Association for Computational Linguistics

Author Index to the Proceedings of ANLP-NAACL 2000 and the Student Research Workshop

April 29—May 4, 2000
Seattle, Washington, USA

Association for Computational Linguistics

6th Applied Natural Language Processing Conference

1st Meeting of the North American Chapter of the Association for Computational Linguistics

Proceedings of the Conferences

and

Proceedings of the ANLP-NAACL 2000 Student Research Workshop

April 29—May 4, 2000
Seattle, Washington, USA

Published by the Association for Computational Linguistics
http://www.acl.web.org/

Distributed by Morgan Kaufmann Publishers
http://www.mkp.com

Production and Manufacturing by
Omipress, Inc.
Post Office Box 7214
Madison, WI 53707-7314

Copyright © 2000 Association of Computational Linguistics

Order copies of this and other ACL publications from:

Morgan Kaufmann Publishers
San Francisco

http://www.mkp.com
orders@mkp.com
1-800-745-7323
1-800-814-6418 (fax)

ANLP-NAACL 2000
ISBN: 1-55860-704-8

Author Index

Section 1: Applied Natural Language Processing Conference (ANLP)

Section 2: North American Chapter of the Association for Computational Linguistics (NAACL)

Section 3: ANLP-NAACL 2000 Student Research Workshop (SRW)

Preface

The year 2000 marks the launch of the North American chapter of the Association for Computational Linguistics (NAACL). The ACL Executive committee under the leadership of Phil Cohen used this as an opportunity to bring industry and researchers together to explore the full spectrum of computational linguistics and natural language processing by planning a joint conference, combining the 6th Applied Natural Language Processing Conference (ANLP) and the 1st Conference of NAACL. ANLP-NAACL2000 marks not only the first major ACL sponsored conference of the new millennium, but reflects a significant change in the field, as we see speech and language applications leave the laboratory and enter every day use.

The technical program covers all facets of the field, from theory and methodology to its application in commercial software. We chose to have two independent program chairs so that papers from each of these two perspectives could be evaluated independently and they appear in separate sections of the proceedings. Janyce Wiebe, New Mexico State University, was the NAACL Program Chair and Sergei Nirenburg, also from New Mexico State University, was the ANLP Program Chair. I would like to thank them and their program committees and teams of reviewers, whose hard work produced an exciting program with three parallel sessions, reflecting the breadth of quality research and application development being done throughout the world.

We were honored to have as invited speakers two industry leaders: Dave Nagel, President, AT&T Laboratories and AT&T Chief Technical Officer, speaking on "Voice access to information", and Rick Rashid, Vice President, Microsoft Research, speaking on "The Future - It isn't what it used to be". From the academic community, Leonard Talmy, State University of New York at Buffalo, an internationally recognized researcher in cognitive science, speaking on "How Language Structures Concepts". These plenary sessions have shown us how far the community has come in bringing speech and language to the marketplace and to remind us of the complexity of language and its relationship to human cognition.

As is traditional at ACL conferences, the program goes beyond the presentation of technical papers and plenary sessions. There are also tutorials aimed at helping participants broaden their knowledge, workshops for exploring an area in depth, demonstrations to show the software implementations behind the theories, and exhibits to display the commercial products that have come out of the research, much of which was presented at past ACLs. I would like to thank the Tutorial Chair: Jennifer Chu Carroll (Lucent Technologies), Workshop Chair: Scott Miller (BBN), Demonstrations Chair: Jeff Reynar (Microsoft) and Exhibits Chair: Deborah Dahl (Unisys), all of whom composed quality programs.

One of the most valuable aspects of ACL conferences is the opportunity it offers to students, who can get a broad perspective on the field and meet the people whose papers they read in their courses. They also have the opportunity to present their own research, both in the main sessions and in special student sessions designed to help those in the early stages of their research. This year, we tried a new approach with a Student Research Workshop held before the conference with panelists to provide feedback to students, for which we received NSF funding from the Knowledge and Cognitive Systems Program under Ephraim Glinert. I would like to thank Committee members Donna Byron (University of Rochester) and Peter Vanderheyden (University of Waterloo), Co-Chairs, and Mary Harper (Purdue University), Faculty Advisor and NSF Principal Investigator and all of the reviewers who contributed to an excellent program and provided feedback to the students on their research. I would also like to thank the students who submitted papers and attended the workshop. They are the future of our organization.

