



# Human Language

Christopher Manning

Stanford University

<http://nlp.stanford.edu/~manning/>

CS300 talk, 2009



---

# Simple Natural Language Processing

---

- Now that our computers are so fast and have so much memory, we should be putting some NLP into programs!
- Simple e.g.: The **file** command
  - `> file ~/current/NLP-notes`
  - `/user/manning/current/NLP-notes: ASCII English text`
  - `> file proposal.txt`
  - `proposal.txt: data`
    - Uh oh! This is also English text. Just has a couple of funny characters (Microsoft “smart quotes”) in it somewhere....
  - Are these “documents” English text to **file**?
    1. I am going to go visit Sonoma county this weekend.
    2. THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.



---

# Simple NLP

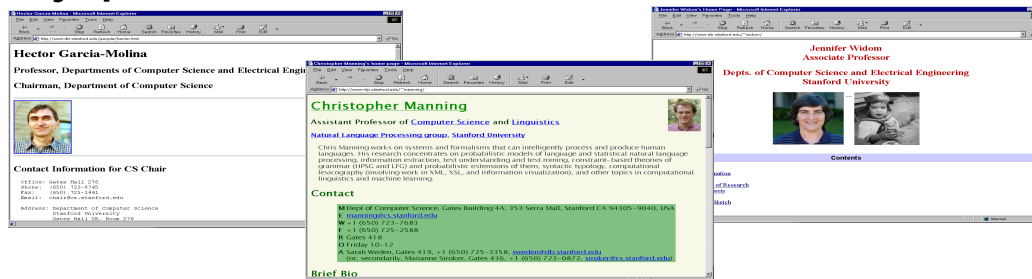
---

- Are these “documents” English text to the **file** command?
    - I am going to go visit Sonoma county this weekend.
    - THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
  - No! Pure ASCII with the word ‘the’ or ‘The’
  - We can do better than this with probabilistic language models (markov models over character sequences)!
    - Character-sequence probabilistic language models are **very** effective language recognizers.
    - Far more robust: **file** is an overextended set of hacks....
      - Even Java vs. C++ can easily confuse it
  - Now happening a little ... at Microsoft, Google, ...
-



# Where's the natural language?

- Everywhere.
- Most of the information in most organizations is material in human languages (reports, customer email, discussion papers, specifications, etc.)
- Lots of unstructured text/web information that we'd like to turn into usable knowledge
  - employs(stanfordUniversity, chrisManning)
- We'd like computer agents to see meanings and be able to intelligently process text





---

## Jeremy Zawodny [random blogger] sez...

---

Back in the late 90s when I was occasionally building things that passed for *knowledge management tools* at [Marathon Oil](#), there was all this talk about *knowledge workers*. These were people who'd have vast quantities of information knowledge at their fingertips. All they needed was a way to organize, classify, index, search, and collaborate with it.

I think we've made it. But the information isn't organized like I had envisioned a few years ago. It's just this big ugly mess known as **The Web**. Lots of pockets of information from mailing lists, weblogs, software projects, communities, and company web sites are loosely tied together by hyperlinks. **There's no grand schema or centralized database. There's little structure or quality control. No global vocabulary.**

But even with all that going against it, it's all indexed and easily searchable thanks largely to [Google](#) and the companies that preceded it (Altavista, Yahoo, etc.). Most of the time it actually works.

Amazing!

---

5900 SUPER  
English عربي واضح صوت بشري  
\$199

الاثنين 23/10/1430 هـ - الموافق 12/10/2009 م (آخر تحديث) الساعة 5:34 (مكة المكرمة)، 2:34 (غرينتش)

الصفحة الرئيسية: أخبار

## إسرائيل تخرق حقوق تلاميذ غزة



إسرائيل دمرت في حربها على غزة 18 مدرسة من أصل 641 مدرسة بالقطاع (الفرنسية-أرشيف)

طالبت منظمة **هيومن رايتس ووتش** السلطات الإسرائيلية بأن ترفع فوراً القيود التي تحرم تلاميذ المدارس العامة في **قطاع غزة** من الكتب واحتياجات المدارس الأساسية مثل الأوراق والأقلام.

وقالت المنظمة إن إسرائيل تشدد منع الواردات إلى غزة من مختلف السلع الأساسية من الأغذية إلى أدوات البناء، مما أدى إلى عجز بالغ تمثل في عدم قدرة التلاميذ على الحصول على احتياجاتهم الدراسية مع بدء العام الدراسي منذ أكثر من شهر، وجعلهم يضطرون للمشاركة في الكتب الدراسية مع بعضهم البعض أو أن يدور الكتاب الواحد على أكثر من طالب.

