ACL 2012

The 50th Annual Meeting of the Association for Computational Linguistics

July 8(Sun) - July 14(Sat), 2012 / ICC JEJU, Jeju island, Korea

Conference Handbook

Handbook Production:
Jung-jae Kim, Nanyang Technological University, Singapore

Message from the General Chair

Welcome to Jeju Island — where ACL makes a return to Asia!

As General Chair, I am indeed honored to pen the first words of ACL 2012 proceedings. In the past year, research in computational linguistics has continued to thrive across Asia and all over the world. On this occasion, I share with you the excitement of our community as we gather again at our annual meeting. On behalf of the organizing team, it is my great pleasure to welcome you to Jeju Island and ACL 2012.

In 2012, ACL turns 50. I feel privileged to chair the conference that marks such an important milestone for our community. We have prepared special programs to commemorate the 50th anniversary, including 'Rediscovering 50 Years of Discovery', a main conference workshop chaired by **Rafael Banchs** with a program on 'the People, the Contents, and the Anthology', which recollects some of the great moments in ACL history, and 'ACL 50th Anniversary Lectures' by **Mark Johnson**, **Aravind K. Joshi** and a Lifetime Achievement Award Recipient.

A large number of people have worked hard to bring this annual meeting to fruition. It has been an unforgettable experience for everyone involved. My deepest thanks go to the authors, reviewers, volunteers, participants, and all members and chairs of the organizing committees. It is your participation that makes a difference.

Program Chairs, **Chin-Yew Lin** and **Miles Osborne**, deserve our gratitude for putting an immense amount of work to ensure that each of the 940 submissions was taken care of. They put together a superb technical program like nobody else. Publication Chairs, **Maggie Li** and **Michael White**, extended the publishing tools to take care of every detail and compiled all the books within an impossible schedule. Tutorial Chair, **Michael Strube**, put together six tutorials that you can never miss. Workshop Chairs, **Massimo Poesio** and **Satoshi Sekine**, working with their EACL and NAACL counterparts, selected 11 quality workshops, many of which are new editions in their popular workshop series. Demo Chair, **Min Zhang**, started a novel review process and selected 29 quality system demos. Faculty Advisors, **Kentaro Inui**, **Greg Kondrak**, and **Yang Liu**, and Student Chairs, **Jackie Cheung**, **Jun Hatori**, **Carlos Henriquez** and **Ann Irvine**, assembled an excellent program for the Student Research Workshop with 12 accepted papers. Mentoring Chair, **Joyce Chai**, coordinated the mentorship of 13 papers. Publicity Chairs, **Jung-jae Kim** and **Youngjoong Ko**, developed the website, newsletters, and conference handbook that kept us updated all the time. Exhibition Chair, **Byeongchang Kim**, coordinated more than 10 exhibitors with a strong

industry presence. All the events are now brought to us on Jeju Island by the Local Arrangements Chairs, **Gary Lee** and **Jong Park**, and their team. I can never thank them enough for all the preparations they have made to host us in such a spectacular place!

I would like to express my gratitude and appreciation to **Kevin Knight**, Chair of the ACL Conference Coordination Committee, **Dragomir Radev**, ACL Secretary, and **Priscilla Rasmussen**, ACL Business Manager, for their advice and guidance throughout the process.

The financial sponsors generously supported ACL 2012 in a meaningful way despite a challenging economic outlook. We are honored to have Baidu as the Platinum Sponsor, Elsevier and Google as Gold Sponsors, Microsoft, KAIST and SK as Sliver Sponsors, 7 Bronze Sponsors, and 3 Supporters. The Donald and Betty Walker Student Scholarship Fund and Asian Federation of Natural Language Processing have supported our student travel grants. The sponsorship program was made possible by the ACL sponsorship committee: **Eiichiro Sumita**, **Haifeng Wang**, **Michael Gamon**, **Patrick Pantel**, **Massimiliano Ciaramita**, and **Idan Szpektor**.

Finally, I do hope that you have an enjoyable and productive time on Jeju Island, and that you will leave with fond memories of ACL's 50th Anniversary. With my best wishes for a successful conference!

Haizhou Li
Institute for Infocomm Research
ACL 2012 General Chair

Message from the Program Committee Co-Chairs

This year we received 571 valid long paper submissions and 369 short paper submissions. 19% of the long papers and 20% of the short papers were accepted. As usual, some are presented orally and some as posters. Taking unigram counts from accepted long paper titles, and ignoring function words, the most popular word were:

entity 5
evaluation 5
hierarchical 5
information 5
joint 5
syntactic 5
topic 5
discriminative 6
lexical 6
statistical 6
chinese 7
dependency 7
machine 8
modeling 8
models 8
language 10
word 10
parsing 11
model 12
learning 14
translation 15

Some areas have grown over time and some have diminished. The most popular area for submissions (as expected) was Machine Translation. We promoted Social Media as a new area.

Twenty nine Area Chairs worked with 665 reviewers, producing 1830 long paper reviews and 1187 short paper reviews. Everything ran to a tight schedule and there were no slippages. This would not have been possible without our wonderful and diligent Area Chairs and Reviewers. Thanks!

We are delighted to have two keynote speakers, both of whom are very well known to the language community: Aravind Joshi and Mark Johnson. They will give coordinated talks addressing the 50th ACL anniversary: “Remembrance of ACLs past” and “Computational linguistics: Where do we go from here?” The ACL Lifetime Achievement Award will be announced on the last day of the conference.

Of the many papers, we selected two as being outstanding:

Bayesian Symbol-Refined Tree Substitution Grammars for Syntactic Parsing
Hiroyuki Shindo, Yusuke Miyao, Akinori Fujino, Masaaki Nagata

String Re-writing Kernel
Fan Bu, Hang Li, Xiaoyan Zhu

They will be presented as best papers in a dedicated session.

We thank the General Conference Chair Haizhou Li, the Local Arrangements Committee headed by Gary Geunbae Lee, Michael White and Maggie Li, the Publication Co-Chairs for coordinating and putting the proceedings together and all other committee chairs for their work. MO is especially thankful to Steve Clark for helpful tips on how to manage and run the whole process.

We hope you enjoy the conference!

Chin-Yew Lin, Microsoft Research Asia
Miles Osborne, University of Edinburgh

ACL 2012 Organizing Committee

General Chair
Haizhou Li (Institute for Infocomm Research, Singapore)

Program Chairs
Chin-Yew Lin (Microsoft Research Asia, China)
Miles Osborne (University of Edinburgh, UK)

Local Arrangement Chairs
Gary Geunbae Lee (POSTECH, Korea)
Jong C. Park (KAIST, Korea)

Workshop Chairs
Massimo Poesio (University of Essex, UK)
Satoshi Sekine (New York University, USA)

Tutorial Chair
Michael Strube (HITS gGmbH, Germany)

Publication Chairs
Maggie Li (The Hong Kong Polytechnic University, Hong Kong)
Michael White (The Ohio State University, USA)

Mentoring Service Chair
Joyce Chai (Michigan State University, USA)

Faculty Advisors (Student Research Workshop)
Kentaro Inui (Tohoku University, Japan)
Greg Kondrak (University of Alberta, Canada)
Yang Liu (University of Texas at Dallas, USA)

Student Chairs (Student Research Workshop)
Jackie Cheung (University of Toronto, Canada)
Jun Hatori (University of Tokyo, Japan)
Carlos Henriquez (Technical University of Catalonia, Spain)
Ann Irvine (Johns Hopkins University, USA)

Publicity Chairs
Jung-jae Kim (Nanyang Technological University, Singapore)
Youngjoong Ko (Dong-A University, Korea)

Demo Chair
Min Zhang (Institute for Infocomm Research, Singapore)

Exhibit Chair
Byeongchang Kim (Catholic University of Daegu, Korea)

Special Session Chair
Rafael E. Banchs (Institute for Infocomm Research, Singapore)

Sponsorship Chairs

[Asia]
Eiichiro Sumita (NICT, Japan)
Haifeng Wang (Baidu, China)

[Europe]
Massimiliano Ciaramita (Google Zurich, Switzerland)
Idan Szpektor (Yahoo! Research Israel, Israel)

[North America]
Michael Gamon (Microsoft, USA)
Patrick Pantel (Microsoft, USA)

Business Manager
Priscilla Rasmussen (ACL)

ACL 2012 Program Committee

Program Co-Chairs

Chin-Yew Lin (Microsoft Research Asia, China)
Miles Osborne (University of Edinburgh, UK)

Area Chairs

[Question Answering]
Hongyuan Zha (Georgia Institute of Technology, USA)
Hsin-Hsi Chen (National Taiwan University, Taiwan)

[Word Segmentation, Morphology and Phonology]
Kemal Oflazer (Carnegie Mellun University - Qatar, Qatar)

[Discourse and Dialogue]
Mikio Nakano (Honda Research Institute, Japan)
Andrei Popescu-Belis (Idiap Research Institute, Switzerland)

[Lexicon and Ontology]
Eneko Agirre (University of the Basque Country, Spain)

[NLP Applications]
Jason Baldridge (The University of Texas at Austin, USA)
David Weir (University of Sussex, UK)

[Machine Learning]
Trevor Cohn (University of Sheffield, UK)
Mark Dredze (Johns Hopkins University, USA)

[Social Media and Sentiment Analysis]
Noah Smith (Carnegie Mellon University, USA)
Jan Wiebe (University of Pittsburgh, USA)

[Multilinguality]
Daniel M. Bikel (Google Research, USA)
Mona T. Diab (Center for Computational Learning Systems, Columbia University, USA)

[Language Resources]
Fei Xia (University of Washington, USA)

[Machine Translation]
Marcello Federico (Fondazione Bruno Kessler, Italy)
David Talbot (Google Research, USA)
Adam Lopez (Johns Hopkins University, USA)
Hua Wu (Baidu, China)

[Lexical Semantics]
Evgeniy Gabrilovich (Google, USA)
Naoaki Okazaki (Tohoku University, Japan)

[Syntax/Parsing]
Stephen Clark (University of Cambridge, UK)
Liang Huang (University Southern California, USA)

[Summarisation and Generation]
Anja Belz (University of Brighton, UK)
Ani Nenkova (University of Pennsylvania, USA)

[Speech]
Chung-Hsien Wu (National Cheng Kung University, Taiwan)

[Information Extraction and Information Retrieval]
Jian Su (Institute for Infocomm Research, Singapore)
Vincent Ng (University of Texas at Dallas, USA)
Jianfeng Gao (Microsoft Research, Redmond, USA)

ACL 2012 Local Organizing Committee

Local Organizing Chairs
Gary Geunbae Lee (POSTECH, Korea)
Jong C. Park (KAIST, Korea)

Social Activities
Jeong-Won Cha (Changwon National University, Korea)

Local Sponsorship
Hanmin Jung (KISTI, Korea)

Local Finance
Seungshik Kang (Kookmin University, Korea)

Local Exhibit
Byeongchang Kim (Catholic University of Daegu, Korea)

Government Sponsorship
Harksoo Kim (Kangwon National University, Korea)

Conference Handbook
Jung-jae Kim (Nanyang Technological University, Singapore)

Web Site & Flyer
Youngjoong Ko (Dong-A University, Korea)

Student Volunteer Management
Heuiseok Lim (Korea University, Korea)

Internet/Wifi/Equipments
Seong-Bae Park (Kyungpook National University, Korea)

The ACL 2012 very gratefully acknowledge the following commitments in sponsorship:

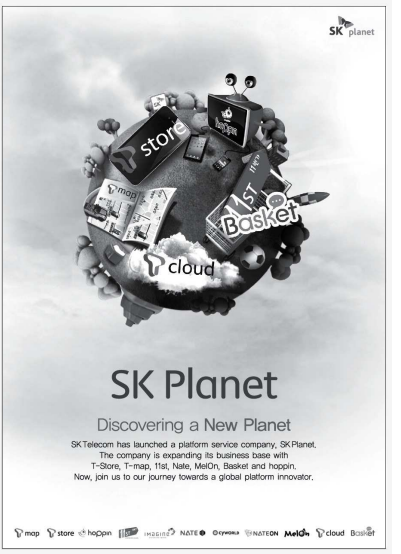
KAIST

DEPARTMENT OF
COMPUTER SCIENCE

http://cs.kaist.ac.kr

The Computer Science Department of KAIST was initially established as a graduate-school only program by the Korean government, and an undergraduate program was later added. KAIST is generally considered as the leader in the computer science education and research in Korea. More important, an ABET evaluation of the department found its education to be comparable to that of the top 10% of institutions in U.S.

The department offers B.S., M.S., and Ph.D. degrees in computer science. Currently, the department consists of 49 full-time professors, 8 research professors, 9 affiliated professors, and 8 post-doctoral scholars, about 300 undergraduates, and more than 390 graduate students. The department has contributed significantly to meeting the nation's needs to educate institutes, industry, and academia. The fact that nearly 20% of all computer science faculty members in Korea earned their advanced degrees at KAIST illustrates the dominant role the CS department has played in Korea.



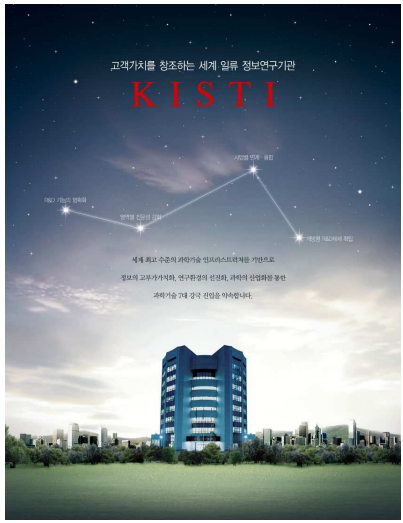
Microsoft®
Research

SHORTENED VERSION

Appen Butler Hill provides high-quality, cost-effective solutions for expanding language technologies into global markets. We support the development of NLP systems, annotating data sets and performing data collections, transcriptions and search engine relevance evaluations.
Please visit us at: http://www.appenbutlerhill.com

LONG VERSION

Appen Butler Hill provides high-quality, cost-effective solutions for developing and expanding Speech, Text and Search technologies into global markets. Combining a deep understanding of language-specific challenges with project management excellence and flexible staffing models, we offer a wide range of services for each phase of the product development. Services include supporting the development, tuning and testing of NLP systems, analyzing and annotating large data sets and performing data collections, transcriptions and search engine relevance evaluations.
With a professional resource pool of over 20,000 linguists and language specialists covering over 120 languages, we provide the highest quality language services and solutions.
Please visit us at: http://www.appenbutlerhill.com



ETS®

Listening. Learning. Leading.®

Contents

1

A Short Introduction to Korea

The Korean peninsula extends southward from the eastern end of the Asian continent. It is roughly 1,000 km (621 miles) long and 216 km (134 miles) wide at its narrowest point. Mountains cover 70% of the land mass, making it one of the most mountainous regions in the world. The peninsula is divided just slightly north of the 38th parallel. South Korea has a population of 48.6 million and consists of nine provinces, capital Seoul, six metropolitan cities, and an additional 77 cities and 88 counties. For more information, please visit http://english.visitkorea.or.kr/enu/index.kto

Climate

Korea lies in a temperate zone and has four distinct seasons.

SPRING

In late March or early April, the trees burst into leafy splendor to mark the beginning of spring. Mostly sunny days can be expected from March to May.

SUMMER

During the relatively hot and rainy summer season, the vegetation is lush. By June the average temperature is over 20 °C(68 °F). Monsoon rains usually begin around the end of June and last until mid-to-late July. August is hot and humid.

AUTUMN

The coming of autumn in late September brings continental winds and clear, dry weather, making these months perhaps the most pleasant time of year. October's vivid gold and vibrant reds create a colorful panorama.

WINTER

December to February are cold and dry with occasional snow. During the winter months, three or four days of cold weather are often followed by a few warmer days.

Flag of South Korea

The flag of South Korea is called taegeukgi. The design symbolizes the principles of the yin and yang in Asian philosophy. The circle in the center of the flag is divided into two equal parts. The red half represents the proactive cosmic forces of the yang. Conversely, the blue half represents the responsive cosmic forces of the yin. The two forces together embody the concepts of continual movement, balance, and harmony that characterize the sphere of infinity. The circle is surrounded by four trigrams, one in each corner. Each trigram symbolizes one of the four universal elements: Heaven, Earth, Fire and Water.

National Flower

The national flower of South Korea is the mugunghwa (Rose of Sharon). Every year from July to October, a profusion of mugunghwa blossoms graces the entire country. The flower's symbolic significance stems from the Korean word mugung (immortality).

Language

The Korean language, like Hungarian, Turkish, Mongolian and Finnish, is classified into the Ural-Altaic language group. Hangeul (the Korean alphabet) is composed of 10 simple vowels and 14 consonants. A group of scholars under the leadership of King Sejong the Great developed this systematic rendition of spoken sound in 1443. It is widely acclaimed by linguists as an ingenious invention. In 2009, the town of Bau-Bau, in Sulawesi, Indonesia, adopted Hangeul as their official written language.

Travel Hotline

Need travel information? Just call 1330!

2 Conference Information

Venues

The ACL 2012 activities will be held at the following two venues:

(Shilla) Tutorials and the welcome reception will be held at the Shilla Jeju on July 8, 2012.

(ICC) Main conference, workshops, and EMNLP-CoNLL will be held at the International Convention Center Jeju from July 9 to July 14, 2012.

Instruction for Presenters

Oral Presentation

A. Presentation File

Please bring your PowerPoint presentation file with you on a CD or USB memory stick and submit presentation material to the operator in session room at least 20 minutes prior to the scheduled start of the session. The operator will load the presentation files to the notebook PC.

B. Preview Room

- Location: Room 304, 3F, ICC Jeju
- Operation Hours: 7:00am – 6:00pm

Please go to the preview room and check your presentation file before your session starts to ensure your presentation file appears properly. If your presentation file contains animations or movies, then you are advised to check over the technical matters for your session. Please also note that network is connected between the session room and the preview room, so it is possible to transfer any of your files between the session room and the preview room.

C. Audio Visual Equipment

Each session room will be equipped with the following equipments:

- Laptop computer running MS-Office PowerPoint 2007 operated in Windows XP, equipped with compact disk reader (CD & DVD) and USB drive
- One switcher with four RGB inputs (except Tutorials and Workshops)
- One smart pointer and mouse
- One bean projector (RGB Port)
- One Screen

The standard electricity supply is 220-volts AC at 60Hz.

To avoid frequently occurring technical problems in the presentation, it is discouraged to bring your own computer (especially Macintosh laptop) unless your presentation requires it for special software and/or hardware. If it is unavoidable to use your own laptop, you should bring all the necessary adaptors that are compatible with our beam projector (RGB port).

Poster Presentation

- Poster size is 840mm (width) by 1200mm (height).
- Poster presenters are required to post the presentation 30 minutes prior to the beginning of the session.
- Please mount your presentation on the partition with your poster number.
- Please remove your poster within 30 minutes after your session is ended. All remaining posters will be removed and disposed of.
- Pushpins, clips, double-sided tape, etc. are provided in the poster session area.

Local Information

Emergency Phone Numbers

Dial for the police, 119 for the fire department, or 1339 for medical emergencies (though most Korean operators speak little or no English). A hotel staff member or hotel manager can arrange a doctor or an ambulance.

International Call

For international calls, first dial any of the following international call company numbers below, then the country code, area code, and finally the number you are calling (e.g., +001+82-(0)64-000-0000).

- Regular International Phone Call Carriers: 001, 002, 005, 006, 008, 00700

Korea Travel

Wherever you are, Korea Travel Phone will help with all your travel needs. Simply dial 1330 from either your cell phone or public telephone to have all your travel questions answered.

First-aid Medicine

First-aid medicine (Aspirin, digestive medicine, antiseptic) is available in the secretary office. Please contact the office if anyone needs one.

- Secretary Office: Room 303, ICC JEJU

Bank (Money Exchange)

Participants of ACL 2012 can exchange their money into KRW here at the bank in the center lobby. The bank is open between 9:00am $\sim$ 4:00pm.

- Location: 1st floor, ICC JEJU

Duty Free Shop

In ICC JEJU, Duty Free Shop provides convenient shopping facilities where you can purchase world-famous brand items and the most up-to-date fashion items at reasonable prices.

- Location: 1st floor, ICC JEJU

Money & Currency

In Korea, bills and coins are used and the unit of the Korean currency is 'Won' (KRW). Coin denominations are 10, 50, 100 and 500. Banknotes are 1,000, 5,000, 10,000 and 50,000. New bills have come out recently, but old bills and new bills are used at the same time.

USD 1 is equivalent to KRW 1,170, and Euro 1 is equivalent to KRW 1,470 as of June 2012.

Tipping

Tipping is not necessary in Korea. Instead, expensive restaurants and luxury hotels may add a service charge of 10%.

Electricity

In Korea, 220 volt outlets for a plug with two round prongs are most common. Some hotels provide 110 volt outlets for shavers.

Restaurants

Inside ICC JEJU

Name	Delizia
Capacity	350
Price (USD)	12 - 60
Location	ICC JEJU 3F
Type / Menu	Korean, Japanese, Chinese, Western dishes

Name	PEARL FOOD
Capacity	60
Price (USD)	5 - 20
Location	ICC JEJU 1F
Type / Menu	Korean dishes, Coffee, etc.

Name	Dunkin' Donuts
Capacity	30
Price (USD)	2 - 5
Location	ICC JEJU 1F
Type / Menu	Café

Name	KRAZE BURGERS
Capacity	30
Price (USD)	6 - 20
Location	ICC JEJU 1F
Type / Menu	Burgers and beverages

Jungmun Tourism Complex (1/5)

Name	Daegijung
Capacity	50
Price (USD)	5 - 50
Location	8 minutes on foot (EAST)
Type / Menu	Korean Seafood

Name	Shangri La
Capacity	1,000
Price (USD)	33.8
Location	15 minutes on foot (West)
Type / Menu	Sea food Buffet

Name	KAOKAO
Capacity	2,000
Price (USD)	15 - 32
Location	15 minutes on foot (West)
Type / Menu	Korean/Chinese Buffet

Name	Han's Family
Capacity	100
Price (USD)	16.3 - 25
Location	15 minutes on foot (North)
Type / Menu	Korean, Japanese, Chinese Buffet

Jungmun Tourism Complex (2/5)

Name	Greenjung
Capacity	80
Price (USD)	5 - 57
Location	15 minutes on foot (North)
Type / Menu	Korean Seafood & Pork

Name	Deok-Seong Won
Capacity	100
Price (USD)	4.29 - 60
Location	15 minutes on foot (North)
Type / Menu	Chinese

Name	Dol Hareubang Noodle
Capacity	80
Price (USD)	4.29 - 8
Location	15 minutes on foot (North)
Type / Menu	Korean Noodle

Name	Garam Dolsotbab
Capacity	100
Price (USD)	8 - 17
Location	15 minutes on foot (North)
Type / Menu	Korean

Jungmun Tourism Complex (3/5)

Name	Galchi Myeong Ga
Capacity	120
Price (USD)	3 – 17
Location	15 minutes on foot (East)
Type / Menu	Korean Seafood

Name	Sahebang
Capacity	50
Price (USD)	3 – 17
Location	15 minutes on foot (East)
Type / Menu	Chinese

Name	chalet
Capacity	30
Price (USD)	7 - 21
Location	15 minutes on foot (East)
Type / Menu	Western

Name	Haesong
Capacity	60
Price (USD)	8 - 25
Location	15 minutes on foot (East)
Type / Menu	Korean Seafood

Jungmun Tourism Complex (4/5)

Name	Jeju Mat Sarang
Capacity	300
Price (USD)	6 - 21
Location	15 minutes on foot (North)
Type / Menu	Korean Pork

Name	Sinwoosung Town
Capacity	1,000
Price (USD)	6 - 21
Location	20 minutes on foot (North)
Type / Menu	Korean Pork & Seafood

Name	Lailai
Capacity	80
Price (USD)	3 – 17
Location	20 minutes on foot (North)
Type / Menu	Chinese

Name	Starbucks Coffee
Capacity	80
Price (USD)	4 - 7
Location	20 minutes on foot (North)
Type / Menu	Café (Inside & Take out)

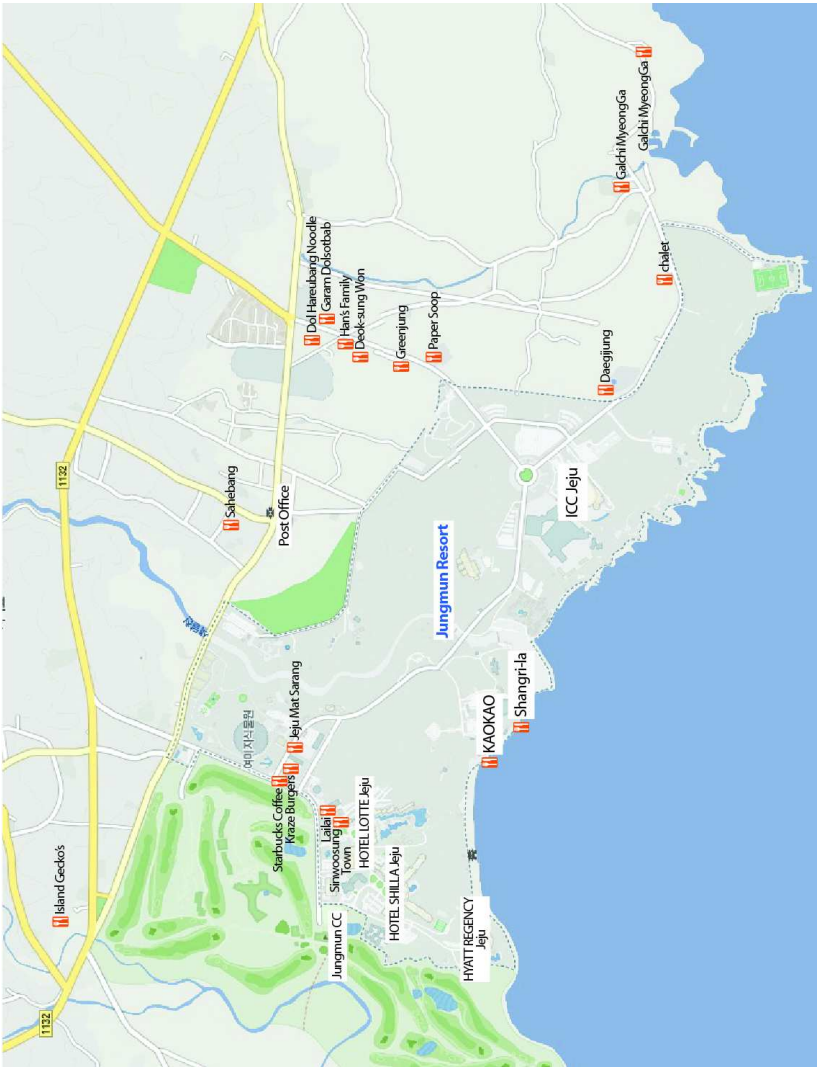
Jungmun Tourism Complex (5/5)

Name	Kraze Burgers
Capacity	30
Price (USD)	6 - 20
Location	20 minutes on foot (North)
Type / Menu	Western (Inside & Take out)

Name	Paper Soop
Capacity	30
Price (USD)	6 - 21
Location	20 minutes on foot (North)
Type / Menu	Western

Name	Island Gecko's
Capacity	150
Price (USD)	6 - 38
Location	10 minutes by car (West)
Type / Menu	Western

Restaurant Locations



Social Activities

1. Welcome Reception

Time Sunday, July 8, 6:00pm - 9:00pm

Location Halla hall of the Shilla Jeju (Tutorial venue)



2. Banquet

Time Tuesday, July 10th, 6:00pm - late

Location Shangri La

The shuttle bus will operate between the conference venue (ICC Jeju) and Shangri La before the banquet, and Shangri La to all the accommodations after the banquet.



3. Student Lunch

Time Tuesday, July 10, 12:30pm - 2:00pm

Location Ocean view of ICC JEJU

Private Tour

1. Jeju Jungmun Daepo Haean Jusang Jeollidae

If you follow the farm road 600 m to the southwest from Depo-dong, at the end of the pine tree forest is a cliff. At the base of this cliff is the Jisatgae Coast along with the imposing stone pillars. The Jusangjeolli are stone pillars piled up along the coast and is a designated cultural monument of Jejudo Island. The Jusangjeolli was formed when the lava from Mt. Hallasan erupted into the sea of Jungmun. They are rock pillars shaped like cubes or hexagons of various sizes and almost seem as if stonemasons had carved them out.

Location 5 minutes on foot from ICC Jeju

Admission Fees Adult 2,000 KRW (Group 1,600 KRW) / Youth and Child 1,000 KRW

2. Jungmun Beach

Jungmun Beach has a sandy shore so fine that it is called Jinmosal. The white sand beach is 560m long and the water has an average depth of 1.2m. Different shades of black, red, and grey are beautifully mixed into the white sand of the beach. This helps to accentuate Jeju-do Island's black Hyeonmuam stones. To the right side of the beach is a natural sea cave called Haesikgul Cave. This area is often used for TV commercials, and the turbulent water is great for windsurfing and other water sports.

Location 25 minutes on foot from ICC Jeju

Admission Fees Free

3. Yakcheonsa

The architectural style of Yakcheonsa Temple is reminiscent of that of Buddhist temples of the early Joseon Dynasty. Measuring an impressive 30m high and spanning a total area of 3,305 meters squared, it is the largest temple in the East.

Location 11 minutes by taxi from ICC Jeju

Admission Fees Free

4. Cheonjeyeon Waterfalls

Cheonjeyeon Falls, named ‘The pond of God’, consists of 3 parts. Around the falls, a variety of plant life thrives, such as the rare ‘solipnan’ reeds. To the east, there is a cave where cold water pours from the ceiling to create a waterfall. The water from the first waterfall becomes the second and third waterfalls and flows into the sea. In Cheonjeyeon Valley, there is the “Seonimgyo Bridge” (an arch bridge that has 7 nymphs carved on the side) and the octagonal “Cheonjeru tower”.

Location 15 minutes on foot from ICC Jeju

Admission Fees Adult 2,500 KRW (Individuals) / Youth and Child 1,350 KRW

5. Yeomiji Botanical Gardens

Located within Jungmun Resort, Yeomiji Botanical Garden is the best botanical garden in Asia that has a unique southern atmosphere. The indoor garden contains approximately 2,000 rare tropical and subtropical plants, including Flower Garden, Water Lily Garden, Jungle Garden, Subtropical Fruits Garden, Cactus and Succulent Garden, and Observation Platform. The outdoor garden houses, Jeju Native Garden, as well as Korean, Japanese, Italian and French folk gardens.

Location 15 minutes on foot from ICC Jeju

Admission Fees Adult (age 19-64) 7,000 KRW / Youth (age 13-19) 4,500 KRW / Child (age 3-13) 3,500 KRW / Senior (age over 65) 4,000 KRW

6. Teddy Bear Museum

The Teddy Bear Museum lives up to its name, boasting quite an impressive variety that have been loved for more than a hundred years all over the world. Inside the two galleries you can view the teddy bears from various countries.

Location 15 minutes on foot from ICC Jeju

Admission Fees Adult 7,000 KRW / Youth 6,000 KRW / Child, senior and disabled visitors 5,000 KRW

7. Pacific Land

Pacific Land, located at the entrance of Jungmun Beach, is a sea resort offering a variety of performances such as a monkey show, sea lion show, and dolphin show. With the cost of admission, visitors are able to enjoy all three performances along with a mini-aquarium full of Jeju's fish and other species of fish that are on display. After the performances, visitors may see the animals and take pictures with them.

Location 20 minutes on foot from ICC Jeju

Admission Fees 20 years and above 12,000 KRW / Age 13 to 19 10,000 KRW

8. Jeju Olle 7th

“Olle” is the Jeju word for a narrow pathway that is connected from the street to the front gate of a house. Hence, “olle” is a path coming out from a secret room to an open space and a gateway to the world. If the road is connected, it is linked to the whole island and the rest of the world as well. It has the same sound as “Would you come?” in Korean, so Jeju’s “olle” sounds the same as “Would you come to Jeju”.

