

# **WS'02: Preworkshop Activities**

**Sanjeev Khudanpur**

**July 15, 2002.**

**WS'02 Opening Day Presentations**

Center for Language and Speech Processing  
The Johns Hopkins University

# The Annual Workshop Cycle:

## It Really Is a Year-Round Activity for CLSP!

September	October	November	December
Call for Proposals	Proposal Deadline Preliminary Screening Invitations to Meeting	Peer Review Meeting Select Team Leaders Form Skeletal Teams	\$\$ from NSF to JHU Advertise for UGs (Deadline Feb 15)
January	February	March	April
Finalize Sr. Members Data Requirements Computing Req's	Select Grad Students Data Contracts Given Computers Ordered	Select Undergrads First T team Meetings Prelim. Data Arrives	Computers Installed Refine Data Specs Perform Pilot Expts
May	June	July	August
2 <sup>nd</sup> Team Meetings Most Data Arrives Start Baseline Expts	Computing Fine-tuned All the Data Arrives Finish Baseline Expts	Summer School <b>Workshop Begins</b> Team Updates	<b>Workshop Ends</b> Write Final Reports



## The Road to the WS'02 Final Four

---

- Open call for proposals issued Sept 3, 2001.
  - Electronically advertised to 100's of researchers
- 15 proposals received; well past Oct 8 deadline.
  - Initial screening reduced it to 12 viable proposals
- 2-day peer-review panel met Nov 9-11, 2001.
  - Authors, peers, additional wise-persons, government program managers and JHU faculty were invited
  - Single transferable voting scheme was used
  - 3 winning proposals were sought, 4 were selected



# The November Meeting Attendees

## (and Last-Minute Drop-Outs)

Breck Baldwin	Dragomir Radev	Pat O'Neill Brown	Bill Byrne
Jeff Bilmes	Doug Reynolds	George Doddington	Jason Eisner
Everhard Ditters	Paul Roochnik	C Van Ess Dykema	Frederick Jelinek
Louise Guthrie	Anoop Sarkar	Kathleen Egan	Sanjeev Khudanpur
Jan Hajic	Richard Schwartz	Jack Godfrey	David Yarowsky
John Henderson	Mark Steedman	Lynette Hirschman	
Katrin Kirchoff	Petr Zemanek	Charles Wayne	
Mari Ostendorf	Zhiping Zheng	Mark Liberman	
Barbara Peskin	Roberto Basili	Gary Strong	

JHU Faculty: 15%


Other Researchers: 60%

Government PMs: 25%



# Authors (+ Co-Authors) of Proposals Debated on Nov 9-10

- Baldwin : Recognizing **Coreference** Across Languages and Data Formats
- Ditters : Tools and Resources for **Arabic Natural Language Processing**
- Guthrie : Semi-supervised **Semantic Tagging** for Information Extraction
- Hajic : **Generation** from Deep Syntactic Representation for **MT**
- Henderson : **Spoken Language Translation** for Unwritten Languages
- Kirchhoff + : **Low Complexity Speech Processing**
- Ostendorf + : **Multi-linear Speech Recognition** for Arabic
- Radev : Using the Web for **Question Answering**
- Reynolds : Exploiting High-level Info for High-perf **Speaker Recognition**
- Schwartz : Techniques for Improving **Arabic Speech Recognition**
- Steedman + : Co-Training for Wide Coverage **Statistical Parsing**
- Zheng : Open Domain **Question Answering**



# Team Leaders and Projects

## Selected on Nov 11

---

Baldwin : [Recognizing Coreference Across Languages and Data Formats](#)

Ditters : [Tools and Resources for Arabic Natural Language Processing](#)

Guthrie : [Semi-supervised Semantic Tagging for Information Extraction](#)

Hajic : [Generation from Deep Syntactic Representation for MT](#)

Henderson : [Spoken Language Translation for Unwritten Languages](#)

Kirchhoff + : [Low Complexity Speech Processing](#)

Kirchhoff + : [Multi-linear Speech Recognition for Arabic](#)

Radev : [Using the Web for Question Answering](#)

Reynolds : [Exploiting High-level Info for High-perf Speaker Recognition](#)

Schwartz : [Techniques for Improving Arabic Speech Recognition](#)

Steedman : [Co-Training for Wide Coverage Statistical Parsing](#)

Zheng: [Open Domain Question Answering](#)