Rick Wojcik (Boeing) as Local Arrangements Chair took care of the myriad of arrangements that go into an event of this size. Kathy McCoy, Secretary-Treasurer of the ACL, and Priscilla Rasmussen, ACL Business Manager, provided unending support in the planning, decision making, and execution of all of the arrangements. Mark Maybury (Mitre), Publicity Chair, expanded the number of places we advertise the conference to help widen our audience. June Santeusanio and Russell Kenyon from GTE created and updated a terrific web site, bringing the information about the conference to a central place. Russell developed the on-line registration form and provided support throughout the registration process.

I would especially like to thank Gary Coen (Boeing), who was the Sponsorship Chair, and all of the commercial companies who contributed to the conference. We had the largest level of industry support of any ACL conference in the past. Microsoft, Boeing, Sun, Intel (China), Xerox PARC, Logos, Conversa, and General Electric all contributed funds to support various events, allowing us to have a commercial grade conference at a hotel facility while keeping the conference fees at their usual level to accommodate academic budgets. GTE (BBN), Answerlogic, IBM, and Inxight provided scholarships to allow more students to attend the conference.

There are many more people who deserve thanks for contributing in one way or another. ACL is an all volunteer organization and it is a credit to everyone who participates that we can put on an event of this size and scope, benefiting everyone who attends through the exchange of ideas and inspiration we get from seeing the leading edge of the field.

Marie Meteer
General Chair, ANLP-NAACL 2000
April 2000

CONFERENCE PROGRAM

TUTORIALS, Saturday 29 April 2000, 8:30AM-5:00PM

Information Retrieval, 8:30AM-12:00PM
James Allan, University of Massachusetts, Amherst

The State of the Art in Language Modeling, 8:30AM-12:00PM
Joshua Goodman, Microsoft Research

Finite-State Morphology/Phonology: Theory, Applications, and Recent Developments, 1:30PM-5:00PM
George Kiraz, Lucent Technologies Bell Labs

Machine Translation, 1:30PM-5:00PM
Kevin Knight, USC/ISI

Monday, May 1	**ANLP/NAACL TECHNICAL SESSIONS**	**PRELIMINARY PROGRAM**	
	NAACL	*ANLP*	
8:45	**Introductory Remarks**		
	Session 1	**Session 2**	**Session 3**
9:00	**Modelling Grounding and Discourse Obligations Using Update Rules** Colin Matheson, David Traum, Massimo Poesio	**BusTUC-A Natural Language Bus Route Oracle** Amble Tore	
9:25	**The Automatic Translation of Discourse Structures** Daniel Marcu, Lynn Carlson, Maki Watanabe	**Machine Translation Between Very Close Languages** Jan Hajic, Jan Hric, Vladislav Kubon	
9:50	**A Probabilistic Genre-independent Model of Pronominalization** Michael Strube, Maria Wolters	**Cross-Language Multimedia Information Retrieval** Sharon Flank	
10:15	**Advances in domain independent linear text segmentation** Freddy Y.Y. Choi	**Automatic Construction of Parallel English-Chinese Corpus for Cross-language Information Retrieval** Jiang Chen, Jian-Yun Nie	
10:40	*BREAK*		
	NAACL	*ANLP*	*NAACL*