- متملقات
- شح القرباسية يشوش التعليم بفترة
  - موسم دراسي غير عادي في غزة
  - منظمة إسرائيلية: الاحتلال ينكل بطلاب الخليل كل صباح
  - مؤسسات تحذر من تدهور التعليم بفترة
- أهم أخبار الصفحة الرئيسية
- مشعل يرفض تبرير عباس بشأن التقرير
  - منظمة التحرير تتمسك بموعد المصالحة
  - الأمم للتحدة تدعو لهبة إنسانية باليمن
  - حظر تجول بالرمادي بعد ثلاثة تفجيرات





طالبت منظمة هيومن رايتس ووتش السلطات الإسرائيلية بأن ترفع فوراً القيود التي تحرم تلاميذ المدارس العامة في قطاع غزة من الكتب واحتياجات المدارس الأساسية مثل الأوراق والأقلام.

Google translate

Called Human Rights Watch, the Israeli authorities to immediately lift restrictions that prohibit public school students in the Gaza Strip of books and basic school needs such as paper and pens.



# The early history of NLP/MT: 1950s







طالبت منظمة هيومن رايتس ووتش السلطات الإسرائيلية بأن ترفع فوراً القيود التي تحرم تلاميذ المدارس العامة في قطاع غزة من الكتب واحتياجات المدارس الأساسية مثل الأوراق والأقلام.

Called on organization Human Rights Watch the Israeli authorities to immediately lift restrictions that deny public school students in the Gaza Strip books

Google translate

Called Human Rights Watch, the Israeli authorities to immediately lift restrictions that prohibit public school students in the Gaza Strip of books and basic school needs such as paper and pens.



“Also knowing nothing official about, but having guessed and inferred considerable about, the powerful new mechanized methods in cryptography—methods which I believe succeed even when one does not know what language has been coded—one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’ ” – Warren Weaver, March 1947





“When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’ ”  
– Warren Weaver, March 1947



“... as to the problem of mechanical translation, I frankly am afraid that the [semantic] boundaries of words in different languages are too vague ... to make any quasi-mechanical translation scheme very hopeful.”  
– Norbert Wiener, April 1947



# Phrase-Based Translation

这 7人 中包括 来自 法国 和 俄罗斯 的 航 员 .

the	7 people	including	by some	and	the russian	the	the astronauts	,
it	7 people included		by france	and the	the russian		international astronautical	of rapporteur .
this	7 out	including the	from	the french	and the russian	the fifth		.
these	7 among	including from		the french and	of the russian	of	space	members .
that	7 persons	including from the		of france	and to	russian	of the aerospace	members .
	7 include		from the	of france and	russian		astronauts	. the
	7 numbers include		from france		and russian		of astronauts who	."
	7 populations include		those from france		and russian		astronauts .	
	7 deportees included		come from	france	and russia	in	astronautical	personnel ;
	7 philtrum	including those from		france and	russia	a space		member
		including representatives from		france and the	russia		astronaut	
		include	came from	france and russia			by cosmonauts	
		include representatives from		french	and russia		cosmonauts	
		include	came from france		and russia 's		cosmonauts .	
		includes	coming from	french and	russia 's		cosmonaut	
				french and russian		's	astronautical	member .
				french	and russia		astronauts	
					and russia 's			special rapporteur
					, and	russia		rapporteur
					, and russia			rapporteur .
					, and russia			
					or	russia 's		

Table 1: #11# the seven - member crew includes astronauts from france and russia .



# Phrase-Based Translation

这 7人 中包括 来自 法国 和 俄罗斯 的 航 员 .

the	7 people	including	by some	and	the russian	the	the astronauts	,
it	7 people	included	by france	and the	the russian		international astronautical	of rapporteur .
this	7 out	including the	from	the french	and the russian	the fifth		.
these	7 among	including from		the french and	of the russian	of	space	members .
that	7 persons	including from the		of france	and to	russian	of the aerospace	members .
	7 include	from the		of france and	russian		astronauts	. the
	7 numbers include	from france		and russian			of astronauts who	."
	7 populations include	those from france		and russian			astronauts .	
	7 deportees included	come from	france	and russia		in	astronautical	personnel ;
	7 philtrum	including those from	france and	russia		a space		member
		including representatives from	france and the	russia			astronaut	
		include	came from	france and russia			by cosmonauts	
		include representatives from	french	and russia			cosmonauts	
		include	came from france	and russia 's			cosmonauts .	
		includes	coming from	french and	russia 's		cosmonaut	
				french and russian	's		astronavigation	member .
				french	and russia		astronauts	
				and russia 's				special rapporteur
				, and	russia			rapporteur
				, and russia				rapporteur .
				, and russia				
				or	russia 's			

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.  
 Try to output a sentence with frequent English word sequences.