# Team Formation and Spring Planning Meetings

- Senior members of the teams selected over Christmas, graduate students added in Jan-Feb, 2002, and undergraduates added in March, 2002.
  - 37 full-time participants: highest since 1993
    - 20 senior researchers, 17 students
    - 12 academic, 4 industry, 4 govt, 9 grad students, 8 undergrads
    - 29 male, 8 female; Also (tentatively) 22 U.S., 15 international
  - Significant number (at least 8) of explicit “affiliate” members
- Each team has had two face-to-face meetings.
  - Generation: March 9-10 and May 12-13
  - SuperSID: March 10-11 and May 18-19
  - Arabic ASR: April 13-14 and June 15-16
  - Parsing: May 5-6 and June 7



# That Brings Us to July 1, 2002

September 2001	October 2001	November 2001	December 2001
Call for Proposals	Proposal Deadline Preliminary Screening Invitations to Meeting	Peer Review Meeting Select Team Leaders Form Skeletal Teams	Advertise for UGs
January 2002	February 2002	March 2002	April 2002
Finalize Sr. Members	Select Grad Students	Select Undergrads MT Data Contract \$\$ Leave DARPA	1 <sup>st</sup> Team Meetings SSID Data Contract
May 2002	June 2002	July 2002	August 2002
1 <sup>st</sup> Team Meetings \$\$ Arrive @ NSF 2 <sup>nd</sup> Team Meetings	2 <sup>nd</sup> Team Meetings SSID Data Done MT Data Enroute (?)	Summer School Workshop Begins \$\$ Arrive JHU (?)	Workshop Ends Final Reports (?)



# The JHU Summer School on Human Language Technology

- A two week tutorial on all aspects of human language technology by expert academics and practitioners.
  - ASR, TTS, IR, MT, text analysis, summarization, generation.
- Started in 1998 to bring undergraduate students up to speed on human language technology.
  - Attended by DoD personnel looking for a short comprehensive overview of language technology
  - Additional participants this year from NAACL
- Another record attendance!
  - 9 undergraduates (8 WS02, 1 NAACL)
  - 23 graduate students (9 WS02, 8 NAACL, 6 Others)
  - 5 researchers (2 NSA, 1 DLI, 1 Army, 1 San Francisco State U.)



# Summer School Calendar

## July 1<sup>st</sup> to 12<sup>th</sup>, 2002

Mon Jul 1	Tue Jul 2	Wed Jul 3	Thu Jul 4	Fri Jul 5
Basic NLP	Basic IR	Basic MT	Independence Day Holiday	Generation
E. Charniak	Jim Mayfield	K. Papineni	Shankar Kumar Paola Virga	Jan Hajic
Al Gorin	Donna Harman	Daniel Marcu		Hatzivassiloglou
Jason Eisner	Phillip Resnik	David Yarowsky	Drago Radev	Jan Hajic
Noah Smith	Elliott Drabek	Charles Schafer		

Mon Jul 8	Tue Jul 9	Wed Jul 10	Thu Jul 11	Fri Jul 12
Basic ASR	Arabic ASR	Speaker ID	Parsing	Dialog Sys.
F. Kubala	Katrin Kirchhoff	Doug Reynolds	Mark Steedman	Marilyn Walker
Bill Byrne	Owen Kimball	Larry Heck	Mark Johnson	R. Pieraccini
Bill Byrne	Gang Ji	Walt Andrews	Anoop Sarkar	T. J. Hazen
S. Tsakalidis	Sourin Das	Andre Adami	Paul Ruhlén	

# The Workshop Calendar

<http://www.clsp.jhu.edu/ws2002>

Mon Jul 15	Tue Jul 16	Wed Jul 17	Thu Jul 18	Fri Jul 19
Opening Presentation Steering Comm Mtg		McCallum Seminar	UG Lunch Group Dinner	
Mon Jul 22	Tue Jul 23	Wed Jul 24	Thu Jul 25	Fri Jul 26
Steering Comm Mtg		Pereira Seminar Team Updates	UG Lunch Group Dinner	Townhall Meeting
Mon Jul 29	Tue Jul 30	Wed Jul 31	Thu Aug 1	Fri Aug 2
Steering Comm Mtg		Bengio Seminar	UG Lunch Group Dinner	
Mon Aug 5	Tue Aug 6	Wed Aug 7	Thu Aug 8	Fri Aug 9
Steering Comm Mtg		Seminar TBA Team Updates	UG Lunch Group Dinner	Townhall Meeting
Mon Aug 12	Tue Aug 13	Wed Aug 14	Thu Aug 15	Fri Aug 16
Steering Comm Mtg		Seminar TBA	UG Lunch Group Dinner	
Mon Aug 19	Tue Aug 20	Wed Aug 21	Thu Aug 22	Fri Aug 23
Steering Comm Mtg		Final Presentation	Final Presentation Dinner Party	Lab Clean-Up Day Farewell Tea