11:05	**Bagging and boosting a Treebank Parser** John C. Henderson, Eric Brill	**Providing Customer Service Using Natural Language Dialog** Amit Bagga, Tomek Strzalkowski	**Encoding information on adjectives in an Italian semantic net for computational applications** Antonietta Alonge, Francesca Bertagna, Nicoletta Calzolari, Adriana Roventini
11:30	**Noun Phrase Recognition by System Combination** Erik F. Tjong Kim Sang	**Translation Using Dialogue Participants' Information** Setsuo Yamada, Eiichiro Sumita, Hideki Kashioka	**The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure** Christopher Johnson, Charles Fillmore
11:55	**A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation** Ted Pedersen	**Distilling Dialogues: A Method Using Natural Dialogue Corpora for Dialogue Systems Development** Nils Dahlback, Arne Jonsson	**Generation in the Lambek Calculus Framework: an Approach with Semantic Proof Nets** Sylvain Pogodalla
12:20	*LUNCH*		
	NAACL	*ANLP*	*ANLP*
1:50	**Word-for-Word Glossing with Contextually Similar Words** Patrick Pantel, Dekang Lin	**Plan-Based Dialogue Management in a Physics Tour** Reva Freedman	**A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Transformations** Benoit Lavoie
2:15	**Arabic Morphology Generation Using a Concatenative Strategy** Violetta Cavalli-Sforza, Abdelhadi Soudi, Teruko Mitamura	**Talk'n'Travel: A Conversational System for Air Travel Planning** David Stallard	**REES: A Large-Scale Relation and Event Extraction System** Chinatsu Aone, Mila Ramos-Santacruz
2:40	**Morphological Tagging: Data vs. Dictionaries** Jan Hajic	**Analysing Speech Transcriptions** Mark Stevenson, Robert Gaizauskas	**DIP: Detector for Incorrect Presuppositions in Survey Questions** Katja Wiemer-Hastings, Peter Wiemer-Hastings, Sonya Rajan, Arthur C. Graesser, Roger J. Kreuz, Ashish Karnavat
3:05	*BREAK*		
3:30	**Voice Access to Information** **Dave Nagel,** *Invited Speaker*		
4:30	*BREAK*		
	NAACL	*ANLP*	
4:40	**The Effectiveness of Corpus-Induced Dependency Grammars for Post-processing Speech** Mary P. Harper, Christopher M. White, Wen Wang, Michael T. Johnson, Randall A. Helzerman	**MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries** Jennifer Chu-Carroll	

5:05	**Analyzing Dependencies of Japanese Subordinate Clauses based on Statistics of Scope Embedding Preference** Takehito Utsuro, Shigeyuki Nishiokayama, Masakazu Fujio, Yuji Matsumoto	**Javox: A Toolkit for Building Speech Enabled Applications** Michael Fuklerson, Alan W. Biermann	
5:30	**Rapid Parser Development: A Machine Learning Approach for Korean** Ulf Hermjakob	**A Compact Architecture for Dialogue Management Based on Scripts** Manny Rayner, Beth Ann Hockey, Frankie James	
5:55	*Close*		

Tuesday, May 2	**ANLP/NAACL TECHNICAL SESSIONS**	**PRELIMINARY PROGRAM**	
	Session 1	**Session 2**	**Session 3**
	NAACL	*ANLP*	
9:00	**A Classification Approach to Word Prediction** Yair Even-Zohar, Dan Roth	**A Representation for Complex and Evolving Data Dependencies in Generation** Roger Evans, Mike Reape, Lynne Cahill, Christine Doran, Chris Mellish, Daniel Paiva, Donia Scott, Niel Tipper	
9:25	**A Maximum-Entropy-Inspired Parser** Eugene Charniak	**An Automatic Reviser: The TransCheck System** Jean-Marc Jutras	
9:50	**Detecting Grammatical Errors without Negative Evidence** Martin Chodorow, Claudia Leacock	**Unit Completion for a Computer-Aided Translation Typing System** Phillippe Langlais, George Foster, Guy Lapalme	
10:15	**Automatic Corpus Correction with Anomaly Detection** Eleazar Eskin	**Multilingual Coreference Resolution** Sanda Harabagiu Steven J. Maiorano	
10:40	*BREAK*		
	NAACL	*ANLP*	*ANLP*
11:05	**Exploiting Auxiliary distributions in stochastic unification-based grammars** Mark Johnson, Stefan Riezler	**Using Predictive Annotation to Rank Suspected Answers** Dragomir Radev, John Pranger, Varerie Sam	**Message Classification in the Call Center** Stephan Busemann, Sven Schmeier, Roman Georg Arens

