Cognitive Computation Group
University of Illinois at Urbana-Champaign

ACL
The Annual Meeting of the Association of Computational Linguistics

Dan Roth

University of Illinois
Urbana-Champaign

Program co-chair ACL 2003 (with Erhard Hinrichs)

Hwee Tou Ng, Program co-chair ACL 2005 will give the ACL-2 presentation.

ACL

- The main conference in Computational Linguistics – Natural Language Processing.
  - ACL 2003
  - Trends in the general research area

- Many other conferences in this research area:

ACL

North American ACL; European ACL; Asian ACL (IJCNLP)

EMNLP: Conference onf Empirical Methods in NLP

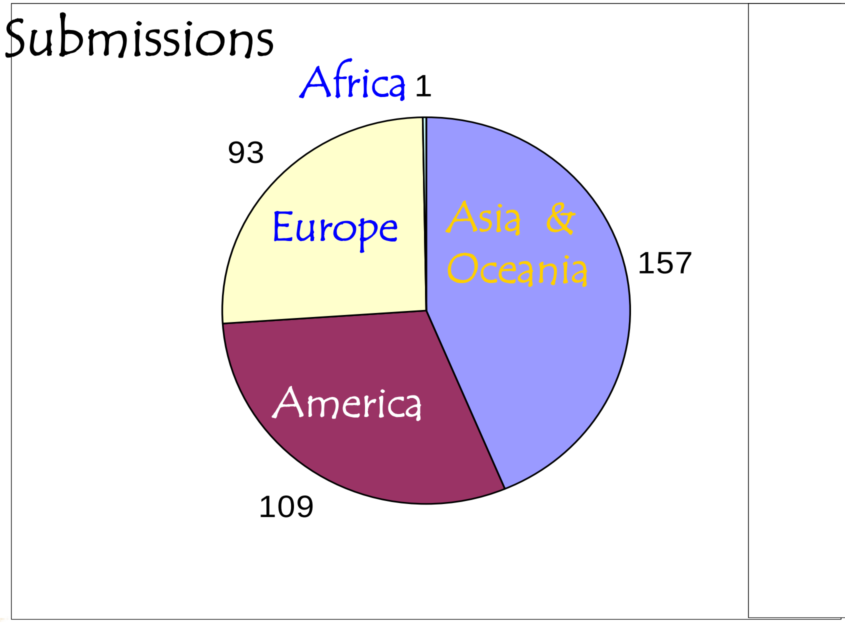
CoNLL: Conference on Natural Language Learning

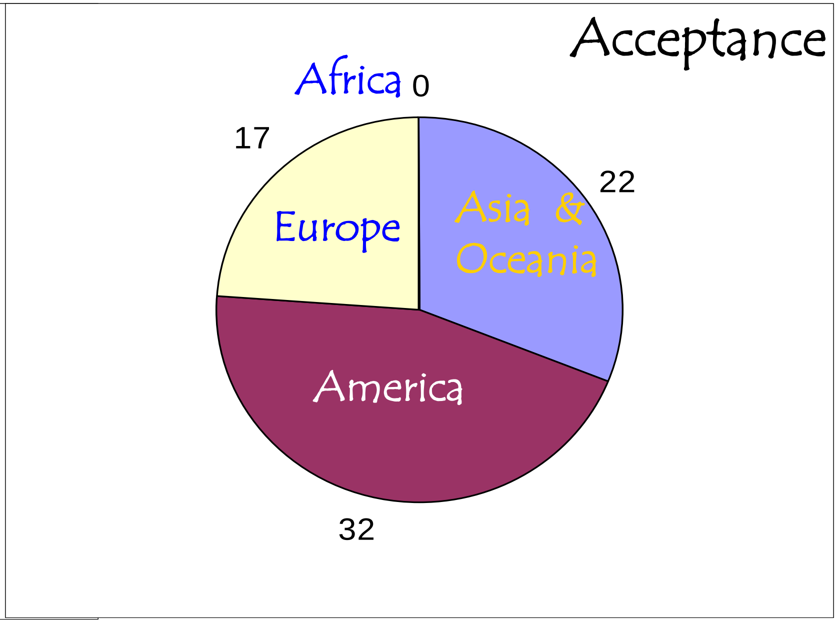
Many other smaller meetings/workshops

ACL 2003

2003: Sapporo, Japan ~700 participants

Submitted: 376→ 360 Accepted: 71 (19.73%)