Location 5 minutes on foot from ICC Jeju

Admission Fees Free

3

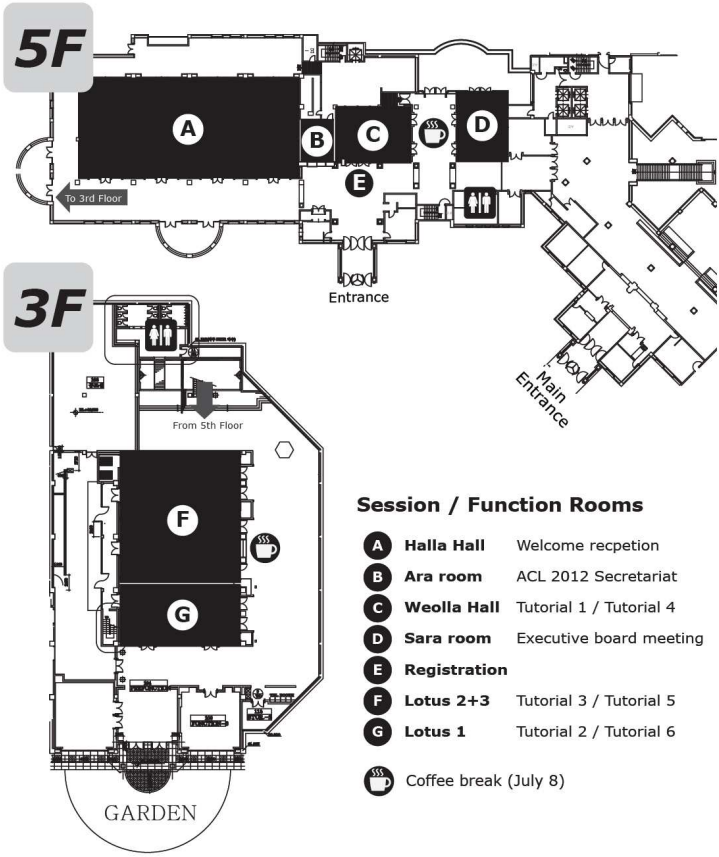
Maps and Floor Plans

Jeju Island

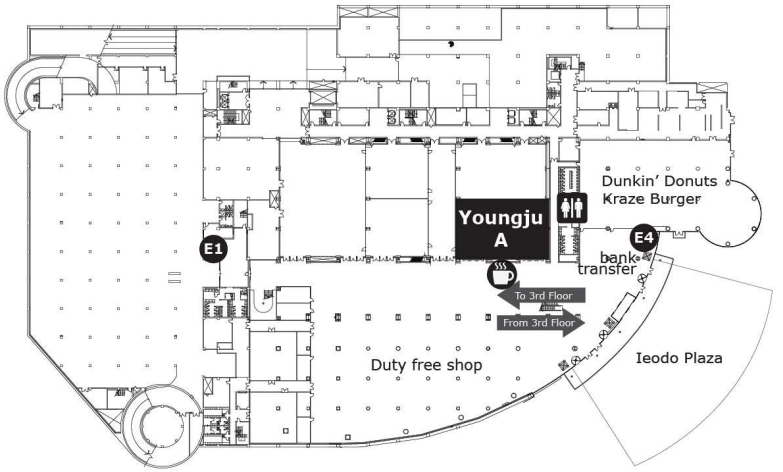
Jungmun Resort

- Conference venues and surrounding area -

Floor Plan - The Shilla Jeju



Floor Plan - ICC JEJU 1F



Session Rooms

Youngju A

EMNLP Plenary, Parallel 1 (July 12-14)

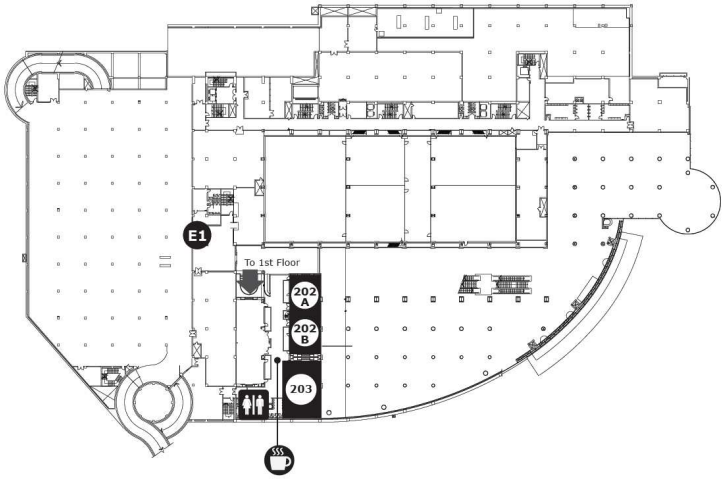
Function spaces

E1 Elevator on the 1st , 2nd , 3rd , 5th Floors

E4 Elevator on the 1st , 3rd , 4th , 5th , 6th Floors

Coffee break (July 12-14)

Floor Plan - ICC JEJU 2F



Session Rooms

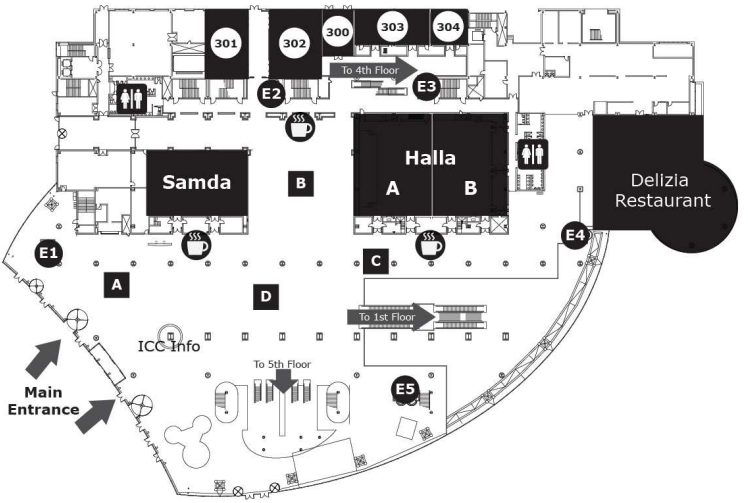
202 A · 202 B · 203
EMNLP Parallel (July 12-14)

Function spaces

E1 Elevator on the 1st, 2nd, 3rd, 5th Floors

Coffee break (July 12-14)

Floor Plan - ICC JEJU 3F



Session Rooms

Samda · Halla A · Halla B
Main conference (July 9-11)

301 W3 (July 12-13)

302 W4 (July 13)

Function Rooms

300 ACL 2012 sercretariat

303 ACL secretary office

304 Preview room

Function spaces

A Registration Desk

B ACL Poster & SRW (July 9), Demo (July 10)
EMNLP Poster (July 12)

C Exhibition (July 9-11)

D Internet Rounge & Rest Area

Delizia Lunch

E1 Elevator on the 1st, 2nd, 3rd, 5th Floors

E2 Elevator on the 1st, 3rd, 4th, 5th Floors

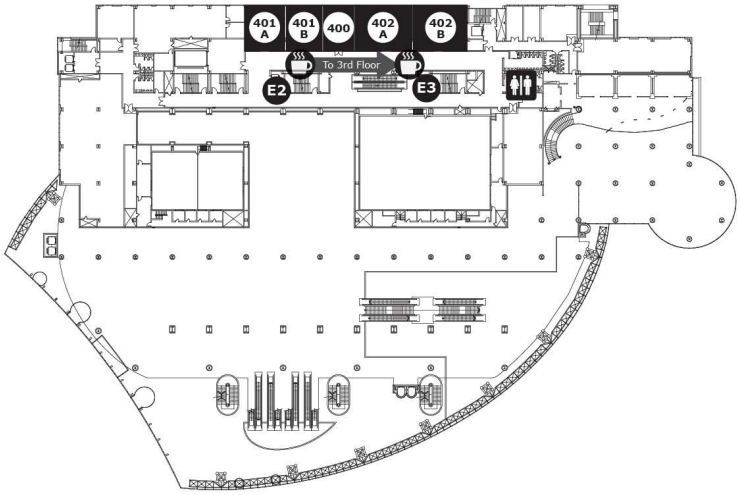
E3 Elevator on the 1st, 3rd, 4th, 5th, 6th Floors

E4 Elevator on the 1st, 3rd, 4th, 5th, 6th Floors

E5 Elevator on the 3rd, 5th, 6th Floors

Coffee break (July 9-13)
(in front of Samda and Halla: July 9-11)

Floor Plan - ICC JEJU 4F



Session Rooms

401 · 402
Main conference (July 9-11)

401A	W1 (July 12), W7 (July 13)
401B	W2 (July 12), W8 (July 13)
400	W11 (July 12)
402A	W9 (July 12), W5 (July 13)
402B	W10 (July 12), W6 (July 13)

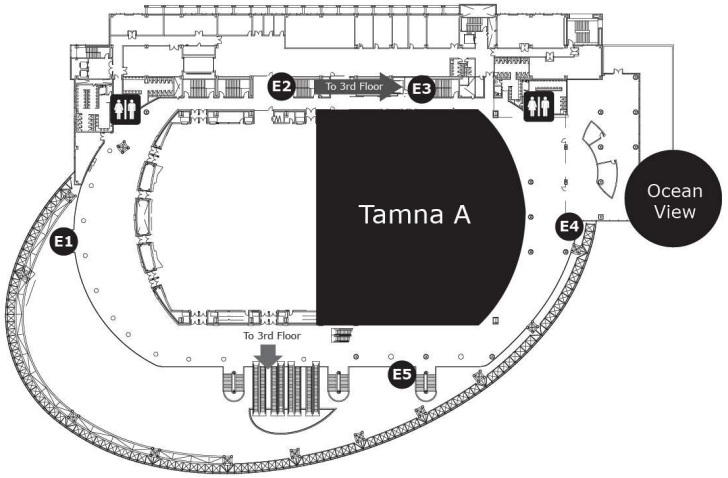
Function spaces

E2 Elevator on the 1st, 3rd, 4th, 5th Floors

E3 Elevator on the 1st, 3rd, 4th, 5th, 6th Floors

Coffee break (July 9-13)

Floor Plan - ICC JEJU 5F



Session Rooms

Tamna A

Opening & Invited talk Ⅰ (July 9)
Best Paper talk (July 10)
Lifetime Achievement Award (July 11)
Invited talk Ⅱ (July 11)
ACL Business meeting (July 11)
Closing (July 11)

Function spaces

Ocean View Student lunch

E1 Elevator on the 1st, 2nd, 3rd, 5th Floors
E2 Elevator on the 1st, 3rd, 4th, 5th Floors
E3 Elevator on the 1st, 3rd, 4th, 5th, 6th Floors
E4 Elevator on the 1st, 3rd, 4th, 5th, 6th Floors
E5 Elevator on the 3rd, 5th, 6th Floors

4

Sunday, July 8, 2012: Tutorials

Overview

7:00am – 6:00pm	**Registration**	**3F Lobby (Shilla)**	
9:00am – 12:30pm	**Morning Tutorials:**	**T1:** Weolla Room (Shilla) **T2:** Lotus 1 (Shilla) **T3:** Lotus 2+3 (Shilla)	
10:30am – 11:00am	**Coffee Break**		
12:30pm – 2:00pm	**Lunch Break**		
2:00pm – 5:30pm	**Afternoon Tutorials:**	**T4:** Weolla Room (Shilla) **T5:** Lotus 2+3 (Shilla) **T6:** Lotus 1 (Shilla)	
3:30pm – 4:00pm	**Coffee Break**		
6:00pm – 9:00pm	**Welcome Reception**	**Halla Hall (Shilla)**	

Tutorial 1

Qualitative Modeling of Spatial Prepositions and Motion Expressions

Inderjeet Mani and James Pustejovsky

Sunday, July 8, 2012, 9:00am – 12:30pm

Weolla Room (Shilla)

Abstract: The ability to understand spatial prepositions and motion in natural language will enable a variety of new applications involving systems that can respond to verbal directions, map travel guides, display incident reports, etc., providing for enhanced information extraction, question-answering, information retrieval, and more principled text to scene rendering. Until now, however, the semantics of spatial relations and motion verbs has been highly problematic. This tutorial presents a new approach to the semantics of spatial descriptions and motion expressions based on linguistically interpreted qualitative reasoning. Our approach allows for formal inference from spatial descriptions in natural language, while leveraging annotation schemes for time, space, and motion, along with machine learning from annotated corpora. We introduce a compositional semantics for motion expressions that integrates spatial primitives drawn from qualitative calculi.

No previous exposure to the semantics of spatial prepositions or motion verbs is assumed. The tutorial will sharpen cross-linguistic intuitions about the interpretation of spatial prepositions and motion constructions. The attendees will also learn about qualitative reasoning schemes for static and dynamic spatial information, as well as three annotation schemes: TimeML, SpatialML, and ISO-Space, for time, space, and motion, respectively.

While both cognitive and formal linguistics have examined the meaning of motion verbs and spatial prepositions, these earlier approaches do not yield precise computable representations that are expressive enough for natural languages. However, the previous literature makes it clear that communication of motion relies on imprecise and highly abstract geometric descriptions, rather than Euclidean ones that specify the coordinates and shapes of every object. This property makes these expressions a fit target for the field of qualitative spatial reasoning in AI, which has developed a rich set of geometric primitives for representing time, space (including distance, orientation, and topological relations), and motion. The results of such research have yielded a wide variety of spatial and temporal reasoning logics and tools. By reviewing these calculi and resources, this tutorial aims to systematically connect qualitative reasoning to natural language.

Presenters:

Inderjeet Mani
Children's Organization of Southeast Asia
Thailand
Inderjeet.Mani@gmail.com

James Pustejovsky
Department of Computer Science
Brandeis University
jamesp@cs.brandeis.edu

Tutorial 2

State-of-the-Art Kernels for Natural Language Processing

Alessandro Moschitti

Sunday, July 8, 2012, 9:00am – 12:30pm

Lotus 1 (Shilla)

Abstract: In recent years, machine learning (ML) has been used more and more to solve complex tasks in different disciplines, ranging from Data Mining to Information Retrieval or Natural Language Processing (NLP). These tasks often require the processing of structured input, e.g., the ability to extract salient features from syntactic/semantic structures is critical to many NLP systems. Mapping such structured data into explicit feature vectors for ML algorithms requires large expertise, intuition and deep knowledge about the target linguistic phenomena. Kernel Methods (KM) are powerful ML tools, which can alleviate the data representation problem. They substitute feature-based similarities with similarity functions, i.e., kernels, directly defined between training/test instances, e.g., syntactic trees. Hence feature vectors are not needed any longer. Additionally, kernel engineering, i.e., the composition or adaptation of several prototype kernels, facilitates the design of effective similarities required for new tasks.

Unfortunately, at the moment, there is neither comprehensive documentation describing the engineering techniques above nor application-oriented descriptions nor clear organization and explanation of the many successful kernels in NLP. Typically, what can be found is a documentation reporting complex theories, which obscure important practical aspects; or research papers reporting KM applications, which tend to be very specific with rather diverse notation and/or background. As a direct consequence KM technology results unappealing for most NLP researchers. This is unfortunate since KM can be easily used to speed up the design of machine learning systems for NLP (especially when using syntactic/semantic structures).

The tutorial aims at addressing the problems above: firstly, it will introduce essential and simplified theory of Support Vector Machines and KM with the only aim of motivating practical procedures and interpreting the results. Secondly, it will simply describe the current best practices for designing applications based on effective kernels. For this purpose, it will survey state-of-the-art kernels for diverse NLP applications, reconciling the different approaches with a uniform and global notation/theory. Such survey will benefit from practical expertise acquired from directly working on many natural language applications, ranging from Text Categorization to Syntactic/Semantic Parsing. Moreover, practical demonstrations using SVM-Light-TK toolkit will nicely support the application-oriented perspective of the tutorial. The latter will lead NLP researchers with heterogeneous background to the acquisition of the KM know-how, which can be used to design any target NLP application.

Finally, the tutorial will propose interesting new best practices, e.g., some recent methods for large-scale learning with structural kernels, structural lexical similarities and reverse kernel engineering.

Presenter:

Alessandro Moschitti
Department of Computer Science and Engineering
University of Trento
Via Sommarive 5
38123 POVO (TN) - Italy
moschitti@disi.unitn.it
http://disi.unitn.it/moschitti

Tutorial 3

Topic Models, Latent Space Models, Sparse Coding, and All That: A Systematic Understanding of Probabilistic Semantic Extraction in Large Corpora

Eric Xing

Sunday, July 8, 2012, 9:00am – 12:30pm

Lotus 2+3 (Shilla)

Abstract: Probabilistic topic models have recently gained much popularity in information retrieval and related areas. Via such models, one can project high-dimensional objects such as text documents into a low dimensional space where their latent semantics are captured and modeled; can integrate multiple sources of information – to "share statistical strength" among components of a hierarchical probabilistic model; and can structurally display and classify the otherwise unstructured object collections. However, to many practitioners, how topic models work, what to and not to expect from a topic model, how is it different from and related to classical matrix algebraic techniques such as LSI, NMF in NLP, how to empower topic models to deal with complex scenarios such as multimodal data, contractual text in social media, evolving corpus, or presence of supervision such as labeling and rating, how to make topic modeling computationally tractable even on web-scale data, etc., in a principled way, remain unclear. In this tutorial, I will demystify the conceptual, mathematical, and computational issues behind all such problems surrounding the topic models and their applications by presenting a systematic overview of the mathematical foundation of topic modeling, and its connections to a number of related methods popular in other fields such as the LDA, admixture model, mixed membership model, latent space models, and sparse coding. I will offer a simple and unifying view of all these techniques under the framework multi-view latent space embedding, and outline the roadmap of model extension and algorithmic design toward different applications in IR and NLP. A main theme of this tutorial that tie together a wide range of issues and problems will build on the "probabilistic graphical model" formalism, a formalism that exploits the conjoined talents of graph theory and probability theory to build complex models out of simpler pieces. I will use this formalism as a main aid to discuss both the mathematical underpinnings for the models and the related computational issues in a unified, simplistic, transparent, and actionable fashion.

Presenter:

Eric P. Xing, PhD, PhD
8101 Gates-Hillman Center (GHC)
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
epxing@cs.cmu.edu
http://www.cs.cmu.edu/∼epxing/

Tutorial 4

Multilingual Subjectivity and Sentiment Analysis

Rada Mihalcea, Carmen Banea and Janyce Wiebe

Sunday, July 8, 2012, 2:00pm – 5:30pm

Weolla Room (Shilla)

Abstract: Subjectivity and sentiment analysis focuses on the automatic identification of private states, such as opinions, emotions, sentiments, evaluations, beliefs, and speculations in natural language. While subjectivity classification labels text as either subjective or objective, sentiment classification adds an additional level of granularity, by further classifying subjective text as either positive, negative or neutral.

While much of the research work in this area has been applied to English, research on other languages is growing, including Japanese, Chinese, German, Spanish, Romanian. While most of the researchers in the field are familiar with the methods applied on English, few of them have closely looked at the original research carried out in other languages. For example, in languages such as Chinese, researchers have been looking at the ability of characters to carry sentiment information. In Romanian, due to markers of politeness and additional verbal modes embedded in the language, experiments have hinted that subjectivity detection may be easier to achieve. These additional sources of information may not be available across all languages, yet, various articles have pointed out that by investigating a synergistic approach for detecting subjectivity and sentiment in multiple languages at the same time, improvements can be achieved not only in other languages, but in English as well. The development and interest in these methods is also highly motivated by the fact that only 27% of Internet users speak English (www.internetworldstats.com/stats.htm, Oct 11, 2011), and that number diminishes further every year, as more people across the globe gain Internet access.

The aim of this tutorial is to familiarize the attendees with the subjectivity and sentiment research carried out on languages other than English in order to enable and promote cross-fertilization. Specifically, we will review work along three main directions. First, we will present methods where the resources and tools have been specifically developed for a given target language. In this category, we will also briefly overview the main methods that have been proposed for English, but which can be easily ported to other languages. Second, we will describe cross-lingual approaches, including several methods that have been proposed to leverage on the resources and tools available in English by using cross-lingual projections. Finally, third, we will show how the expression of opinions and polarity pervades language boundaries, and thus methods that holistically explore multiple languages at the same time can be effectively considered.

Presenters:

Rada Mihalcea
Department of Computer Science and Engineering
University of North Texas
P.O.Box 311366
Denton, TX 76203
rada@cs.unt.edu

Carmen Banea
Department of Computer Science and Engineering
University of North Texas

P.O.Box 311366
Denton, TX 76203
carmenbanea@my.unt.edu

Janyce Wiebe
Department of Computer Science
Sennott Square Building, Room 5409
210 S. Bouquet St.
Pittsburgh, PA 15260, U.S.A.
wiebe@cs.pitt.edu

Tutorial 5

Deep Learning for NLP (without Magic)

Richard Socher, Yoshua Bengio and Christopher D. Manning

Sunday, July 8, 2012, 2:00pm – 5:30pm

Lotus 2+3 (Shilla)

Abstract: Machine learning is everywhere in today's NLP, but by and large machine learning amounts to numerical optimization of weights for human designed representations and features. The goal of deep learning is to explore how computers can take advantage of data to develop features and representations appropriate for complex interpretation tasks.

This tutorial aims to cover the basic motivation, ideas, models and learning algorithms in deep learning for natural language processing. Recently, these methods have been shown to perform very well on various NLP tasks such as language modeling, POS tagging, named entity recognition, sentiment analysis and paraphrase detection, among others. The most attractive quality of these techniques is that they can perform well without any external hand-designed resources or time-intensive feature engineering. Despite these advantages, many researchers in NLP are not familiar with these methods. Our focus is on insight and understanding, using graphical illustrations and simple, intuitive derivations. The goal of the tutorial is to make the inner workings of these techniques transparent, intuitive and their results interpretable, rather than black boxes labeled "magic here".

The first part of the tutorial presents the basics of neural networks, neural word vectors, several simple models based on local windows and the math and algorithms of training via backpropagation. In this section applications include language modeling and POS tagging.

In the second section we present recursive neural networks which can learn structured tree outputs as well as vector representations for phrases and sentences. We cover both equations as well as applications. We show how training can be achieved by a modified version of the backpropagation algorithm introduced before. These modifications allow the algorithm to work on tree structures. Applications include sentiment analysis and paraphrase detection. We also draw connections to recent work in semantic compositionality in vector spaces. The principle goal, again, is to make these methods appear intuitive and interpretable rather than mathematically confusing. By this point in the tutorial, the audience members should have a clear understanding of how to build a deep learning system for word-, sentence- and document-level tasks.

The last part of the tutorial gives a general overview of the different applications of deep learning in NLP, including bag of words models. We will provide a discussion of NLP-oriented issues in modeling, interpretation, representational power, and optimization.

Presenters:

Richard Socher
Stanford University
353 Serra Mall Rm 228
Stanford, CA 94305-9040, USA
richard@socher.org
http://www.socher.org

Yoshua Bengio
Département d'Informatique et de recherche opérationnelle
Université de Montréal, P.O. Box 6128, Centre-Ville Branch
Montréal (QC), H3C 3J7, Canada

Yoshua.Bengio@umontreal.ca
http://www.iro.umontreal.ca/∼bengioy/yoshua_en/index.html

Christopher D. Manning
Stanford University
353 Serra Mall Rm 158
Stanford, CA 94305-9040, USA
manning@cs.stanford.edu
http://nlp.stanford.edu/∼manning/

Tutorial 6

Graph-based Semi-Supervised Learning Algorithms for NLP

Amarnag Subramanya and Partha Pratim Talukdar

Sunday, July 8, 2012, 2:00pm – 5:30pm

Lotus 1 (Shilla)

Abstract: While labeled data is expensive to prepare, ever increasing amounts of unlabeled linguistic data are becoming widely available. In order to adapt to this phenomenon, several semi-supervised learning (SSL) algorithms, which learn from labeled as well as unlabeled data, have been developed. In a separate line of work, researchers have started to realize that graphs provide a natural way to represent data in a variety of domains. Graph-based SSL algorithms, which bring together these two lines of work, have been shown to outperform the state-of-the-art in many applications in speech processing, computer vision and NLP. In particular, recent NLP research has successfully used graph-based SSL algorithms for PoS tagging, semantic parsing, knowledge acquisition, sentiment analysis, and text categorization.

Recognizing this promising and emerging area of research, this tutorial focuses on graph-based SSL algorithms (e.g., label propagation methods). The tutorial is intended to be a sequel to the ACL 2008 SSL tutorial, focusing exclusively on graph-based SSL methods and recent advances in this area, which were beyond the scope of the previous tutorial.

The tutorial is divided in two parts. In the first part, we will motivate the need for graph-based SSL methods, introduce some standard graph-based SSL algorithms, and discuss connections between these approaches. We will also discuss how linguistic data can be encoded as graphs and show how graph-based algorithms can be scaled to large amounts of data (e.g., web-scale data).

Part 2 of the tutorial will focus on how graph-based methods can be used to solve several critical NLP tasks, including basic problems such as PoS tagging, semantic parsing, coreference resolution and more downstream tasks such as text categorization and information acquisition, and sentiment analysis. We will conclude the tutorial with some exciting avenues for future work.

Familiarity with semi-supervised learning and graph-based methods will not be assumed, and the necessary background will be provided. Examples from NLP tasks will be used throughout the tutorial to convey the necessary concepts. At the end of this tutorial, the attendee will walk away with the following:

An in-depth knowledge of the current state-of-the-art in graph-based SSL algorithms, and the ability to implement them. The ability to decide on the suitability of graph-based SSL methods for a problem. Familiarity with different NLP tasks where graph-based SSL methods have been successfully applied.

In addition to the above goals, we hope that this tutorial will better prepare the attendee to conduct exciting research at the intersection of NLP and other emerging areas with natural graph-structured data (e.g., Computation Social Science).

Please visit http://graph-ssl.wikidot.com/ for details.

Presenters:

Amarnag Subramanya
Google Research
1600 Amphitheater Pkwy.
Mountain View, CA 94043

asubram@google.com
http://sites.google.com/site/amarsubramanya

Partha Pratim Talukdar
GHC 8133, Machine Learning Department
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
partha.talukdar@cs.cmu.edu
http://www.talukdar.net

5

Monday, July 9, 2012: Main Conference

Overview

7:30am – 5:00pm	**Registration**	**(3F Lobby (ICC))**
8:45am – 9:00am	**Welcome**	**(Tamna A (ICC))**
9:00am – 10:30am	**Invited Talk: Aravind K. Joshi**	**(Tamna A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
12:30pm – 2:00pm	**Lunch Break**	
2:00pm – 3:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
6:00pm – 8:30pm	**Posters & Demos (with buffet dinner)**	**(3F Lobby (ICC))**

Schedule

7:30am – 5:00pm	Registration (3F Lobby (ICC))

8:45am – 9:00am	Welcome (Tamna A (ICC))				
9:00am – 10:30am	Invited Talk (Tamna A (ICC)) ***Aravind K. Joshi:* "Remembrance of ACLs past"** **Chair:** *Mark Steedman*				
10:30am – 11:00am	Coffee Break (3F Lobby (ICC))				
Parallel Sessions	**Halla A (ICC)**	**Halla B (ICC)**	**Samda (ICC)**	**402 (ICC)**	**401 (ICC)**
11:00am – 12:30pm	**Machine Translation I** **Chair:** *David Chiang*	**Speech I** **Chair:** *Sanjeev Khudanpur*	**Discourse I** **Chair:** *Bonnie Webber*	**Time** **Chair:** *Jochen Leidner*	**Meaning I** **Chair:** *Alessandro Moschitti*
11:00am – 11:30am	Learning to Translate with Multiple Objectives *K. Duh, K. Sudoh, X. Wu, H. Tsukada, and M. Nagata* ..... **p. 53**	Probabilistic Integration of Partial Lexical Information for Noise Robust Haptic Voice Recognition *K. C. Sim* ..... **p. 53**	Text-level Discourse Parsing with Rich Linguistic Features *V. W. Feng and G. Hirst* ..... **p. 54**	Extracting Narrative Timelines as Temporal Dependency Structures *O. Kolomiyets, S. Bethard, and M.-F. Moens* ... **p. 55**	Efficient Tree-based Approximation for Entailment Graph Learning *J. Berant, I. Dagan, M. Adler, and J. Goldberger* ..... **p. 56**
11:30am – 12:00pm	Joint Feature Selection in Distributed Stochastic Learning for Large-Scale Discriminative Training in SMT *P. Simianer, S. Riezler, and C. Dyer* ..... **p. 53**	A Nonparametric Bayesian Approach to Acoustic Model Discovery *C.-y. Lee and J. Glass* ..... **p. 54**	PDTB-style Discourse Annotation of Chinese Text *Y. Zhou and N. Xue* . **p. 54**	Labeling Documents with Timestamps: Learning from their Time Expressions *N. Chambers* ..... **p. 55**	Learning High-Level Planning from Text *S. Branavan, N. Kushman, T. Lei, and R. Barzilay* ..... **p. 56**
12:00pm – 12:30pm	Prediction of Learning Curves in Machine Translation *P. Kolachina, N. Cancedda, M. Dymetman, and S. Venkatapathy* .... **p. 53**	Automated Essay Scoring Based on Finite State Transducer: towards ASR Transcription of Oral English Speech *X. Peng, D. Ke, and B. Xu* ..... **p. 54**	SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations *V.-A. Nguyen, J. Boyd-Graber, and P. Resnik* ..... **p. 55**	Temporally Anchored Relation Extraction *G. Garrido, A. Peñas, B. Cabaleiro, and A. Rodrigo* ..... **p. 55**	Distributional Semantics in Technicolor *E. Bruni, G. Boleda, M. Baroni, and N. K. Tran* ..... **p. 56**
12:30pm – 2:00pm	**Lunch Break**				

Parallel Sessions	Halla A (ICC)	Samda (ICC)	Halla B (ICC)	402 (ICC)	401 (ICC)
2:00pm – 3:30pm	**Machine Translation II** **Chair:** *Kevin Duh*	**Speech II** **Chair:** *Khe Chai Sim*	**Parsing I** **Chair:** *Dominick Ng*	**Chinese Lexical Processing** **Chair:** *Jun Hatori*	**Lexical Semantics** **Chair:** *Lillian Lee*
2:00pm – 2:30pm	A Class-Based Agreement Model for Generating Accurately Inflected Translations *S. Green and J. DeNero* . . . . . **p. 56**	Fast Syntactic Analysis for Statistical Language Modeling via Substructure Sharing and Uptraining *A. Rastrow, M. Dredze, and S. Khudanpur* . . . . . **p. 57**	Discriminative Strategies to Integrate Multiword Expression Recognition and Parsing *M. Constant, A. Sigogne, and P. Watrin* . . . . . **p. 58**	Reducing Approximation and Estimation Errors for Chinese Lexical Processing with Heterogeneous Annotations *W. Sun and X. Wan* . . **p. 59**	Verb Classification using Distributional Similarity in Syntactic and Semantic Structures *D. Croce, A. Moschitti, R. Basili, and M. Palmer* . . . . . **p. 60**
2:30pm – 3:00pm	Deciphering Foreign Language by Combining Language Models and Context Vectors *M. Nuhn, A. Mauser, and H. Ney* . . . . . **p. 57**	Bootstrapping a Unified Model of Lexical and Phonetic Acquisition *M. Elsner, S. Goldwater, and J. Eisenstein* . . . **p. 57**	Utilizing Dependency Language Models for Graph-based Dependency Parsing Models *W. Chen, M. Zhang, and H. Li* . . . . . **p. 58**	Capturing Paradigmatic and Syntagmatic Lexical Relations: Towards Accurate Chinese Part-of-Speech Tagging *W. Sun and H. Uszkoreit* . . . . . **p. 59**	Word Sense Disambiguation Improves Information Retrieval *Z. Zhong and H. T. Ng* . . . . . **p. 60**
3:00pm – 3:30pm	Machine Translation without Words through Substring Alignment *G. Neubig, T. Watanabe, S. Mori, and T. Kawahara* . . . . . **p. 57**	Discriminative Pronunciation Modeling: A Large-Margin, Feature-Rich Approach *H. Tang, J. Keshet, and K. Livescu* . . . . . **p. 58**	Spectral Learning of Latent-Variable PCFGs *S. B. Cohen, K. Stratos, M. Collins, D. P. Foster, and L. Ungar* . . . . . **p. 59**	Fast Online Training with Frequency-Adaptive Learning Rates for Chinese Word Segmentation and New Word Detection *X. Sun, H. Wang, and W. Li* . . . . . **p. 59**	Efficient Search for Transformation-based Inference *A. Stern, R. Stern, I. Dagan, and A. Felner* . . . . . **p. 60**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**				
Parallel Sessions	**Halla A (ICC)**	**Halla B (ICC)**	**Samda (ICC)**	**402 (ICC)**	**401 (ICC)**
4:00pm – 5:30pm	**Machine Translation III** **Chair:** *Spence Green*	**Aspect** **Chair:** *Jing Jiang*	**Generation and Machine Reading** **Chair:** *Kristian Woodsend*	**Dialogue and Discourse** **Chair:** *Michael Strube*	**Lexicon** **Chair:** *Hwee Tou Ng*
4:00pm – 4:30pm	Maximum Expected BLEU Training of Phrase and Lexicon Translation Models *X. He and L. Deng* . . **p. 60**	Modeling Review Comments *A. Mukherjee and B. Liu* . . . . . **p. 61**	Learning to "Read Between the Lines" using Bayesian Logic Programs *S. Raghavan, R. Mooney, and H. Ku* . . . . . **p. 62**	A Discriminative Hierarchical Model for Fast Coreference at Large Scale *M. Wick, S. Singh, and A. McCallum* . . . . . **p. 63**	Cross-Domain Co-Extraction of Sentiment and Topic Lexicons *F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu* **p. 63**

4:30pm – 5:00pm	Learning Translation Consensus with Structured Label Propagation *S. Liu, C.-H. Li, M. Li, and M. Zhou* ..... **p. 60**	A Joint Model for Discovery of Aspects in Utterances *A. Celikyilmaz and D. Hakkani-Tur* .... **p. 61**	Collective Generation of Natural Image Descriptions *P. Kuznetsova, V. Ordonez, A. Berg, T. Berg, and Y. Choi* ..... **p. 62**	Coreference Semantics from Web Features *M. Bansal and D. Klein* ..... **p. 63**	Learning Syntactic Verb Frames using Graphical Models *T. Lippincott, A. Korhonen, and D. Ó Séaghdha* . **p. 64**
5:00pm – 5:30pm	Smaller Alignment Models for Better Translations: Unsupervised Word Alignment with the l0-norm *A. Vaswani, L. Huang, and D. Chiang* ..... **p. 61**	Aspect Extraction through Semi-Supervised Modeling *A. Mukherjee and B. Liu* ..... **p. 61**	Concept-to-text Generation via Discriminative Reranking *I. Konstas and M. Lapata* ..... **p. 62**	Subgroup Detection in Ideological Discussions *A. Abu-Jbara, P. Dasigi, M. Diab, and D. Radev* ..... **p. 63**	Fast Online Lexicon Learning for Grounded Language Acquisition *D. Chen* ..... **p. 64**
6:00pm – 8:30pm	**— 3F Lobby (ICC) —** **Poster Session and Demos (with Buffet Dinner)**				

Keynote Address: Aravind K. Joshi

"Remembrance of ACLs past"

Aravind K. Joshi

Monday, July 9, 2012, 9:00am – 10.30am

Tamna A (ICC)

Abstract: Besides briefly covering some highlights of the past 50 years of ACL from my perspective, I will try to comment on (1) why some directions of research were pursued for a while and then dropped, sometimes for a good reason and sometimes apparently for no reason, (2) why the relationship to Linguistics, Psycholinguistics, and AI goes up and down, and (3) are there any leftovers that have the possibility of being turned into delicious contributions!