# Phrase-Based Translation

这	7人	中包括	来自	法国	和	俄罗斯	的	航	员	.
the	7 people	including	by some	and	the russian	the	the astronauts	,		
it	7 people included	by france	and the	the russian	international astronautical	of rapporteur .				
this	7 out	including the	from	the french	and the russian	the fifth	.			
these	7 among	including from	the french and	of the russian	of	space	members	.		
that	7 persons	including from the	of france	and to	russian	of the	aerospace	members		
	7 include	from the	of france and	russian	astronauts	.	the			
	7 numbers include	from france	and russian	of astronauts who	.					
	7 populations include	those from france	and russian	astronauts .						
	7 deportees included	come from	france	and russia	in	astronautical	personnel	;		
	7 philtrum	including those from	france and	russia	a space	member				
		including representatives from	france and the	russia	astronaut					
		include	came from	france and russia	by cosmonauts					
		include representatives from	french	and russia	cosmonauts					
		include	came from france	and russia 's	cosmonauts .					
		includes	coming from	french and	russia 's	cosmonaut				
				french and russian	's	astronavigation	member .			
				french	and russia	astronauts				
				and russia 's			special rapporteur			
				, and	russia		rapporteur			
				, and russia			rapporteur .			
				, and russia						
				or	russia 's					

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.  
 Try to output a sentence with frequent English word sequences.



# Phrase-Based Translation

这	7人	中包括	来自	法国	和	俄罗斯	的	航	员	.
the	7 people	including	by some	and	the russian	the	the astronauts			,
it	7 people included	by france		and the	the russian		international astronautical	of rapporteur		.
this	7 out	including the	from	the french	and the russian	the fifth				.
these	7 among	including from		the french	and	of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the	aerospace	members	.
	7 include	from the		of france and	russian		astronauts			.
	7 numbers include	from france		and russian		of astronauts who				.
	7 populations include	those from france		and russian		astronauts				.
	7 deportees included	come from	france	and russia		in	astronautical	personal		;
	7 philtrum	including those from	france and	russia		a space		member		.
		including representatives from	france and the	russia		astronaut				.
		include	came from	france and russia		by cosmonauts				.
		include representatives from	french	and russia		cosmonauts				.
		include	came from france	and russia 's		cosmonauts				.
		includes	coming from	french and	russia 's	cosmonaut				.
				french and russian		's	astronavigation	member		.
				french	and russia	astronauts				.
				and russia 's				special rapporteur		.
				, and	russia			rapporteur		.
				, and russia				rapporteur		.
				, and russia						.
				or	russia 's					.

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.  
 Try to output a sentence with frequent English word sequences.

# Why NLP is difficult: Newspaper headlines



- Minister Accused Of Having 8 Wives In Jail
- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- China to Orbit Human on Oct. 15
- Local High School Dropouts Cut in Half
- Red Tape Holds Up New Bridges
- Clinton Wins on Budget, but
- Hospitals Are Sued by 7 Families
- Police: Crack Found in Maryland

## Minister Accused Of Having 8 Wives In Jail

May 21, 2007 06:49 AM



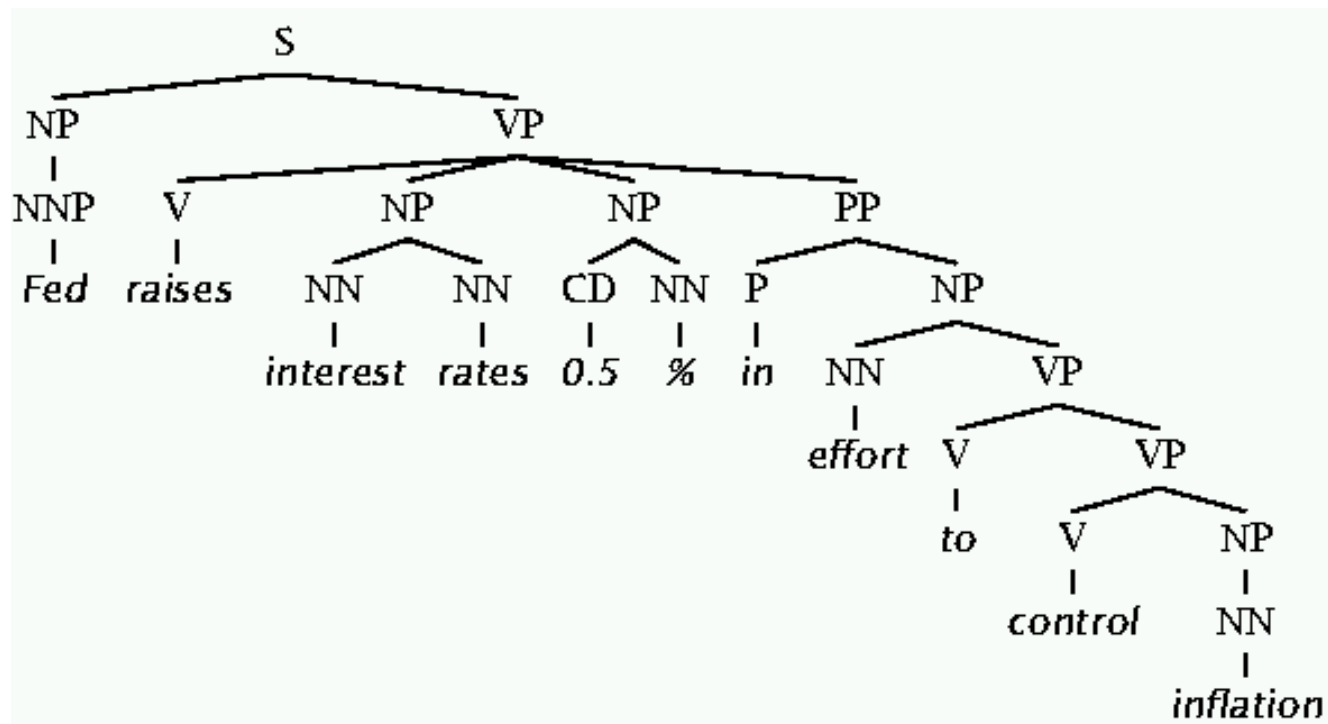
**ATLANTA (AP)** -- A man who served two years in prison has been jailed again for marrying more women.