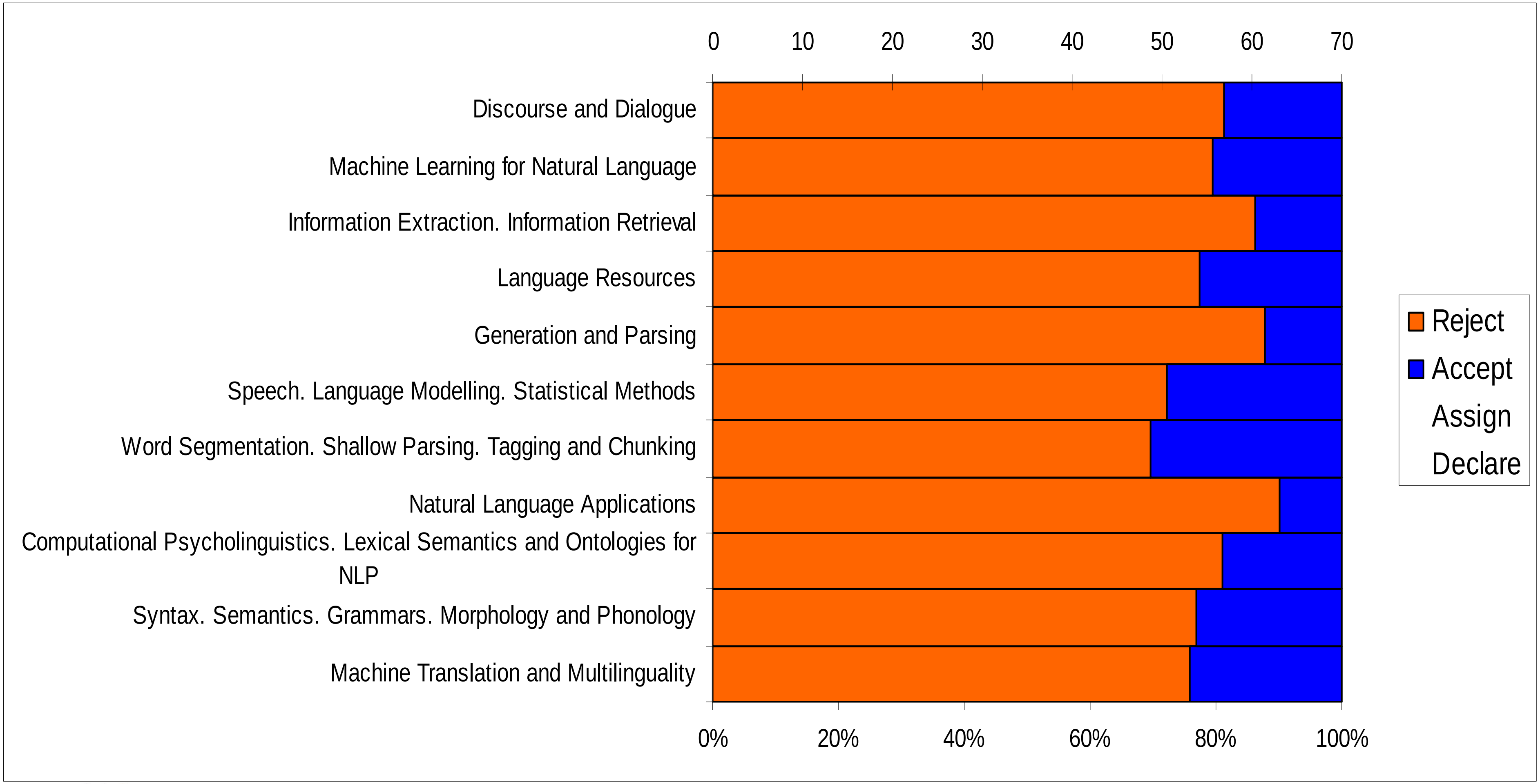


Major Topics Studied

- Discourse and Dialog
- Information Extraction; Information Retrieval
- Summarization and Generation
- Speech; language modeling
- Shallow parsing; tagging and chunking
- Lexical Semantics and Ontologies
- Computational Psycholinguistics
- Syntax; Semantics; morphology and phonology
- Machine Translation and Multilinguality
- Machine Learning for Natural Language

ACL 2003

Growth in: IE; Machine learning; Generation& parsing



Major Trends

- Is it all statistical?
- No, but there is a lot of emphasis on evaluation

- Almost all papers in 2003, and all papers in 2005 had an experimental evaluation.