Short Bio: After completing his undergraduate work in Electrical and Communication Engineering in India, Aravind Joshi came to the University of Pennsylvania and obtained his Ph.D. in Electrical Engineering in 1960. At present, he is the Henry Salvatori Professor of Computer and Cognitive Science at the University of Pennsylvania.

Joshi has worked on several problems that overlap computer science and linguistics. More specifically, he has worked on topics in mathematical linguistics as they relate to formal and linguistic adequacy of different formalisms and their processing implications. He has also worked on several aspects of theories of representation and inference in natural language, especially as they relate to discourse.

Joshi was the President of ACL in 1975 and was appointed as a Founding Fellow of ACL in 2011.

He was awarded the Lifetime Achievement Award of ACL in 2002, the David Rumelhart Prize of the Cognitive Science Society in 2003 and the Franklin Medal for Computer and Cognitive Science, Franklin Institute, Philadelphia, in 2005.

Short Paper, Poster, and Student Research Workshop Sessions

Short Papers

Posters

Posters from the Student Research Workshop

Long Paper Abstracts

Learning to Translate with Multiple Objectives

Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada, and Masaaki Nagata

Monday 11:00am–11:30pm — Halla A (ICC)

We introduce an approach to optimize a machine translation (MT) system on multiple metrics simultaneously. Different metrics (e.g. BLEU, TER) focus on different aspects of translation quality; our multi-objective approach leverages these diverse aspects to improve overall quality.

Our approach is based on the theory of Pareto Optimality. It is simple to implement on top of existing single-objective optimization methods (e.g. MERT, PRO) and outperforms ad hoc alternatives based on linear-combination of metrics. We also discuss the issue of metric tunability and show that our Pareto approach is more effective in incorporating new metrics from MT evaluation for MT optimization.

Joint Feature Selection in Distributed Stochastic Learning for Large-Scale Discriminative Training in SMT

Patrick Simianer, Stefan Riezler, and Chris Dyer

Monday 11:30am–12:00pm — Halla A (ICC)

With a few exceptions, discriminative training in statistical machine translation (SMT) has been content with tuning weights for large feature sets on small development data. Evidence from machine learning indicates that increasing the training sample size results in better prediction. The goal of this paper is to show that this common wisdom can also be brought to bear upon SMT. We deploy local features for SCFG-based SMT that can be read off from rules at runtime, and present a learning algorithm that applies l1/l2 regularization for joint feature selection over distributed stochastic learning processes. We present experiments on learning on 1.5 million training sentences, and show significant improvements over tuning discriminative models on small development sets.

Prediction of Learning Curves in Machine Translation

Prasanth Kolachina, Nicola Cancedda, Marc Dymetman, and Sriram Venkatapathy

Monday 12:00pm–12:30pm — Halla A (ICC)

Parallel data in the domain of interest is the key resource when training a statistical machine translation (SMT) system for a specific purpose. Since ad-hoc manual translation can represent a significant investment in time and money, a prior assesment of the amount of training data required to achieve a satisfactory accuracy level can be very useful. In this work, we show how to predict what the learning curve of an SMT system would look like if we were to manually translate increasing amounts of data.

We consider two scenarios, 1) Monolingual samples in the source and target languages are available and 2) An additional small amount of parallel corpus is also available. We propose methods for predicting learning curves for both these scenarios.

Probabilistic Integration of Partial Lexical Information for Noise Robust Haptic Voice Recognition

Khe Chai Sim

Monday 11:00am–11:30am — Halla B (ICC)

This paper presents a probabilistic framework that combines multiple knowledge sources for Haptic Voice Recognition (HVR), a multi-modal input method designed to provide efficient text entry on modern mobile devices. HVR extends the conventional voice input by allowing users to provide complementary partial lexical information via touch input to improve the efficiency and accuracy of voice recognition. This paper investigates the use of the initial letter of the words in the utterance as the partial lexical information. In addition to the acoustic and language models used in automatic speech recognition systems, HVR uses the haptic and partial lexical models as additional knowledge sources to reduce the recognition search space and suppress confusions.

Experimental results show that both the word error rate and runtime factor can be re- duced by a factor of two using HVR.

A Nonparametric Bayesian Approach to Acoustic Model Discovery
Chia-ying Lee and James Glass
Monday 11:30am–12:00pm — Halla B (ICC)

We investigate the problem of acoustic modeling in which prior language-specific knowledge and transcribed data are unavailable. We present an unsupervised model that simultaneously segments the speech, discovers a proper set of sub-word units (e.g., phones) and learns a Hidden Markov Model (HMM) for each induced acoustic unit. Our approach is formulated as a Dirichlet process mixture model in which each mixture is an HMM that represents a sub-word unit. We apply our model to the TIMIT corpus, and the results demonstrate that our model discovers sub-word units that are highly correlated with English phones and also produces better segmentation than the state-of-the-art unsupervised baseline. We test the quality of the learned acoustic models on a spoken term detection task. Compared to the baselines, our model improves the relative precision of top hits by at least 22.1% and outperforms a language-mismatched acoustic model.

Automated Essay Scoring Based on Finite State Transducer: towards ASR Transcription of Oral English Speech
Xingyuan Peng, Dengfeng Ke, and Bo Xu
Monday 12:00pm–12:30pm — Halla B (ICC)

Conventional Automated Essay Scoring (AES) measures may cause severe problems when directly applied in scoring Automatic Speech Recognition (ASR) transcription as they are error sensitive and unsuitable for the characteristic of ASR transcription. Therefore, we introduce a framework of Finite State Transducer (FST) to avoid the shortcomings. Compared with the Latent Semantic Analysis with Support Vector Regression (LSA-SVR) method (stands for the conventional measures), our FST method shows better performance especially towards the ASR transcription. In addition, we apply the synonyms similarity to expand the FST model. The final scoring performance reaches an acceptable level of 0.80 which is only 0.07 lower than the correlation (0.87) between human raters.

Text-level Discourse Parsing with Rich Linguistic Features
Vanessa Wei Feng and Graeme Hirst
Monday 11:00am–11:30am — Samda (ICC)

In this paper, we develop an RST-style text-level discourse parser, based on the HILDA discourse parser (Hernault et al., 2010b). We significantly improve its tree-building step by incorporating our own rich linguistic features. We also analyze the difficulty of extending traditional sentence-level discourse parsing to text-level parsing by comparing discourse parsing performance under different discourse conditions.

PDTB-style Discourse Annotation of Chinese Text
Yuping Zhou and Nianwen Xue
Monday 11:30am–12:00pm — Samda (ICC)

We describe a discourse annotation scheme for Chinese and report on the preliminary results. Our scheme, inspired by the Penn Discourse TreeBank (PDTB), adopts the lexically grounded approach; at the same time, it makes adaptations based on the linguistic and statistical characteristics of Chinese text. Annotation results show that these adaptations work well in practice. Our scheme, taken together with other PDTB-style schemes (e.g. for English, Turkish, Hindi, and Czech), affords a broader perspective on how the generalized lexically grounded approach can flesh itself out in the context of cross-linguistic annotation of discourse relations.

SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations

Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik

Monday 12:00pm–12:30pm — Samda (ICC)

One of the key tasks for analyzing conversational data is segmenting it into coherent topic segments. However, most models of topic segmentation ignore the social aspect of conversations, focusing only on the words used. We introduce a hierarchical Bayesian nonparametric model, Speaker Identity for Topic Segmentation (SITS), that discovers (1) the topics used in a conversation, (2) how these topics are shared across conversations, (3) when these topics shift, and (4) a person-specific tendency to introduce new topics. We evaluate against current unsupervised segmentation models to show that including person-specific information improves segmentation performance on meeting corpora and on political debates. Moreover, we provide evidence that SITS captures an individual's tendency to introduce new topics in political contexts, via analysis of the 2008 US presidential debates and the television program Crossfire.

Extracting Narrative Timelines as Temporal Dependency Structures

Oleksandr Kolomiyets, Steven Bethard, and Marie-Francine Moens

Monday 11:00am–11:30am — 402 (ICC)

We propose a new approach to characterizing the timeline of a text: temporal dependency structures, where all the events of a narrative are linked via partial ordering relations like BEFORE, AFTER, OVERLAP and IDENTITY. We annotate a corpus of children's stories with temporal dependency trees, achieving agreement (Krippendorff's Alpha) of 0.856 on the event words, 0.822 on the links between events, and of 0.700 on the ordering relation labels. We compare two parsing models for temporal dependency structures, and show that a deterministic non-projective dependency parser outperforms a graph-based maximum spanning tree parser, achieving labeled attachment accuracy of 0.647 and labeled tree edit distance of 0.596. Our analysis of the dependency parser errors gives some insights into future research directions.

Labeling Documents with Timestamps: Learning from their Time Expressions

Nathanael Chambers

Monday 11:30am–12:00pm — 402 (ICC)

Temporal reasoners for document understanding typically assume that a document's creation date is known. Algorithms to ground relative time expressions and order events often rely on this timestamp to assist the learner. Unfortunately, the timestamp is not always known, particularly on the Web. This paper addresses the task of automatic document timestamping, presenting two new models that incorporate rich linguistic features about time. The first is a discriminative classifier with new features extracted from the text's time expressions (e.g., 'since 1999'). This model alone improves on previous generative models by 77%. The second model learns probabilistic constraints between time expressions and the unknown document time. Imposing these learned constraints on the discriminative model further improves its accuracy. Finally, we present a new experiment design that facilitates easier comparison by future work.

Temporally Anchored Relation Extraction

Guillermo Garrido, Anselmo Peñas, Bernardo Cabaleiro, and Álvaro Rodrigo

Monday 12:00pm–12:30pm — 402 (ICC)

Although much work on relation extraction has aimed at obtaining static facts, many of the target relations are actually fluents, as their validity is naturally anchored to a certain time period. This paper proposes a methodological approach to temporally anchored relation extraction. Our proposal performs distant supervised learning to extract a set of relations from a natural language corpus, and anchors each of them to an interval of temporal validity, aggregating evidence from documents supporting the relation. We use a rich graph-based document-level representation to generate novel features for this task. Results show that our implementation for temporal anchoring is able to achieve a 69% of the upper bound performance imposed by the relation extraction step. Compared to the state of the art, the overall system achieves the highest precision

reported.

Efficient Tree-based Approximation for Entailment Graph Learning
Jonathan Berant, Ido Dagan, Meni Adler, and Jacob Goldberger
Monday 11:00am–11:30am — 401 (ICC)

Learning entailment rules is fundamental in many semantic-inference applications and has been an active field of research in recent years. In this paper we address the problem of learning transitive graphs that describe entailment rules between predicates (termed entailment graphs). We first identify that entailment graphs exhibit a "tree-like" property and are very similar to a novel type of graph termed forest-reducible graph. We utilize this property to develop an iterative efficient approximation algorithm for learning the graph edges, where each iteration takes linear time. We compare our approximation algorithm to a recently-proposed state-of-the-art exact algorithm and show that it is more efficient and scalable both theoretically and empirically, while its output quality is close to that given by the optimal solution of the exact algorithm.

Learning High-Level Planning from Text
S.R.K. Branavan, Nate Kushman, Tao Lei, and Regina Barzilay
Monday 11:30am–12:00pm — 401 (ICC)

Comprehending action preconditions and effects is an essential step in modeling the dynamics of the world. In this paper, we express the semantics of precondition relations extracted from text in terms of planning operations. The challenge of modeling this connection is to ground language at the level of relations. This type of grounding enables us to create high-level plans based on language abstractions. Our model jointly learns to predict precondition relations from text and to perform high-level planning guided by those relations. We implement this idea in the reinforcement learning framework using feedback automatically obtained from plan execution attempts. When applied to a complex virtual world and text describing that world, our relation extraction technique performs on par with a supervised baseline, yielding an F-measure of 66% compared to the baseline's 65%. Additionally, we show that a high-level planner utilizing these extracted relations significantly outperforms a strong, text unaware baseline – successfully completing 80% of planning tasks as compared to 69% for the baseline.

Distributional Semantics in Technicolor
Elia Bruni, Gemma Boleda, Marco Baroni, and Nam Khanh Tran
Monday 12:00pm–12:30pm — 401 (ICC)

Following a recent trend, our research aims at building computational models of word meaning that are perceptually grounded. Using computer vision techniques, we build visual and multimodal distributional models and compare them to standard textual models. Our results show that, while visual models with state-of-the-art computer vision techniques perform worse than textual models in general semantic tasks (accounting for word similarity and relatedness), they are as good or better models of the meaning of words with visual correlates such as color terms, even in a non-trivial task that involves non-literal uses of such words. Moreover, we show that visual and textual information are tapping on different aspects of meaning, such that they are complementary sources of information. We also show that models entirely based on the text that labels images are a powerful surrogate of visual information at least for simple tasks.

A Class-Based Agreement Model for Generating Accurately Inflected Translations
Spence Green and John DeNero
Monday 2:00pm–2:30pm — Halla A (ICC)

When automatically translating from a weakly inflected source language like English to a target language with richer grammatical features such as gender and dual number, the output commonly contains morpho-syntactic agreement errors. To address this issue, we present a target-side, class-based agreement model. Agreement is promoted by scoring a sequence of fine-grained

morpho-syntactic classes that are predicted during decoding for each translation hypothesis. For English-to-Arabic translation, our model yields a +1.04 BLEU average improvement over a state-of-the-art baseline. The model does not require bitext or phrase table annotations and can be easily implemented as a feature in many phrase-based decoders.

Deciphering Foreign Language by Combining Language Models and Context Vectors
Malte Nuhn, Arne Mauser, and Hermann Ney
Monday 2:30pm–3:00pm — Halla A (ICC)

In this paper we show how to train statistical machine translation systems on real-life tasks using only non-parallel monolingual data from two languages. We present a modification of the method shown in (Ravi and Knight, 2011) that is scalable to vocabulary sizes of several thousand words. On the task shown in (Ravi and Knight, 2011) we obtain better results with only 5% of the computational effort when running our method with an n-gram language model. The efficiency improvement of our method allows us to run experiments with vocabulary sizes of around 5,000 words, such as a non-parallel version of the Verbmobil corpus. We also report results using data from the monolingual French and English Gigaword corpora.

Machine Translation without Words through Substring Alignment
Graham Neubig, Taro Watanabe, Shinsuke Mori, and Tatsuya Kawahara
Monday 3:00pm–3:30pm — Halla A (ICC)

In this paper, we demonstrate that accurate machine translation is possible without the concept of "words," treating MT as a problem of transformation between character strings. We achieve this result by applying phrasal inversion transduction grammar alignment techniques to character strings to train a character-based translation model, and using this in the phrase-based MT framework. We also propose a look-ahead parsing algorithm and substring-informed prior probabilities to achieve more effective and efficient alignment. In an evaluation, we demonstrate that character-based translation can achieve results that compare to word-based systems while effectively translating unknown and uncommon words over several language pairs.

Fast Syntactic Analysis for Statistical Language Modeling via Substructure Sharing and Uptraining
Ariya Rastrow, Mark Dredze, and Sanjeev Khudanpur
Monday 2:00pm–2:30pm — Samda (ICC)

Long-span features, such as syntax, can improve language models for tasks such as speech recognition and machine translation. However, these language models can be difficult to use in practice because of the time required to generate features for rescoring a large hypothesis set. In this work, we propose substructure sharing, which saves duplicate work in processing hypothesis sets with redundant hypothesis structures. We apply substructure sharing to a dependency parser and part of speech tagger to obtain significant speedups, and further improve the accuracy of these tools through up-training. When using these improved tools in a language model for speech recognition, we obtain significant speed improvements with both N-best and hill climbing rescoring, and show that up-training leads to WER reduction.

Bootstrapping a Unified Model of Lexical and Phonetic Acquisition
Micha Elsner, Sharon Goldwater, and Jacob Eisenstein
Monday 2:30pm–3:00pm — Samda (ICC)

During early language acquisition, infants must learn both a lexicon and a model of phonetics that explains how lexical items can vary in pronunciation—for instance "the" might be realized as [Di] or [D@]. Previous models of acquisition have generally tackled these problems in isolation, yet behavioral evidence suggests infants acquire lexical and phonetic knowledge simultaneously. We present a Bayesian model that clusters together phonetic variants of the same lexical item while learning both a language model over lexical items and a log-linear model of pronunciation

variability based on articulatory features. The model is trained on transcribed surface pronunciations, and learns by bootstrapping, without access to the true lexicon. We test the model using a corpus of child-directed speech with realistic phonetic variation and either gold standard or automatically induced word boundaries. In both cases modeling variability improves the accuracy of the learned lexicon over a system that assumes each lexical item has a unique pronunciation.

Discriminative Pronunciation Modeling: A Large-Margin, Feature-Rich Approach
Hao Tang, Joseph Keshet, and Karen Livescu
Monday 3:00pm–3:30pm — Samda (ICC)

We address the problem of learning the mapping between words and their possible pronunciations in terms of sub-word units. Previous approaches have largely involved generative modeling of the distribution of pronunciations, usually trained to maximize likelihood. We propose a discriminative, feature-rich approach using large-margin learning. This approach allows us to optimize an objective closely related to a discriminative task, to incorporate a large number of complex features, and still do inference efficiently. We test the approach on the task of lexical access; that is, the prediction of a word given a phonetic transcription. In experiments on a subset of the Switchboard conversational speech corpus, our models thus far improve classification error rates from a previously published result of 29.1% to about 15%. We find that large-margin approaches outperform CRF learning, and that the Passive-Aggressive (PA) algorithm for large-margin learning is faster to converge than the Pegasos algorithm.

Discriminative Strategies to Integrate Multiword Expression Recognition and Parsing
Matthieu Constant, Anthony Sigogne, and Patrick Watrin
Monday 2:00pm–2:30pm — Halla B (ICC)

The integration of multiword expressions in a parsing procedure has been shown to improve accuracy in an artificial context where such expressions have been perfectly pre-identified.

This paper evaluates two empirical strategies to integrate multiword units in a real constituency parsing context and shows that the results are not as promising as has sometimes been suggested. Firstly, we show that pre-grouping multiword expressions before parsing with a state-of-the-art recognizer improves multiword recognition accuracy and unlabeled attachment score. However, it has no statistically significant impact in terms of F-score as incorrect multiword expression recognition has important side effects on parsing. Secondly, integrating multiword expressions in the parser grammar followed by a reranker specific to such expressions slightly improves all evaluation metrics.

Utilizing Dependency Language Models for Graph-based Dependency Parsing Models
Wenliang Chen, Min Zhang, and Haizhou Li
Monday 2:30pm–3:00pm — Halla B (ICC)

Most previous graph-based dependency parsing models increase decoding complexity when they use high-order features as a result of exact-inference decoding. In this paper, we present an approach to enriching high-order feature representations for graph-based dependency parsing models using a dependency language model and beam search. The dependency language model is built on a large-amount of additional auto-parsed data that is processed by a baseline parser. Based on the dependency language model, we represent a set of features for the parsing model. Finally, the features are efficiently integrated into the parsing model during decoding using beam search. Our approach has two advantages. Firstly we utilize rich high-order features defined over a view of large scope and additional large raw corpus. Secondly our approach does not increase the decoding complexity. We evaluate the proposed approach on English and Chinese data. The experimental results show that our new parser achieves the best accuracy on the Chinese data and comparable accuracy with the best known systems on the English data.

Spectral Learning of Latent-Variable PCFGs
Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster, and Lyle Ungar
Monday 3:00pm–3:30pm — Halla B (ICC)

We introduce a spectral learning algorithm for latent-variable PCFGs (Petrov et al., 2006; Matsuzaki et al., 2005). Under a separability (singular value) condition, we prove that the method provides consistent parameter estimates. Our result rests on three theorems: the first gives a tensor form of the inside-outside algorithm for PCFGs; the second shows that the required tensors can be estimated directly from training examples where hidden-variable values are missing; the third gives a PAC-style convergence bound for the estimation method.

Reducing Approximation and Estimation Errors for Chinese Lexical Processing with Heterogeneous Annotations
Weiwei Sun and Xiaojun Wan
Monday 2:00pm–2:30pm — 402 (ICC)

We address the issue of consuming heterogeneous annotation data for Chinese word segmentation and part-of-speech tagging. We empirically analyze the diversity between two representative corpora, i.e. Penn Chinese Treebank (CTB) and PKU's People's Daily (PPD), on manually mapped data, and show that their linguistic annotations are systematically different and highly compatible. The analysis is further exploited to improve processing accuracy by (1) integrating systems that are respectively trained on heterogeneous annotations to reduce the approximation error, and (2) re-training models with high quality automatically converted data to reduce the estimation error. Evaluation on the CTB and PPD data shows that our novel model achieves a relative error reduction of 11% over the best reported result in the literature.

Capturing Paradigmatic and Syntagmatic Lexical Relations: Towards Accurate Chinese Part-of-Speech Tagging
Weiwei Sun and Hans Uszkoreit
Monday 2:30pm–3:00pm — 402 (ICC)

From the perspective of structural linguistics, we explore paradigmatic and syntagmatic lexical relations for Chinese POS tagging, an important and challenging task for Chinese language processing. Paradigmatic lexical relations are explicitly captured by word clustering on large-scale unlabeled data and are used to design new features to enhance a discriminative tagger. Syntagmatic lexical relations are implicitly captured by constituent parsing and are utilized via system combination. Experiments on the Penn Chinese Treebank demonstrate the importance of both paradigmatic and syntagmatic relations. Our linguistically motivated approaches yield a relative error reduction of 18% in total over a state-of-the-art baseline.

Fast Online Training with Frequency-Adaptive Learning Rates for Chinese Word Segmentation and New Word Detection
Xu Sun, Houfeng Wang, and Wenjie Li
Monday 3:00pm–3:30pm — 402 (ICC)

We present a joint model for Chinese word segmentation and new word detection. We present high dimensional new features, including word-based features and enriched edge (label-transition) features, for the joint modeling. As we know, training a word segmentation system on large-scale datasets is already costly. In our case, adding high dimensional new features will further slow down the training speed. To solve this problem, we propose a new training method, adaptive online gradient descent based on feature frequency information, for very fast online training of the parameters, even given large-scale datasets with high dimensional features. Compared with existing training methods, our training method is an order magnitude faster in terms of training time, and can achieve equal or even higher accuracies. The proposed fast training method is a general purpose optimization method, and it is not limited in the specific task discussed in this paper.

Verb Classification using Distributional Similarity in Syntactic and Semantic Structures
Danilo Croce, Alessandro Moschitti, Roberto Basili, and Martha Palmer
Monday 2:00pm–2:30pm — 401 (ICC)

In this paper, we propose innovative representations for automatic classification of verbs according to mainstream linguistic theories, namely VerbNet and FrameNet. First, syntactic and semantic structures capturing essential lexical and syntactic properties of verbs are defined. Then, we design advanced similarity functions between such structures, i.e., semantic tree kernel functions, for exploiting distributional and grammatical information in Support Vector Machines. The extensive empirical analysis on VerbNet class and frame detection shows that our models capture meaningful syntactic/semantic structures, which allows for improving the state-of-the-art.

Word Sense Disambiguation Improves Information Retrieval
Zhi Zhong and Hwee Tou Ng
Monday 2:30pm–3:00pm — 401 (ICC)

Previous research has conflicting conclusions on whether word sense disambiguation (WSD) systems can improve information retrieval (IR). In this paper, we propose a method to estimate sense distributions for short queries. Together with the senses predicted for words in documents, we incorporate word senses into the language modeling approach to IR and also exploit the integration of synonym relations. Our experimental results on TREC collections show that, using the word senses tagging by a supervised WSD system, we can obtain significant improvements over a state-of-the-art IR system.

Efficient Search for Transformation-based Inference
Asher Stern, Roni Stern, Ido Dagan, and Ariel Felner
Monday 3:00pm–3:30pm — 401 (ICC)

This paper addresses the search problem in textual inference, where systems need to infer one piece of text from another. A prominent approach to this task is attempts to transform one text into the other through a sequence of inference-preserving transformations, a.k.a. a proof, while estimating the proof's validity. This raises a search challenge of finding the best possible proof. We explore this challenge through a comprehensive investigation of prominent search algorithms and propose two novel algorithmic components specifically designed for textual inference: a gradient-style evaluation function, and a local-lookahead node expansion method. Evaluations, using the open-source system, BIUTEE, show the contribution of these ideas to search efficiency and proof quality.

Maximum Expected BLEU Training of Phrase and Lexicon Translation Models
Xiaodong He and Li Deng
Monday 4:00pm–4:30pm — Halla A (ICC)

This paper proposes a new discriminative training method in constructing phrase and lexicon translation models. In order to reliably learn a myriad of parameters in these models, we propose an expected BLEU score-based utility function with KL regularization as the objective, and train the models on a large parallel dataset. For training, we derive growth transformations for phrase and lexicon translation probabilities to iteratively improve the objective. The proposed method, evaluated on the Europarl German-to-English dataset, leads to a 1.1 BLEU point improvement over a state-of-the-art baseline translation system. In IWSLT 2011 Benchmark, our system using the proposed method achieves the best Chinese-to-English translation result on the task of translating TED talks.

Learning Translation Consensus with Structured Label Propagation
Shujie Liu, Chi-Ho Li, Mu Li, and Ming Zhou
Monday 4:30pm–5:00pm — Halla A (ICC)

In this paper, we address the issue for learning better translation consensus in machine translation (MT) research, and explore the search of translation consensus from similar, rather than the same,

source sentences or their spans. Unlike previous work on this topic, we formulate the problem as structured labeling over a much smaller graph, and we propose a novel structured label propagation for the task. We convert such graph-based translation consensus from similar source strings into useful features both for n-best output re-ranking and for decoding algorithm. Experimental results show that, our method can significantly improve machine translation performance on both IWSLT and NIST data, compared with a state-of-the-art baseline.

Smaller Alignment Models for Better Translations: Unsupervised Word Alignment with the l0-norm

Ashish Vaswani, Liang Huang, and David Chiang

Monday 5:00pm–5:30pm — Halla A (ICC)

Two decades after their invention, the IBM word-based translation models, widely available in the GIZA++ toolkit, remain the dominant approach to word alignment and an integral part of many statistical translation systems. Although many models have surpassed them in accuracy, none have supplanted them in practice. In this paper, we propose a simple extension to the IBM models: an l0 prior to encourage sparsity in the word-to-word translation model. We explain how to implement this extension efficiently for large-scale data (also released as a modification to GIZA++) and demonstrate, in experiments on Czech, Arabic, Chinese, and Urdu to English translation, significant improvements over IBM Model 4 in both word alignment (up to +6.7 F1) and translation quality (up to +1.4 Bleu).

Modeling Review Comments

Arjun Mukherjee and Bing Liu

Monday 4:00pm–4:30pm — Halla B (ICC)

Writing comments about news articles, blogs, or reviews have become a popular activity in social media. In this paper, we analyze reader comments about reviews. Analyzing review comments is important because reviews only tell the experiences and evaluations of reviewers about the reviewed products or services. Comments, on the other hand, are readers' evaluations of reviews, their questions and concerns. Clearly, the information in comments is valuable for both future readers and brands. This paper proposes two latent variable models to simultaneously model and extract these key pieces of information. The results also enable classification of comments accurately. Experiments using Amazon review comments demonstrate the effectiveness of the proposed models.

A Joint Model for Discovery of Aspects in Utterances

Asli Celikyilmaz and Dilek Hakkani-Tur

Monday 4:30pm–5:00pm — Halla B (ICC)

We describe a joint model for understanding user actions in natural language utterances. Our multi-layer generative approach uses both labeled and unlabeled utterances to jointly learn aspects regarding utterance's target domain (e.g. movies), intention (e.g., finding a movie) along with other semantic units (e.g., movie name). We inject information extracted from unstructured web search query logs as prior information to enhance the generative process of the natural language utterance understanding model. Using utterances from five domains, our approach shows up to 4.5% improvement on domain and dialog act performance over cascaded approach in which each semantic component is learned sequentially and a supervised joint learning model (which requires fully labeled data).

Aspect Extraction through Semi-Supervised Modeling

Arjun Mukherjee and Bing Liu

Monday 5:00pm–5:30pm — Halla B (ICC)

Aspect extraction is a central problem in sentiment analysis. Current methods either extract aspects without categorizing them, or extract and categorize them using unsupervised topic mod-

eling. By categorizing, we mean the synonymous aspects should be clustered into the same category. In this paper, we solve the problem in a different setting where the user provides some seed words for a few aspect categories and the model extracts and clusters aspect terms into categories simultaneously. This setting is important because categorizing aspects is a subjective task. For different application purposes, different categorizations may be needed. Some form of user guidance is desired. In this paper, we propose two statistical models to solve this seeded problem, which aim to discover exactly what the user wants. Our experimental results show that the two proposed models are indeed able to perform the task effectively.

Learning to "Read Between the Lines" using Bayesian Logic Programs
Sindhu Raghavan, Raymond Mooney, and Hyeonseo Ku
Monday 4:00pm–4:30pm — Samda (ICC)

Most information extraction (IE) systems identify facts that are explicitly stated in text. However, in natural language, some facts are implicit, and identifying them requires "reading between the lines". Human readers naturally use common sense knowledge to infer such implicit information from the explicitly stated facts. We propose an approach that uses Bayesian Logic Programs (BLPs), a statistical relational model combining first-order logic and Bayesian networks, to infer additional implicit information from extracted facts. It involves learning uncertain commonsense knowledge (in the form of probabilistic first-order rules) from natural language text by mining a large corpus of automatically extracted facts. These rules are then used to derive additional facts from extracted information using BLP inference. Experimental evaluation on a benchmark data set for machine reading demonstrates the efficacy of our approach.

Collective Generation of Natural Image Descriptions
Polina Kuznetsova, Vicente Ordonez, Alexander Berg, Tamara Berg, and Yejin Choi
Monday 4:30pm–5:00pm — Samda (ICC)

We present a holistic data-driven approach to image description generation, exploiting the vast amount of (noisy) parallel image data and associated natural language descriptions available on the web. More specifically, given a query image, we retrieve existing human-composed phrases used to describe visually similar images, then selectively combine those phrases to generate a novel description for the query image. We cast the generation process as constraint optimization problems, collectively incorporating multiple interconnected aspects of language composition for content planning, surface realization and discourse structure. Evaluation by human annotators indicates that our final system generates more semantically correct and linguistically appealing descriptions than two nontrivial baselines.

Concept-to-text Generation via Discriminative Reranking
Ioannis Konstas and Mirella Lapata
Monday 5:00pm–5:30pm — Samda (ICC)

This paper proposes a data-driven method for concept-to-text generation, the task of automatically producing textual output from non-linguistic input. A key insight in our approach is to reduce the tasks of content selection ("what to say") and surface realization ("how to say") into a common parsing problem. We define a probabilistic context-free grammar that describes the structure of the input (a corpus of database records and text describing some of them) and represent it compactly as a weighted hypergraph. The hypergraph structure encodes exponentially many derivations, which we rerank discriminatively using local and global features. We propose a novel decoding algorithm for finding the best scoring derivation and generating in this setting. Experimental evaluation on the ATIS domain shows that our model outperforms a competitive discriminative system both using BLEU and in a judgment elicitation study.

A Discriminative Hierarchical Model for Fast Coreference at Large Scale
Michael Wick, Sameer Singh, and Andrew McCallum
Monday 4:00pm–4:30pm — 402 (ICC)

Methods that measure compatibility between mention pairs are currently the dominant approach to coreference. However, they suffer from a number of drawbacks including difficulties scaling to large numbers of mentions and limited representational power. As these drawbacks become increasingly restrictive, the need to replace the pairwise approaches with a more expressive, highly scalable alternative is becoming urgent. In this paper we propose a novel discriminative hierarchical model that recursively partitions entities into trees of latent sub-entities. These trees succinctly summarize the mentions providing a highly compact, information-rich structure for reasoning about entities and coreference uncertainty at massive scales. We demonstrate that the hierarchical model is several orders of magnitude faster than pairwise, allowing us to perform coreference on six million author mentions in under four hours on a single CPU.

Coreference Semantics from Web Features
Mohit Bansal and Dan Klein
Monday 4:30pm–5:00pm — 402 (ICC)

To address semantic ambiguities in coreference resolution, we use Web n-gram features that capture a range of world knowledge in a diffuse but robust way. Specifically, we exploit short-distance cues to hypernymy, semantic compatibility, and semantic context, as well as general lexical co-occurrence. When added to a state-of-the-art coreference baseline, our Web features give significant gains on multiple datasets (ACE 2004 and ACE 2005) and metrics (MUC and B3), resulting in the best results reported to date for the end-to-end task of coreference resolution.

Subgroup Detection in Ideological Discussions
Amjad Abu-Jbara, Pradeep Dasigi, Mona Diab, and Dragomir Radev
Monday 5:00pm–5:30pm — 402 (ICC)

The rapid and continuous growth of social networking sites has led to the emergence of many communities. Many of these groups discuss ideological and political topics. It is not uncommon that the participants in such discussions split into two or more subgroups. The members of each subgroup share the same opinion toward the discussion topic and are more likely to agree with members of the same subgroup and disagree with members from opposing subgroups. In this paper, we propose an unsupervised approach for automatically detecting discussant subgroups in online communities. We analyze the text exchanged between the participants of a discussion to identify the attitude they carry not only towards each other but also towards the various aspects of the discussion topic. We use attitude predictions to construct an attitude vector for each discussant. We use clustering techniques to group these vectors in order to determine the subgroup membership of each participant. We compare our methods to text clustering and other baselines, and show that our method achieves promising results.