Bishop Anthony Owens, Ga., is in a Gwinnett County jail after four women claimed he had sex with them after being released from prison.



# Inverse problems: Ambiguity close up

- Tree for: *Fed raises interest rates 0.5% in effort to control inflation* (NYT headline 5/17/00)





# Where are the ambiguities?

## Part of speech ambiguities

		VB				
	VBZ	VBP	VBZ			
NNP	NNS	NN	NNS	CD	NN	
<i>Fed</i>	<i>raises</i>	<i>interest</i>	<i>rates</i>	<i>0.5</i>	<i>%</i>	

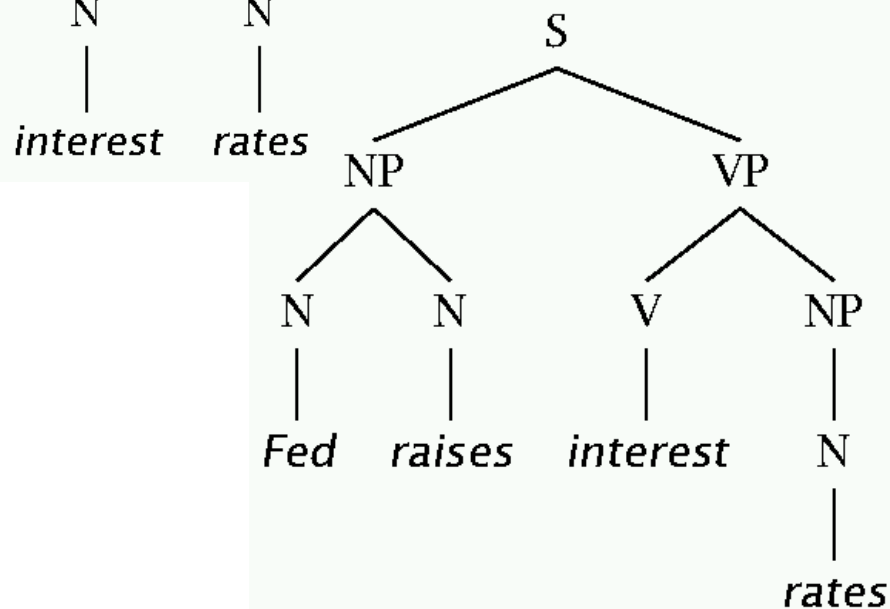
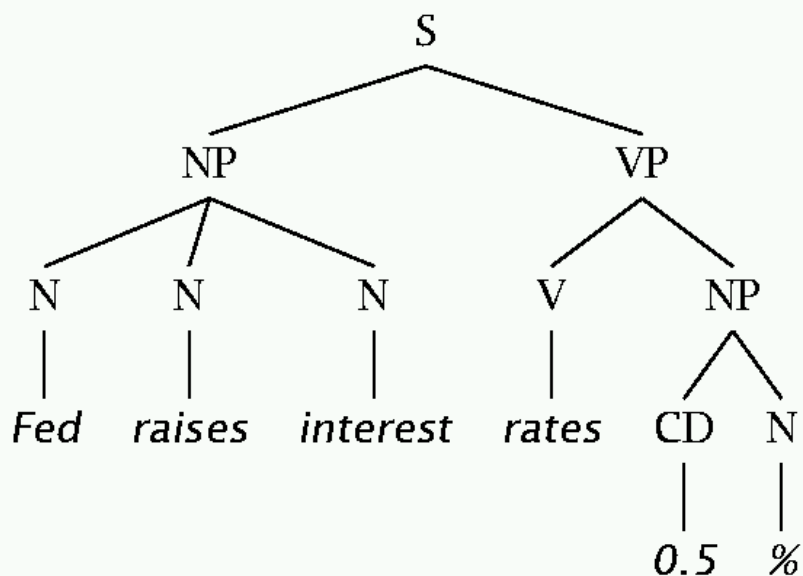
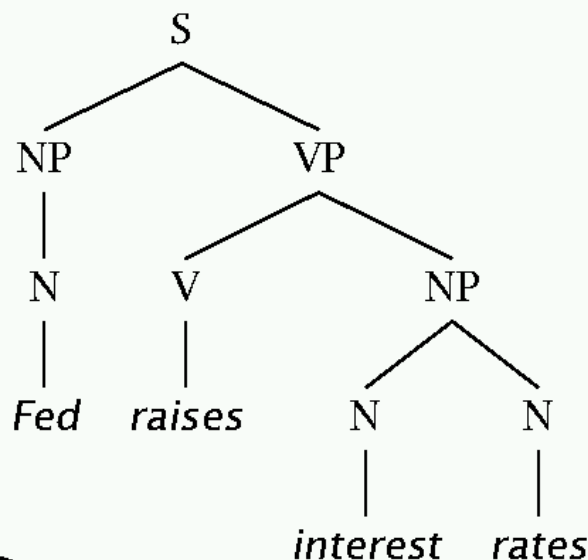
*Syntactic  
attachment  
ambiguities*