- While Machine Learning methods are very dominant, things have changed considerably in the last 5 years or so –
  - No longer (only) simple counting
  - More sophisticated learning methods.
  - More structure; global constraints; more inference.

Not Simple Counting

- Early statistical work in NLP has roots in speech recognition

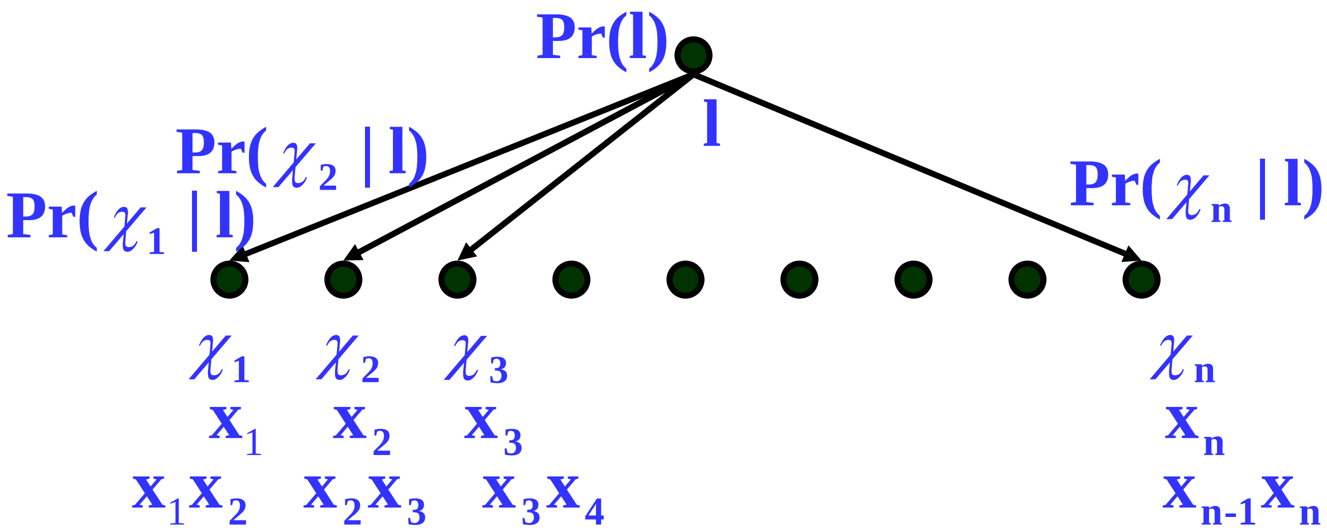
- Based on generative models.
- Estimate most likely parameters and use Bayes rule to obtain a classifier. Involves making simplifying assumptions.

- Learning Theory and algorithmic Machine Learning

➔

- Better understanding of the relations between probabilistic models and discriminative models; had a significant effect on work in NLP
- Key observation: everything is linear

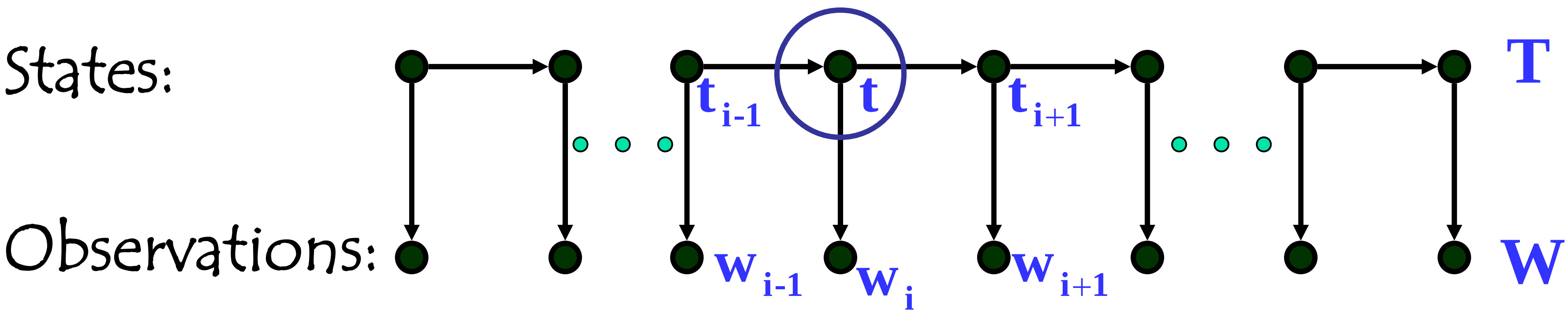
Example: naive Bayes



$$\text{prediction}(\mathbf{x}) = \operatorname{argmax}_{\{l=0.1\}} \sum_{i=0}^{n} c_{[x_i, l]} \chi_i(\mathbf{x})$$