Cross-Domain Co-Extraction of Sentiment and Topic Lexicons
Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang, and Xiaoyan Zhu
Monday 4:00pm–4:30pm — 401 (ICC)

Extracting sentiment and topic lexicons is important for opinion mining. Previous works have showed that supervised learning methods are superior for this task. However, the performance of supervised methods highly relies on manually labeled training data. In this paper, we propose a domain adaptation framework for sentiment- and topic- lexicon co-extraction in a domain of interest where we do not require any labeled data, but have lots of labeled data in another related domain. The framework is twofold. In the first step, we generate a few high-confidence sentiment and topic seeds in the target domain. In the second step, we propose a novel Relational Adaptive bootstraPping (RAP) algorithm to expand the seeds in the target domain by exploiting the labeled source domain data and the relationships between topic and sentiment words. Experimental results show that our domain adaptation framework can extract precise lexicons in the target domain without any annotation.

Learning Syntactic Verb Frames using Graphical Models
Thomas Lippincott, Anna Korhonen, and Diarmuid Ó Séaghdha
Monday 4:30pm–5:00pm — 401 (ICC)

We present a novel approach for building verb subcategorization lexicons using a simple graphical model. In contrast to previous methods, we show how the model can be trained without parsed input or a predefined subcategorization frame inventory. Our method outperforms the state-of-the-art on a verb clustering task, and is easily trained on arbitrary domains. This quantitative evaluation is complemented by a qualitative discussion of verbs and their frames. We discuss the advantages of graphical models for this task, in particular the ease of integrating semantic information about verbs and arguments in a principled fashion. We conclude with future work to augment the approach.

Fast Online Lexicon Learning for Grounded Language Acquisition
David Chen
Monday 5:00pm–5:30pm — 401 (ICC)

Learning a semantic lexicon is often an important first step in building a system that learns to interpret the meaning of natural language. It is especially important in language grounding where the training data usually consist of language paired with an ambiguous perceptual context. Recent work by Chen and Mooney (2011) introduced a lexicon learning method that deals with ambiguous relational data by taking intersections of graphs. While the algorithm produced good lexicons for the task of learning to interpret navigation instructions, it only works in batch settings and does not scale well to large datasets. In this paper we introduce a new online algorithm that is an order of magnitude faster and surpasses the state-of-the-art results. We show that by changing the grammar of the formal meaning representation language and training on additional data collected from Amazon's Mechanical Turk we can further improve the results. We also include experimental results on a Chinese translation of the training data to demonstrate the generality of our approach.

Short Paper Abstracts

Movie-DiC: a Movie Dialogue Corpus for Research and Development

Rafael E. Banchs

Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper describes Movie-DiC a Movie Dialogue Corpus recently collected for research and development purposes. The collected dataset comprises 132,229 dialogues containing a total of 764,146 turns that have been extracted from 753 movies. Details on how the data collection has been created and how it is structured are provided along with its main statistics and characteristics.

Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions

Elena Cabrio and Serena Villata

Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Blogs and forums are widely adopted by online communities to debate about various issues. However, a user that wants to cut in on a debate may experience some difficulties in extracting the current accepted positions, and can be discouraged from interacting through these applications. In our paper, we combine textual entailment with argumentation theory to automatically extract the arguments from debates and to evaluate their acceptability.

Towards the Unsupervised Acquisition of Discourse Relations

Christian Chiarcos

Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper describes a novel approach towards the empirical approximation of discourse relations between different utterances in texts. Following the idea that every pair of events comes with preferences regarding the range and frequency of discourse relations connecting both parts, the paper investigates whether these preferences are manifested in the distribution of relation words (that serve to signal these relations).

The paper presents experiments on two large-scale English web corpora, showing that significant correlations between pairs of adjacent events and relation words exist, that they are reproducible on different data sets, and for three relation words, that their distribution corresponds to theory-based assumptions.

These findings serve to support the methodological validity of the approach. On this basis, subsequent research will address the relationship between relation words and discourse relations.

Arabic Retrieval Revisited: Morphological Hole Filling

Kareem Darwish and Ahmed Ali

Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Due to Arabic's morphological complexity, Arabic retrieval benefits greatly from morphological analysis — particularly stemming. However, the best known stemming does not handle linguistic phenomena such as broken plurals and malformed stems. In this paper we propose a model of character-level morphological transformation that is trained using Wikipedia hypertext to page title links. The use of our model yields statistically significant improvements in Arabic retrieval over the use of the best statistical stemming technique. The technique can potentially be applied to other languages.

Extracting and modeling durations for habits and events from Twitter

Jennifer Williams and Graham Katz

Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We seek to automatically estimate typical durations for events and habits described in Twitter tweets. A corpus of more than 14 million tweets containing temporal duration information was

collected. These tweets were classified as to their habituality status using a bootstrapped, decision tree. For each verb lemma, associated duration information was collected for episodic and habitual uses of the verb. Summary statistics for 483 verb lemmas and their typical habit and episode durations has been compiled and made available. This automatically generated duration information is broadly comparable to hand-annotation.

Event Linking: Grounding Event Reference in a News Archive
Joel Nothman, Matthew Honnibal, Ben Hachey, and James R. Curran
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Interpreting news requires identifying its constituent events. Events are complex linguistically and ontologically, so disambiguating their reference is challenging. We introduce event linking, which canonically labels an event reference with the article where it was first reported. This implicitly relaxes coreference to co-reporting, and will practically enable augmenting news archives with semantic hyperlinks. We annotate and analyse a corpus of 150 documents, extracting 501 links to a news archive with reasonable inter-annotator agreement.

Coupling Label Propagation and Constraints for Temporal Fact Extraction
Yafang Wang, Maximilian Dylla, Marc Spaniol, and Gerhard Weikum
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

The Web and digitized text sources contain a wealth of information about named entities such as politicians, actors, companies, or cultural landmarks. Extracting this information has enabled the automated construction of large knowledge bases, containing hundred millions of binary relationships or attribute values about these named entities. However, in reality most knowledge is transient, i.e. changes over time, requiring a temporal dimension in fact extraction. In this paper we develop a methodology that combines label propagation with constraint reasoning for temporal fact extraction. Label propagation aggressively gathers fact candidates, and an Integer Linear Program is used to clean out false hypotheses that violate temporal constraints. Our method is able to improve on recall while keeping up with precision, which we demonstrate by experiments with biography-style Wikipedia pages and a large corpus of news articles.

Using Search-Logs to Improve Query Tagging
Kuzman Ganchev, Keith Hall, Ryan McDonald, and Slav Petrov
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Syntactic analysis of search queries is important for a variety of information-retrieval tasks; however, the lack of annotated data makes training query analysis models difficult. We propose a simple, efficient procedure in which part-of-speech tags are transferred from retrieval-result snippets to queries at training time. Unlike previous work, our final model does not require any additional resources at run-time. Compared to a state-of-the-art approach, we achieve more than 20% relative error reduction. Additionally, we annotate a corpus of search queries with part-of-speech tags, providing a resource for future work on syntactic query analysis.

Toward Automatically Assembling Hittite-Language Cuneiform Tablet Fragments into Larger Texts
Stephen Tyndall
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper presents the problem within Hittite and Ancient Near Eastern studies of fragmented and damaged cuneiform texts, and proposes to use well-known text classification metrics, in combination with some facts about the structure of Hittite-language cuneiform texts, to help classify a number of fragments of clay cuneiform-script tablets into more complete texts. In particular, I propose using Sumerian and Akkadian ideogrammatic signs within Hittite texts to improve the performance of Naive Bayes and Maximum Entropy classifiers. The performance in some cases is improved, and in some cases very much not, suggesting that the variable frequency

of occurrence of these ideograms in individual fragments makes considerable difference in the ideal choice for a classification method. Further, complexities of the writing system and the digital availability of Hittite texts complicate the problem.

A Corpus of Textual Revisions in Second Language Writing
John Lee and Jonathan Webster
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper describes the creation of the first large-scale corpus containing drafts and final versions of essays written by non-native speakers, with the sentences aligned across different versions. Furthermore, the sentences in the drafts are annotated with comments from teachers. The corpus is intended to support research on textual revision by language learners, and how it is influenced by feedback. This corpus has been converted into an XML format conforming to the standards of the Text Encoding Initiative (TEI).

Coarse Lexical Semantic Annotation with Supersenses: An Arabic Case Study
Nathan Schneider, Behrang Mohit, Kemal Oflazer, and Noah A. Smith
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

"Lightweight" semantic annotation of text calls for a simple representation, ideally without requiring a semantic lexicon to achieve good coverage in the language and domain. In this paper, we repurpose WordNet's supersense tags for annotation, developing specific guidelines for nominal expressions and applying them to Arabic Wikipedia articles in four topical domains. The resulting corpus has high coverage and was completed quickly with reasonable inter-annotator agreement.

Word Epoch Disambiguation: Finding How Words Change Over Time
Rada Mihalcea and Vivi Nastase
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

In this paper we introduce the novel task of "word epoch disambiguation," defined as the problem of identifying changes in word usage over time. Through experiments run using word usage examples collected from three major periods of time (1800, 1900, 2000), we show that the task is feasible, and significant differences can be observed between occurrences of words in different periods of time.

Authorship Attribution with Author-aware Topic Models
Yanir Seroussi, Fabian Bohnert, and Ingrid Zukerman
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Authorship attribution deals with identifying the authors of anonymous texts. Building on our earlier finding that the Latent Dirichlet Allocation (LDA) topic model can be used to improve authorship attribution accuracy, we show that employing a previously-suggested Author-Topic (AT) model outperforms LDA when applied to scenarios with many authors. In addition, we define a model that combines LDA and AT by representing authors and documents over two disjoint topic sets, and show that our model outperforms LDA, AT and support vector machines on datasets with many authors.

Information-theoretic Multi-view Domain Adaptation
Pei Yang, Wei Gao, Qi Tan, and Kam-Fai Wong
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We use multiple views for cross-domain document classification. The main idea is to strengthen the views' consistency for target data with source training data by identifying the correlations of domain-specific features from different domains. We present an Information-theoretic Multi-view Adaptation Model (IMAM) based on a multi-way clustering scheme, where word and link clusters can draw together seemingly unrelated domain-specific features from both sides and

iteratively boost the consistency between document clusterings based on word and link views. Experiments show that IMAM significantly outperforms state-of-the-art baselines.

Efficient Tree-Based Topic Modeling
Yuening Hu and Jordan Boyd-Graber
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Topic modeling with a tree-based prior has been used for a variety of applications because it can encode correlations between words that traditional topic modeling cannot. However, its expressive power comes at the cost of more complicated inference. In this paper, we extend the SparseLDA (Yao et al., 2009) inference scheme for latent Dirichlet allocation (LDA) to tree-based topic models. This sampling scheme computes the exact conditional distribution for Gibbs sampling much more quickly than enumerating all possible latent variable assignments. We further improve performance by iteratively refining the sampling distribution only when needed. Experiments show that the proposed techniques dramatically improve the computation time.

Learning Better Rule Extraction with Translation Span Alignment
Jingbo Zhu, Tong Xiao, and Chunliang Zhang
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

To address a practical issue that even one spurious word alignment link can rule out some useful rules for syntax-based translation, we propose an unsupervised approach to learning translation span alignments from parallel data, and improves syntactic rule extraction by deleting spurious word alignment links and adding new valuable links based on bilingual translation span correspondences. Experiments on Chinese-English translation demonstrate improvements over standard methods for tree-to-string and tree-to-tree translation.

Enhancing Statistical Machine Translation with Character Alignment
Ning Xi, Guangchao Tang, Xinyu Dai, Shujian Huang, and Jiajun Chen
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

The dominant practice of statistical machine translation (SMT) uses the same Chinese word segmentation specification in both alignment and translation rule induction steps in building Chinese-English SMT system, which may suffer from a suboptimal problem that word segmentation better for alignment is not necessarily better for translation. To tackle this, we propose a framework that uses two different segmenta-tion specifications for alignment and translation respectively: we use Chinese character as the basic unit for alignment, and then convert this alignment to conventional word alignment for translation rule induction. Experimentally, our approach outperformed two baselines: fully word-based system (using word for both alignment and translation) and fully charac-ter-based system, in terms of alignment quality and translation performance.

Translation Model Size Reduction for Hierarchical Phrase-based Statistical Machine Translation
Seung-Wook Lee, Dongdong Zhang, Mu Li, Ming Zhou, and Hae-Chang Rim
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

In this paper, we propose a novel method of reducing the size of translation model for hierarchical phrase-based machine translation systems. Previous approaches try to prune infrequent entries or unreliable entries based on statistics, but cause a problem of reducing the translation coverage. On the contrary, the proposed method try to prune only ineffective entries based on the estimation of the information redundancy encoded in phrase pairs and hierarchical rules, and thus preserve the search space of SMT decoders as much as possible. Experimental results on Chinese-to-English machine translation tasks show that our method is able to reduce almost the half size of the translation model with very tiny degradation of translation performance.

Heuristic Cube Pruning in Linear Time
Andrea Gesmundo, Giorgio Satta, and James Henderson
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We propose a novel approximation algorithm for Cube Pruning running in linear time in the beam size. In an empirical setting, we show a gain in running time of a standard machine translation system, at a small loss in accuracy.

Combining Word-Level and Character-Level Models for Machine Translation Between Closely-Related Languages
Preslav Nakov and Jörg Tiedemann
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We propose several techniques for improving statistical machine translation between closely-related languages with scarce resources. We use character-level translation trained on n-gram-character-aligned bitexts and tuned using word-level BLEU, which we further augment with character-based transliteration at the word level and combine with a word-level translation model. The evaluation on Macedonian-Bulgarian movie subtitles shows an improvement of 2.84 BLEU points over a phrase-based word-level baseline.

Improving the IBM Alignment Models Using Variational Bayes
Darcey Riley and Daniel Gildea
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Bayesian approaches have been shown to reduce the amount of overfitting that occurs when running the EM algorithm, by placing prior probabilities on the model parameters. We apply one such Bayesian technique, variational Bayes, to the IBM models of word alignment for statistical machine translation. We show that using variational Bayes improves the performance of the widely used GIZA++ software, as well as improving the overall performance of the Moses machine translation system in terms of BLEU score.

Post-ordering by Parsing for Japanese-English Statistical Machine Translation
Isao Goto, Masao Utiyama, and Eiichiro Sumita
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Reordering is a difficult task in translating between widely different languages such as Japanese and English. We employ the post-ordering framework proposed by (Sudoh et al., 2011b) for Japanese to English translation and improve upon the reordering method. The existing post-ordering method reorders a sequence of target language words in a source language word order via SMT, while our method reorders the sequence by: 1) parsing the sequence to obtain syntax structures similar to a source language structure, and 2) transferring the obtained syntax structures into the syntax structures of the target language.

An Exploration of Forest-to-String Translation: Does Translation Help or Hurt Parsing?
Hui Zhang and David Chiang
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Syntax-based translation models that operate on the output of a source-language parser have been shown to perform better if allowed to choose from a set of possible parses. In this paper, we investigate whether this is because it allows the translation stage to overcome parser errors or to override the syntactic structure itself. We find that it is primarily the latter, but that under the right conditions, the translation stage does correct parser errors, improving parsing accuracy on the Chinese Treebank.

Unsupervised Morphology Rivals Supervised Morphology for Arabic MT
David Stallard, Jacob Devlin, Michael Kayser, Yoong Keok Lee, and Regina Barzilay
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

If unsupervised morphological analyzers could approach the effectiveness of supervised ones, they would be a very attractive choice for improving MT performance on low-resource inflected languages. In this paper, we compare performance gains for state-of-the-art supervised vs. unsupervised morphological analyzers, using a state-of-the-art Arabic-to-English MT system. We apply maximum marginal decoding to the unsupervised analyzer, and show that this yields the best published segmentation accuracy for Arabic, while also making segmentation output more stable. Our approach gives an 18% relative BLEU gain for Levantine dialectal Arabic. Furthermore, it gives higher gains for Modern Standard Arabic (MSA), as measured on NIST MT-08, than does MADA, a leading *supervised* MSA segmenter.

A Meta Learning Approach to Grammatical Error Correction
Hongsuck Seo, Jonghoon Lee, Seokhwan Kim, Kyusong Lee, Sechun Kang, and Gary Geunbae Lee
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We introduce a novel method for grammatical error correction with a number of small corpora. To make the best use of several corpora with different characteristics, we employ a meta-learning with several base classifiers trained on different corpora. This research focuses on a grammatical error correction task for article errors. A series of experiments is presented to show the effectiveness of the proposed approach on two different grammatical error tagged corpora.

Fine Granular Aspect Analysis using Latent Structural Models
Lei Fang and Minlie Huang
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

In this paper, we present a structural learning model for joint sentiment classification and aspect analysis of text at various levels of granularity. Our model aims to identify highly informative sentences that are aspect-specific in online custom reviews. The primary advantages of our model are two-fold: first, it performs document-level and sentence-level sentiment polarity classification jointly; second, it is able to find informative sentences that are closely related to some respects in a review, which may be helpful for aspect-level sentiment analysis such as aspect-oriented summarization. The proposed method was evaluated with 9,000 Chinese restaurant reviews. Preliminary experiments demonstrate that our model obtains promising performance.

Identifying High-Impact Sub-Structures for Convolution Kernels in Document-level Sentiment Classification
Zhaopeng Tu, Yifan He, Jennifer Foster, Josef van Genabith, Qun Liu, and Shouxun Lin
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

Convolution kernels support the modeling of complex syntactic information in machine-learning tasks. However, such models are highly sensitive to the type and size of syntactic structure used. It is therefore an important challenge to automatically identify high impact sub-structures relevant to a given task. In this paper we present a systematic study investigating (combinations of) sequence and convolution kernels using different types of sub-structures in document-level sentiment classification. We show that minimal sub-structures extracted from constituency and dependency trees guided by a polarity lexicon show 1.45 point absolute improvement in accuracy over a bag-of-words classifier on a widely used sentiment corpus.

Exploiting Latent Information to Predict Diffusions of Novel Topics on Social Networks
Tsung-Ting Kuo, San-Chuan Hung, Wei-Shih Lin, Nanyun Peng, Shou-De Lin, and Wei-Fen Lin
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper brings a marriage of two seemly unrelated topics, natural language processing (NLP) and social network analysis (SNA). We propose a new task in SNA which is to predict the dif-

fusion of a new topic, and design a learning-based framework to solve this problem. We exploit the latent semantic information among users, topics, and social connections as features for prediction. Our framework is evaluated on real data collected from public domain. The experiments show 18% AUC improvement over baseline methods. The source code and dataset are available at http://www.csie.ntu.edu.tw/ d97944007/diffusion/

Sentence Compression with Semantic Role Constraints
Katsumasa Yoshikawa, Ryu Iida, Tsutomu Hirao, and Manabu Okumura
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

For sentence compression, we propose new semantic constraints to directly capture the relations between a predicate and its arguments, whereas the existing approaches have focused on relatively shallow linguistic properties, such as lexical and syntactic information. These constraints are based on semantic roles and superior to the constraints of syntactic dependencies. Our empirical evaluation on the Written News Compression Corpus (Clarke and Lapata 2008) demonstrates that our system achieves state-of-the-art results.

Fully Abstractive Approach to Guided Summarization
Pierre-Etienne Genest and Guy Lapalme
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper shows that full abstraction can be accomplished in the context of guided summarization. We describe a work in progress that relies on Information Extraction, statistical content selection and Natural Language Generation. Early results already demonstrate the effectiveness of the approach.

Assessing the Effect of Inconsistent Assessors on Summarization Evaluation
Karolina Owczarzak, Peter A. Rankel, Hoa Trang Dang, and John M. Conroy
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We investigate the consistency of human assessors involved in summarization evaluation to understand its effect on system ranking and automatic evaluation techniques. Using Text Analysis Conference data, we measure annotator consistency based on human scoring of summaries for Responsiveness, Readability, and Pyramid scoring. We identify inconsistencies in the data and measure to what extent these inconsistencies affect the ranking of machine summarization systems. Finally, we examine the stability of automatic metrics (ROUGE and CLASSY) with respect to the inconsistent marking.

Fast and Robust Part-of-Speech Tagging Using Dynamic Model Selection
Jinho D. Choi and Martha Palmer
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper presents a novel way of improving POS tagging on heterogeneous data. First, two separate models are trained (generalized and domain-specific) from the same data set by controlling lexical items with different document frequencies. During decoding, one of the models is selected dynamically given the cosine similarity between each sentence and the training data. This dynamic model selection approach, coupled with a one-pass, left-to-right POS tagging algorithm, is evaluated on corpora from seven different genres. Even with this simple tagging algorithm, our system shows comparable results against other state-of-the-art systems, and gives higher accuracies when evaluated on a mixture of the data. Furthermore, our system is able to tag about 31,914 tokens per second. We believe that this model selection approach can be applied to more sophisticated tagging algorithms and improve their robustness even further.

Lemmatisation as a Tagging Task
Andrea Gesmundo and Tanja Samardzic
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We present a novel approach to the task of word lemmatisation. We formalise lemmatisation as a category tagging task, by describing how a word-to-lemma transformation rule can be encoded in a single label and how a set of such labels can be inferred for a specific language. In this way, a lemmatisation system can be trained and tested using any supervised tagging model. In contrast to previous approaches, the proposed technique allows us to easily integrate relevant contextual information. We test our approach on eight languages reaching a new state-of-the-art level for the lemmatisation task.

How Are Spelling Errors Generated and Corrected? A Study of Corrected and Uncorrected Spelling Errors Using Keystroke Logs
Yukino Baba and Hisami Suzuki
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper presents a comparative study of spelling errors that are corrected as you type, vs. those that remain uncorrected. First, we generate naturally occurring online error correction data by logging users' keystrokes, and by automatically deriving pre- and post-correction strings from them. We then perform an analysis of this data against the errors that remain in the final text as well as across languages. Our analysis shows a clear distinction between the types of errors that are generated and those that remain uncorrected, as well as across languages.

Tokenization: Returning to a Long Solved Problem — A Survey, Contrastive Experiment, Recommendations, and Toolkit —
Rebecca Dridan and Stephan Oepen
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

We examine some of the frequently disregarded subtleties of tokenization in Penn Treebank style, and present a new rule-based pre-processing toolkit that not only reproduces the Treebank tokenization with unmatched accuracy, but also maintains exact stand-off pointers to the original text and allows flexible configuration to diverse use cases (eg to genre- or domain-specific idiosyncrasies).

Unsupervized Word Segmentation: the Case for Mandarin Chinese
Pierre Magistry and Benoît Sagot
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

In this paper, we present an unsupervised segmentation system tested on Mandarin Chinese. Following Harris Hypothesis in Kempe (1999) and Tanaka-Ishii's (2005) reformulation, we base our work on the Variation of Branching Entropy. We improve on (2006) by adding normalization and viterbi-decoding. This enable us to remove most of the thresholds and parameters from their model and to reach near state-of-the-art results (Wang et al., 2011) with a simpler system. We provide evaluation on different corpora available from the Segmentation bake-off II (Emerson, 2005) and define a more precise topline for the task using cross-trained supervised system available off-the-shelf (Zhang and Clark, 2010; Zhao and Kit, 2008; Huang and Zhao, 2007)

Grammar Error Correction Using Pseudo-Error Sentences and Domain Adaptation
Kenji Imamura, Kuniko Saito, Kugatsu Sadamitsu, and Hitoshi Nishikawa
Monday 6:00pm–8:30pm — 3F Lobby (ICC)

This paper presents grammar error correction for Japanese particles that uses discriminative sequence conversion, which corrects erroneous particles by substitution, insertion, and deletion. The error correction task is hindered by the difficulty of collecting large error corpora. We tackle this problem by using pseudo-error sentences generated automatically. Furthermore, we apply domain adaptation, the pseudo-error sentences are from the source domain, and the real-error sentences are from the target domain. Experiments show that stable improvement is achieved by

using domain adaptation.

Abstracts from the Student Research Workshop

A Broad Evaluation of Techniques for Automatic Acquisition of Multiword Expressions

Carlos Ramisch, Vitor De Araujo, and Aline Villavicencio

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Several approaches have been proposed for the automatic acquisition of multiword expressions from corpora. However, there is no agreement about which of them presents the best cost-benefit ratio, as they have been evaluated on distinct datasets and/or languages. To address this issue, we investigate these techniques analysing the following dimensions: expression type (compound nouns, phrasal verbs), language (English, French) and corpus size. Results show that these techniques tend to extract similar candidate lists with high recall ($\sim$ 80%) for nominals and high precision ($\sim$ 70%) for verbals. The use of association measures for candidate filtering is useful but some of them are more onerous and not significantly better than raw counts. We finish with an evaluation of flexibility and an indication of which technique is recommended for each language-type-size context.

Detecting Power Relations from Written Dialog

Vinodkumar Prabhakaran

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

In my thesis I propose a data-oriented study on how social power relations between participants manifest in the language and structure of online written dialogs. I propose that there are different types of power relations and they are different in the ways they are expressed and revealed in dialog and across different languages, genres and domains. So far, I have defined four types of power and annotated them in corporate email threads in English and found support that they in fact manifest differently in the threads. Using dialog and language features, I have built a system to predict participants possessing these types of power within email threads. I intend to extend this system to other languages, genres and domains and to improve it's performance using deeper linguistic analysis.

Active Learning with Transfer Learning

Chunyong Luo, Yangsheng Ji, Xinyu Dai, and Jiajun Chen

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

In sentiment classification, unlabeled user reviews are often free to collect for new products, while sentiment labels are rare. In this case, active learning is often applied to build a high-quality classifier with as small amount of labeled instances as possible. However, when the labeled instances are insufficient, the performance of active learning is limited. In this paper, we aim at enhancing active learning by employing the labeled reviews from a different but related (source) domain. We propose a framework Active Vector Rotation (AVR), which adaptively utilizes the source domain data in the active learning procedure. Thus, AVR gets benefits from souce domain when it is helpful, and avoids the negative affects when it is harmful. Extensive experiments on toy data and review texts show our success, compared with other state-of-the-art active learning approaches, as well as approaches with domain adaptation.

Query classification using topic models and support vector machine

Dieu-Thu Le and Raffaella Bernardi

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper describes a query classification system for a specialized domain. We take as a case study queries asked to a search engine of an art, cultural and history library and classify them against the library cataloguing categories. We show how click-through links, i.e., the links that a user clicks after submitting a query, can be exploited for extracting information useful to enrich the query as well as for creating the training set for a machine learning based classifier. Moreover, we show how Topic Model can be exploited to further enrich the query with hidden topics induced from the library meta-data. The experimental evaluations show that this system

considerably outperforms a matching and ranking classification approach, where queries (and categories) were also enriched with similar information.

Evaluating Unsupervised Ensembles when applied to Word Sense Induction
Keith Stevens
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Ensembles combine together knowledge from diverse machine learning approaches into a general flexible system. While supervised ensembles frequently show great benefit, unsupervised ensembles prove to be more challenging. We propose evaluating the value of various unsupervised ensembles when applied to the unsupervised task of Word Sense Induction with a framework for combining diverse feature spaces and clustering algorithms. We evaluate our system using standard shared tasks and also introduce new automated semantic evaluations and supervised baselines, both of which highlight the current limitations of existing Word Sense Induction evaluations.

Topic Extraction based on Prior Knowledge obtained from Target Documents
Kayo Tatsukawa
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper investigates the relation between prior knowledge and latent topic classification. There are many cases where the topic classification done by Latent Dirichlet Allocation results in the different classification that humans expects. To improve this problem, several studies using Dirichlet Forest prior instead of Dirichlet distribution have been studied in order to provide constraints on words so as they are classified into the same or not the same topics. However, in many cases, the prior knowledge is constructed from a subjective view of humans, but is not constructed based on the properties of objective documents. In this study, we construct prior knowledge based on the words extracted from objective documents and provide it as constraints for topic classification. We discuss the result of topic classification with the constraints.

TopicTiling: A Text Segmentation Algorithm based on LDA
Martin Riedl and Chris Biemann
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This work presents a Text Segmentation algorithm called TopicTiling. This algorithm is based on the well-known TextTiling algorithm, and segments documents using the Latent Dirichlet Allocation (LDA) topic model. We show that using the mode topic ID assigned during the inference method of LDA, used to annotate unseen documents, improves performance by stabilizing the obtained topics. We show significant improvements over state of the art segmentation algorithms on two standard datasets. As an additional benefit, TopicTiling performs the segmentation in linear time and thus is computationally less expensive than other LDA-based segmentation methods.

Domain Adaptation of a Dependency Parser with a Class-Class Selectional Preference Model
Raphael Cohen, Yoav Goldberg, and Michael Elhadad
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

When porting parsers to a new domain, many of the errors are related to wrong attachment of out-of-vocabulary words. Since there is no available annotated data to learn the attachment preferences of the target domain words, we attack this problem using a model of selectional preferences based on domain-specific word classes. Our method uses Latent Dirichlet Allocations (LDA) to learn a domain-specific Selectional Preference model in the target domain using unannotated data. The model provides features that model the affinities among pairs of words in the domain. To incorporate these new features in the parsing model, we adopt the co-training approach and retrain the parser with the selectional preferences features. We apply this method for adapting Easy First, a fast non-directional parser trained on WSJ, to the biomedical domain

(Genia Treebank). The Selectional Preference features reduce error by 4.5% over the co-training baseline.

Extracting fine-grained durations for verbs from Twitter

Jennifer Williams

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper describes recent work for extracting durations associated with common verbs from the Twitter microblogging website. Twitter is a rich resource for temporal information about everyday events because people post their 'tweets' to Twitter publicly in real-time as they conduct their activities throughout the day. Implicit information about temporal durations is crucial to any natural language processing task involving temporal understanding and reasoning (Pan et al., 2011; Kozareva and Hovy, 2011; Gusev et al., 2011). Two objectives were examined in this paper (1) how to extract fine-grain durations for verbs using data from Twitter, and (2) to discover duration distributions for verbs.

Using a corpus of more than 14 million tweets (tagged for part-of-speech), regular expression patterns were applied to extract duration information for verbs in tweets. Four types of extraction frames were used to associate each event with its duration, achieving better than 90% accuracy. The resulting extracted corpus contains more than 350,000 tweets and typical durations were found for 486 verb lemmas. The mean durations for verb lemmas were compared with the work of Gusev et al. (2011) and it was found that there is a small positive correlation. All of the data and materials is made publicly available. Automatic extraction of event durations is faster than hand-annotation of events. Results show that verbs and durations can be extracted from Twitter using regular expressions with high accuracy. Some events have a short typical duration (such as 'snooze') while some events have a longer typical duration (such as 'exist').

Discourse Structure in Simultaneous Spoken Turkish

Isin Demirsahin

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

The current debate regarding the data structure necessary to represent discourse structure, specifically whether tree-structure is sufficient to represent discourse structure or not, is mainly focused on written text. This paper reviews some of the major claims about the structure in discourse and proposes an investigation of discourse structure for simultaneous spoken Turkish by focusing on tree-violations and exploring ways to explain them away by non-structural means.

A Generic Framework for Multiword Expressions Treatment: from Acquisition to Applications

Carlos Ramisch

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper presents an open and flexible methodological framework for the automatic acquisition of multiword expressions (MWEs) from monolingual textual corpora. This research is motivated by the importance of MWEs for NLP applications. After briefly presenting the modules of the framework, the paper reports extrinsic evaluation results considering two applications: computer-aided lexicography and statistical machine translation. Both applications can benefit from automatic MWE acquisition and the expressions acquired automatically from corpora can both speed up and improve their quality. The promising results of previous and ongoing experiments encourage further investigation about the optimal way to integrate MWE treatment into these and many other applications.

Towards Automatic Construction of Knowledge Bases from Chinese Online Resources

Liwei Chen, Yansong Feng, Yidong Chen, Lei Zou, and Dongyan Zhao

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Automatically constructing knowledge bases from online resources has become a crucial task in many research areas. Most existing knowledge bases are built from English resources, while few

efforts have been made for other languages. Building knowledge bases for Chinese is of great importance on its own right. However, simply adapting existing tools from English to Chinese yields inferior results.In this paper, we propose to create Chinese knowledge bases from online resources with less human involvement.This project will be formulated in a self-supervised framework which requires little manual work to extract knowledge facts from online encyclopedia resources in a probabilistic view.In addition, this framework will be able to update the constructed knowledge base with knowledge facts extracted from up-to-date newswire.Currently, we have obtained encouraging results in our pilot experiments that extracting knowledge facts from infoboxes can achieve a high accuracy of around 95%, which will be then used as training data for the extraction of plain webpages.

Poster Abstracts

Chinese Comma Disambiguation for Discourse Analysis
Yaqin Yang and Nianwen Xue
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

The Chinese comma signals the boundary of discourse units and also anchors discourse relations between adjacent text spans. In this work, we propose a discourse structure-oriented classification of the comma that can be automatically extracted from the Chinese Treebank based on syntactic patterns. We then experimented with two supervised learning methods that automatically disambiguate the Chinese comma based on this classification. The first method integrates comma classification into parsing, and the second method adopts a "post-processing" approach that extracts features from automatic parses to train a classifier to classify the comma. The experimental results show that the second approach compares favorably against the first approach.

Collective Classification for Fine-grained Information Status
Katja Markert, Yufang Hou, and Michael Strube
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Previous work on classifying information status (Nissim, 2006; Rahman and Ng, 2011) is restricted to coarse-grained classification and focuses on conversational dialogue. We here introduce the task of classifying fine-grained information status and work on written text. We add a fine-grained information status layer to the Wall Street Journal portion of the OntoNotes corpus. We claim that the information status of a mention depends not only on the mention itself but also on other mentions in the vicinity and solve the task by collectively classifying the information status of all mentions. Our approach strongly outperforms reimplementations of previous work.