*in effort  
to control  
inflation*

*Word sense ambiguities: Fed → “federal agent”  
interest → a feeling of wanting to know or learn more*

*Semantic interpretation ambiguities above the word level*

# The bad effects of V/N ambiguities



---

# Research areas of interest: NLP/Comp Ling/text/ML

---



- Statistical NLP models: Combining sophisticated linguistics and machine learning models
  - Unsupervised lexicon and grammar induction
  - Learning of knowledge from reading (large amounts of) text data
  - Statistical machine translation
  - Learning about interdisciplinarity in science, politics, etc. from text
  - NLP for other languages
  - Statistical parsing, name recognition, joint learning, etc.
-



# Extracting Structured Knowledge

**Lawrence Livermore National Laboratory**

From Wikipedia, the free encyclopedia

Coordinates: 37°6′024″N 121°7′09547″W﻿ / ﻿﻿ / ﻿

The **Lawrence Livermore National Laboratory** (LLNL) in Livermore, California is a scientific research laboratory founded by the University of California in 1952. It is funded by the United States Department of Energy (DOE) and managed by Lawrence Livermore National Security, LLC (LLNS), a partnership of the University of California, Bechtel Corporation, Babcock and Wilcox, the URS Corporation, and Battelle Memorial Institute. On October 1, 2007 LLNS assumed management of LLNL from the University of California, which had exclusively managed and operated the Laboratory since its inception 55 years before.

Lawrence Livermore National Laboratory	
<b>Motto</b>	"Science in the national interest"
<b>Established</b>	1952 by the University of California
<b>Research Type</b>	National security, nuclear science
<b>Budget</b>	US\$1.6 billion
<b>Director</b>	George H. Miller
<b>Staff</b>	6,800
<b>Location</b>	Livermore, California
<b>Campus</b>	3.2 km² (800 acres)
<b>Operating Agency</b>	Lawrence Livermore National Security, LLC
<b>Website</b>	<a href="http://www.llnl.gov">www.llnl.gov</a>

**Contents** [hide]

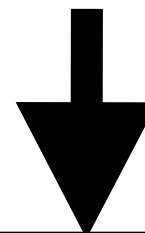
- Background
- Origins
- Weapons projects
- Plutonium research
- National Ignition Facility and photon science
- Global security program
- Other programs
- Key accomplishments
- Unique facilities
- World-class computers
- Sponsors
- Directors
- Organization
- Footnotes
- References
- External links and sources

**Background**

LLNL is self-described as "a premier research and development institution for science and technology applied to national security."<sup>[1]</sup> Its principal responsibility is ensuring the safety, security and reliability of the nation's nuclear weapons through the application of advanced science, engineering and technology. The Laboratory also applies its special expertise and multidisciplinary capabilities to preventing the proliferation and use of weapons of mass destruction, bolstering homeland security and solving other nationally important problems, including energy and environmental security, basic science and economic competitiveness.

