper understanding would be necessary.<sup>3</sup>

### Our contributions:

1. We have proposed a *new kind of grammar* – computable construction grammars, which are neither semantic, nor syntactic. Instead, their "productions" combine lexical, syntactic, semantic and pragmatic information.<sup>4</sup>
2. We have described *data structures for constructions*, and have shown that they can be effectively used by the parser. Note that the same data structure is used to encode the lexicon and the "syntactic" forms.
3. We have shown *how to parse with constructions*. We have implemented a simple chart parsing algorithm, which can be easily extended to an Eearly-like parser, as long as the construction grammar remains L-attributed. We have found that even a simple parser of construction can be quite efficient. This is partly due to the fact that it does not require copying of all syntactic and semantic information from daughters to mothers; the goal of parsing consists in producing an interpretation, and structural information can be discarded once an interpretation of a phrase is produced. It is also worth emphasizing

<sup>3</sup>We have also thought about another possibility, that is, enhancing an IR system, e.g. with the understanding of date expressions.

<sup>4</sup>In what sense are they "computable"? Although this adjective might suggest a formal model with computational complexity results, etc., what we have in mind is pretty trivial: (1) the system actually computes the messages of grammatical construction; (2) the grammars and constructions are well defined data structures, and parsing (combining all associated constructions in all possible ways) is decidable.

that invoking domain semantics drastically reduces the number of parses constructed.

4. We have proposed a *modular architecture for NL interfaces* based on the division between linguistic knowledge, domain knowledge base, and application knowledge base. Based on our experience, we believe that this architecture should work in general for speech-enabled interfaces for restricted domains.

## References

- [1] *Computational Linguistics*. Special Issue on the Lexicon, 1987. 13: 3-4.
- [2] Nelson Correa. An extension of earley's algorithm for s- and l- attributed grammars. In *Proc. Intl. Conf On Current Issues in Computational Linguistics*, pages 248-261. Penang, Malaysia, 1991.
- [3] Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. Regularity and idiomaticity in grammatical constructions. *Language*, 64(3):501-538, 1988.
- [4] D. Jurafsky. *An On-line Computational Model of Sentence Interpretation*. PhD thesis, University of California, Berkeley, 1992. Report No. UCB/CSD 92/676.
- [5] Paul Kay. Even. *Linguistics and Philosophy*, 13(1):59-111, 1990.
- [6] M. McCord, Arendse Bernth, Shalom Lappin, and Wlodek Zadrozny. Natural language processing technology based on slot grammar. *International Journal on Artificial Intelligence Tools*, 1(2):229-297, 1992.
- [7] I. Melcuk. *Dependency Syntax: Theory and Practice*. State University of New York Press, Albany, NY, 1988.
- [8] Wlodek Zadrozny. Reasoning with background knowledge – a three-level theory. *Computational Intelligence*, 10(2), 1994. To appear.
- [9] Wlodek Zadrozny and Karen Jensen. Semantics of paragraphs. *Computational Linguistics*, 17(2):171-210, 1991.
- [10] Wlodek Zadrozny and Alexis Manaster-Ramer. The significance of constructions. *Submitted to Computational Linguistics*, 1994.

**©COLING 94 Organizing Committee**

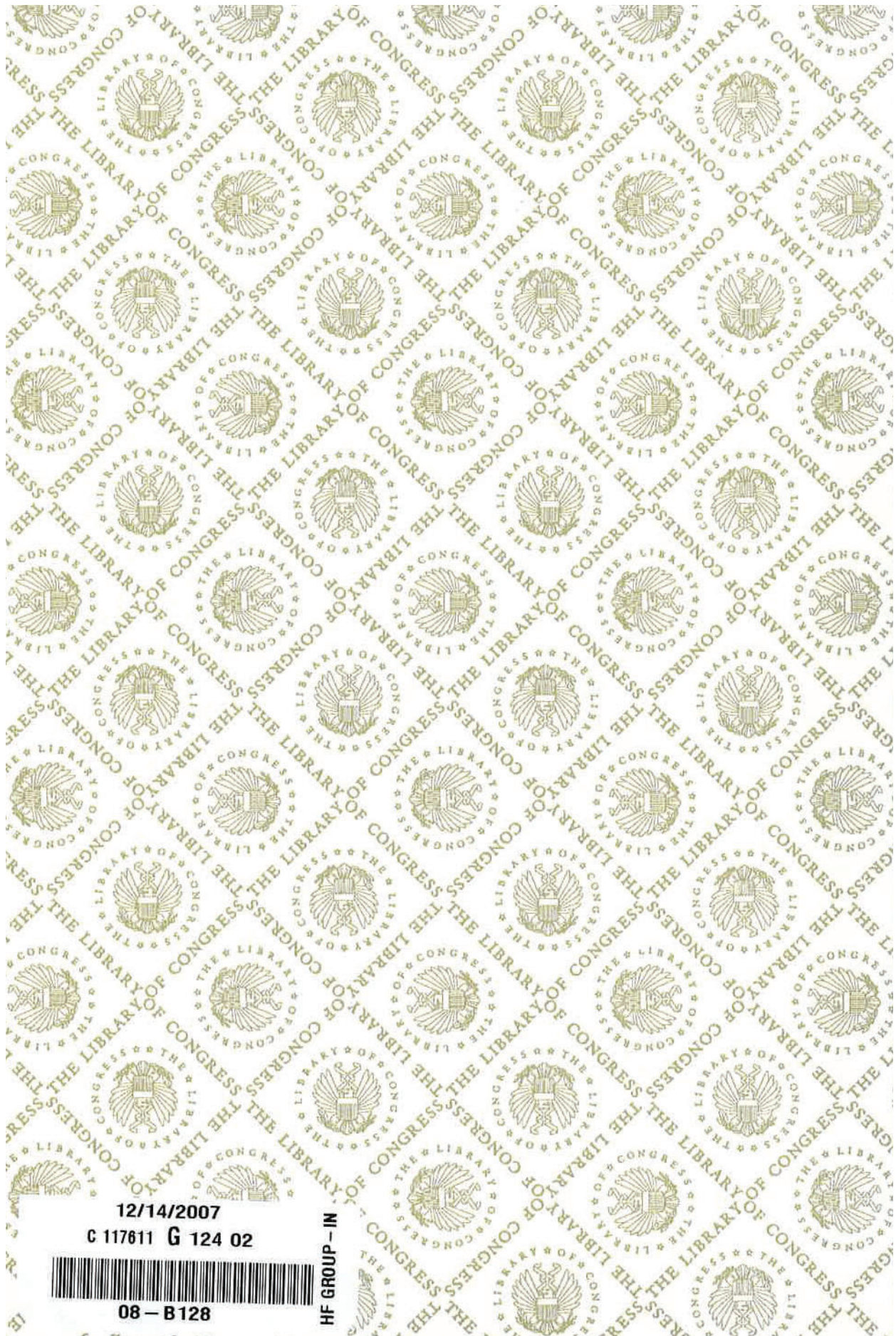
All rights reserved. No part of this publication may be reproduced, stored in retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior notice.

無断転載を禁ず。

LC ACQUISITIONS



0 014 794 944 3



12/14/2007

C 117611 G 124 02



08 - B128

HF GROUP - IN