Structuring E-Commerce Inventory
Karin Mauge, Khash Rohanimanesh, and Jean-David Ruvini
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Large e-commerce enterprises feature millions of items entered daily by a large variety of sellers. While some sellers provide rich, structured descriptions of their items, a vast majority of them provide unstructured natural language descriptions. In the paper we present a 2 steps method for structuring items into descriptive properties. The first step consists in unsupervised property discovery and extraction. The second step involves supervised property synonym discovery using a maximum entropy based clustering algorithm. We evaluate our method on a year worth of e-commerce data and show that it achieves excellent precision with good recall.

Named Entity Disambiguation in Streaming Data
Alexandre Davis, Adriano Veloso, Altigran Soares, Alberto Laender, and Wagner Meira Jr.
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

The named entity disambiguation task is to resolve the many-to-many correspondence between ambiguous names and the unique real world entity. This task can be modeled as a classification problem, provided that positive and negative examples are available for learning binary classifiers. High-quality sense-annotated data, however, is hard to be obtained in streaming environments, since the training corpus would have to be constantly updated in order to accommodate the fresh data coming on the stream. On the other hand, few positive examples plus a large amount of unlabeled data may be easily acquired. Producing binary classifiers directly from this data, however, leads to poor disambiguation performance, and thus, we propose to enhance the quality of the classifiers using finer-grained variations of the well-known Expectation-Maximization (EM) algorithm. We conducted a systematic evaluation using Twitter streaming data, and the results show that our classifiers are extremely effective, providing improvements ranging from 1% to 20%, when compared to the current state-of-the-art biased SVMs, being 120 times faster.

Big Data versus the Crowd: Looking for Relationships in All the Right Places
Ce Zhang, Feng Niu, Christopher Ré, and Jude Shavlik
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Classically, training relation extractors relies on high-quality, manually annotated training data, which can be expensive to obtain. To mitigate this cost, NLU researchers have considered two newly available sources of less expensive (but potentially lower quality) labeled data from distant supervision and crowd sourcing. There is, however, no study comparing the relative impact of these two sources on the precision and recall of post-learning answers. To fill this gap, we empirically study how state-of-the-art techniques are affected by scaling these two sources. We use corpus sizes of up to 100 million documents and tens of thousands of crowd-source labeled examples. Our experiments show that increasing the corpus size for distant supervision has a statistically significant, positive impact on quality (F1 score). In contrast, human feedback has a positive and statistically significant, but lower, impact on precision and recall.

Automatic Event Extraction with Structured Preference Modeling
Wei Lu and Dan Roth
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper presents a novel sequence labeling model based on the latent-variable semi-Markov conditional random fields for jointly extracting argument roles of events from texts. The model takes in coarse mention and type information and predicts argument roles for a given event template.

This paper addresses the event extraction problem in a primarily unsupervised setting, where no labeled training instances are available. Our key contribution is a novel learning framework called structured preference modeling (PM), that allows arbitrary preference to be assigned to certain structures during the learning procedure. We establish and discuss connections between this framework and other existing works. We show empirically that the structured preferences are crucial to the success of our task. Our model, trained without annotated data and with a small number of structured preferences, yields performance competitive to some baseline supervised approaches.

Discriminative Learning for Joint Template Filling
Einat Minkov and Luke Zettlemoyer
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper presents a joint model for template filling, where the goal is to automatically specify the fields of target relations such as seminar announcements or corporate acquisition events. The approach models mention detection, unification and field extraction in a flexible, feature-rich model that allows for joint modeling of interdependencies at all levels and across fields. Such an approach can, for example, learn likely event lengths and the fact that start times should come before end times. While the joint inference space is large, we demonstrate effective learning with a Perceptron-style approach that uses simple, greedy beam decoding. Empirical results in two benchmark domains demonstrate consistently strong performance on both mention detection and template filling tasks.

Classifying French Verbs Using French and English Lexical Resources
Ingrid Falk, Claire Gardent, and Jean-Charles Lamirel
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We present a novel approach to the automatic acquisition of a Verbnet like classification of French verbs which involves the use (i) of a neural clustering method which associates clusters with features, (ii) of several supervised and unsupervised evaluation metrics and (iii) of various existing syntactic and semantic lexical resources. We evaluate our approach on an established test set and show that it outperforms previous related work with an F-measure of 0.70.

Modeling Sentences in the Latent Space
Weiwei Guo and Mona Diab
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Sentence Similarity is the process of computing a similarity score between two sentences. Previous sentence similarity work finds that latent semantics approaches to the problem do not perform well due to insufficient information in single sentences. In this paper, we show that by carefully handling words that are not in the sentences (missing words), we can train a reliable latent variable model on sentences. In the process, we propose a new evaluation framework for sentence similarity: Concept Definition Retrieval. The new framework allows for large scale tuning and testing of Sentence Similarity models. Experiments on the new task and previous data sets show significant improvement of our model over baselines and other traditional latent variable models. Our results indicate comparable and even better performance than current state of the art systems addressing the problem of sentence similarity.

Improving Word Representations via Global Context and Multiple Word Prototypes
Eric Huang, Richard Socher, Christopher Manning, and Andrew Ng
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Unsupervised word representations are very useful in NLP tasks both as inputs to learning algorithms and as extra word features in NLP systems. However, most of these models are built with only local context and one representation per word. This is problematic because words are often polysemous and global context can also provide useful information for learning word meanings. We present a new neural network architecture which 1) learns word embeddings that better capture the semantics of words by incorporating both local and global document context, and 2) accounts for homonymy and polysemy by learning multiple embeddings per word. We introduce a new dataset with human judgments on pairs of words in sentential context, and evaluate our model on it, showing that our model outperforms competitive baselines and other neural language models.

Exploiting Social Information in Grounded Language Learning via Grammatical Reduction
Mark Johnson, Katherine Demuth, and Michael Frank
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

This paper describes an unsupervised model of grounded language acquisition. The input to the model consists of (orthographically transcribed) child-directed utterances accompanied by the set of objects present in the non-linguistic context. Each object is annotated by social cues, indicating e.g., whether the caregiver is looking at or touching the object. We show how to model the task of inferring which objects are being talked about (and which words refer to which objects) as standard grammatical inference, and describe PCFG-based unigram models and adaptor grammar-based collocation models for the task. Exploiting social cues improves the performance of all models. Our models learn the relative importance of each social cue jointly with word-object mappings and collocation structure, consistent with the idea that children could discover the importance of particular social information sources during word learning.

You Had Me at Hello: How Phrasing Affects Memorability
Cristian Danescu-Niculescu-Mizil, Justin Cheng, Jon Kleinberg, and Lillian Lee
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Understanding the ways in which information achieves widespread public awareness is a research question of significant interest. We consider whether, and how, the way in which the information is phrased — the choice of words and sentence structure — can affect this process. To this end, we develop an analysis framework and build a corpus of movie quotes, annotated with memorability information, in which we are able to control for both the speaker and the setting of the quotes. We find that there are significant differences between memorable and non-memorable quotes in several key dimensions, even after controlling for situational and contextual factors. One is lexical distinctiveness: in aggregate, memorable quotes use less common word choices,

but at the same time are built upon a scaffolding of common syntactic patterns. Another is that memorable quotes tend to be more general in ways that make them easy to apply in new contexts — that is, more portable. We also show how the concept of "memorable language" can be extended across domains.

Modeling the Translation of Predicate-Argument Structure for SMT

Deyi Xiong, Min Zhang, and Haizhou Li

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Predicate-argument structure contains rich semantic information of which statistical machine translation hasn't taken full advantage. In this paper, we propose two discriminative, feature-based models to exploit predicate-argument structures for statistical machine translation: 1) a predicate translation model and 2) an argument reordering model. The predicate translation model explores lexical and semantic contexts surrounding a verbal predicate to select desirable translations for the predicate. The argument reordering model automatically predicts the moving direction of an argument relative to its predicate after translation using semantic features. The two models are integrated into a state-of-the-art phrase-based machine translation system and evaluated on Chinese-to-English translation tasks with large-scale training data. Experimental results demonstrate that the two models significantly improve translation accuracy.

A Ranking-based Approach to Word Reordering for Statistical Machine Translation

Nan Yang, Mu Li, Dongdong Zhang, and Nenghai Yu

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Long distance word reordering is a major challenge in statistical machine translation research. Previous work has shown using source syntactic trees is an effective way to tackle this problem between two languages with substantial word order difference. In this work, we further extend this line of exploration and propose a novel but simple approach, which utilizes a ranking model based on word order precedence in the target language to reposition nodes in the syntactic parse tree of a source sentence. The ranking model is automatically derived from word aligned parallel data with a syntactic parser for source language based on both lexical and syntactical features. We evaluated our approach on large-scale Japanese-English and English-Japanese machine translation tasks, and show that it can significantly outperform the baseline phrase-based SMT system.

Character-Level Machine Translation Evaluation for Languages with Ambiguous Word Boundaries

Chang Liu and Hwee Tou Ng

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

In this work, we introduce the TESLA-CELAB metric (Translation Evaluation of Sentences with Linear-programming-based Analysis — Character-level Evaluation for Languages with Ambiguous word Boundaries) for automatic machine translation evaluation. For languages such as Chinese where words usually have meaningful internal structure and word boundaries are often fuzzy, TESLA-CELAB acknowledges the advantage of character-level evaluation over word-level evaluation. By reformulating the problem in the linear programming framework, TESLA-CELAB addresses several drawbacks of the character-level metrics, in particular the modeling of synonyms spanning multiple characters. We show empirically that TESLA-CELAB significantly outperforms character-level BLEU in the English-Chinese translation evaluation tasks.

PORT: a Precision-Order-Recall MT Evaluation Metric for Tuning

Boxing Chen, Roland Kuhn, and Samuel Larkin

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Many machine translation (MT) evaluation metrics have been shown to correlate better with human judgment than BLEU. In principle, tuning on these metrics should yield better systems than tuning on BLEU. However, due to various factors, such as speed, requirement for linguistic

resources, and optimization difficulty, they have not been widely adopted for tuning. This paper presents PORT , a new MT evaluation metric which combines precision, recall and an ordering metric and which is primarily designed for tuning MT systems. PORT does not require external resources and is quick to compute. It has a better correlation with human judgment than BLEU. We compare PORT-tuned MT systems to BLEU-tuned baselines in five experimental conditions involving four language pairs, tuning with minimum error rate training. PORT tuning achieves consistently better performance than BLEU tuning, according to four automated metrics (include BLEU) and to human evaluation: in comparisons of outputs from 300 source sentences, human judges preferred the PORT-tuned output 45.3% of the time (vs. 32.7% BLEU tuning preferences and 22.0% ties).

Mixing Multiple Translation Models in Statistical Machine Translation

Majid Razmara, George Foster, Baskaran Sankaran, and Anoop Sarkar

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Statistical machine translation is often faced with the problem of combining training data from many diverse sources into a single translation model which then has to translate sentences in a new domain. We propose a novel approach, ensemble decoding, which com- bines a number of translation systems dynamically at the decoding step. In this paper, we evaluate performance on a domain adaptation setting where we translate sentences from the medical domain. Our experimental results show that ensemble decoding outperforms various strong baselines including mixture models, the current state of the art for do- main adaptation in machine translation.

Hierarchical Chunk-to-String Translation

Yang Feng, Dongdong Zhang, Mu Li, and Qun Liu

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We present a hierarchical chunk-to-string translation model, which can be seen as a compromise between the hierarchical phrase-based model and the tree-to-string model, to combine the merits of the two models. With the help of shallow parsing, our model learns rules consisting of words and *chunks* and meanwhile introduce syntax cohesion. Under the weighed synchronous context-free grammar defined by these rules, our model searches for the best translation derivation and yields target translation simultaneously. Our experiments show that our model significantly outperforms the hierarchical phrase-based model and the tree-to-string model on English-Chinese Translation tasks.

Large-Scale Syntactic Language Modeling with Treelets

Adam Pauls and Dan Klein

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We propose a simple generative, syntactic language model that conditions on overlapping windows of tree context (or treelets) in the same way that n-gram language models condition on overlapping windows of sentence context. We estimate the parameters of our model by collecting counts from automatically parsed text using standard n-gram language model estimation techniques, allowing us to train a model on over one billion tokens of data using a single machine in a matter of hours. We evaluate on perplexity and a range of grammaticality tasks, and find that we perform as well or better than n-gram models and other generative baselines. Our model even competes with state-of-the-art discriminative models hand-designed for the grammaticality tasks, despite training on positive data alone. We also show fluency improvements in a preliminary machine translation experiment.

Text Segmentation by Language Using Minimum Description Length

Hiroshi Yamaguchi and Kumiko Tanaka-Ishii

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

The problem addressed in this paper is to segment a given multilingual document into segments for each language and then identify the language of each segment. The problem was motivated

by an attempt to collect a large amount of linguistic data for non-major languages from the web. The problem is formulated in terms of obtaining the minimum description length of a text, and the proposed solution finds the segments and their languages through dynamic programming. Empirical results demonstrating the potential of this approach are presented for experiments using texts taken from the Universal Declaration of Human Rights and Wikipedia, covering more than 200 languages.

Improve SMT Quality with Automatically Extracted Paraphrase Rules

Wei He, Hua Wu, Haifeng Wang, and Ting Liu

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We propose a novel approach to improve SMT via paraphrase rules which are automatically extracted from the bilingual training data. Without using extra paraphrase resources, we acquire the rules by comparing the source side of the parallel corpus with the target-to-source translations of the target side. Besides the word and phrase paraphrases, the acquired paraphrase rules mainly cover the structured paraphrases on the sentence level. These rules are employed to enrich the SMT inputs for translation quality improvement. The experimental results show that our proposed approach achieves significant improvements of 1.6 3.6 points of BLEU in the oral domain and 0.5 1 points in the news domain.

Ecological Evaluation of Persuasive Messages Using Google AdWords

Marco Guerini, Carlo Strapparava, and Oliviero Stock

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

In recent years there has been a growing interest in crowdsourcing methodologies to be used in experimental research for NLP tasks. In particular, evaluation of systems and theories about persuasion is difficult to accommodate within existing frameworks. In this paper we present a new cheap and fast methodology that allows fast experiment building and evaluation with fully-automated analysis at a low cost. The central idea is exploiting existing commercial tools for advertising on the web, such as Google AdWords, to measure message impact in an ecological setting. The paper includes a description of the approach, tips for how to use AdWords for scientific research, and results of pilot experiments on the impact of affective text variations which confirm the effectiveness of the approach.

Polarity Consistency Checking for Sentiment Dictionaries

Eduard Dragut, Hong Wang, Clement Yu, Prasad Sistla, and Weiyi Meng

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Polarity classification of words is important for applications such as Opinion Mining and Sentiment Analysis. A number of sentiment word/sense dictionaries have been manually or (semi)automatically constructed. The dictionaries have substantial inaccuracies. Besides obvious instances, where the same word appears with different polarities in different dictionaries, the dictionaries exhibit complex cases, which cannot be detected by mere manual inspection. We introduce the concept of polarity consistency of words/senses in sentiment dictionaries in this paper. We show that the consistency problem is NP-complete. We reduce the polarity consistency problem to the satisfiability problem and utilize a fast SAT solver to detect inconsistencies in a sentiment dictionary. We perform experiments on four sentiment dictionaries and WordNet.

Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation

Ziheng Lin, Chang Liu, Hwee Tou Ng, and Min-Yen Kan

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

An ideal summarization system should produce summaries that have high content coverage and linguistic quality. Many state-of-the-art summarization systems focus on content coverage by extracting content-dense sentences from source articles. A current research focus is to process these

sentences so that they read fluently as a whole. The current AESOP task encourages research on evaluating summaries on content, readability, and overall responsiveness. In this work, we adapt a machine translation metric to measure content coverage, apply an enhanced discourse coherence model to evaluate summary readability, and combine both in a trained regression model to evaluate overall responsiveness. The results show significantly improved performance over AESOP 2011 submitted metrics.

Sentence Simplification by Monolingual Machine Translation
Sander Wubben, Antal van den Bosch, and Emiel Krahmer
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

In this paper we describe a method for simplifying sentences using Phrase Based Machine Translation, augmented with a re-ranking heuristic based on dissimilarity, and trained on a monolingual parallel corpus. We compare our system to a word-substitution baseline and two state-of-the-art systems, all trained and tested on paired sentences from the English part of Wikipedia and Simple Wikipedia. Human test subjects judge the output of the different systems. Analysing the judgements shows that by relatively careful phrase-based paraphrasing our model achieves similar simplification results to state-of-the-art systems, while generating better formed output. We also argue that text readability metrics such as the Flesch-Kincaid grade level should be used with caution when evaluating the output of simplification systems.

A Cost Sensitive Part-of-Speech Tagging: Differentiating Serious Errors from Minor Errors
Hyun-Je Song, Jeong-Woo Son, Tae-Gil Noh, Seong-Bae Park, and Sang-Jo Lee
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

All types of part-of-speech (POS) tagging errors have been equally treated by existing taggers. However, the errors are not equally important, since some errors affect the performance of subsequent natural language processing (NLP) tasks seriously while others do not. This paper aims to minimize these serious errors while retaining the overall performance of POS tagging. Two gradient loss functions are proposed to reflect the different types of errors. They are designed to assign a larger cost to serious errors and a smaller one to minor errors. Through a set of POS tagging experiments, it is shown that the classifier trained with the proposed loss functions reduces serious errors compared to state-of-the-art POS taggers. In addition, the experimental result on text chunking shows that fewer serious errors help to improve the performance of subsequent NLP tasks.

A Broad-Coverage Normalization System for Social Media Language
Fei Liu, Fuliang Weng, and Xiao Jiang
Monday 6:00pm–9:00pm — 3F Lobby (ICC)

Social media language contains huge amount and wide variety of nonstandard tokens, created both intentionally and unintentionally by the users. It is of crucial importance to normalize the noisy nonstandard tokens before applying other NLP techniques. A major challenge facing this task is the system coverage, i.e., for any user-created nonstandard term, the system should be able to restore the correct word within its top n output candidates. In this paper, we propose a cognitively-driven normalization system that integrates different human perspectives in normalizing the nonstandard tokens, including the enhanced letter transformation, visual priming, and string/phonetic similarity. The system was evaluated on both word- and message-level using four SMS and Twitter data sets. Results show that our system achieves over 90% word-coverage across all data sets (a 10% absolute increase compared to state-of-the-art); the broad word-coverage can also successfully translate into message-level performance gain, yielding 6% absolute increase compared to the best prior approach.

Incremental Joint Approach to Word Segmentation, POS Tagging, and Dependency Parsing in Chinese

Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, and Jun'ichi Tsujii

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We propose the first joint model for word segmentation, POS tagging, and dependency parsing for Chinese. Based on an extension of the incremental joint model for POS tagging and dependency parsing (Hatori et al., 2011), we propose an efficient character-based decoding method that can combine features from state-of-the-art segmentation, POS tagging, and dependency parsing models. We also describe our method to align comparable states in the beam, and how we can combine features of different characteristics in our incremental framework. In experiments using the Chinese Treebank (CTB), we show that the accuracies of the three tasks can be improved significantly over the baseline models, particularly by 0.6% for POS tagging and 2.4% for dependency parsing. We also perform comparison experiments with the partially joint models.

Exploring Deterministic Constraints: from a Constrained English POS Tagger to an Efficient ILP Solution to Chinese Word Segmentation

Qiuye Zhao and Mitch Marcus

Monday 6:00pm–9:00pm — 3F Lobby (ICC)

We show for both English POS tagging and Chinese word segmentation that with proper representation, large number of deterministic constraints can be learned from training examples, and these are useful in constraining probabilistic inference. For tagging, learned constraints are directly used to constrain Viterbi decoding. For segmentation, character-based tagging constraints can be learned with the same templates. However, they are better applied to a word-based model, thus an integer linear programming (ILP) formulation is proposed. For both problems, the corresponding constrained solutions have advantages in both efficiency and accuracy.

6

Tuesday, July 10, 2012: Main Conference

Overview

7:30am – 5:00pm	**Registration**	**(3F Lobby (ICC))**
9:00am – 10:30am	**Best Paper Awards Session**	**(Tamna A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
10:30am – 4:00pm	**Demonstrations**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
12:30pm – 2:00pm	**Lunch Break**	
12:30pm – 2:00pm	**Student Lunch (students only!)**	**(Ocean View (ICC))**
2:00pm – 3:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:30pm	**Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
6:00pm – late	**Banquet**	**(Shangri La)**

Schedule

7:30am – 5:00pm	**Registration (3F Lobby (ICC))**				
9:00am – 10:30am	**Best Paper Awards (Tamna A (ICC))** **Chair:** *Slav Petrov*)				
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**				
Parallel Sessions	**Samda (ICC)**	**Halla A (ICC)**	**Halla B (ICC)**	**402 (ICC)**	**401 (ICC)**
11:00am – 12:30pm	**R50: The People** **Chair:** *Ulrich Schäfer*	**Machine Translation IV** **Chair:** *Malte Nuhn*	**Parsing II** **Chair:** *Emily Pitler*	**Social Media** **Chair:** *Miles Osborne*	**User Generated Content** **Chair:** *Jordan Boyd-Graber*
11:00am – 11:30am	Rediscovering ACL Discoveries Through the Lens of ACL Anthology Network Citing Sentences *D. Radev and A. Abu-Jbara* ..... **p. 94**	Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information *J. Su, H. Wu, H. Wang, Y. Chen, X. Shi, H. Dong, and Q. Liu* ..... **p. 98**	Semantic Parsing with Bayesian Tree Transducers *B. Jones, M. Johnson, and S. Goldwater* ..... **p. 98**	Tweet Recommendation with Graph Co-Ranking *R. Yan, M. Lapata, and X. Li* ..... **p. 99**	Spice it up? Mining Refinements to Online Instructions from User Generated Content *G. Druck and B. Pang* ..... **p. 100**
11:30am – 12:00pm	Towards a Computational History of the ACL: 1980-2008 *A. Anderson, D. Jurafsky, and D. A. McFarland* **p. 94**	A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining *H. Sajjad, A. Fraser, and H. Schmid* ..... **p. 98**	Dependency Hashing for n-best CCG Parsing *D. Ng and J. R. Curran* ..... **p. 99**	Joint Inference of Named Entity Recognition and Normalization for Tweets *X. Liu, M. Zhou, X. Zhou, Z. Fu, and F. Wei* ... **p. 99**	Sentence Dependency Tagging in Online Question Answering Forums *Z. Qu and Y. Liu* ... **p. 100**
12:00pm – 12:30pm	Discovering Factions in the Computational Linguistics Community *Y. Sim, N. A. Smith, and D. A. Smith* ..... **p. 94** He Said, She Said: Gender in the ACL Anthology *A. Vogel and D. Jurafsky* ..... **p. 94**	Modified Distortion Matrices for Phrase-Based Statistical Machine Translation *A. Bisazza and M. Federico* ..... **p. 98**	Strong Lexicalization of Tree Adjoining Grammars *A. Maletti and J. Engelfriet* ..... **p. 99**	Finding Bursty Topics from Microblogs *Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim* ..... **p. 100**	Mining Entity Types from Query Logs via User Intent Modeling *P. Pantel, T. Lin, and M. Gamon* ..... **p. 100**
12:30pm – 2:00pm	**Lunch Break & Student Lunch (students only!) (Ocean View (ICC))**				

Parallel Sessions	Samda (ICC)	401 (ICC)	Halla B (ICC)	402 (ICC)	Halla A (ICC)
2:00pm – 3:30pm	**R50: The Contents** **Chair:** *Min Yen Kan*	**NLP Apps I** **Chair:** *Katja Filippova*	**Machine Learning I** **Chair:** *Sebastian Riedel*	**Multilinguality I** **Chair:** *Julia Hockenmair*	**Parsing III** **Chair:** *Bevan Jones*
2:00pm – 2:30pm	Discourse Structure and Computation: Past, Present and Future *B. Webber and A. Joshi* . . . . . **p. 95**	Cross-Lingual Mixture Model for Sentiment Classification *X. Meng, F. Wei, X. Liu, M. Zhou, G. Xu, and H. Wang* . . . . . **p. 101**	Computational Approaches to Sentence Completion *G. Zweig, J. C. Platt, C. Meek, C. J. Burges, A. Yessenalina, and Q. Liu* . . . . . **p. 102**	Selective Sharing for Multilingual Dependency Parsing *T. Naseem, R. Barzilay, and A. Globerson* . . . . . **p. 102**	Head-driven Transition-based Parsing with Top-down Prediction *K. Hayashi, T. Watanabe, M. Asahara, and Y. Matsumoto* . . . . . **p. 103**
2:30pm – 3:00pm	Extracting glossary sentences from scholarly articles: A comparative evaluation of pattern bootstrapping and deep analysis *M. Reiplinger, U. Schäfer, and M. Wolska* . . . . . **p. 95**	Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization *W. Chan, X. Zhou, W. Wang, and T.-S. Chua* . . . . **p. 101**	Iterative Viterbi A* Algorithm for K-Best Sequential Decoding *Z. Huang, Y. Chang, B. Long, J.-F. Crespo, A. Dong, S. Keerthi, and S.-L. Wu* . . . . . **p. 102**	The Creation of a Corpus of English Metalanguage *S. Wilson* . . . . . **p. 103**	MIX Is Not a Tree-Adjoining Language *M. Kanazawa and S. Salvati* . . . . . **p. 103**
3:00pm – 3:30pm	Applying Collocation Segmentation to the ACL Anthology Reference Corpus *V. Daudaravicius* . . . **p. 95** Text Reuse with ACL: (Upward) Trends *P. Gupta and P. Rosso* . . . . . **p. 95**	Error Mining on Dependency Trees *C. Gardent and S. Narayan* . . . . . **p. 101**	Bootstrapping via Graph Propagation *M. Whitney and A. Sarkar* . . . . . **p. 102**	Crosslingual Induction of Semantic Roles *I. Titov and A. Klementiev* . . . . . **p. 103**	Exploiting Multiple Treebanks for Parsing with Quasi-synchronous Grammars *Z. Li, T. Liu, and W. Che* . . . . . **p. 103**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**				
Parallel Sessions	**Samda (ICC)**	**401 (ICC)**	**Halla B (ICC)**	**402 (ICC)**	**Halla A (ICC)**
4:00pm – 5:30pm	**R50: The Anthology** **Chair:** *Deyi Xiong*	**Names** **Chair:** *Xiaohua Liu*	**Relations** **Chair:** *Guillermo Garrido*	**Topics** **Chair:** *Tadashi Nomoto*	**Parsing IV** **Chair:** *Andreas Maletti*
4:00pm – 4:30pm	Integrating User-Generated Content in the ACL Anthology *P. Bysani and M.-Y. Kan* . . . . . **p. 96**	A Probabilistic Model for Canonicalizing Named Entity Mentions *D. Yogatama, Y. Sim, and N. A. Smith* . . . . . **p. 104**	Unsupervised Relation Discovery with Sense Disambiguation *L. Yao, S. Riedel, and A. McCallum* . . . . . **p. 104**	Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model *W. Y. Wang, E. Mayfield, S. Naidu, and J. Dittmar* . . . . . **p. 105**	Attacking Parsing Bottlenecks with Unlabeled Data and Relevant Factorizations *E. Pitler* . . . . . **p. 106**

	Towards an ACL Anthology Corpus with Logical Document Structure. An Overview of the ACL 2012 Contributed Task *U. Schäfer, J. Read, and S. Oepen* ..... **p. 96**				
4:30pm – 5:00pm	Towards High-Quality Text Stream Extraction from PDF. Technical Background to the ACL 2012 Contributed Task *O. R. Berg, S. Oepen, and J. Read* ..... **p. 96** Combining OCR Outputs for Logical Document Structure Markup. Technical Background to the ACL 2012 Contributed Task *U. Schäfer and B. Weitz* ..... **p. 96**	Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia *S. Kim, K. Toutanova, and H. Yu* ..... **p. 104**	Reducing Wrong Labels in Distant Supervision for Relation Extraction *S. Takamatsu, I. Sato, and H. Nakagawa* ..... **p. 105**	A Topic Similarity Model for Hierarchical Phrase-based Translation *X. Xiao, D. Xiong, M. Zhang, Q. Liu, and S. Lin* ..... **p. 105**	Semi-supervised Dependency Parsing using Lexical Affinities *S. A. Mirroshandel, A. Nasr, and J. Le Roux* .... **p. 106**
5:00pm – 5:30pm	Linking Citations to their Bibliographic references *H. D. Hoang Nhat and P. Bysani* ..... **p. 97**	A Computational Approach to the Automation of Creative Naming *G. Ozbal and C. Strapparava* .... **p. 104**	Finding Salient Dates for Building Thematic Timelines *R. Kessler, X. Tannier, C. Hagège, V. Moriceau, and A. Bittar* ..... **p. 105**	Modeling Topic Dependencies in Hierarchical Text Categorization *A. Moschitti, Q. Ju, and R. Johansson* ..... **p. 106**	
6:00pm – late	**Banquet (Shangri La)**				

Best Paper and Demo Sessions

Best Paper

Demonstrations

Best Paper Abstract

Bayesian Symbol-Refined Tree Substitution Grammars for Syntactic Parsing

Hiroyuki Shindo, Yusuke Miyao, Akinori Fujino, and Masaaki Nagata

Tuesday 9:00am–9:45am — Tamna A (ICC)

We propose Symbol-Refined Tree Substitution Grammars (SR-TSGs) for syntactic parsing. An SR-TSG is an extension of the conventional TSG model where each nonterminal symbol can be refined (subcategorized) to fit the training data. We aim to provide a unified model where TSG rules and symbol refinement are learned from training data in a fully automatic and consistent fashion. We present a novel probabilistic SR-TSG model based on the hierarchical Pitman-Yor Process to encode backoff smoothing from a fine-grained SR-TSG to simpler CFG rules, and develop an efficient training method based on Markov Chain Monte Carlo (MCMC) sampling. Our SR-TSG parser achieves an F1 score of 92.4% in the Wall Street Journal (WSJ) English Penn Treebank parsing task, which is a 7.7 point improvement over a conventional Bayesian TSG parser, and better than state-of-the-art discriminative reranking parsers.

String Re-writing Kernel

Fan Bu, Hang Li, and Xiaoyan Zhu

Tuesday 9:45am–10:30am — Tamna A (ICC)

Learning for sentence re-writing is a fundamental task in natural language processing and information retrieval. In this paper, we propose a new class of kernel functions, referred to as string re-writing kernel, to address the problem. A string re-writing kernel measures the similarity between two pairs of strings, each pair representing re-writing of a string. It can capture the lexical and structural similarity between two pairs of sentences without the need of constructing syntactic trees. We further propose an instance of string re-writing kernel which can be computed efficiently. Experimental results on benchmark datasets show that our method can achieve comparable results with state-of-the-art methods on two sentence re-writing learning tasks: paraphrase identification and recognizing textual entailment.

Abstracts from "Rediscovering 50 Years of Discoveries"

Rediscovering ACL Discoveries Through the Lens of ACL Anthology Network Citing Sentences

Dragomir Radev and Amjad Abu-Jbara

Tuesday 11:00am–11:20am — Samda (ICC)

The ACL Anthology Network (AAN) is a comprehensive manually curated networked database of citations and collaborations in the field of Computational Linguistics. Each citation edge in AAN is associated with one or more citing sentences. A citing sentence is one that appears in a scientific article and contains an explicit reference to another article. In this paper, we shed the light on the usefulness of AAN citing sentences for understanding research trends and summarizing previous discoveries and contributions. We also propose and motivate several different uses and applications of citing sentences.

Towards a Computational History of the ACL: 1980-2008

Ashton Anderson, Dan Jurafsky, and Daniel A. McFarland

Tuesday 11:20am–11:40am — Samda (ICC)

We develop a people-centered computational history of science that tracks authors over topics and apply it to the history of computational linguistics. We present four findings in this paper. First, we identify the topical subfields authors work on by assigning automatically generated topics to each paper in the ACL Anthology from 1980 to 2008. Next, we identify four distinct research epochs where the pattern of topical overlaps are stable and different from other eras: an early NLP period from 1980 to 1988, the period of US government-sponsored MUC and ATIS evaluations from 1989 to 1994, a transitory period until 2001, and a modern integration period from 2002 onwards. Third, we analyze the flow of authors across topics to discern how some subfields flow into the next, forming different stages of ACL research. We find that the government-sponsored bakeoffs brought new researchers to the field, and bridged early topics to modern probabilistic approaches. Last, we identify steep increases in author retention during the bakeoff era and the modern era, suggesting two points at which the field became more integrated.

Discovering Factions in the Computational Linguistics Community

Yanchuan Sim, Noah A. Smith, and David A. Smith

Tuesday 11:40am–12:00pm — Samda (ICC)

We present a joint probabilistic model of who cites whom in computational linguistics, and also of the words they use to do the citing. The model reveals latent factions, or groups of individuals whom we expect to collaborate more closely within their faction, cite within the faction using language distinct from citation outside the faction, and be largely understandable through the language used when cited from without. We conduct an exploratory data analysis on the ACL Anthology. We extend the model to reveal changes in some authors' faction memberships over time.

He Said, She Said: Gender in the ACL Anthology

Adam Vogel and Dan Jurafsky

Tuesday 12:00pm–12:20pm — Samda (ICC)

Studies of gender balance in academic computer science are typically based on statistics on enrollment and graduation. Going beyond these coarse measures of gender participation, we conduct a fine-grained study of gender in the field of Natural Language Processing. We use topic models (Latent Dirichlet Allocation) to explore the research topics of men and women in the ACL Anthology Network. We find that women publish more on dialog, discourse, and sentiment, while men publish more than women in parsing, formal semantics, and finite state models. To conduct our study we labeled the gender of authors in the ACL Anthology mostly manually, creating a useful resource for other gender studies. Finally, our study of historical

patterns in female participation shows that the proportion of women authors in computational linguistics has been continuously increasing, with approximately a 50% increase in the three decades since 1980.