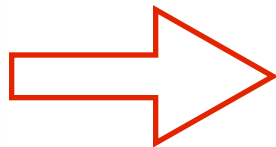
LLNL is home to many unique facilities and a number of the most powerful computer systems in the world, according to the TOP500 list, including Blue Gene/L, the world's fastest computer from 2004 until Los Alamos National Laboratory's Roadrunner supercomputer surpassed it in 2008. The Lab is a leader in technical innovation: since 1978, LLNL has received a total of 118 prestigious R&D 100 Awards, including

“The Lawrence Livermore National Laboratory (LLNL) in Livermore, California is a scientific research laboratory founded by the University of California in 1952.”



LLNL EQ Lawrence Livermore National Laboratory  
LLNL LOC-IN California  
Livermore LOC-IN California  
LLNL IS-A scientific research laboratory  
LLNL FOUNDED-BY University of California  
LLNL FOUNDED-IN 1952

# Goal: Machine-readable summaries



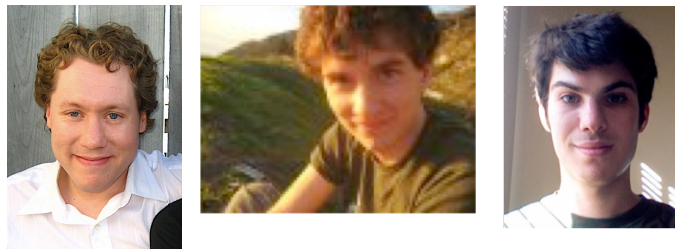
Subject	Relation	Object
p53	is_a	protein
Bax	is_a	protein
p53	has_function	apoptosis
Bax	has_function	induction
apoptosis	involved_in	cell_death
Bax	is_in	mitochondrial outer membrane
Bax	is_in	cytoplasm
apoptosis	related_to	caspase activation
...	...	...

Textual abstract:  
Summary for human

Structured knowledge  
extraction: Summary for  
machine



# Machine Reading with Distant Supervision



Corpus

 WIKIPEDIA

2.3 million articles  
31.5 million sentences

Training Set

 Freebase

2700 relations > 10 instances  
3.7 million entities  
5.2 million instances

Mintz, Bills, Snow, Jurafsky (2009) Distant supervision for relation extraction without labeled data. ACL-2009.



# Frequent Freebase Relations

Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

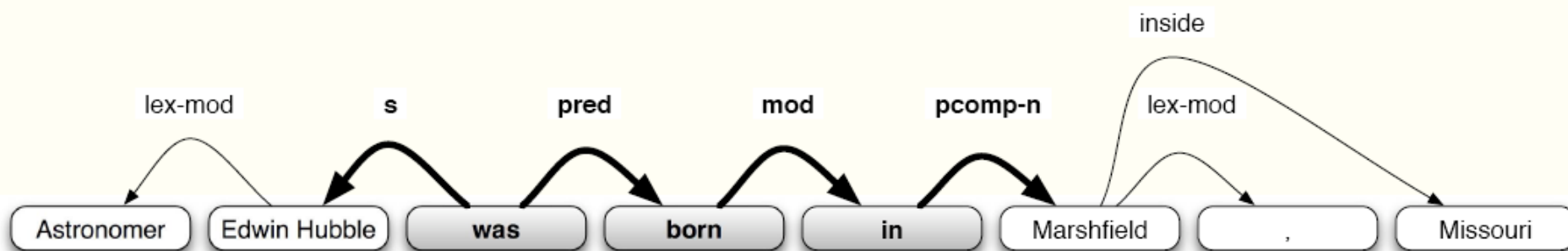




# Lexical and Syntactic Features

Astronomer Edward Hubble was born in Marshfield, Missouri

Feature type	Left window	NE1	Middle	NE2	Right window
Lexical	[]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[]
Lexical	[Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[.]
Lexical	[#PAD#, Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[, Missouri]
Syntactic	[]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[]
Syntactic	[Edwin Hubble ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[]
Syntactic	[Astronomer ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[]
Syntactic	[]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>lex-mod</sub> ,]
Syntactic	[Edwin Hubble ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>lex-mod</sub> ,]
Syntactic	[Astronomer ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>lex-mod</sub> ,]
Syntactic	[]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>inside</sub> Missouri]
Syntactic	[Edwin Hubble ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>inside</sub> Missouri]
Syntactic	[Astronomer ↓ <sub>lex-mod</sub> ]	PER	[↑ <sub>s</sub> was ↓ <sub>pred</sub> born ↓ <sub>mod</sub> in ↓ <sub>pcomp-n</sub> ]	LOC	[↓ <sub>inside</sub> Missouri]





---

# New relations learned

---

- Montmartre **IS-IN** Paris
  - Fort Erie **IS-IN** Ontario
  - Fyoder Kamesnky **DIED-IN** Clearwater
  - Upton Sinclair **WROTE** Lanny Budd
  - Vince McMahon **FOUNDED** WWE
  - Thomas Mellon **HAS-PROFESSION** Judge
-