Example: Markov Models

Input: $x = ((w_1 : t_1), \ldots (w_{i-1} : t_{i-1}), \underline{(w_i : ?)}, (w_{i+1} : t_{i+1}) \ldots)$



$$\text{prediction}(x) = \text{argmax}_{\{t \in T\}} \sum_{\chi \in X} c_{[x_i, t]} \chi(x)$$

Everything is Linear

$$\mathbf{prediction(x) = argmax_{\{t \in T\}} \sum_{\chi \in X} c_{[x_i, t]} \chi(x)}$$

- Probabilistic classifiers make use of a linear representation over some feature space.
- Possible to use the same representation, develop other ways of parameter estimation, driven directly by the eventual goal, to support better predictions.

  [Roth'98,Roth'99,Collins'01;Ng&Jordan'01,….]
- Work today is dominated by discriminative methods:

  Perceptron/Winnow, SVM, other optimization techniques.
- This insight can be extended to learning structure, via dynamic programming or other inference methods

Identifying Phrase Structure

[$_{NP}$ He] [$_{VP}$ reckons] [$_{NP}$ the current account deficit] [$_{VP}$ will narrow] [$_{PP}$ to] [$_{NP}$ only # 1.8 billion] [$_{PP}$ in] [$_{NP}$ September]

- Classifiers
  1. Recognizing "The beginning of NP"
  2. Recognizing "The end of NP"
     (or: word based classifiers: BIO representation)

  Also for other kinds of phrases…
- Some Constraints
  1. Phrases do not overlap
  2. Order of phrases
  3. Length of phrases
- Goal: infer a coherent set of phrases
- Multiple ways to train learners for these problems.

Predicate-Argument Representation

- For each predicate in a sentence
  Identify and classify all constituents that fill a semantic role
  - Core Arguments, e.g., Agent, Patient or Instrument
  - Their adjuncts, e.g., Locative, Temporal or Manner
- Multiple decisions; many interdependencies

A0 : leaver

A2 : benefactor

I **left** my pearls to my daughter-in-law in my will.

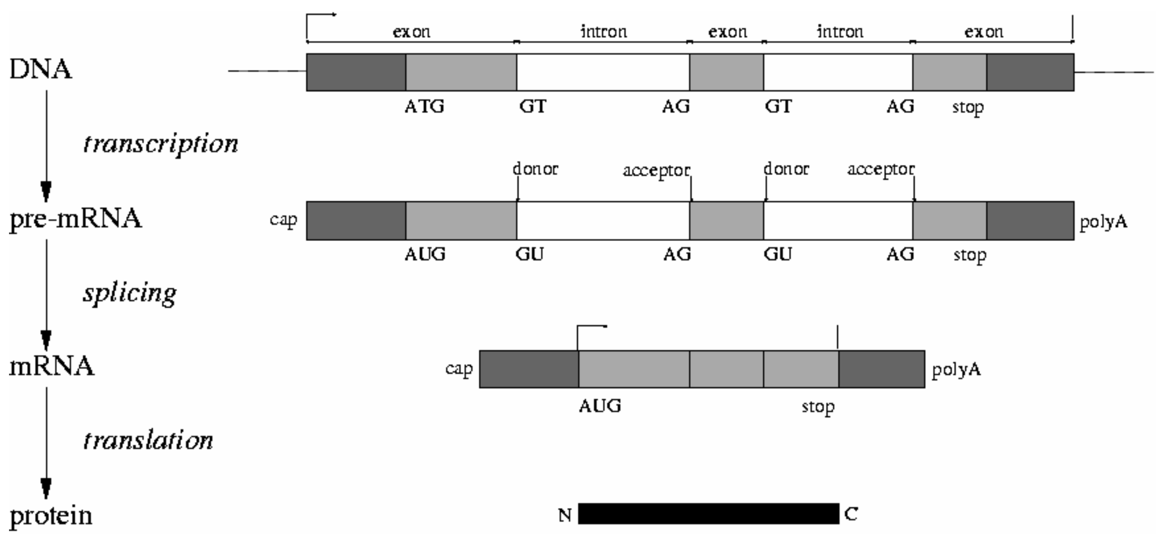
A1 : thing left

AM-LOC

Many Other Applications

- Shallow parsing; Full Parsing
- Morphology (Semitic languages)
- Information Extraction: Named Entity Recognition; Identifying Relations; Identifying Document Structure