Discourse Structure and Computation: Past, Present and Future
Bonnie Webber and Aravind Joshi
Tuesday 2:00pm–2:20pm — Samda (ICC)

The discourse properties of text have long been recognized as critical to language technology, and over the past 40 years, our understanding of and ability to exploit the discourse properties of text has grown in many ways. This essay briefly recounts these developments, the technology they employ, the applications they support, and the new challenges that each subsequent development has raised. We conclude with the challenges faced by our current understanding of discourse, and the applications that meeting these challenges will promote.

Extracting glossary sentences from scholarly articles: A comparative evaluation of pattern bootstrapping and deep analysis
Melanie Reiplinger, Ulrich Schäfer, and Magdalena Wolska
Tuesday 2:20pm–2:40pm — Samda (ICC)

The paper reports on a comparative study of two approaches to extracting definitional sentences from a corpus of scholarly discourse: one based on bootstrapping lexico-syntactic patterns and another based on deep analysis. Computational Linguistics was used as the target domain and the ACL Anthology as the corpus. Definitional sentences extracted for a set of well-defined concepts were rated by domain experts. Results show that both methods extract high-quality definition sentences intended for automated glossary construction.

Applying Collocation Segmentation to the ACL Anthology Reference Corpus
Vidas Daudaravicius
Tuesday 2:40pm–3:00pm — Samda (ICC)

Collocation is a well-known linguistic phenomenon which has a long history of research and use. In this study I employ collocation segmentation to extract terms from the large and complex ACL Anthology Reference Corpus, and also briefly research and describe the history of the ACL. The results of the study show that until 1986, the most significant terms were related to formal/rule based methods. Starting in 1987, terms related to statistical methods became more important. For instance, language model, similarity measure, text classification. In 1990, the terms Penn Treebank, Mutual Information , statistical parsing, bilingual corpus, and dependency tree became the most important, showing that newly released language resources appeared together with many new research areas in computational linguistics. Although Penn Treebank was a significant term only temporarily in the early nineties, the corpus is still used by researchers today. The most recent significant terms are Bleu score and semantic role labeling. While machine translation as a term is significant throughout the ACL ARC corpus, it is not significant for any particular time period. This shows that some terms can be significant globally while remaining insignificant at a local level.

Text Reuse with ACL: (Upward) Trends
Parth Gupta and Paolo Rosso
Tuesday 3:00pm–3:20pm — Samda (ICC)

With rapidly increasing community, a plethora of conferences related to Natural Language Processing and easy access to their proceedings make it essential to check the integrity and novelty of the new submissions. This study aims to investigate the trends of text reuse in the ACL submissions, if any. We carried a set of analyses on two spans of five years papers (the past and the present) of ACL using a publicly available text reuse detection application to notice the behaviour. In our study, we found some strong reuse cases which can be an indicator to establish a clear policy to handle text reuse for the upcoming editions of ACL. The results are anonymised.

Integrating User-Generated Content in the ACL Anthology

Praveen Bysani and Min-Yen Kan

Tuesday 4:00pm–4:20pm — Samda (ICC)

The ACL Anthology was revamped in 2012 to its second major version, encompassing faceted navigation, social media use, as well as author- and reader-generated content and comments on published work as part of the revised front-end user interface. At the back-end, the Anthology was updated to incorporate its publication records into a database. We describe the ACL Anthology's previous legacy, redesign and revamp process and technologies, and its resulting functionality.

Towards an ACL Anthology Corpus with Logical Document Structure. An Overview of the ACL 2012 Contributed Task

Ulrich Schäfer, Jonathon Read, and Stephan Oepen

Tuesday 4:20pm–4:40pm — Samda (ICC)

The ACL 2012 Contributed Task is a community effort aiming to provide the full ACL Anthology as a high-quality corpus with rich markup, following the TEI P5 guidelines—a new resource dubbed the ACL Anthology Corpus (AAC). The goal of the task is threefold: (a) to provide a shared resource for experimentation on scientific text; (b) to serve as a basis for advanced search over the ACL Anthology, based on textual content and citations; and, by combining the aforementioned goals, (c) to present a showcase of the benefits of natural language processing to a broader audience. The Contributed Task extends the current Anthology Reference Corpus (ARC) both in size, quality, and by aiming to provide tools that allow the corpus to be automatically extended with new content—be they scanned or born-digital.

Towards High-Quality Text Stream Extraction from PDF. Technical Background to the ACL 2012 Contributed Task

Øyvind Raddum Berg, Stephan Oepen, and Jonathon Read

Tuesday 4:40pm–5:00pm — Samda (ICC)

Extracting textual content and document structure from PDF presents a surprisingly (depressingly, to some, in fact) difficult challenge, owing to the purely display-oriented design of the PDF document standard. While a variety of lower-level PDF extraction toolkits exist, none fully support the recovery of original text (in reading order) and relevant structural elements, even for so-called born-digital PDFs, i.e. those prepared electronically using typesetting systems like LaTeX, OpenOffice, and the like. This short paper summarizes a new tool for high-quality extraction of text and structure from PDFs, combining state-of-the-art PDF parsing, font interpretation, layout analysis, and TEI-compliant output of text and logical document markup.

Combining OCR Outputs for Logical Document Structure Markup. Technical Background to the ACL 2012 Contributed Task

Ulrich Schäfer and Benjamin Weitz

Tuesday 5:00pm–5:20pm — Samda (ICC)

We describe how paperxml, a logical document structure markup for scholarly articles, is generated on the basis of OCR tool outputs. Paperxml has been initially developed for the ACL Anthology Searchbench. The main purpose was to robustly provide uniform access to sentences in ACL Anthology papers from the past 46 years, ranging from scanned, typewriter-written conference and workshop proceedings papers, up to recent high-quality typeset, born digital journal articles, with varying layouts. Paperxml markup includes additional useful information on page and paragraph breaks, section headings, footnotes, tables, captions, boldface and italics character styles as well as bibliographic and publication metadata. The role of paperxml in the ACL Contributed Task 'Rediscovering 50 Years of Discoveries' is to serve as fall-back source (1) for older, scanned papers (mostly published before the year 2000), for which born digital PDF sources are not available, (2) for born digital PDF papers on which the pdfextract method failed, (3) for document parts where pdfextract does not output useful markup such as currently for tables. We sketch transformation of paperxml into the Contributed Task's TEI P5 XML.

Linking Citations to their Bibliographic references

Huy Do Hoang Nhat and Praveen Bysani

Tuesday 5:20pm–5:40pm — Samda (ICC)

In this paper we describe our participation in the contributed task of ACL Special workshop 2012. We contribute to the goal of enriching the textual content in ACL Anthology by identifying the citation contexts in a paper and linking them to their corresponding references in the bibliography section. We use Parscit, to process the references in each paper. Pattern matching heuristics are then used to connect the citations with their references. Furthermore, we prepared a small evaluation dataset, to test the efficiency of our method. We achieved 95% precision and 80% recall on this dataset.

Long Paper Abstracts

Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information

Jinsong Su, Hua Wu, Haifeng Wang, Yidong Chen, Xiaodong Shi, Huailin Dong, and Qun Liu

Tuesday 11:00am–11:30am — Halla A (ICC)

To adapt a translation model trained from the data in one domain to another, previous works paid more attention to the studies of parallel corpus while ignoring the in-domain monolingual corpora which can be obtained more easily. In this paper, we propose a novel approach for translation model adaptation by utilizing in-domain monolingual topic information instead of the in-domain bilingual corpora, which incorporates the topic information into translation probability estimation. Our method establishes the relationship between the out-of-domain bilingual corpus and the in-domain monolingual corpora via topic mapping and topic-phrase distribution probability estimation from in-domain monolingual corpora. Experimental result on the NIST Chinese-English translation task shows that our approach significantly outperforms the baseline system.

A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining

Hassan Sajjad, Alexander Fraser, and Helmut Schmid

Tuesday 11:30am–12:00pm — Halla A (ICC)

We propose a novel model to automatically extract transliteration pairs from parallel corpora. Our model is efficient, language pair independent and mines transliteration pairs in a consistent fashion in both unsupervised and semi-supervised settings. We model transliteration mining as an interpolation of transliteration and non-transliteration sub-models. We evaluate on NEWS 2010 shared task data and on parallel corpora with competitive results.

Modified Distortion Matrices for Phrase-Based Statistical Machine Translation

Arianna Bisazza and Marcello Federico

Tuesday 12:00pm–12:30pm — Halla A (ICC)

This paper presents a novel method to suggest long word reorderings to a phrase-based SMT decoder. We address language pairs where long reordering concentrates on few patterns, and use fuzzy chunk-based rules to predict likely reorderings for these phenomena. Then we use reordered n-gram LMs to rank the resulting permutations and select the n-best for translation. Finally we encode these reorderings by modifying selected entries of the distortion cost matrix, on a per-sentence basis. In this way, we expand the search space by a much finer degree than if we simply raised the distortion limit. The proposed techniques are tested on Arabic-English and German-English using well-known SMT benchmarks.

Semantic Parsing with Bayesian Tree Transducers

Bevan Jones, Mark Johnson, and Sharon Goldwater

Tuesday 11:00am–11:30am — Halla B (ICC)

Many semantic parsing models use tree transformations to map between natural language and meaning representation. However, while tree transformations are central to several state-of-the-art approaches, little use has been made of the rich literature on tree automata. This paper makes the connection concrete with a tree transducer based semantic parsing model and suggests that other models can be interpreted in a similar framework, increasing the generality of their contributions. In particular, this paper further introduces a variational Bayesian inference algorithm that is applicable to a wide class of tree transducers, producing state-of-the-art semantic parsing results while remaining applicable to any domain employing probabilistic tree transducers.

Dependency Hashing for n-best CCG Parsing
Dominick Ng and James R. Curran
Tuesday 11:30am–12:00pm — Halla B (ICC)

Optimising for one grammatical representation, but evaluating over a different one is a particular challenge for parsers and n-best CCG parsing. We find that this mismatch causes many n-best CCG parses to be semantically equivalent, and describe a hashing technique that eliminates this problem, improving oracle n-best F-score by 0.7% and reranking accuracy by 0.4%. We also present a comprehensive analysis of errors made by the C&C CCG parser, providing the first breakdown of the impact of implementation decisions, such as supertagging, on parsing accuracy.

Strong Lexicalization of Tree Adjoining Grammars
Andreas Maletti and Joost Engelfriet
Tuesday 12:00pm–12:30pm — Halla B (ICC)

Recently, it was shown (Kuhlmann, Satta: Tree-adjoining grammars are not closed under strong lexicalization. Comput. Linguist., 2012) that finitely ambiguous tree adjoining grammars cannot be transformed into a normal form (preserving the generated tree language), in which each production contains a lexical symbol. A more powerful model, the simple context-free tree grammar, admits such a normal form. It can be effectively constructed and the maximal rank of the nonterminals only increases by 1. Thus, simple context-free tree grammars strongly lexicalize tree adjoining grammars and themselves.

Tweet Recommendation with Graph Co-Ranking
Rui Yan, Mirella Lapata, and Xiaoming Li
Tuesday 11:00am–11:30am — 402 (ICC)

As one of the most popular micro-blogging services, Twitter attracts millions of users, producing millions of tweets daily. Shared information through this service spreads faster than would have been possible with traditional sources; however the proliferation of user-generation content poses challenges to browsing and finding valuable information. In this paper we propose a graph-theoretic model for tweet recommendation that presents users with items they may have an interest in. Our model ranks tweets and their authors simultaneously using several networks: the social network connecting the users, the network connecting the tweets, and a third network that ties the two together. Tweet and author entities are ranked following a co-ranking algorithm based on the intuition that that there is a mutually reinforcing relationship between tweets and their authors that could be reflected in the rankings. We show that this framework can be parametrized to take into account user preferences, the popularity of tweets and their authors, and diversity. Experimental evaluation on a large dataset shows that our model outperforms competitive approaches by a large margin.

Joint Inference of Named Entity Recognition and Normalization for Tweets
Xiaohua Liu, Ming Zhou, Xiangyang Zhou, Zhongyang Fu, and Furu Wei
Tuesday 11:30am–12:00pm — 402 (ICC)

Tweets represent a critical source of fresh information, in which named entities occur frequently with rich variations. We study the problem of named entity normalization (NEN) for tweets. Two main challenges are the errors propagated from named entity recognition (NER) and the lack of information in a single tweet. We propose a novel graphical model to simultaneously conduct NER and NEN on multiple tweets to address these challenges. Particularly, our model introduces a binary random variable for each pair of words with the same lemma across similar tweets, whose value indicates whether the related two words belongs to the mentions of the same entity. We evaluate our method on a manually annotated data set, and show that our method outperforms the baseline that handles these two tasks separately, boosting the F1 from 80.2% to 83.6% for NER, and the Accuracy from 79.4% to 82.6% for NEN, respectively.

Finding Bursty Topics from Microblogs
Qiming Diao, Jing Jiang, Feida Zhu, and Ee-Peng Lim
Tuesday 12:00pm–12:30pm — 402 (ICC)

Microblogs such as Twitter reflect the general public's reactions to major events. Bursty topics from microblogs reveal what events have attracted the most online attention. Although bursty event detection from text streams has been studied before, previous work may not be suitable for microblogs because compared with other text streams such as news articles and scientific publications, microblog posts are particularly diverse and noisy. To find topics that have bursty patterns on microblogs, we propose a topic model that simultaneously captures two observations: (1) posts published around the same time are more likely to have the same topic, and (2) posts published by the same user are more likely to have the same topic. The former helps find event-driven posts while the latter helps identify and filter out "personal" posts. Our experiments on a large Twitter dataset show that there are more meaningful and unique bursty topics in the top-ranked results returned by our model than an LDA baseline and two degenerate variations of our model. We also show some case studies that demonstrate the importance of considering both the temporal information and users' personal interests for bursty topic detection from microblogs.

Spice it up? Mining Refinements to Online Instructions from User Generated Content
Gregory Druck and Bo Pang
Tuesday 11:00am–11:30am — 401 (ICC)

There are a growing number of popular web sites where users submit and review instructions for completing tasks as varied as building a table and baking a pie. In addition to providing their subjective evaluation, reviewers often provide actionable refinements. These refinements clarify, correct, improve, or provide alternatives to the original instructions. However, reading all relevant reviews is a daunting task for a user. In this paper, we propose a generative model that jointly identifies user-proposed refinements in instruction reviews at multiple granularities, and aligns them to the appropriate steps in the original instruction. In unsupervised experiments in the recipe domain, our model provides 90.2% F_1 for predicting refinements at the review level, 76.7% F_1 for predicting refinements segments within reviews, and useful alignments with the recipe being reviewed.

Sentence Dependency Tagging in Online Question Answering Forums
Zhonghua Qu and Yang Liu
Tuesday 11:30am–12:00pm — 401 (ICC)

Online forums are becoming a popular resource in the state of the art question answering (QA) systems. Because of its nature as an online community, it contains more updated knowledge than other places. However, going through tedious and redundant posts to look for answers could be very time consuming. Most prior work focused on extracting only question answering sentences from user conversations. In this paper, we introduce the task of sentence dependency tagging. Finding dependency structure can not only help find answer quickly but also allow users to trace back how the answer is concluded through user conversations. We use linear-chain conditional random fields (CRF) for sentence type tagging, and a 2D CRF to label the dependency relation between sentences. Our experimental results show that our proposed approach performs well for sentence dependency tagging. This dependency information can benefit other tasks such as thread ranking and answer summarization in online forums.

Mining Entity Types from Query Logs via User Intent Modeling
Patrick Pantel, Thomas Lin, and Michael Gamon
Tuesday 12:00pm–12:30pm — 401 (ICC)

We predict entity type distributions in Web search queries via probabilistic inference in graphical models that capture how entity bearing queries are generated. We jointly model the interplay between latent user intents that govern queries and unobserved entity types, leveraging observed signals from query formulations and document clicks. We apply the models to resolve entity types in new queries and to assign prior type distributions over an existing knowledge base.

Our models are efficiently trained using maximum likelihood estimation over millions of real-world Web search queries. We show that modeling user intent significantly improves entity type resolution for head queries over the state of the art, on several metrics, without degradation in tail query performance.

Cross-Lingual Mixture Model for Sentiment Classification
Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang
Tuesday 2:00pm–2:30pm — 401 (ICC)

The amount of labeled sentiment data in English is much larger than that in other languages. Such a disproportion arouse interest in cross-lingual sentiment classification, which aims to conduct sentiment classification in the target language (e.g. Chinese) using labeled data in the source language (e.g. English). Most existing work relies on machine translation engines to directly adapt labeled data from the source language to the target language. This approach suffers from the limited coverage of vocabulary in the machine translation results. In this paper, we propose a generative cross-lingual mixture model (CLMM) to leverage unlabeled bilingual parallel data. By fitting parameters to maximize the likelihood of the bilingual parallel data, the proposed model learns previously unseen sentiment words from the large bilingual parallel data and improves vocabulary coverage significantly. Experiments on multiple data sets show that CLMM is consistently effective in two settings: (1) labeled data in the target language are unavailable; and (2) labeled data in the target language are also available.

Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization
Wen Chan, Xiangdong Zhou, Wei Wang, and Tat-Seng Chua
Tuesday 2:30pm–3:00pm — 401 (ICC)

We present a novel answer summarization method for community Question Answering services (cQAs) to address the problem of "incomplete answer", i.e., the "best answer" of a complex multi-sentence question misses valuable information that is contained in other answers. In order to automatically generate a novel and non-redundant community answer summary, we segment the complex original multi-sentence question into several sub questions and then propose a general Conditional Random Field (CRF) based answer summary method with group L1 regularization. Various textual and non-textual QA features are explored. Specifically, we explore four different types of contextual factors, namely, the information novelty and non-redundancy modeling for local and non-local sentence interactions under question segmentation. To further unleash the potential of the abundant cQA features, we introduce the group L1 regularization for feature learning. Experimental results on a Yahoo! Answers dataset show that our proposed method significantly outperforms state-of-the-art methods on cQA summarization task.

Error Mining on Dependency Trees
Claire Gardent and Shashi Narayan
Tuesday 3:00pm–3:30pm — 401 (ICC)

In recent years, error mining approaches were developed to help identify the most likely sources of parsing failures in parsing systems using handcrafted grammars and lexicons. However the techniques they use (e.g., suffix arrays) to enumerate and count n-grams builds on the sequential nature of a text corpus and do not easily extend to structured data. In this paper, we propose an algorithm for mining trees and apply it to detect the most likely sources of generation failure. We show that the tree mining algorithm we propose permits identifying not only errors in the generation system (grammar, lexicon, algorithm) but also mismatches between the structures contained in the input and the input structures expected by our generator.

Computational Approaches to Sentence Completion
Geoffrey Zweig, John C. Platt, Christopher Meek, Christopher J.C. Burges, Ainur Yessenalina, and Qiang Liu
Tuesday 2:00pm–2:30pm — Halla B (ICC)

This paper studies the problem of sentence-level semantic coherence by answering SAT-style sentence completion questions. These questions test the ability of algorithms to distinguish sense from nonsense based on a variety of sentence-level phenomena. We tackle the problem with two approaches: methods that use local lexical information, such as the n-grams of a classical language model; and methods that evaluate global coherence, such as latent semantic analysis. We evaluate these methods on a suite of practice SAT questions, and on a recently released sentence completion task based on data taken from five Conan Doyle novels. We find that by fusing local and global information, we can exceed 50% on this task (chance baseline is 20%), and we suggest some avenues for further research.

Iterative Viterbi A* Algorithm for K-Best Sequential Decoding
Zhiheng Huang, Yi Chang, Bo Long, Jean-Francois Crespo, Anlei Dong, Sathiya Keerthi, and Su-Lin Wu
Tuesday 2:30pm–3:00pm — Halla B (ICC)

Sequential modeling has been widely used in a variety of important applications including named entity recognition and shallow parsing. However, as more and more real time large-scale tagging applications arise, decoding speed has become a bottleneck for existing sequential tagging algorithms. In this paper we propose 1-best A*, 1-best iterative A*, k-best A* and k-best iterative Viterbi A* algorithms for sequential decoding. We show the efficiency of these proposed algorithms for five NLP tagging tasks. In particular, we show that iterative Viterbi A* decoding can be several times or orders of magnitude faster than the state-of-the-art algorithm for tagging tasks with a large number of labels. This algorithm makes real-time large-scale tagging applications with thousands of labels feasible.

Bootstrapping via Graph Propagation
Max Whitney and Anoop Sarkar
Tuesday 3:00pm–3:30pm — Halla B (ICC)

Bootstrapping a classifier from a small set of seed rules can be viewed as the propagation of labels between examples via features shared between them. This paper introduces a novel variant of the Yarowsky algorithm based on this view. It is a bootstrapping learning method which uses a graph propagation algorithm with a well defined objective function. The experimental results show that our proposed bootstrapping algorithm achieves state of the art performance or better on several different natural language data sets.

Selective Sharing for Multilingual Dependency Parsing
Tahira Naseem, Regina Barzilay, and Amir Globerson
Tuesday 2:00pm–2:30pm — 402 (ICC)

We present a novel algorithm for multilingual dependency parsing that uses annotations from a diverse set of source languages to parse a new unannotated language. Our motivation is to broaden the advantages of multilingual learning to languages that exhibit significant differences from existing resource-rich languages. The algorithm learns which aspects of the source languages are relevant for the target language and ties model parameters accordingly. The model factorizes the process of generating a dependency tree into two steps: selection of syntactic dependents and their ordering. Being largely language-universal, the selection component is learned in a supervised fashion from all the training languages. In contrast, the ordering decisions are only influenced by languages with similar properties. We systematically model this cross-lingual sharing using typological features. In our experiments, the model consistently outperforms a state-of-the-art multilingual parser. The largest improvement is achieved on the non Indo-European languages yielding a gain of 14.4%.

The Creation of a Corpus of English Metalanguage
Shomir Wilson
Tuesday 2:30pm–3:00pm — 402 (ICC)

Metalanguage is an essential linguistic mechanism which allows us to communicate explicit information about language itself. However, it has been underexamined in research in language technologies, to the detriment of the performance of systems that could exploit it. This paper describes the creation of the first tagged and delineated corpus of English metalanguage, accompanied by an explicit definition and a rubric for identifying the phenomenon in text. This resource will provide a basis for further studies of metalanguage and enable its utilization in language technologies.

Crosslingual Induction of Semantic Roles
Ivan Titov and Alexandre Klementiev
Tuesday 3:00pm–3:30pm — 402 (ICC)

We argue that multilingual parallel data provides a valuable source of indirect supervision for induction of shallow semantic representations. Specifically, we consider unsupervised induction of semantic roles from sentences annotated with automatically-predicted syntactic dependency representations and use a state-of-the-art generative Bayesian non-parametric model. At inference time, instead of only seeking the model which explains the monolingual data available for each language, we regularize the objective by introducing a soft constraint penalizing for disagreement in argument labeling on aligned sentences. We propose a simple approximate learning algorithm for our set-up which results in efficient inference. When applied to German-English parallel data, our method obtains a substantial improvement over a model trained without using the agreement signal, when both are tested on non-parallel sentences.

Head-driven Transition-based Parsing with Top-down Prediction
Katsuhiko Hayashi, Taro Watanabe, Masayuki Asahara, and Yuji Matsumoto
Tuesday 2:00pm–2:30pm — Halla A (ICC)

This paper presents a novel top-down data-driven dependency parsing algorithm. Our algorithm is a top-down transition-based parser combined with a weighted prediction from a statistical model inspired by the Earley prediction. Experiments on the English Penn Treebank data and the Chinese CoNLL-06 data show that the proposed algorithm achieves comparable results with other data-driven dependency parsing algorithms, while showing different characteristics from pure bottom-up dependency parsers.

MIX Is Not a Tree-Adjoining Language
Makoto Kanazawa and Sylvain Salvati
Tuesday 2:30pm–3:00pm — Halla A (ICC)

The language MIX consists of all strings over the three-letter alphabet a,b,c that contain an equal number of occurrences of each letter. We prove Joshi's (1985) conjecture that MIX is not a tree-adjoining language.

Exploiting Multiple Treebanks for Parsing with Quasi-synchronous Grammars
Zhenghua Li, Ting Liu, and Wanxiang Che
Tuesday 3:00pm–3:30pm — Halla A (ICC)

We present a simple and effective framework for exploiting multiple monolingual treebanks with different annotation guidelines for parsing. Several types of transformation patterns (TP) are designed to capture the systematic annotation inconsistencies among different treebanks. Based on such TPs, we design quasisynchronous grammar features to augment the baseline parsing models. Our approach can significantly advance the state-of-the-art parsing accuracy on two widely used target treebanks (Penn Chinese Treebank 5.1 and 6.0) using the Chinese Dependency Treebank as the source treebank. The improvements are respectively 1.37% and 1.10% with automatic part-of-speech tags. Moreover, an indirect comparison indicates that our approach

also outperforms previous work based on treebank conversion.

A Probabilistic Model for Canonicalizing Named Entity Mentions
Dani Yogatama, Yanchuan Sim, and Noah A. Smith
Tuesday 4:00pm–4:30pm — 401 (ICC)

We present a statistical model for canonicalizing named entity mentions into a table whose rows represent entities and whose columns are attributes (or parts of attributes). The model is novel in that it incorporates entity context, surface features, first-order dependencies among attribute-parts, and a notion of noise. Transductive learning from a few seeds and a collection of mention tokens combines Bayesian inference and conditional estimation. We evaluate our model and its components on two datasets collected from political blogs and sports news, finding that it outperforms a simple agglomerative clustering approach and previous work

Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia
Sungchul Kim, Kristina Toutanova, and Hwanjo Yu
Tuesday 4:30pm–5:00pm — 401 (ICC)

In this paper we propose a method to automatically label multi-lingual data with named entity tags. We build on prior work utilizing Wikipedia metadata and show how to effectively combine the weak annotations stemming from Wikipedia metadata with information obtained through English-foreign language parallel Wikipedia sentences. The combination is achieved using a novel semi-CRF model for foreign sentence tagging in the context of a parallel English sentence. The model outperforms both standard annotation projection methods and methods based solely on Wikipedia metadata.

A Computational Approach to the Automation of Creative Naming
Gozde Ozbal and Carlo Strapparava
Tuesday 5:00pm–5:30pm — 401 (ICC)

In this paper, we propose a computational approach to generate neologisms consisting of homophonic puns and metaphors based on the category of the service to be named and the properties to be underlined. We describe all the linguistic resources and natural language processing techniques that we have exploited for this task. Then, we analyze the performance of the system that we have developed. The empirical results show that our approach is generally effective and it constitutes a solid starting point for the automation of the naming process.

Unsupervised Relation Discovery with Sense Disambiguation
Limin Yao, Sebastian Riedel, and Andrew McCallum
Tuesday 4:00pm–4:30pm — Halla B (ICC)

To discover relation types from text, most methods cluster shallow or syntactic patterns of relation mentions, but consider only one possible sense per pattern. In practice this assumption is often violated. In this paper we overcome this issue by inducing clusters of pattern senses from feature representations of patterns. In particular, we employ a topic model to partition entity pairs associated with patterns into sense clusters using local and global features. We merge these sense clusters into semantic relations using hierarchical agglomerative clustering. We compare against several baselines: a generative latent-variable model, a clustering method that does not disambiguate between path senses, and our own approach but with only local features. Experimental results show our proposed approach discovers dramatically more accurate clusters than models without sense disambiguation, and that incorporating global features, such as the document theme, is crucial.

Reducing Wrong Labels in Distant Supervision for Relation Extraction
Shingo Takamatsu, Issei Sato, and Hiroshi Nakagawa
Tuesday 4:30pm–5:00pm — Halla B (ICC)

In relation extraction, distant supervision seeks to extract relations between entities from text by using a knowledge base, such as Freebase, as a source of supervision. When a sentence and a knowledge base refer to the same entity pair, this approach heuristically labels the sentence with the corresponding relation in the knowledge base. However, this heuristic can fail with the result that some sentences are labeled wrongly. This noisy labeled data causes poor extraction performance. In this paper, we propose a method to reduce the number of wrong labels. We present a novel generative model that directly models the heuristic labeling process of distant supervision. The model predicts whether assigned labels are correct or wrong via its hidden variables. Our experimental results show that this model detected wrong labels with higher performance than baseline methods. In the experiment, we also found that our wrong label reduction boosted the performance of relation extraction.

Finding Salient Dates for Building Thematic Timelines
Rémy Kessler, Xavier Tannier, Caroline Hagège, Véronique Moriceau, and André Bittar
Tuesday 5:00pm–5:30pm — Halla B (ICC)

We present an approach for detecting salient (important) dates in texts in order to automatically build event timelines from a search query (e.g. the name of an event or person, etc.). This work was carried out on a corpus of newswire texts in English provided by the Agence France Presse (AFP). In order to extract salient dates that warrant inclusion in an event timeline, we first recognize and normalize temporal expressions in texts and then use a machine-learning approach to extract salient dates that relate to a particular topic. We focused only on extracting the dates and not the events to which they are related.

Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model
William Yang Wang, Elijah Mayfield, Suresh Naidu, and Jeremiah Dittmar
Tuesday 4:00pm–4:30pm — 402 (ICC)

We propose a latent variable model to enhance historical analysis of large corpora. This work extends prior work in topic modelling by incorporating metadata, and the interaction between the components in metadata, in a general way. To test this, we collect a corpus of slavery-related United States property law judgements sampled from the years 1730 to 1866. We study the language use in these legal cases, with a special focus on shifts in opinions on controversial topics across different regions. Because this is a longitudinal data set, we are also interested in understanding how these opinions change over the course of decades. We show that the joint learning scheme of our sparse mixed-effects model improves on other state-of-the-art generative and discriminative models on the region and time period identification tasks. Experiments show that our sparse mixed-effects model is more accurate quantitatively and qualitatively interesting, and that these improvements are robust across different parameter settings.

A Topic Similarity Model for Hierarchical Phrase-based Translation
Xinyan Xiao, Deyi Xiong, Min Zhang, Qun Liu, and Shouxun Lin
Tuesday 4:30pm–5:00pm — 402 (ICC)

Previous work using topic model for statistical machine translation (SMT) explore topic information at the word level. However, SMT has been advanced from word-based paradigm to phrase/rule-based paradigm. We therefore propose a topic similarity model to exploit topic information at the synchronous rule level for hierarchical phrase-based translation. We associate each synchronous rule with a topic distribution, and select desirable rules according to the similarity of their topic distributions with given documents. We show that our model significantly improves the translation performance over the baseline on NIST Chinese-to-English translation experiments. Our model also achieves a better performance and a faster speed than previous approaches that work at the word level.

Modeling Topic Dependencies in Hierarchical Text Categorization
Alessandro Moschitti, Qi Ju, and Richard Johansson
Tuesday 5:00pm–5:30pm — 402 (ICC)

n this paper, we encode topic dependencies in hierarchical multi-label Text Categorization (TC) by means of rerankers. We represent reranking hypotheses with several innovative kernels considering both the structure of the hierarchy and the probability of nodes. Additionally, to better investigate the role of category relationships, we consider two interesting cases: (i) traditional schemes in which node-fathers include all the documents of their child-categories; and (ii) more general schemes, in which children can include documents not belonging to their fathers. The extensive experimentation on Reuters Corpus Volume 1 shows that our rerankers inject effective structural semantic dependencies in multi-classifiers and significantly outperform the state-of-the-art.

Attacking Parsing Bottlenecks with Unlabeled Data and Relevant Factorizations
Emily Pitler
Tuesday 4:00pm–4:30pm — Halla A (ICC)

Prepositions and conjunctions are two of the largest remaining bottlenecks in parsing. Across various existing parsers, these two categories have the lowest accuracies, and mistakes made have consequences for downstream applications. Prepositions and conjunctions are often assumed to depend on lexical dependencies for correct resolution. As lexical statistics based on the training set only are sparse, unlabeled data can help ameliorate this sparsity problem. By including unlabeled data features into a factorization of the problem which matches the representation of prepositions and conjunctions, we achieve a new state-of-the-art for English dependencies with 93.55% correct attachments on the current standard. Furthermore, conjunctions are attached with an accuracy of 90.8%, and prepositions with an accuracy of 87.4%.

Semi-supervised Dependency Parsing using Lexical Affinities
Seyed Abolghasem Mirroshandel, Alexis Nasr, and Joseph Le Roux
Tuesday 4:30pm–5:00pm — Halla A (ICC)

Treebanks are not large enough to reliably model precise lexical phenomena. This deficiency provokes attachment errors in the parsers trained on such data. We propose in this paper to compute lexical affinities, on large corpora, for specific lexico-syntactic configurations that are hard to disambiguate and introduce the new information in a parser. Experiments on the French Treebank showed a relative decrease of the error rate of 7.1 % Labeled Accuracy Score, yielding the best parsing results on this treebank.

Demo Abstracts

Applications of GPC Rules and Character Structures in Games for Learning Chinese Characters

Wei-Jie Huang, Chia-Ru Chou, Yu-Lin Tzeng, Chia-Ying Lee, and Chao-Lin Liu

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We demonstrate applications of psycholinguistic and sublexical information for learning Chinese characters. The knowledge about the grapheme-phoneme conversion (GPC) rules of languages has been shown to be highly correlated to the ability of reading alphabetic languages and Chinese. We build and will demo a game platform for strengthening the association of phonological components in Chinese characters with the pronunciations of the characters. Results of a preliminary evaluation of our games indicated significant improvement in learners' response times in Chinese naming tasks. In addition, we construct a Web-based open system for teachers to prepare their own games to best meet their teaching goals. Techniques for decomposing Chinese characters and for comparing the similarity between Chinese characters were employed to recommend lists of Chinese characters for authoring the games. Evaluation of the authoring environment with 20 subjects showed that our system made the authoring of games more effective and efficient.