---

# Human evaluation: precision using Mechanical Turk labelers

---



<i>Feature</i>	<i>Precision</i>
Syntactic	.67
Lexical	.66
Both	.69



---

# Where syntactic knowledge helps

---

- **Back Street** is a 1932 film made by Universal Pictures, **directed** by **John M. Stahl**, and produced by Carl Laemmle Jr.
- **Back Street** and **John M. Stahl** are very far apart in surface string
- But are close together in dependency parse



# Textual inference

---

- Does premise  $P$  justify inference to hypothesis  $H$ ?
  - An informal, intuitive notion of inference: not strict logic
  - Focus on local inference steps, not long chains of deduction
  - Emphasis on variability of linguistic expression
- Robust, accurate textual inference would enable:
  - Semantic search: H: *lobbyists attempting to bribe U.S. legislators*  
P: *The A.P. named two more senators who received contributions engineered by lobbyist Jack Abramoff in return for political favors.*
  - Question answering: H: *Who bought J.D. Edwards?*  
P: *Thanks to its recent acquisition of J.D. Edwards, Oracle will soon be able...*
  - Customer email response
  - Document summarization

# An example (MacCartney & Manning 2008)

P *Jimmy Dean refused to move without blue jeans.*

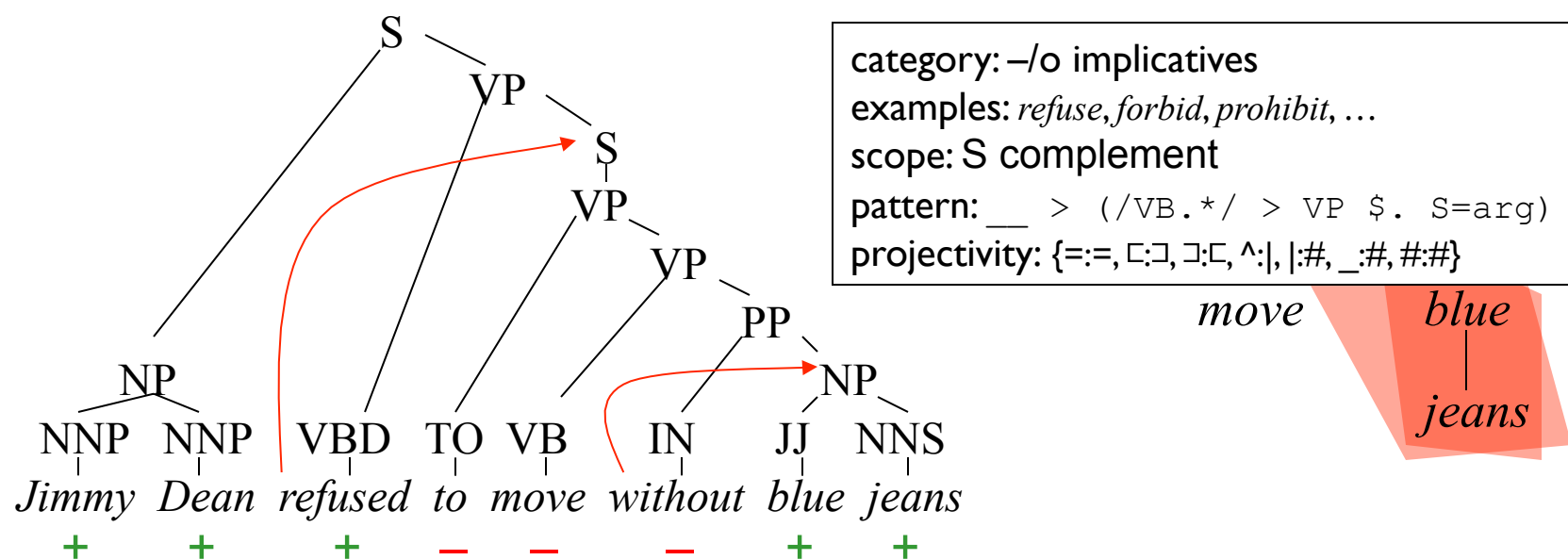
H *James Dean didn't dance without pants*  
*yes*

OK, the example is contrived, but it compactly exhibits containment, exclusion, and implicativity



# Step I: Linguistic analysis

- Tokenize & parse, etc. input sentences
- Identify items w/ special projectivity & determine scope
- Problem: PTB-style parse tree  $\neq$  semantic structure!