11:30	**Ambiguity Packing in HPSG --- Practical Results** John Carroll, Stephan Oepen	**A Question Answering System Supported by Information Extraction** Rohini Srihari, Wei Li	**Categorizing Unknown Words: Using Decision Trees to Differentiate Names and Misspellings** Janine Toole
11:55	**Forest-Based Statistical Sentence Generation** Irene Langkilde	**Examining the Role of Statistical and Linguistic Knowledge Sources in a General-Knowledge Question Answering System** Claire Cardie, Vincent Ng, David Pierce, Chris Buckley	**Extracting Molecular Binding Relationships from Biomedical Text** Thomas Rindflesch, Jayant Rajan, Lawrence Hunter
12:20	*LUNCH*		
	NAACL	*ANLP*	*ANLP*
1:50	**Cut-and-Paste Based Text Summarization** Hongyan Jing, Kathleen R. McKeown	**Korean Compound Noun Segmentation Based on Lexical Data Extracted from Corpus** Juntae Yoon	**Experiments with Corpus-Based LFG Specialization** Nicola Cancedda, Christer Samuelsson
2:15	**Minimizing Word Error Rate in Textual Summaries of Spoken Language** Klaus Zechner, Alex Waibel	**A Tool for Automated Revision of Context-Free Grammars** Nanda Kambhatla, Wlodek Zadrozny	**Aggressive Morphology for Robust Lexical Coverage** William Woods
2:40	**Trainable methods for Surface natural Language Generation** Adwait Ratnaparkhi	**TnT-A Statistical Part-of-Speech Tagger** Thorsten Brants	**Language Independent Morphological Analysis** Tatuo Yamastia, Yuji Matsumoto
3:05	*BREAK*		
3:30	**How Language Structures Concepts** **Leonard Talmy,** *Invited Speaker*		
4:30	*BREAK*		
4:40	**Evaluating automatic dialogue stategy adaptation for a spoken dialogue system** Jennifer Chu-Carroll, Jill Suzanne Nickerson	**A Divide-and-Conquer Strategy for Shallow Parsing of German Free Texts** Christian Braun, Guenter Neumann, Jakub Piskorski	
5:05	**Learning to recognize probabilistic Situations in a Spoken Dialogue system: Experiments with 'How May I Help You?'** Marilyn Walker, Irene Langkilde, Jerry Wright, Allen Gorin, Diane Litman	**A Hybrid Approach for Named Entity and Sub-type Tagging** Rohini Srihari, Cheng Niu, Wei Li	

5:30	**Predicting Automatic Speech Recognition Performance Using Prosodic Cues** Diane J. Litman, Julia B. Hirschberg, Marc Swerts	**Spelling and Grammar Correction for Danish in SCARRIE** Patrizia Paggio	
5:55	*CLOSE*		

Wednesday, May 3	**ANLP/NAACL TECHNICAL SESSIONS**	**PRELIMINARY PROGRAM**	
	Session 1	**Session 2**	**Session 3**
9:00	Plenary (TBA)		
	NAACL	*ANLP*	
9:50	**A Novel Use of Statistical Parsing to Extract Information from Text** Scott Miller, Heidi Fox, Lance Ramshaw, Ralph Weischedel	**Linguistic Knowledge can Improve Information Retrieval** William Woods, Lawrence A. Bookman, Ann Houston, Robert J. Kuhns, Paul Martin	
10:15	**Assigning Function Tags to Parsed Text** Don Blaheta, Eugene Charniak	**Domain-Specific Knowledge Acquisition Using Word Net** Dan Moldovan, Roxana Girju, Vasile Rus	
10:40	*BREAK*		
	NAACL	*NAACL*	*NAACL*
11:05	**Mostly-Unsupervised Statistical Segmentation of Japanese** Rie Kubota Ando, Lillian Lee	**Removing Left Recursion from Context-Free Grammars** Robert C. Moore	**Using Semantic Preferences to Identify Verbal Participation in Role Switching Alternations** Diana McCarthy
11:30	**Tagging Sentence Boundaries** Andrei Mikheev	**Left-To-Right Parsing and BiLexical Context-Free Grammars** Mark-Jan Nederhof, Giorgio Satta	**Acknowledgments in Human-Computer Interaction** Karen Ward, Peter A. Heeman
11:55	*NAACL BUSINESS MEETING*		
12:20	*LUNCH*		
	NAACL	*ANLP*	*ANLP*