Specifying Viewpoint and Information Need with Affective Metaphors: A System Demonstration of the Metaphor-Magnet Web App/Service

Tony Veale and Guofu Li

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Metaphors pervade our language because they are elastic enough to allow a speaker to express an affective viewpoint on a topic without committing to a specific meaning. This balance of expressiveness and indeterminism means that metaphors are just as useful for eliciting information as they are for conveying information. We explore here, via a demonstration of a system for metaphor interpretation and generation called Metaphor Magnet, the practical uses of metaphor as a basis for formulating affective information queries. We also consider the kinds of deep and shallow stereotypical knowledge that are needed for such a system, and demonstrate how they can be acquired from corpora and from the web.

Social Event Radar: A Bilingual Context Mining and Sentiment Analysis Summarization System

Wen-Tai Hsieh, Chen-Ming Wu, Tsun Ku, and Seng-cho T. Chou

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Social Event Radar is a new social networking-based service platform, that aim to alert as well as monitor any merchandise flaws, food-safety related issues, unexpected eruption of diseases or campaign issues towards to the Government, enterprises of any kind or election parties, through keyword expansion detection module, using bilingual sentiment opinion analysis tool kit to conclude the specific event social dashboard and deliver the outcome helping authorities to plan "risk control" strategy. With the rapid development of social network, people can now easily publish their opinions on the Internet. On the other hand, people can also obtain various opinions from others in a few seconds even though they do not know each other. A typical approach to obtain required information is to use a search engine with some relevant keywords. We thus take the social media and forum as our major data source and aim at collecting specific issues efficiently and effectively in this work.

UWN: A Large Multilingual Lexical Knowledge Base

Gerard de Melo and Gerhard Weikum

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present UWN, a large multilingual lexical knowledge base that describes the meanings and

relationships of words in over 200 languages. This paper explains how link prediction, information integration and taxonomy induction methods have been used to build UWN based on WordNet and extend it with millions of named entities from Wikipedia. We additionally introduce extensions to cover lexical relationships, frame-semantic knowledge, and language data. An online interface provides human access to the data, while a software API enables applications to look up over 16 million words and names.

langid.py: An Off-the-shelf Language Identification Tool
Marco Lui and Timothy Baldwin
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present langid.py, an off-the-shelf language identification tool. We discuss the design and implementation of langid.py, and provide an empirical comparison on 5 long-document datasets, and 2 datasets from the microblog domain. We find that langid.py maintains consistently high accuracy across all domains, making it ideal for end-users that require language identification without wanting to invest in preparation of in-domain training data.

Personalized Normalization for a Multilingual Chat System
Ai Ti Aw and Lian Hau Lee
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

This paper describes the personalized normalization of a multilingual chat system that supports chatting in user defined short-forms or abbreviations. One of the major challenges for multilingual chat realized through machine translation technology is the normalization of non-standard, self-created short-forms in the chat message to standard words before translation. Due to the lack of training data and the variations of short-forms used among different social communities, it is hard to normalize and translate chat messages if user uses vocabularies outside the training data and create short-forms freely. We develop a personalized chat normalizer for English and integrate it with a multilingual chat system, allowing user to create and use personalized short-forms in multilingual chat.

A Graphical Interface for MT Evaluation and Error Analysis
Meritxell Gonzàlez, Jesús Giménez, and Lluís Màrquez
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Error analysis in machine translation is a necessary step in order to investigate the strengths and weaknesses of the MT systems under development and allow fair comparisons among them. This work presents an application that shows how a set of heterogeneous automatic metrics can be used to evaluate a test bed of automatic translations. To do so, we have set up an online graphical interface for the ASIYA toolkit, a rich repository of evaluation measures working at different linguistic levels. The current implementation of the interface shows constituency and dependency trees as well as shallow syntactic and semantic annotations, and word alignments. The intelligent visualization of the linguistic structures used by the metrics, as well as a set of navigational functionalities, may lead towards advanced methods for automatic error analysis.

LetsMT!: Cloud-Based Platform for Do-It-Yourself Machine Translation
Andrejs Vasiljevs, Raivis Skadiņš, and Jörg Tiedemann
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

To facilitate the creation and usage of custom SMT systems we have created a cloud-based platform for do-it-yourself MT. The platform is developed in the EU collaboration project LetsMT!. This system demonstration paper presents the motivation in developing the LetsMT! platform, its main features, architecture, and an evaluation in a practical use case.

A Web-based Evaluation Framework for Spatial Instruction-Giving Systems
Srinivasan Janarthanam, Oliver Lemon, and Xingkun Liu
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We demonstrate a web-based environment for development and testing of different pedestrian route instruction-giving systems. The environment contains a City Model, a TTS interface, a game-world, and a user GUI including a simulated street-view. We describe the environment and components, the metrics that can be used for the evaluation of pedestrian route instruction-giving systems, and the shared challenge which is being organised using this environment.

DOMCAT: A Bilingual Concordancer for Domain-Specific Computer Assisted Translation
Ming-Hong Bai, Yu-Ming Hsieh, Keh-Jiann Chen, and Jason S. Chang
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

In this paper, we propose a web-based bilingual concordancer, DOMCAT , for domain-specific computer assisted translation. Given a multi-word expression as a query, the system involves retrieving sentence pairs from a bilingual corpus, identifying translation equivalents of the query in the sentence pairs (translation spotting) and ranking the retrieved sentence pairs according to the relevance between the query and the translation equivalents. To provide high-precision translation spotting for domain-specific translation tasks, we exploited a normalized correlation method to spot the translation equivalents. To ranking the retrieved sentence pairs, we propose a correlation function modified from the Dice coefficient for assessing the correlation between the query and the translation equivalents. The performances of the translation spotting module and the ranking module are evaluated in terms of precision-recall measures and coverage rate respectively.

The OpenGrm open-source finite-state grammar software libraries
Brian Roark, Richard Sproat, Cyril Allauzen, Michael Riley, Jeffrey Sorensen, and Terry Tai
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

In this paper, we present a new collection of open-source software libraries that provides command line binary utilities and library classes and functions for compiling regular expression and context-sensitive rewrite rules into finite-state transducers, and for n-gram language modeling. The OpenGrm libraries use the OpenFst library to provide an efficient encoding of grammars and general algorithms for building, modifying and applying models.

Multilingual WSD with Just a Few Lines of Code: the BabelNet API
Roberto Navigli and Simone Paolo Ponzetto
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

In this paper we present an API for programmatic access to BabelNet - a wide-coverage multilingual lexical knowledge base - and multilingual knowledge-rich Word Sense Disambiguation (WSD). Our aim is to provide the research community with easy-to-use tools to perform multilingual lexical semantic analysis and foster further research in this direction.

BIUTEE: A Modular Open-Source System for Recognizing Textual Entailment
Asher Stern and Ido Dagan
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

This paper introduces BIUTEE, an open-source system for recognizing textual entailment. Its main advantages are its ability to utilize various types of knowledge resources, and its extensibility by which new knowledge resources and inference components can be easily integrated. These abilities make BIUTEE an appealing RTE system for two research communities: (1) researchers of end applications, that can benefit from generic textual inference, and (2) RTE researchers, who can integrate their novel algorithms and knowledge resources into our system, saving the time and effort of developing a complete RTE system from scratch. Notable assistance for these researchers is provided by a visual tracing tool, by which researchers can refine and “debug” their knowledge resources and inference components.

Entailment-based Text Exploration with Application to the Health-care Domain
Meni Adler, Jonathan Berant, and Ido Dagan
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present a novel text exploration model, which extends the scope of state-of-the-art technologies by moving from standard concept-based exploration to statement-based exploration. The proposed scheme utilizes the textual entailment relation between statements as the basis of the exploration process. A user of our system can explore the result space of a query by drilling down/up from one statement to another, according to entailment relations specified by an entailment graph and an optional concept taxonomy. As a prominent use case, we apply our exploration system and illustrate its benefit on the health-care domain. To the best of our knowledge this is the first implementation of an exploration system at the statement level that is based on the textual entailment relation.

CSNIPER - Annotation-by-query for Non-canonical Constructions in Large Corpora
Richard Eckart de Castilho, Sabine Bartsch, and Iryna Gurevych
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present CSNIPER (Corpus Sniper), a tool that implements (i) a web-based multi-rater scenario for identifying and annotating non-canonical grammatical constructions in large corpora based on linguistic queries and (ii) evaluation of annotation quality by measuring inter-rater agreement. This annotation-by-query approach efficiently harnesses expert knowledge to identify instances of linguistic phenomena that are hard to identify by means of existing automatic annotation tools.

ACCURAT Toolkit for Multi-Level Alignment and Information Extraction from Comparable Corpora
Mārcis Pinnis, Radu Ion, Dan Ştefănescu, Fangzhong Su, Inguna Skadiņa, Andrejs Vasiļjevs, and Bogdan Babych
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

The lack of parallel corpora and linguistic resources for many languages and domains is one of the major obstacles for the further advancement of automated translation. A possible solution is to exploit comparable corpora (non-parallel bi- or multi-lingual text resources) which are much more widely available than parallel translation data. Our presented toolkit deals with parallel content extraction from comparable corpora. It consists of tools bundled in two workflows: (1) alignment of comparable documents and extraction of parallel sentences and (2) extraction and bilingual mapping of terms and named entities. The toolkit pairs similar bilingual comparable documents and extracts parallel sentences and bilingual terminological and named entity dictionaries from comparable corpora. This demonstration focuses on the English, Latvian, Lithuanian, and Romanian languages.

Demonstration of IlluMe: Creating Ambient According to Instant Message Logs
Lun-Wei Ku, Cheng-Wei Sun, and Ya-Hsin Hsueh
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present IlluMe, a software tool pack which creates a personalized ambient using the music and lighting. IlluMe includes an emotion analysis software, the small space ambient lighting, and a multimedia controller. The software analyzes emotional changes from instant message logs and corresponds the detected emotion to the best sound and light settings. The ambient lighting can sparkle with different forms of light and the smart phone can broadcast music respectively according to different atmosphere. All settings can be modified by the multimedia controller at any time and the new settings will be feedback to the emotion analysis software. The IlluMe system, equipped with the learning function, provides a link between residential situation and personal emotion. It works in a Chinese chatting environment to illustrate the language technology in life.

INPRO_iSS: A Component for Just-In-Time Incremental Speech Synthesis
Timo Baumann and David Schlangen
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present a component for incremental speech synthesis (iSS) and a set of applications that demonstrate its capabilities. This component can be used to increase the responsivity and naturalness of spoken interactive systems. While iSS can show its full strength in systems that generate output incrementally, we also discuss how even otherwise unchanged systems may profit from its capabilities.

WizIE: A Best Practices Guided Development Environment for Information Extraction
Yunyao Li, Laura Chiticariu, Huahai Yang, Frederick Reiss, and Arnaldo Carreno-fuentes
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Information extraction (IE) is becoming a critical building block in many enterprise applications. In order to satisfy the increasing text analytics demands of enterprise applications, it is crucial to enable developers with general computer science background to develop high quality IE extractors. In this demonstration, we present WizIE, an IE development environment intended to reduce the development life cycle and enable developers with little or no linguistic background to write high quality IE rules. WizIE provides an integrated wizard-like environment that guides IE developers step-by-step throughout the entire development process, based on best practices synthesized from the experience of expert developers. In addition, WizIE reduces the manual effort involved in performing key IE development tasks by offering automatic result explanation and rule discovery functionality. Preliminary results indicate that WizIE is a step forward towards enabling extractor development for novice IE developers.

A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle
Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

This paper describes a system for real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election as expressed on Twitter, a micro-blogging service. Twitter has become a central site where people express their opinions and views on political parties and candidates. Emerging events or news are often followed almost instantly by a burst in Twitter volume, providing a unique opportunity to gauge the relation between expressed public sentiment and electoral events. In addition, sentiment analysis can help explore how these events affect public opinion. While traditional content analysis takes days or weeks to complete, the system demonstrated here analyzes sentiment in the entire Twitter traffic about the election, delivering results instantly and continuously. It offers the public, the media, politicians and scholars a new and timely perspective on the dynamics of the electoral process and public opinion.

Building Trainable Taggers in a Web-based, UIMA-Supported NLP Workbench
Rafal Rak, BalaKrishna Kolluru, and Sophia Ananiadou
Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Argo is a web-based NLP and text mining workbench with a convenient graphical user interface for designing and executing processing workflows of various complexity. The workbench is intended for specialists and non-technical audiences alike, and provides the ever expanding library of analytics compliant with the Unstructured Information Management Architecture, a widely adopted interoperability framework. We explore the flexibility of this framework by demonstrating workflows involving three processing components capable of performing self-contained machine learning-based tagging. The three components are responsible for the three distinct tasks of 1) generating observations or features, 2) training a statistical model based on the generated features, and 3) tagging unlabelled data with the model. The learning and tagging components are based on an implementation of conditional random fields (CRF); whereas the feature generation component is an analytic capable of extending basic token information to a comprehensive set of features. Users define the features of their choice directly from Argo's graphical interface,

without resorting to programming (a commonly used approach to feature engineering). The experimental results performed on two tagging tasks, chunking and named entity recognition, showed that a tagger with a generic set of features built in Argo is capable of competing with task-specific solutions.

Akamon: An Open Source Toolkit for Tree/Forest-Based Statistical Machine Translation

Xianchao Wu, Takuya Matsuzaki, and Jun'ichi Tsujii

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We describe Akamon, an open source toolkit for tree and forest-based statistical machine translation (Liu et al., 2006; Mi et al., 2008; Mi and Huang, 2008). Akamon implements all of the algorithms required for tree/forest-to-string decoding using tree-to-string translation rules: multiple-thread forest-based decoding, n-gram language model integration, beam- and cube-pruning, k-best hypotheses extraction, and minimum error rate training. In terms of tree-to-string translation rule extraction, the toolkit implements the traditional maximum likelihood algorithm using PCFG trees (Galley et al., 2004) and HPSG trees/forests (Wu et al., 2010).

Subgroup Detector: A System for Detecting Subgroups in Online Discussions

Amjad Abu-Jbara and Dragomir Radev

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present Subgroup Detector, a system for analyzing threaded discussions and identifying the attitude of discussants towards one another and towards the discussion topic. The system uses attitude predictions to detect the split of discussants into subgroups of opposing views. The system uses an unsupervised approach based on rule-based sentiment analysis and unsupervised clustering techniques. The system is open source and is freely available for download. An online demo of the system is available at: http://clair.eecs.umich.edu/SubgroupDetector/

IRIS: a Chat-oriented Dialogue System based on the Vector Space Model

Rafael E. Banchs and Haizhou Li

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

This system demonstration paper presents IRIS (Informal Response Interactive System), a chat-oriented dialogue system based on the vector space model framework. The system belongs to the class of example-based dialogue systems and builds its chat capabilities on a dual search strategy over a large collection of dialogue samples. Additional strategies allowing for system adaptation and learning implemented over the same vector model space framework are also described and discussed.

Online Plagiarized Detection Through Exploiting Lexical, Syntax, and Semantic Information

Wan-Yu Lin, Nanyun Peng, Chun-Chao Yen, and Shou-de Lin

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

In this paper, we introduce a framework that identifies online plagiarism by exploiting lexical, syntactic and semantic features that includes duplication-gram, reordering and alignment of words, POS and phrase tags, and semantic similarity of sentences. We establish an ensemble framework to combine the predictions of each model. Results demonstrate that our system can not only find considerable amount of real-world online plagiarism cases but also outperforms several state-of-the-art algorithms and commercial software.

FLOW: A First-Language-Oriented Writing Assistant System

MeiHua Chen, ShihTing Huang, HungTing Hsieh, TingHui Kao, and Jason S. Chang

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Writing in English might be one of the most difficult tasks for EFL (English as a Foreign Language) learners. Evidence of a first language assisting learners can be found in a number of

studies. This paper presents FLOW , a writing assistance system. It is built based on first-language-oriented input function and context sensitive approach, aiming at providing immediate and appropriate suggestions including translations, paraphrases, and n-grams during composing and revising processes. FLOW is expected to help EFL writers achieve their writing flow without being interrupted by their insufficient lexical knowledge.

NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation

Tong Xiao, Jingbo Zhu, Hao Zhang, and Qiang Li

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present a new open source toolkit for phrase-based and syntax-based machine translation. The toolkit supports several state-of-the-art models developed in statistical machine translation, including the phrase-based model, the hierachical phrase-based model, and various syntax-based models. The key innovation provided by the toolkit is that the decoder can work with various grammars and offers different choices of decoding algrithms, such as phrase-based decoding, decoding as parsing/tree-parsing and forest-based decoding. Moreover, several useful utilities were distributed with the toolkit, including a discriminative reordering model, a simple and fast language model, and an implementation of minimum error rate training for weight tuning.

QuickView: NLP-based Tweet Search

Xiaohua Liu, Furu Wei, Ming Zhou, and QuickView Team Microsoft

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

Tweets have become a comprehensive repository for real-time information. However, it is often hard for users to quickly get information they are interested in from tweets, owing to the sheer volume of tweets as well as their noisy and informal nature. We present QuickView, an NLP-based tweet search platform to tackle this issue. Specifically, it exploits a series of natural language processing technologies, such as tweet normalization, named entity recognition, semantic role labeling, sentiment analysis, tweet classification, to extract useful information, i.e., named entities, events, opinions, etc., from a large volume of tweets. Then, non-noisy tweets, together with the mined information, are indexed, on top of which two brand new scenarios are enabled, i.e., categorized browsing and advanced search, allowing users to effectively access either the tweets or fine-grained information they are interested in.

Syntactic Annotations for the Google Books NGram Corpus

Yuri Lin, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov

Tuesday 11:00am–5:30pm — 3F Lobby (ICC)

We present a new edition of the Google Books Ngram Corpus, which describes how often words and phrases were used over a period of five centuries, in eight languages; it reflects 6% of all books ever published. This new edition introduces syntactic annotations: words are tagged with their part-of-speech, and head-modifier relationships are recorded. The annotations are produced automatically with statistical models that are specifically adapted to historical text. The corpus will facilitate the study of linguistic trends, especially those related to the evolution of syntax.

7

Wednesday, July 11, 2012: Main Conference

Overview

7:30am – 5:00pm	**Registration**	**(3F Lobby (ICC))**
8:00am – 9:00am	**Lifetime Achievement Award**	**(Tamna A (ICC))**
9:00am – 10:30am	**Invited Talk: Mark Johnson**	**(Tamna A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Short Paper Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
12:30pm – 2:00pm	**Lunch Break**	
2:00pm – 3:30pm	**ACL Business Meeting — open to everyone!**	**(Tamna A (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:20pm	**Short Paper Parallel Sessions**	**(Halla A&B, Samda, 401, 402 (ICC))**
5:30pm – 6:30pm	**Closing**	**(Tamna A (ICC))**

Schedule

7:30am – 5:00pm	Registration (3F Lobby (ICC))

8:00am – 9:00am	Lifetime Achievement Award (Tamna A (ICC))				
9:00am – 10:30am	Invited Talk (Tamna A (ICC)) *Mark Johnson:* "Computational linguistics: Where do we go from here?" Chair: by *Bob Moore*				
10:30am – 11:00am	Coffee Break (3F Lobby (ICC))				
Parallel Sessions	**Halla A (ICC)**	**Halla B (ICC)**	**Samda (ICC)**	**401 (ICC)**	**402 (ICC)**
11:00am – 12:30pm	**Parsing V** **Chair:** *Yusuke Miyao*	**Machine Translation V** **Chair:** *Keh-Yih Su*	**Relations and Events** **Chair:** *Surdeanu Mihai*	**Discourse II** **Chair:** *Marie-Francine Moens*	**Machine Learning II** **Chair:** *Shou-de Lin*
11:00am – 11:20am	Higher-order Constituent Parsing and Parser Combination *X. Chen and C. Kit* **p. 119**	Private Access to Phrase Tables for Statistical Machine Translation *N. Cancedda* ..... **p. 119**	A Novel Burst-based Text Representation Model for Scalable Event Detection *X. Zhao, R. Chen, K. Fan, H. Yan, and X. Li* .. **p. 120**	Self-Disclosure and Relationship Strength in Twitter Conversations *J. Bak, S. Kim, and A. Oh* ..... **p. 121**	Decoding Running Key Ciphers *S. Reddy and K. Knight* ..... **p. 122**
11:20am – 11:40am	Joint Evaluation of Morphological Segmentation and Syntactic Parsing *R. Tsarfaty, J. Nivre, and E. Andersson* ..... **p. 119**	Fast and Scalable Decoding with Language Model Look-Ahead for Phrase-based Statistical Machine Translation *J. Wuebker, H. Ney, and R. Zens* ..... **p. 120**	A Graph-based Cross-lingual Projection Approach for Weakly Supervised Relation Extraction *S. Kim and G. G. Lee* ..... **p. 121**	Genre Independent Subgroup Detection in Online Discussion Threads: A Study of Implicit Attitude using Textual Latent Semantics *P. Dasigi, W. Guo, and M. Diab* ..... **p. 121**	Using Rejuvenation to Improve Particle Filtering for Bayesian Word Segmentation *B. Börschinger and M. Johnson* ..... **p. 122**
11:40am – 12:00pm	A Comparison of Chinese Parsers for Stanford Dependencies *W. Che, V. Spitkovsky, and T. Liu* ..... **p. 119**	Head-Driven Hierarchical Phrase-based Translation *J. Li, Z. Tu, G. Zhou, and J. van Genabith* ... **p. 120**	Pattern Learning for Relation Extraction with a Hierarchical Topic Model *E. Alfonseca, K. Filippova, J.-Y. Delort, and G. Garrido* ..... **p. 121**	Learning to Temporally Order Medical Events in Clinical Text *P. Raghavan, A. Lai, and E. Fosler-Lussier* .. **p. 121**	Baselines and Bigrams: Simple, Good Sentiment and Topic Classification *S. Wang and C. Manning* ..... **p. 122**

12:00pm – 12:20pm	A Feature-Rich Constituent Context Model for Grammar Induction *D. Golland, J. DeNero, and J. Uszkoreit* ..... **p. 119**	Joint Learning of a Dual SMT System for Paraphrase Generation *H. Sun and M. Zhou* **p. 120**		A Context-sensitive, Multi-faceted Model of Lexico-Conceptual Affect *T. Veale* ..... **p. 122**	Automatically Learning Measures of Child Language Development *S. Sahakian and B. Snyder* ..... **p. 123**
12:30pm – 2:00pm	**Lunch Break**				
2:00pm – 3:30pm	**ACL Business Meeting (open to all; Tamna A (ICC))**				
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**				
Parallel Sessions	**Halla A (ICC)**	**Samda (ICC)**	**Halla B (ICC)**	**402 (ICC)**	**401 (ICC)**
4:00pm – 5:20pm	**Parsing and Machine Translation** **Chair:** *Mu Li*	**Multilinguality II** **Chair:** *Hua Wu*	**Meaning II** **Chair:** *Sameer Pradhan*	**NLP Apps and Data** **Chair:** *Ulrich Schäfer*	**NLP Apps II** **Chair:** *Hae-Chang Rim*
4:00pm – 4:20pm	A Comparative Study of Target Dependency Structures for Statistical Machine Translation *X. Wu, K. Sudoh, K. Duh, H. Tsukada, and M. Nagata* ..... **p. 123**	Detecting Semantic Equivalence and Information Disparity in Cross-lingual Documents *Y. Mehdad, M. Negri, and M. Federico* ..... **p. 124**	Learning the Latent Semantics of a Concept from its Definition *W. Guo and M. Diab* ..... **p. 124**	A Comprehensive Gold Standard for the Enron Organizational Hierarchy *A. Agarwal, A. Omuya, A. Harnly, and O. Rambow* ..... **p. 125**	Corpus-based Interpretation of Instructions in Virtual Environments *L. Benotti, M. Villalba, T. Lau, and J. Cerruti* ..... **p. 126**
4:20pm – 4:40pm	Robust Conversion of CCG Derivations to Phrase Structure Trees *J. K. Kummerfeld, D. Klein, and J. R. Curran* .. **p. 123**	Cross-lingual Parse Disambiguation based on Semantic Correspondence *L. Frermann and F. Bond* ..... **p. 124**	Unsupervised Semantic Role Induction with Global Role Ordering *N. Garg and J. Henserdon* ..... **p. 125**	A Two-step Approach to Sentence Compression of Spoken Utterances *D. Wang, X. Qian, and Y. Liu* ..... **p. 125**	Automatically Mining Question Reformulation Patterns from Search Log Data *X. Xue, Y. Tao, D. Jiang, and H. Li* ..... **p. 126**
4:40pm – 5:00pm	Estimating Compact Yet Rich Tree Insertion Grammars *E. Yamangil and S. Shieber* ..... **p. 123**	Learning to Find Transliterations and Transliterations on the Web *J. Z. Chang, J. S. Chang, and R. J.-S. Jang* .. **p. 124**	Humor as Circuits in Semantic Networks *I. Labutov and H. Lipson* ..... **p. 125**	Syntactic Stylometry for Deception Detection *S. Feng, R. Banerjee, and Y. Choi* ..... **p. 126**	Native Language Detection with Tree Substitution Grammars *B. Swanson and E. Charniak* ..... **p. 126**
5:00pm – 5:20pm	Topic Models for Dynamic Translation Model Adaptation *V. Eidelman, J. Boyd-Graber, and P. Resnik* ..... **p. 123**	Beefmoves: Dissemination, Diversity, and Dynamics of English Borrowings in a German Hip Hop Forum *M. Garley and J. Hockenmaier* ... **p. 124**	Crowdsourcing Inference-Rule Evaluation *N. Zeichner, J. Berant, and I. Dagan* ..... **p. 125**	Transforming Standard Arabic to Colloquial Arabic *E. Mohamed, B. Mohit, and K. Oflazer* ..... **p. 126**	Tense and Aspect Error Correction for ESL Learners Using Global Context *T. Tajiri, M. Komachi, and Y. Matsumoto* ..... **p. 127**
5:30pm – 6:30pm	**Closing (Tamna A (ICC))**				

Keynote Address: Mark Johnson

"Computational linguistics: Where do we go from here?"

Mark Johnson

Wednesday, July 11, 2012, 9:00am – 10.30am

West and Center Ballrooms

Abstract: The very fact that we're having a 50th annual meeting means that our field hasn't been a complete failure, but will there still be computational linguistics meetings in 50 years time? How do we fit into the larger intellectual picture, and what would it take to make computational linguistics into a real engineering discipline, or, for that matter, a scientific one? Prognosticating fearlessly (or perhaps just foolishly) I'll draw some lessons from the last 50 years about what the next few might hold.

James W. Pennebaker is a Professor of Language Science (CORE) in the Department of Computing at Macquarie University. He was awarded a BSc (Hons) in 1979 from the University of Sydney, an MA in 1984 from the University of California, San Diego and a PhD in 1987 from Stanford University. He held a postdoctoral fellowship at MIT from 1987 until 1988, and has been a visiting researcher at the University of Stuttgart, the Xerox Research Centre in Grenoble, CSAIL at MIT and the Natural Language group at Microsoft Research. He has worked on a wide range of topics in computational linguistics, but his main research area is parsing and its applications to text and speech processing. He was President of the Association for Computational Linguistics in 2003, and was a professor from 1989 until 2009 in the Departments of Cognitive and Linguistic Sciences and Computer Science at Brown University.

Short Paper Abstracts

Higher-order Constituent Parsing and Parser Combination
Xiao Chen and Chunyu Kit
Wednesday 11:00am–11:20am — Halla A (ICC)

This paper presents a higher-order model for constituent parsing aimed at utilizing more local structural context to decide the score of a grammar rule instance in a parse tree. Experiments on English and Chinese treebanks confirm its advantage over its first-order version. It achieves its best F1 scores of 91.86% and 85.58% on the two languages, respectively, and further pushes them to 92.80% and 85.60% via combination with other high-performance parsers.

Joint Evaluation of Morphological Segmentation and Syntactic Parsing
Reut Tsarfaty, Joakim Nivre, and Evelina Andersson
Wednesday 11:20am–11:40am — Halla A (ICC)

We present novel metrics for parse evaluation in joint segmentation and parsing scenarios where the gold sequence of terminals is not known in advance. The protocol uses distance-based metrics defined for the space of trees over lattices. Our metrics allow us to precisely quantify the performance gap between non-realistic parsing scenarios (assuming gold segmented and tagged input) and realistic ones (not assuming gold segmentation and tags). Our evaluation of segmentation and parsing for Modern Hebrew sheds new light on the performance of the best parsing systems to date in the different scenarios.

A Comparison of Chinese Parsers for Stanford Dependencies
Wanxiang Che, Valentin Spitkovsky, and Ting Liu
Wednesday 11:40am–12:00pm — Halla A (ICC)

Stanford dependencies are widely used in natural language processing as a semantically-oriented representation, commonly generated either by (i) converting the output of a constituent parser, or (ii) predicting dependencies directly. Previous comparisons of the two approaches for English suggest that starting from constituents yields higher accuracies. In this paper, we re-evaluate both methods for Chinese, using more accurate dependency parsers than in previous work. Our comparison of performance and efficiency across seven popular open source parsers (four constituent and three dependency) shows, by contrast, that recent higher-order graph-based techniques can be more accurate, though somewhat slower, than constituent parsers. We demonstrate also that n-way jackknifing is a useful technique for producing automatic (rather than gold) part-of-speech tags to train Chinese dependency parsers. Finally, we analyze the relations produced by both kinds of parsing and suggest which specific parsers to use in practice.

A Feature-Rich Constituent Context Model for Grammar Induction
Dave Golland, John DeNero, and Jakob Uszkoreit
Wednesday 12:00pm–12:20pm — Halla A (ICC)

We present LLCCM, a log-linear variant of the constituent context model (CCM) of grammar induction. LLCCM retains the simplicity of the original CCM but extends robustly to long sentences. On sentences of up to length 40, LLCCM outperforms CCM by 13.9% bracketing F1 and outperforms a right-branching baseline in regimes where CCM does not.

Private Access to Phrase Tables for Statistical Machine Translation
Nicola Cancedda
Wednesday 11:00am–11:20am — Halla B (ICC)

Some Statistical Machine Translation systems never see the light because the owner of the appropriate training data cannot release them, and the potential user of the system cannot disclose what should be translated. We propose a simple and practical encryption-based method addressing this barrier.

Fast and Scalable Decoding with Language Model Look-Ahead for Phrase-based Statistical Machine Translation
Joern Wuebker, Hermann Ney, and Richard Zens
Wednesday 11:20am–11:40am — Halla B (ICC)

In this work we present two extensions to the well-known dynamic programming beam search in phrase-based statistical machine translation, aiming at increased efficiency of decoding by minimizing the number of language model computations and hypothesis expansions. Our results show that language model based pre-sorting yields a small improvement in translation quality and a speedup by a factor of 2. Two look-ahead methods are shown to further increase translation speed by a factor of 2 without changing the search space and a factor of 4 with the side-effect of some additional search errors. We compare our approach with Moses and observe the same performance, but a substantially better trade-off between translation quality and speed. At a speed of roughly 70 words per second, Moses reaches 17.2% BLEU, whereas our approach yields 20.0% with identical models.

Head-Driven Hierarchical Phrase-based Translation
Junhui Li, Zhaopeng Tu, Guodong Zhou, and Josef van Genabith
Wednesday 11:40am–12:00pm — Halla B (ICC)

This paper presents an extension of Chiang's hierarchical phrase-based (HPB) model, called Head-Driven HPB (HD-HPB), which incorporates head information in translation rules to better capture syntax-driven information, as well as improved reordering between any two neighboring non-terminals at any stage of a derivation to explore a larger reordering search space. Experiments on Chinese-English translation on four NIST MT test sets show that the HD-HPB model significantly outperforms Chiang's model with average gains of 1.91 points absolute in BLEU.

Joint Learning of a Dual SMT System for Paraphrase Generation
Hong Sun and Ming Zhou
Wednesday 12:00pm–12:20pm — Halla B (ICC)

SMT has been used in paraphrase generation by translating a source sentence into another (pivot) language and then back into the source. The resulting sentences can be used as candidate paraphrases of the source sentence. Existing work that uses two independently trained SMT systems cannot directly optimize the paraphrase results. Paraphrase criteria especially the paraphrase rate is not able to be ensured in that way. In this paper, we propose a joint learning method of two SMT systems to optimize the process of paraphrase generation. In addition, a revised BLEU score (called $iBLEU$) which measures the adequacy and diversity of the generated paraphrase sentence is proposed for tuning parameters in SMT systems. Our experiments on NIST 2008 testing data with automatic evaluation as well as human judgments suggest that the proposed method is able to enhance the paraphrase quality by adjusting between semantic equivalency and surface dissimilarity.

A Novel Burst-based Text Representation Model for Scalable Event Detection
Xin Zhao, Rishan Chen, Kai Fan, Hongfei Yan, and Xiaoming Li
Wednesday 11:00am–11:20am — Samda (ICC)

Mining retrospective events from text streams has been an important research topic. Classic text representation model (i.e., vector space model) cannot model temporal aspects of documents. To address it, we proposed a novel burst-based text representation model, denoted as BurstVSM. BurstVSM corresponds dimensions to bursty features instead of terms, which can capture semantic and temporal information. Meanwhile, it significantly reduces the number of non-zero entries in the representation. We test it via scalable event detection, and experiments in a 10-year news archive show that our methods are both effective and efficient.

A Graph-based Cross-lingual Projection Approach for Weakly Supervised Relation Extraction

Seokhwan Kim and Gary Geunbae Lee

Wednesday 11:20am–11:40am — Samda (ICC)

Although researchers have conducted extensive studies on relation extraction in the last decade, supervised approaches are still limited because they require large amounts of training data to achieve high performances. To build a relation extractor without significant annotation effort, we can exploit cross-lingual annotation projection, which leverages parallel corpora as external resources for supervision. This paper proposes a novel graph-based projection approach and demonstrates the merits of it by using a Korean relation extraction system based on projected dataset from an English-Korean parallel corpus.