- NLP methods in Computational Biology: Detecting Splice Sites



Learning @ ACL

- Significant work in NLP is done using Machine Learning techniques.
- But: not simple counting.
- Sophisticated techniques; a lot of structure and global constraints.

[CoNLL conferences]

Other Significant Lines of Work

- Parsing
- Discourse
- More Semantics
- More Inference
- Machine Translation

Parsing: Best Paper Award

- Accurate Un-lexicalized Parsing
- Dan Klein and Christophe Manning

- Work in NLP has shown steady improvement in Syntactic Parsing. The technology is being used in many applications.
- Problem: models are very large; not robust on new corpora

- Suggest linguistically motivated design choices for statistical treebank grammars that yield nearly state-of-the-art performance for unlexicalized
- Focus on un-lexicalized parser yield significantly smaller grammars; hopefully, more robust.

Discourse: Best Paper Award

- Towards a Model of Face-to-Face Grounding
  Yukiko Nakano, Gabe Reinstein, Tom Stocky, Justine Cassell

- The use of gaze for grounding in human conversation.
- Aa system that uses non-verbal behavior for grounding in an Embodied Conversational Agent.
- The papers shows:
  - An experiment showing how nonverbal cues can indicate whether a hearer understands an utterance, and
  - A small dialogue system in which some of these findings have been implemented; in particular, when the system detects nonverbal cues indicating that the hearer did not understand the system's last utterance, the likelihood of the system offering an elaboration is increased.

More Semantics

- How to get an handle on semantics?

- Annotate shallow semantics – predicate argument representation
- Two large annotation projects – PropBank and FrameNet
- A lot of machine learning based work to learn how to annotate new text.

- There is also a lot of work on lexical semantics; WSD; verbs; semantic relations;

More Inference

- More inference is done in the context of new problems.
- Most significant effort: on textual entailment

- Determine whether a given sentence implies another statement

WalMart defended itself in court today against claims that its female employees were kept out of jobs in management because they are women

Entails

Subsumed by

$\subseteq$

WalMart was sued for sexual discrimination

Invited Talks

- From Structure to Meaning: Simple Sentence-Structure Cues Guide Sentence Comprehension by Young Children
- Cynthia Fisher

  Psychology and the Beckman Institute, UIUC.

- Economics about Language
- Ariel Rubinstein

  School of Economics, Tel Aviv University and Princeton University

- Layout in NLP: The Case for Document Structure
- Donia Scott

  ITRI, the University of Brighton

Sister Conference Highlights at AAAI-05

Hwee Tou Ng

Department of Computer Science
National University of Singapore
nght@comp.nus.edu.sg

July 2005

ACL-2005

- Program Co-Chair of ACL-2005 (43rd Annual Meeting of the Association for Computational Linguistics)
- 25 – 30 June 2005, Ann Arbor, Michigan, USA
- 423 submissions (record number)
- Accepted 77 papers (18% acceptance rate)
- About 700 attendees

Area Breakdown of Papers

- Segmentation and tagging (2)
- Parsing (10)
- Semantics (12)
- Discourse and Dialog (8)
- Lexical acquisition from corpora (4)
- Generation (3)
- Corpus annotation (2)
- Summarization (3)
- Question answering (2)
- Information Extraction (9)
- Machine Translation (10)
- Speech and language modeling (6)
- Machine Learning and Statistical Methods (6)

Total = 77 papers

Dominant Traits of Papers

- Statistical approach
- Common evaluation data
- Quantitative performance measure
- Robust, broad coverage, somewhat shallow processing