- Solution: specify scope in PTB trees using Tregex [Levy & Andrew 06]

## Step 2: Alignment

<i>P</i>	<i>Jimmy</i>	<i>refused</i>			<i>move</i>	<i>without</i>	<i>blue</i>	<i>jeans</i>
	<i>Dean</i>	<i>to</i>						
<i>H</i>	<i>James</i>		<i>did</i>	<i>n't</i>	<i>dance</i>	<i>without</i>		<i>pants</i>
	<i>Dean</i>							
<i>edit index</i>	1	2	3	4	5	6	7	8
<i>edit type</i>	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB

- Alignment as sequence of *atomic phrase edits*
- Ordering of edits defines path through intermediate forms
  - Need not correspond to sentence order
- Decomposes problem into atomic inference problems





## Step 3: Lexical entailment classification

- Goal: predict entailment relation for each edit, based solely on lexical features, independent of context
- Approach: use lexical resources & machine learning
- Feature representation:
  - WordNet features: synonymy (=), hyponymy ( $\sqsubset/\sqsupset$ ), antonymy (|)
  - Other relatedness features: Jiang-Conrath (WN-based), NomBank
  - Fallback: string similarity (based on Levenshtein edit distance)
  - Also lexical category, quantifier category, implication signature
- Decision tree classifier
  - Trained on 2,449 hand-annotated lexical entailment problems
  - E.g., SUB(*gun*, *weapon*):  $\sqsubset$ , SUB(*big*, *small*): |, DEL(*often*):  $\sqsubset$
  - >99% accuracy on training data — captures relevant distinctions



# Step 3: Lexical entailment classification

<i>P</i>	<i>Jimmy Dean</i>	<i>refused to</i>			<i>move</i>	<i>without</i>	<i>blue</i>	<i>jeans</i>
<i>H</i>	<i>James Dean</i>		<i>did</i>	<i>n't</i>	<i>dance</i>	<i>without</i>		<i>pants</i>
<i>edit index</i>	1	2	3	4	5	6	7	8
<i>edit type</i>	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
<i>lex feats</i>	strsim= 0.67	implic: -/o	cat:aux	cat:neg	hypo			hyper
<i>lex entrel</i>	=		=	^	⊃	=	⊂	⊂

# Step 4: Entailment projection

P	Jimmy Dean	refused to			move	without	blue	jeans
H	James Dean		did	n't	dance	without		pants
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
lex feats	strsim= 0.67	implic: -/o	cat:aux	cat:neg	hypo			hyper
lex entrel	=		=	^	⊃	=	⊃	⊃
projec- tivity	↑	↑	↑	↑	↓	↓	↑	↑
atomic entrel	=		=	^	⊃	=	⊃	⊃

inversion

# Step 5: Entailment composition

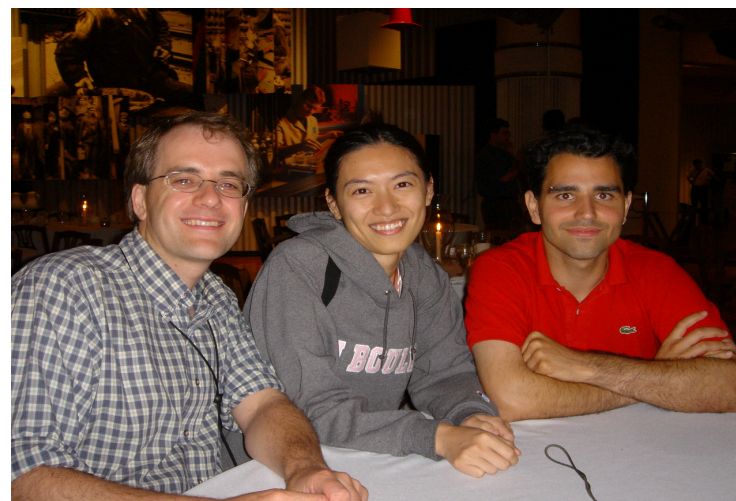
P	Jimmy Dean	refused to			move	without	blue	jeans
H	James Dean		did	n't	dance	without		pants
edit index	1	2	3	4	5	6	7	8
edit type	SUB	DEL	INS	INS	SUB	MAT	DEL	SUB
lex feats	strsim=0.67	implic:-/o	cat:aux	For example:				hyper
lex entrel	=		=	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <p>human <math>\wedge</math> nonhuman</p> <p>fish   human <math>\rightarrow</math> fish &lt; nonhuman</p> </div>				⊂
projectivity	↑	↑	↑					↑
atomic entrel	=		=	^	⊂	=	⊂	⊂
composition	=			⊂	⊂	⊂	⊂	⊂

final answer  
36



# The Stanford NLP Group

- Chris Manning
- Dan Jurafsky
- Computational Linguistics and Speech (“language technology”)
- Enthusiastic, smart, and hard-working – or hungry – students welcome!
- [nlp.stanford.edu](http://nlp.stanford.edu)



---

# The End

---

*Thank  
you!*