1:50	**A Better Algorithm for the Alignment of Phonetic Sequences** Grzegorz Kondrak	**Large-Scale Controlled Vocabulary Indexing for Named Entities** Mark Wasson	**Automatic Discovery of Scenario-Level Pattern for Information Extraction** Roman Yangarber, Ralph Grishman, Pasi Tapanainen, Silja Huttunen
2:15	**Finite-State Reduplication in One-Level Prosodic Morphology** Markus Walther	**Building Lists for Named Entity Recognition** Mark Stevenson, Robert Gaizauskas	**Answer Extraction** Steven Abney, Michael Collins, Amit Singhal
2:40	**A finite-state and data-oriented method for grapheme to phoneme conversion** Gosse Bouma	**Automatic Identification of Phrases for Browsing Electronic Documents** Nina Wacholder, Judith L. Klavans, David Kirk Evans	**Sentence Simplification in Automatic Text Summarization** Hongyan Jing
3:05	*BREAK*		
3:30	**The Future—It isn't what it used to be** **Rick Rashid,** *Invited Speaker*		
4:30	*BREAK*		
	NAACL	***ANLP***	
4:40	**A Framework for Robust Semantic Interpretation** Carolyn Penstein Rose	**Named Entity Extraction from Noisy Input: Speech and OCR** Ralph Weischedel, David Miller, Sean Boisen, Richard Schwartz, Rebecca Stone	
5:05	**Understanding each other** Claire Gardent, Karsten Konrad	**The Use of Instrumentation in Grammar Engineering** Norbert Broeker	
5:30	**An Empirical Assessment of Semantic Interpretation** Martin Romacker, Udo Hahn	**The Efficiency of Multimodal Interaction** Philip R. Cohen, David McGee	
5:55	*CLOSE*		

TUTORIALS

Saturday 29 April 2000, 8:30AM-5:00PM

Information Retrieval, 8:30AM-12:00PM
James Allan, University of Massachusetts, Amherst

This tutorial will examine the role of Natural Language Processing in Information Retrieval (IR). It will start with a historical overview of how NLP has been used and abused over the 40 years of IR research, attempting to illustrate why distrust of NLP was rampant in the IR community. We will next explore reasons that NLP might have been doomed to failure in early IR contexts--where basic IR techniques already captured some of what NLP might have offered, and NLP techniques with even small error rates were therefore unable to offer more. We will then focus on NLP/IR success stories--both past and present--that show how NLP can be (and has been) successfully used in some applications. The last portion of the tutorial will focus on very recent applications of NLP for retrieval and organization purposes. We will discuss the use of NLP and IR as part of Web applications, and as applied to Web pages. The tutorial will also include a discussion of the exciting Question and Answer track from TREC-8 (1999). That task results showed that a combination of IR and NLP was highly effective, creating a great opportunity for future collaboration between the two fields.

The State of the Art in Language Modeling, 8:30AM-12:00PM
Joshua Goodman, Microsoft Research

This tutorial will cover the state-of-the-art in language modeling. The goal of language models is to predict the probability of words; this is directly useful for speech recognition, handwriting recognition, spelling correction, and other areas. Because language modeling is one of the most explored areas of statistical modeling, the state of the art is relatively advanced. Techniques from language modeling may be of interest to anyone pursuing probabilistic modeling, including those interested in statistical parsing, information retrieval, machine translation, and compression. The tutorial should be accessible to anyone with an elementary knowledge of probability.

The most basic language models -- n-gram models -- essentially just count occurrences of words in training data. I will describe six improvements over this simple baseline: smoothing, caching, skipping, sentence-mixture models, clustering, and parsing-based models.

1) Smoothing addresses the problem of data sparsity: there is rarely enough data to accurately estimate the parameters of a language model. Smoothing gives a way to combine less specific, more accurate information with more specific, but noisier data. I will describe two classic techniques: deleted interpolation and Katz (or Good-Turing) smoothing, and one recent technique, Modified Kneser-Ney smoothing, which is the best available.
2) Caching is a widely used technique that uses the observation that recently observed words are likely to occur again. Models from recently observed data can be combined with more general models to improve performance.
3) Skipping models use the observation that even words that are not directly adjacent to the target word contain useful
information.
4) Sentence-mixture models use the observation that there are many different kinds of sentences. By modeling each sentence type separately, performance is improved.
5) Clustering is one of the most useful language modeling techniques. Words can be grouped together into clusters through various automatic techniques; then the probability of a cluster can be predicted instead of the probability of the word. Clustering can be used to make smaller models or better performing ones.
6) All of the previous techniques ignore the structure of language. Recently, Ciprian Chelba has shown that techniques from statistical parsing can be used for improved models.

Finally, I will also talk about some practical aspects of language modeling. I will describe how freely available, off-the-shelf tools can be used to easily build language models, and how to use methods such as count cutoffs to compress language models.

Those who attend the tutorial should walk away with a broad understanding of the current techniques, and the background needed to either build their own language models, or to apply some of these techniques to other fields.