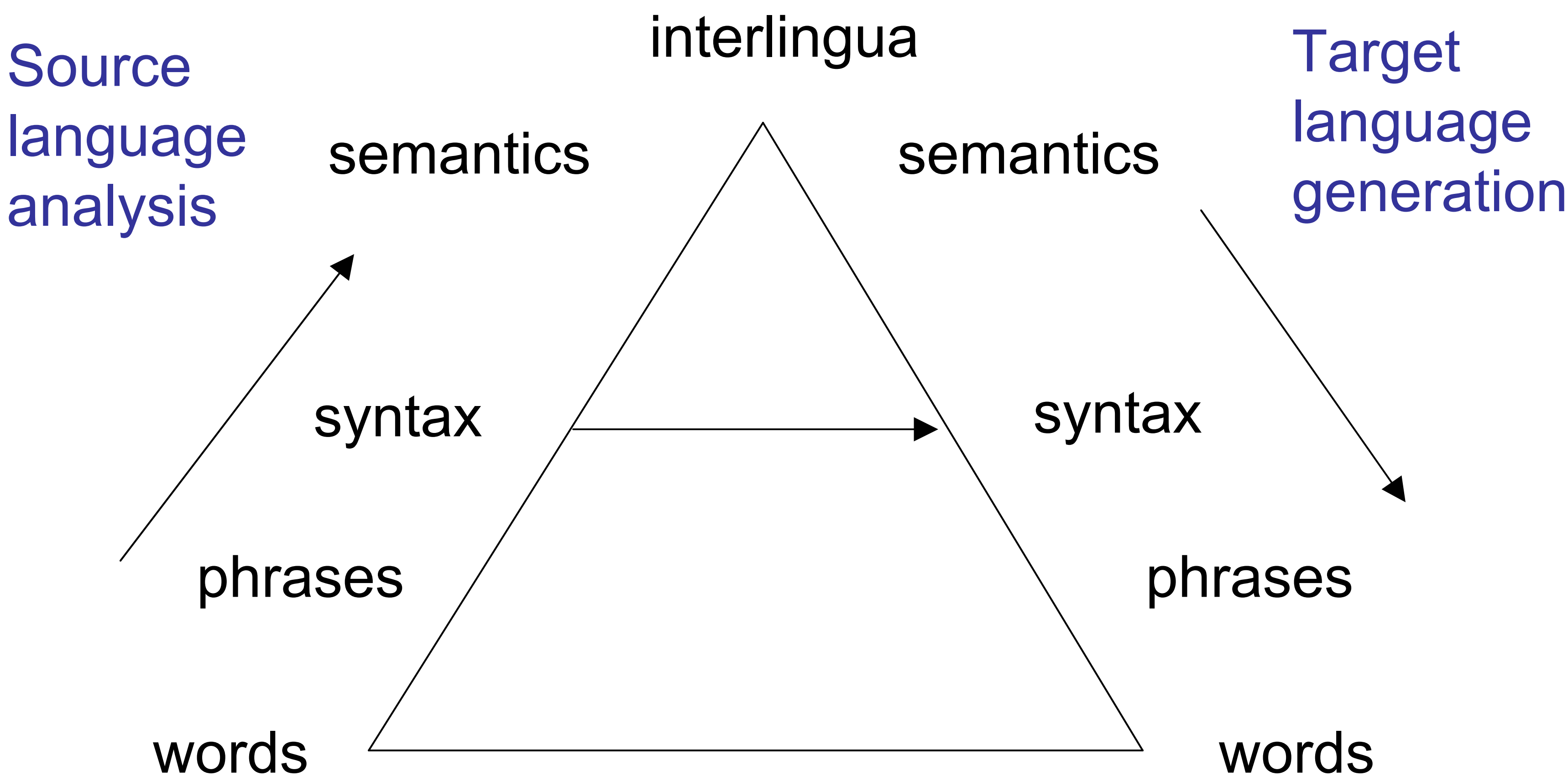
Best Paper Award at ACL-2005

- David Chiang, “A Hierarchical Phrase-Based Model for Statistical Machine Translation”

Machine Translation (MT)

- Annual NIST machine translation evaluation workshop since 2002
- Automatic evaluation metric: BLEU
  - Proposed and adopted in 2002
  - Reference translations and n-gram statistics
  - Reduce costly human evaluation
  - Facilitates rapid feedback on new MT algorithms
- Statistical machine translation (SMT)
  - Large bilingual parallel texts (bitexts)
  - Arabic-English & Chinese-English

MT Pyramid



Source
language
analysis
interlingua
semantics
semantics
Target
language
generation
syntax
syntax
phrases
phrases
words
words