Finite-State Morphology/Phonology:Theory,Applications,and Recent Developments, 1:30PM-5:00PM
George Kiraz, Lucent Technologies Bell Labs

Morphology and phonology are crucial components in any large scale NLP and/or speech system. Finite-state (FS) approaches to morphology/phonology are the state-of-the-art in the field; they aim at modeling the morphology/phonology of language with computational devices whose computational power does not exceed that of FS machines. Such machines are of interest because they are easy to implement, fast to run and mathematically elegant. The purpose of this tutorial is to give a comprehensive introduction to the field of computational morphology, with emphasis on recent developments and practical techniques to build real systems.

The first hour of the tutorial will be introductory, bringing all the participants to a common understanding of the prerequisites involved: (1) morphology from a linguistics perspective, (2) automata theory from a computer-science perspective, and (3) how the two fields make up 'FS computational morphology'.

The second hour will concentrate on the theory of FS morphology. The various popular formalisms and notations will be explained. Algorithms for compiling formalisms into FS machines (i.e., rule compilers) will be detailed with comprehensive step-by-step examples. Algorithms for interpreting the formalisms directly will be outlined. Recent developments in the field will be emphasized.

The third hour will concentrate on building practical applications. An introduction to the available tools (and how to build such tools) will be given. This will be followed by building simple grammars, moving to more complex grammars, including challenging morphological and morphotactic phenomena (e.g., long-distance dependencies, heavy agglutinative morphology, reduplications, Semitic root-and-pattern, etc.). The tutorial will demonstrate how various linguistic descriptions of morphology and phonology (segmental, Autosegmental, prosodic, templatic, optimality theory, etc.) can be modeled with FS approaches.

The tutorial will end with 'Open Questions' detailing some of the issues that await research.

Machine Translation, 1:30PM-5:00PM
Kevin Knight, USC/ISI

Mechanical translation of one human language into another is one of the oldest applications of computer science, with roots going back to the 1940s and 1950s. It did not work very well then, and it does not work very well now. But, people use it all the time! MT programs can translate email, web pages, and online chat; with human post editing, they can translate corporate and government documents. There has also been an explosion in the number of language pairs available. The major force that continues to drive machine translation research is the idea that if it did work, people would use it a lot more. MT also provides a gauge of the breadth and depth of our scientific understanding of human language, including both interpretation and generation aspects.

This tutorial will cover a number of practical and theoretical topics. Most of the tutorial will consist of an in-depth look at several radically different types of MT systems. This will include both special-purpose and general-purpose MT. We will discuss how various MT systems are constructed, and what their strengths and weaknesses are. We will describe the workings of an interlingua-based system, and we will give a gentle yet rigorous description of a corpus-based statistical MT system. We will go through examples during the tutorial and also distribute small-scale (one-hour to eight-hour) assignments that can be done after the tutorial. These assignments demonstrate basic principles without requiring extensive programming.

We will also discuss the evaluation of MT systems, the economics of machine translation (including existing products and services), and the management of both source-language and target-language texts. This last topic includes ways of authoring documents to reduce ambiguity ("controlled language"), ways of preserving unresolved ambiguities in output text, and issues concerning pre-editing and post-editing of documents.

Finally, we will briefly discuss the future of MT, taking special account of current research in integrating MT with other natural language technologies (cross-language information retrieval, speech-to-speech translation, multilingual summarization, etc). We will also look at how current research in NLP is likely to affect MT in the future.

Syllabus:
1. Introduction to MT. Why is MT important? History of MT research and commercial development. General-purpose vs. special-purpose MT. Architectures that have been used for MT.
2. Building MT Systems. Syntactic transfer: building a lexicon, transfer rules. Interlingua-based MT: domain-specific interlingua design, parsing, semantic interpretation, generation. Direct transfer: statistical algorithms trained on bilingual corpora, text alignment, transliteration, example-based MT.
3. Managing the Input, Managing the Output. Pre-editing. Reducing ambiguity through controlled-language source text. Post-editing. Ambiguity preservation.
4. Evaluation of MT Systems. Why it's hard. Metrics that have been proposed and used. Strengths and weaknesses.
5. The Economics of MT. Products and services available.
6. The Future of MT. Integrating MT with other natural language technologies: cross-language information retrieval, speech-to-speech MT, summarizing foreign-language documents. How current NLP research bears on MT.

Table of Contents