Machine Translation

- Current state-of-the-art SMT approach:
  - Phrase-based machine translation (Pharaoh)
- Chiang’s paper
  - Proposed the use of hierarchical phrases (phrases that contain sub-phrases)
  - Weighted synchronous context-free grammar
  - Learns from raw bitexts *without* syntactic annotation

A Hierarchical Phrase-Based Model

澳洲	是	与	北韩	有	邦交
Aozhou	*shi*	*yu*	*Beihan*	*you*	*bangjiao*
Australia	is	with	North Korea	have	diplomatic relations

的	少数	国家	之一
de	*shaoshu*	*guojia*	*zhiyi*
that	few	countries	one of

Australia is one of the few countries that have diplomatic relations with North Korea

<X1 之一, one of X1>
<X1 的 X2, the X2 that X1> relative clause
<与 X1 有 X2, have X2 with X1> prepositional phrase

A Hierarchical Phrase-Based Model

{ S1, S1 }

⇒ …

⇒ { 澳洲 是 X3, Australia is X3 }

⇒ { 澳洲 是 X7 之一, Australia is one of X7 }

⇒ { 澳洲 是 X8 的 X9 之一,
Australia is one of the X9 that X8 }

⇒ { 澳洲 是 与 X10 有 X11 的 X9 之一,
Australia is one of the X9 that have X11 with X10 }

⇒ …

Chiang's Paper

- Bitext size: 16.4 million words
- English text (for language model): 155 million words
- Training process obtains a grammar of 2.2 million rules
- 7.5% relative improvement in BLEU score over Pharaoh (state-of-the-art phrase-based SMT system)
- Statistically significant improvement ($p < 0.01$)

Discourse

- Existing theories of discourse structure and coherence:
  - Rhetorical Structure Theory (Mann & Thompson)
  - Centering Theory (Grosz, Joshi, Weinstein)
- Barzilay & Lapata, ACL 2005, "Modeling Local Coherence: An Entity-based Approach"
  - A statistical approach to discourse coherence

Modeling Local Coherence: An Entity-based Approach

1. [The Justice Department]$_S$ is conducting an [anti-trust trial]$_O$ against [Microsoft Corp.]$_X$ with [evidence]$_X$ that [the company]$_S$ is increasingly attempting to crush [competitors]$_O$.
2. [Microsoft]$_O$ is accused of trying to forcefully buy into [markets]$_X$ where [its own products]$_S$ are not competitive enough to unseat [established brands]$_O$.
3. [The case]$_S$ revolves around [evidence]$_O$ of [Microsoft]$_S$ aggressively pressuring [Netscape]$_O$ into merging [browser software]$_O$.
4. [Microsoft]$_S$ claims [its tactics]$_S$ are commonplace and good economically.

Modeling Local Coherence: An Entity-based Approach

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics
1	S	O	S	X	O	-	-	-	-	-	-	-
2	-	-	O	-	-	X	S	O	-	-	-	-
3	-	-	S	O	-	-	-	-	S	O	O	-
4	-	-	S	-	-	-	-	-	-	-	-	S

Modeling Local Coherence: An Entity-based Approach

- Distribution of entities in locally coherent texts exhibits regularities
- Based on features of entity transitions (<S,S>, <S,O>, …, <-,->), learn parameters that yield a ranking score function

Modeling Local Coherence: An Entity-based Approach

- Evaluate the coherence of:
  - documents with good sentence ordering versus scrambled sentence ordering
  - automatically generated multi-document summaries
- Need to rank a coherent document higher than an incoherent document
- Achieved ranking accuracy of 81% to 90%

Parsing

- Common evaluation data set
  - Wall Street Journal Penn Treebank
  - Training data: Section 02 – 21
  - Test data: Section 23 (2,416 sentences of length $\leq$ 100 words)
- Charniak & Johnson, ACL 2005, “Coarse-to-fine n-best parsing and MaxEnt discriminative reranking”

Progress in Statistical Parsing

Section 23 of Penn Treebank, ≤ 100 words (2,416 sentences)

Parser	Labeled Recall	Labeled Precision	F-score
Magerman 95	84.0	84.3	84.1
Collins 96	85.3	85.7	85.5
Charniak 97	86.7	86.6	86.6
Ratnaparkhi 97	86.3	87.5	86.9
Collins 99	88.0	88.3	88.1
Charniak 00	89.6	89.5	89.5
Charniak & Johnson 05	NA	NA	91.0

Conclusion

- NLP submissions are growing, suggesting increased NLP research activities
- Steady progress made as calibrated by quantitative performance measures