Pattern Learning for Relation Extraction with a Hierarchical Topic Model

Enrique Alfonseca, Katja Filippova, Jean-Yves Delort, and Guillermo Garrido

Wednesday 11:40am–12:00pm — Samda (ICC)

We describe the user of a hierarchical topic model for automatically identifying syntactic and lexical patterns that explicitly state ontological relations. We use distant supervision using relations from FreeBase, but do not require any manual heuristic nor manual seed list selections. Results show that the learned patterns can be used to extract new relations with good precision.

Self-Disclosure and Relationship Strength in Twitter Conversations

JinYeong Bak, Suin Kim, and Alice Oh

Wednesday 11:00am–11:20am — 401 (ICC)

In social psychology, it is generally accepted that one discloses more of his/her personal information to someone in a strong relationship. We present a computational framework for automatically analyzing such self-disclosure behavior in Twitter conversations. Our framework uses text mining techniques to discover topics, emotions, sentiments, lexical patterns, as well as personally identifiable information (PII) and personally embarrassing information (PEI). Our preliminary results illustrate that in relationships with high relationship strength, Twitter users show significantly more frequent behaviors of self-disclosure.

Genre Independent Subgroup Detection in Online Discussion Threads: A Study of Implicit Attitude using Textual Latent Semantics

Pradeep Dasigi, Weiwei Guo, and Mona Diab

Wednesday 11:20am–11:40am — 401 (ICC)

We describe an unsupervised approach to the problem of automatically detecting sub-groups of people holding similar opinions in a discussion thread. An intuitive way of identifying this is to detect the attitudes of discussants towards each other or named entities or topics mentioned in the discussion. Sentiment tags play an important role in this detection, but we also note another dimension to the detection of people's attitudes in a discussion: if two persons share the same opinion, they tend to use similar language content. We consider the latter to be an implicit attitude. In this paper, we investigate the impact of implicit and explicit attitude in two genres of social media discussion data, more formal Wikipedia discussions and a debate discussion forum that is much more informal. Experimental results strongly suggest that implicit attitude is an important complement for explicit attitudes (expressed via sentiment) and it can improve the sub-group detection performance independent of genre.

Learning to Temporally Order Medical Events in Clinical Text

Preethi Raghavan, Albert Lai, and Eric Fosler-Lussier

Wednesday 11:40am–12:00pm — 401 (ICC)

We investigate the problem of ordering medical events in unstructured clinical narratives by learning to rank them based on their time of occurrence. We represent each medical event as a time duration, with a corresponding start and stop, and learn to rank the starts/stops based on their

proximity to the admission date. Such a representation allows us to learn all of Allen's temporal relations between medical events. Interestingly, we observe that this methodology performs better than a classification-based approach for this domain, but worse on the relationships found in the Timebank corpus. This finding has important implications for styles of data representation and resources used for temporal relation learning: clinical narratives may have different language attributes corresponding to temporal ordering relative to Timebank, implying that the field may need to look at a wider range of domains to fully understand the nature of temporal ordering.

A Context-sensitive, Multi-faceted Model of Lexico-Conceptual Affect
Tony Veale
Wednesday 12:00pm–12:20pm — 401 (ICC)

Since we can 'spin' words and concepts to suit our affective needs, context is a major determinant of the perceived affect of a word or concept. We view this re-profiling as a selective emphasis or de-emphasis of the qualities that underpin our shared stereotype of a concept or a word meaning, and construct our model of the affective lexicon accordingly. We show how a large body of affective stereotypes can be acquired from the web, and also show how these are used to create and interpret affective metaphors.

Decoding Running Key Ciphers
Sravana Reddy and Kevin Knight
Wednesday 11:00am–11:20am — 402 (ICC)

There has been recent interest in the problem of decoding letter substitution ciphers using techniques inspired by natural language processing. We consider a different type of classical encoding scheme known as the running key cipher, and propose a search solution using Gibbs sampling with a word language model. We evaluate our method on synthetic ciphertexts of different lengths, and find that it outperforms previous work that uses Viterbi decoding with character-based language models.

Using Rejuvenation to Improve Particle Filtering for Bayesian Word Segmentation
Benjamin Börschinger and Mark Johnson
Wednesday 11:20am–11:40am — 402 (ICC)

We present a novel extension to a recently proposed incremental learning algorithm for the word segmentation problem originally introduced in Goldwater (2006). By adding rejuvenation to a particle filter, we are able to considerably improve its performance, both in terms of finding higher probability and higher accuracy solutions.

Baselines and Bigrams: Simple, Good Sentiment and Topic Classification
Sida Wang and Christopher Manning
Wednesday 11:40am–12:00pm — 402 (ICC)

Variants of Naive Bayes (NB) and Support Vector Machines (SVM) are often used as baseline methods for text classification, but their performance varies greatly depending on the model variant, features used and task/dataset. We show that:

(i) the inclusion of word bigram features gives consistent gains on sentiment analysis tasks; (ii) for short snippet sentiment tasks, NB actually does better than SVMs (while for longer documents the opposite result holds); (iii) a simple but novel SVM variant using NB log-count ratios as feature values consistently performs well across tasks and datasets.

Based on these observations, we identify simple NB and SVM variants which outperform most published results on sentiment analysis datasets, sometimes providing a new state-of-the-art performance level.

Automatically Learning Measures of Child Language Development
Sam Sahakian and Benjamin Snyder
Wednesday 12:00pm–12:20pm — 402 (ICC)

We propose a new approach for the creation of child language development metrics. A set of linguistic features is computed on child speech samples and used as input in two age prediction experiments. In the first experiment, we learn a child-specific metric and predicts the ages at which speech samples were produced. We then learn a more general developmental index by applying our method across children, predicting relative temporal orderings of speech samples. In both cases we compare our results with established measures of language development, showing improvements in age prediction performance.

A Comparative Study of Target Dependency Structures for Statistical Machine Translation
Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, and Masaaki Nagata
Wednesday 4:00pm–4:20pm — Halla A (ICC)

This paper presents a comparative study of target dependency structures yielded by several state-of-the-art linguistic parsers. Our approach is to measure the impact of these non-isomorphic dependency structures to be used for string-to-dependency translation. Besides using traditional dependency parsers, we also use the dependency structures transformed from PCFG trees and predicate-argument structures (PASs) which are generated by an HPSG parser and a CCG parser. The experiments on Chinese-to-English translation show that the HPSG parser's PASs achieved the best dependency and translation accuracies.

Robust Conversion of CCG Derivations to Phrase Structure Trees
Jonathan K. Kummerfeld, Dan Klein, and James R. Curran
Wednesday 4:20pm–4:40pm — Halla A (ICC)

We propose an improved, bottom-up method for converting CCG derivations into PTB-style phrase structure trees. In contrast with past work (Clark and Curran, 2009), which used simple transductions on category pairs, our approach uses richer transductions attached to single categories. Our conversion preserves more sentences under round-trip conversion (51.1% vs. 39.6%) and is more robust. In particular, unlike past methods, ours does not require ad-hoc rules over non-local features, and so can be easily integrated into a parser.

Estimating Compact Yet Rich Tree Insertion Grammars
Elif Yamangil and Stuart Shieber
Wednesday 4:40pm–5:00pm — Halla A (ICC)

We present a Bayesian nonparametric model for estimating tree insertion grammars (TIG), building upon recent work in Bayesian inference of tree substitution grammars (TSG) via Dirichlet processes. Under our general variant of TIG, grammars are estimated via the Metropolis-Hastings algorithm that uses a context free grammar transformation as a proposal, which allows for cubic-time string parsing as well as tree-wide joint sampling of derivations in the spirit of Cohn and Blunsom (2010). We use the Penn treebank for our experiments and find that our proposal Bayesian TIG model not only has competitive parsing performance but also finds compact yet linguistically rich TIG representations of the data.

Topic Models for Dynamic Translation Model Adaptation
Vladimir Eidelman, Jordan Boyd-Graber, and Philip Resnik
Wednesday 5:00pm–5:20pm — Halla A (ICC)

We propose an approach that biases machine translation systems toward relevant translations based on topic-specific contexts, where topics are induced in an unsupervised way using topic models; this can be thought of as inducing subcorpora for adaptation without any human annotation. We use these topic distributions to compute topic-dependent lexical weighting probabilities and directly incorporate them into our translation model as features. Conditioning lexical

probabilities on the topic biases translations toward topic relevant output, resulting in significant improvements of up to 1 BLEU and 3 TER on Chinese to English translation over a strong baseline.

Detecting Semantic Equivalence and Information Disparity in Cross-lingual Documents
Yashar Mehdad, Matteo Negri, and Marcello Federico
Wednesday 4:00pm–4:20pm — Samda (ICC)

We address a core aspect of the multilingual content synchronization task: the identification of novel, more informative or semantically equivalent pieces of information in two documents about the same topic. This can be seen as an application-oriented variant of textual entailment recognition where: i) T and H are in different languages, and ii) entailment relations between T and H have to be checked in both directions. Using a combination of lexical, syntactic, and semantic features to train a cross-lingual textual entailment system, we report promising results on different datasets.

Cross-lingual Parse Disambiguation based on Semantic Correspondence
Lea Frermann and Francis Bond
Wednesday 4:20pm–4:40pm — Samda (ICC)

We present a system for cross-lingual parse disambiguation, exploiting the assumption that the meaning of a sentence remains unchanged during translation and the fact that different languages have different ambiguities. We simultaneously reduce ambiguity in multiple languages in a fully automatic way. Evaluation shows that the system reliably discards dispreferred parses from the raw parser output, which results in a pre-selection that can speed up manual treebanking.

Learning to Find Translations and Transliterations on the Web
Joseph Z. Chang, Jason S. Chang, and Roger Jyh-Shing Jang
Wednesday 4:40pm–5:00pm — Samda (ICC)

In this paper, we present a new method for learning to finding translations and transliterations on the Web for a given term. The approach involves using a small set of terms and translations to obtain mixed-code snippets from a search engine, and automatically annotating the snippets with tags and features for training a conditional random field model. At runtime,the model is used to extracting translation candidates for a given term. Preliminary experiments and evaluation show our method cleanly combining various features, resulting in a system that outperforms previous work.

Beefmoves: Dissemination, Diversity, and Dynamics of English Borrowings in a German Hip Hop Forum
Matt Garley and Julia Hockenmaier
Wednesday 5:00pm–5:20pm — Samda (ICC)

We investigate how novel English-derived words (anglicisms) are used in a German-language Internet hip hop forum, and what factors contribute to their uptake.

Learning the Latent Semantics of a Concept from its Definition
Weiwei Guo and Mona Diab
Wednesday 4:00pm–4:20pm — Halla B (ICC)

In this paper we study unsupervised word sense disambiguation based on sense definition. We learn low-dimensional latent semantic vectors of concept definitions to construct a more robust sense similarity measure *wmfvec*. Experiments on all-words WSD tasks show significant improvement of the sense similarity measure over traditional similarity measures and LDA based vectors.

Unsupervised Semantic Role Induction with Global Role Ordering
Nikhil Garg and James Henserdon
Wednesday 4:20pm–4:40pm — Halla B (ICC)

We propose a probabilistic generative model for unsupervised semantic role induction, which integrates local role assignment decisions and a global role ordering decision in a unified model. The role sequence is divided into intervals based on the notion of primary roles, and each interval generates a sequence of secondary roles and syntactic constituents using local features. The global role ordering consists of the sequence of primary roles only, thus making it a partial ordering.

Humor as Circuits in Semantic Networks
Igor Labutov and Hod Lipson
Wednesday 4:40pm–5:00pm — Halla B (ICC)

This work presents a first step to a general implementation of the Semantic-Script Theory of Humor (SSTH). Of the scarce amount of research in computational humor, no research had focused on humor generation beyond simple puns and punning riddles. We propose an algorithm for mining simple humorous scripts from a semantic network (ConceptNet) by specifically searching for dual scripts that jointly maximize overlap and incongruity metrics in line with Raskin's Semantic-Script Theory of Humor. Initial results show that a more relaxed constraint of this form is capable of generating humor of deeper semantic content than wordplay riddles. We evaluate the said metrics through a user-assessed quality of the generated two-liners.

Crowdsourcing Inference-Rule Evaluation
Naomi Zeichner, Jonathan Berant, and Ido Dagan
Wednesday 5:00pm–5:20pm — Halla B (ICC)

The importance of inference rules to semantic applications has long been recognized and extensive work has been carried out to automatically acquire inference-rule resources. However, evaluating such resources has turned out to be a non-trivial task, slowing progress in the field. In this paper, we suggest a framework for evaluating inference-rule resources. Our framework simplifies a previously proposed "instance-based evaluation" method that involved substantial annotator training, making it suitable for crowdsourcing. We show that our method produces a large amount of annotations with high inter-annotator agreement for a low cost at a short period of time, without requiring training expert annotators.

A Comprehensive Gold Standard for the Enron Organizational Hierarchy
Apoorv Agarwal, Adinoyi Omuya, Aaron Harnly, and Owen Rambow
Wednesday 4:00pm–4:20pm — 402 (ICC)

Many researchers have attempted to predict the Enron corporate hierarchy from the data. This work, however, has been hampered by a lack of data. We present a new, large, and freely available gold-standard hierarchy. Us- ing our new gold standard, we show that a simple lower bound for social network-based systems outperforms an upper bound on the approach taken by current NLP systems.

A Two-step Approach to Sentence Compression of Spoken Utterances
Dong Wang, Xian Qian, and Yang Liu
Wednesday 4:20pm–4:40pm — 402 (ICC)

This paper presents a two-step approach to compress spontaneous spoken utterances. In the first step, we use a sequence labeling method to determine if a word in the utterance can be removed, and generate n-best compressed sentences. In the second step, we use a discriminative training approach to capture sentence level global information from the candidates and rerank them. For evaluation, we compare our system output with multiple human references. Our results show that the new features we introduced in the first compression step improve performance upon the previous work on the same data set, and reranking is able to yield additional gain, especially

when training is performed to take into account multiple references.

Syntactic Stylometry for Deception Detection
Song Feng, Ritwik Banerjee, and Yejin Choi
Wednesday 4:40pm–5:00pm — 402 (ICC)

Most previous studies in computerized deception detection have relied only on shallow lexico-syntactic patterns. This paper investigates syntactic stylometry for deception detection, adding a somewhat unconventional angle to prior literature. Over four different datasets spanning from the product review to the essay domain, we demonstrate that features driven from Context Free Grammar (CFG) parse trees consistently improve the detection performance over several baselines that are based only on shallow lexico-syntactic features. Our results improve the best published result on the hotel review data (Ott et al.,2011) reaching 91.2% accuracy with 14% error reduction.

Transforming Standard Arabic to Colloquial Arabic
Emad Mohamed, Behrang Mohit, and Kemal Oflazer
Wednesday 5:00pm–5:20pm — 402 (ICC)

We present a method for generating Colloquial Egyptian Arabic (CEA) from morphologically disambiguated Modern Standard Arabic (MSA). When used in POS tagging, this process improves the accuracy from 73.24% to 86.84% on unseen CEA text, and reduces the percentage of out-of-vocabulary words from 28.98% to 16.66%. The process holds promise for any NLP task targeting the dialectal varieties of Arabic; e.g., this approach may provide a cheap way to leverage MSA data and morphological resources to create resources for colloquial Arabic to English machine translation. It can also considerably speed up the annotation of Arabic dialects.

Corpus-based Interpretation of Instructions in Virtual Environments
Luciana Benotti, Martin Villalba, Tessa Lau, and Julian Cerruti
Wednesday 4:00pm–4:20pm — 401 (ICC)

Previous approaches to instruction interpretation have required either extensive domain adaptation or manually annotated corpora. This paper presents a novel approach to instruction interpretation that leverages a large amount of unannotated, easy-to-collect data from humans interacting with a virtual world. We compare several algorithms for automatically segmenting and discretizing this data into (utterance, reaction) pairs and training a classifier to predict reactions given the next utterance. Our empirical analysis shows that the best algorithm achieves 70required.

Automatically Mining Question Reformulation Patterns from Search Log Data
Xiaobing Xue, Yu Tao, Daxin Jiang, and Hang Li
Wednesday 4:20pm–4:40pm — 401 (ICC)

Natural language questions have become popular in web search. However, various questions can be formulated to convey the same information need, which poses a great challenge to search systems. In this paper, we automatically mined *5w1h question reformulation patterns* from large scale search log data. The question reformulations generated from these patterns are further incorporated into the retrieval model. Experiments show that using question reformulation patterns can significantly improve the search performance of natural language questions.

Native Language Detection with Tree Substitution Grammars
Benjamin Swanson and Eugene Charniak
Wednesday 4:40pm–5:00pm — 401 (ICC)

We investigate the potential of Tree Substitu- tion Grammars as a source of features for na- tive language detection, the task of inferring an author's native language from text in a dif- ferent language. We compare two state of the art methods for Tree Substitution Grammar induction and show that features from both methods outperform previous state of the art results at native

language detection. Further- more, we contrast these two induction algo- rithms and show that the Bayesian approach produces superior classification results with a smaller feature set.

Tense and Aspect Error Correction for ESL Learners Using Global Context

Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto

Wednesday 5:00pm–5:20pm — 401 (ICC)

As the number of learners of English is constantly growing, automatic error correction of ESL learners' writing is an increasingly active area of research. However, most research has mainly focused on errors concerning articles and prepositions even though tense/aspecterrors are also important. One of the main reasons why tense/aspect error correction is difficult is that the choice of tense/aspect is highly dependent on global context. Previous research on grammatical error correction typically uses pointwise prediction that performs classification on each word independently, and thus fails to capture the information of neighboring labels. In order to take global information into account, we regard the task as sequence labeling: each verb phrase in a document is labeled with tense/aspect depending on surrounding labels. Our experiments show that the global context makes a moderate contribution to tense/aspect error correction.

8

Thursday, July 12, 2012: ACL Workshops and EMNLP-CoNLL

Overview

EMNLP-CoNLL **- Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning**
Chairs: Jun'ichi Tsujii, James Henderson, Marius Paşca

The LAW VI **- The 6th Linguistic Annotation Workshop**
Chairs: Nancy Ide, Fei Xia

SP-Sem-MRL2012 **- Parsing and Semantic Processing of Morphologically Rich Languages 2012**
Chairs: Marianna Apidianaki, Ido Dagan, Jennifer Foster, Yuval Marton, Djamé Seddah, Reut Tsarfaty

SMIAE 2012 **- Speech and Multimodal Interaction in Assistive Environments**
Chairs: Dimitra Anastasiou, Desislava Zhekova, Cui Jian, Robert Ross

WASSA2012 **- 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis**
Chairs: Alexandra Balahur, Andres Montoyo, Patricio Martinez-Barco, Ester Boldrini

SSST-6 **- Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation**
Chairs: Marine Carpuat, Lucia Specia, Dekai Wu

DSSD **- Detecting Structure in Scholarly Discourse**
Chairs: Sophia Ananiadou, Antal van den Bosch, Ágnes Sándor, Hagit Shatkay, Anita de Waard

NEWS2012 **- 4th Named Entities Workshop**
Chairs: Min Zhang, A Kumaran, Haizhou Li

Schedule*

7:30am – 5:00pm	**Registration (3F Lobby (ICC))**

Session **Venue**	**EMNLP-CoNLL Oral A** **Youngju A (ICC)**	**EMNLP-CoNLL Oral B** **203 (ICC)**	**EMNLP-CoNLL Oral C** **202 A (ICC)**	**EMNLP-CoNLL Oral D** **202 B (ICC)**	**Workshops**
9:00am – 10:30am	Opening Remarks & Invited Talk: Eric Xing (Youngju A (ICC))				**Morning Session 1**
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**				
11:00am – 12:30pm	**Machine Translation: Bilingual Lexicons and Alignment**	**Social Media: Author Style and Attribution**	**Dialogue and Interactive Systems**	**Information Extraction: Entity Disambiguation**	**Morning Session 2**
12:30pm – 2:00pm	**Lunch**				
2:00pm – 3:30pm	**Sentiment Analysis**	**Semantics: Nouns, Verbs and Predicates**	**Machine Learning: Latent Models**	**Summarization**	**Afternoon Session 1**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**				
4:00pm – 5:30pm	**Machine Translation**	**Dependency Parsing**	**Phonemes, Words and Speech**	**Question Answering**	**Afternoon Session 2**
5:30pm – 6:00pm	**Break**				
6:00pm – 10:00pm	Poster Session and Reception (3F Lobby (ICC))				

*This is an *approximate* summary. Please refer to the workshop sections for detailed schedules.

9

Friday, July 13, 2012: ACL Workshops and EMNLP-CoNLL

Overview

EMNLP-CoNLL **- Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning**
Chairs: Jun'ichi Tsujii, James Henderson, Marius Paşca

The LAW VI **- The 6th Linguistic Annotation Workshop**
Chairs: Nancy Ide, Fei Xia

ExProM **- Extra-propositional Aspects of Meaning in Computational Linguistics**
Chairs: Roser Morante, Caroline Sporleder

MM **- Multilingual Modeling**
Chairs: Jagadeesh Jagarlamudi, Sujith Ravi, Xiaojun Wan, Hal Daume III

The People's Web meets NLP **- Collaboratively Constructed Semantic Resources and their Applications to NLP**
Chairs: Iryna Gurevych, Nicoletta Calzolari Zamorani, Jungi Kim

TextGraphs-7 **- Graph-based Methods for Natural Language Processing**
Chairs: Ahmed Hassan, Gaël Dias, Irina Matveeva

Schedule*

Session Venue	EMNLP-CoNLL Oral A Youngju A (ICC)	EMNLP-CoNLL Oral B 203 (ICC)	EMNLP-CoNLL Oral C 202 A (ICC)	EMNLP-CoNLL Oral D 202 B (ICC)	Workshops
9:00am – 10:30am	Invited Talk: Patrick Pantel (Youngju A (ICC))				**Morning Session 1**
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**				
11:00am – 12:30pm	**Machine Translation: Role of Syntax**	**Information Extraction: Temporally-Aware Extraction**	**Discourse and Generation**	**CoNLL Shared Task: Oral Presentations**	**Morning Session 2**
12:30pm – 2:00pm	**Lunch & SIGDAT and SIGNLL Business Meeting (Youngju A (ICC))**				
2:00pm – 3:30pm	**Semantics: Words and Topics**	**Machine Translation: Pruning**	**Evaluation**	**CoNLL Shared Task: Poster Presentations**	**Afternoon Session 1**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**				
4:00pm – 5:30pm	**Information Extraction: Relation and Event Extraction**	**Parsing Models and Evaluation**	**Large-Scale NLP Algorithms**	**Machine Learning: Inference**	**Afternoon Session 2**

*This is an *approximate* summary. Please refer to the workshop sections for detailed schedules.

10

Saturday, July 14, 2012: EMNLP-CoNLL

Overview

***EMNLP-CoNLL* - Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning**
Chairs: Jun'ichi Tsujii, James Henderson, Marius Paşca

Schedule

Session Venue	EMNLP-CoNLL Oral A Youngju A (ICC)	EMNLP-CoNLL Oral B 203 (ICC)	EMNLP-CoNLL Oral C 202 A (ICC)	EMNLP-CoNLL Oral D 202 B (ICC)
9:00am – 10:30am	Plenary Session (Youngju A (ICC))			
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**			
11:00am – 12:30pm	**Machine Translation: Decoding**	**Distributional and Compositional Semantics**	**Discourse: Coreference Resolution**	**Information Retrieval**
12:30pm – 2:00pm	**Lunch**			
2:00pm – 3:30pm	**Machine Learning: Transfer and Biases**	**Opinion Mining: Discovering Opinion Expressions**	**Part of Speech Tagging**	**Word Sense Disambiguation**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**			
4:00pm – 5:30pm	**Syntax and Parsing: Joint Parsing Models**	**Social Media**	**NLP Applications**	
5:30pm – 5:45pm	**Closing Remarks**			

11

EMNLP-CoNLL

Preface by General and Program Chairs

It is our pleasure to welcome you to the EMNLP-CoNLL 2012 conference, a joint meeting of the Conference on Empirical Methods in Natural Language Learning (EMNLP) and the Conference on Computational Natural Language Learning (CoNLL). After the successful first collaboration in 2007, EMNLP-CoNLL 2012 is being jointly organized by the SIGDAT and SIGNLL special interest groups of the Association of Computational Linguistics.

This time, EMNLP-CoNLL is co-located with, and immediately after ACL's 50th anniversary conference. The choice of the location is an opportunity for the ACL community to return to the beautiful Jeju Island, Korea, following a seven-year hiatus since the Second International Joint Conference on Natural Language Processing (IJCNLP 2005) was held here.

Out of 606 submissions received by EMNLP-CoNLL this year, a total of 36 submissions were eventually withdrawn or rejected without review. From the remaining submissions, 99 were accepted for oral presentation and 40 for poster presentation, for a combined acceptance rate of 24.8%. As in recent editions of EMNLP, authors were given the opportunity to provide supplementary material in conjunction with their submissions, which the program committee could but was not required to take into account during reviewing. Also as in recent editions, authors of accepted papers were offered an additional page in the camera-ready version of their submissions, so that comments received from reviewers could be more easily addressed.

The papers submitted to the conference were subject to a rigorous reviewing process, made possible by efforts of a team of 525 primary and 66 secondary reviewers, acting under the guidance of 22 area chairs. Presence of unsupported claims, or failure to properly compare with previous work, were likely serious obstacles, on the path from initial submission to acceptance and then publication in our proceedings. Luckily, our preface has a guaranteed placement in the proceedings. Therefore, without access to insider data and impressions from previous editions of the conference, we will still go on a limb here, and make the unsupported claim that our team of area chairs has been the greatest. That their expertise, dedication and willingness to go beyond the call of duty had a positive impact on a timely reviewing process and a high-quality conference program, would be an understatement. It has been a pleasure to interact and work with our area chairs.

The schedule of our conference is strengthened by two invited speakers, Eric Xing and Patrick Pantel, who we were very happy to have accept our invitation; and by the CoNLL Shared Task, an annual tradition for the CoNLL conferences. This year's CoNLL Shared Task is Modeling Multilingual Unrestricted Coreference in OntoNotes, and its proceedings and detailed schedule are available separately.

We would like to thank all authors who submitted to our conference, for their willingness to share their knowledge with the rest of us. It may take a few weeks or many years, for the knowledge distilled into the present proceedings to have a measurable impact on our field and beyond. An impact that would not be possible without countless hours spent by authors, from developing ideas to running experiments to building usable systems - steps that often fail, and sometimes succeed.

We would also like to thank all members of the program committee, for their willingness to offer feedback that sometimes reaches extraordinary levels of detail and value to authors. To recognise some of the most dedicated reviewers, we include the Best Reviewer awards a little later in these proceedings.

Naoaki Okazaki, Publications Chair, deserves our special thanks. He brought in a healthy dose of rigor to the planning and preparation of not only this proceedings but also conference materials, matched only by his dedication to deliver under tight scheduling constraints.

If the combination of oral presentations, posters and invited talks that make up EMNLP-CoNLL 2012 is considered a success, it is because it benefited from the touch of many people. Francesco Figari easily kept tabs on our salvos of large and small requests for updates to our conference website. Rich Gerber, Paolo Gai and the larger team managing the conference submission system were quick to offer answers to all our questions. The publication chairs and local arrangements committee of ACL 2012, including Michael White, Maggie Li, Jong Park and es-

pecially Gary Geunbae Lee, covered significant tasks on behalf of EMNLP-CoNLL, all with a smile. Chin-Yew Lin, Miles Osborne, Eric Fosler-Lussier, Dekang Lin, Rada Mihalcea, Regina Barzilay, Ulrich Germann and David Yarowsky offered high-level advice or answered detailed questions, drawing upon their experience as organizers of previous ACL-sponsored conferences.

We are grateful to our sponsors (Baidu, Google and Microsoft), for their support of best paper awards and support of student travel in particular, and the financial well-being of the conference in general. It has been an honor to be of service to the conference, for which we would like to thank the community and those who offered us this opportunity. We hope that you enjoy the conference, and have a productive and pleasant stay in South Korea!

Jun'ichi Tsujii, General Chair

James Henderson and Marius Paşca, Program Chairs

EMNLP-CoNLL Organizing Committee

General Chair
Jun'ichi Tsujii (Microsoft Research Asia)

Program Chairs
James Henderson (Xerox Research Centre Europe)
Marius Paşca (Google)

Publication Chair
Naoaki Okazaki (Tohoku University)

Area Chairs
Eugene Agichtein (Emory University)
Yejin Choi (Stony Brook University)
Hal Daumé III (University of Maryland)
Doug Downey (Northwestern University)
Chris Dyer (Carnegie Mellon University)
Adrià de Gispert (University of Cambridge)
Julia Hirschberg (Columbia University)
Anna Korhonen (University of Cambridge)
Bing Liu (University of Illinois at Chicago)
Hwee Tou Ng (National University of Singapore)
Patrick Pantel (Microsoft Research)
Marco Pennacchiotti (Yahoo! Labs)
Slav Petrov (Google)
Simone Paolo Ponzetto (Sapienza University of Rome)
John Prager (IBM Research)
Chris Quirk (Microsoft Research)
Sebastian Riedel (University of Massachusetts at Amherst)
Hiroya Takamura (Tokyo Institute of Technology)
Partha Talukdar (Carnegie Mellon University)
Kristina Toutanova (Microsoft Research)
Reut Tsarfaty (Uppsala University)
Marilyn Walker (University of California at Santa Cruz)

Program

Day 1 (July 12, 2012)

7:30am – 5:00pm	**Registration**	**(3F Lobby (ICC))**
9:00am – 9:15am	**Opening Remarks**	**(Youngju A (ICC))**
9:15am – 10:30am	**Invited Talk: Eric Xing**	**(Youngju A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Machine Translation: Bilingual Lexicons and Alignment**	**(Youngju A (ICC))**
	Social Media: Author Style and Attribution	**(203 (ICC))**
	Dialogue and Interactive Systems	**(202 A (ICC))**
	Information Extraction: Entity Disambiguation	**(202 B (ICC))**
12:30pm – 2:00pm	**Lunch**	
2:00pm – 3:30pm	**Sentiment Analysis**	**(Youngju A (ICC))**
	Semantics: Nouns, Verbs and Predicates	**(203 (ICC))**
	Machine Learning: Latent Models	**(202 A (ICC))**
	Summarization	**(202 B (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:30pm	**Machine Translation**	**(Youngju A (ICC))**
	Dependency Parsing	**(203 (ICC))**
	Phonemes, Words and Speech	**(202 A (ICC))**
	Question Answering	**(202 B (ICC))**
5:30pm – 6:00pm	**Break**	
6:00pm – 10:00pm	**Poster Session and Reception**	**(3F Lobby (ICC))**

Schedule (July 12, 2012)

7:30am – 5:00pm	**Registration (3F Lobby (ICC))**

9:00am – 9:15am	**Opening Remarks (Youngju A (ICC))**			
9:15am – 10:30am	**Invited Talk (Youngju A (ICC))** ***Eric Xing:* "On Learning Sparse Structured Input-Output Models"** **Chair:** *James Henderson*			
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
11:00am – 12:30pm	**Machine Translation: Bilingual Lexicons and Alignment** **Chair:** *David Chiang*	**Social Media: Author Style and Attribution** **Chair:** *Janyce Wiebe*	**Dialogue and Interactive Systems** **Chair:** *Michael White*	**Information Extraction: Entity Disambiguation** **Chair:** *Raymond Mooney*
11:00am – 11:30am	Syntactic Transfer Using a Bilingual Lexicon *G. Durrett, A. Pauls, and D. Klein* . . . . . **p. 158**	Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls *K. Hong, C. G. Kohler, M. E. March, A. A. Parker, and A. Nenkova* . . . . . **p. 158**	Generative Goal-Driven User Simulation for Dialog Management *A. Eshky, B. Allison, and M. Steedman* . . . . . **p. 159**	An Entity-Topic Model for Entity Linking *X. Han and L. Sun* . . . . . **p. 160**
11:30am – 12:00pm	Regularized Interlingual Projections: Evaluation on Multilingual Transliteration *J. Jagarlamudi and H. Daume III* . . . . . **p. 158**	Streaming Analysis of Discourse Participants *B. Van Durme* . . . . . **p. 159**	Optimising Incremental Dialogue Decisions Using Information Density for Interactive Systems *N. Dethlefs, H. Hastie, V. Rieser, and O. Lemon* . . . . . **p. 160**	Linking Named Entities to Any Database *A. Sil, E. Cronin, P. Nie, Y. Yang, A.-M. Popescu, and A. Yates* . . . . . **p. 160**
12:00pm – 12:30pm	Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation *A. Tamura, T. Watanabe, and E. Sumita* . . . . . **p. 158**	Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants *A. Hassan, A. Abu-Jbara, and D. Radev* . . . . . **p. 159**	Mixed Membership Markov Models for Unsupervised Conversation Modeling *M. J. Paul* . . . . . **p. 160**	Towards Efficient Named-Entity Rule Induction for Customizability *A. Nagesh, G. Ramakrishnan, L. Chiticariu, R. Krishnamurthy, A. Dharkar, and P. Bhattacharyya* . . . . . **p. 161**
12:30pm – 2:00pm	**Lunch Break**			

Parallel Sessions	Youngju A (ICC)	203 (ICC)	202 A (ICC)	202 B (ICC)
2:00pm – 3:30pm	**Sentiment Analysis** **Chair:** *Bing Liu*	**Semantics: Nouns, Verbs and Predicates** **Chair:** *Rada Mihalcea*	**Machine Learning: Latent Models** **Chair:** *Alexandre Klementiev*	**Summarization** **Chair:** *Dragomir Radev*
2:00pm – 2:30pm	Active Learning for Imbalanced Sentiment Classification *S. Li, S. Ju, G. Zhou, and X. Li* ..... **p. 161**	Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering *M. Roth and A. Frank* .... **p. 162**	Spectral Dependency Parsing with Latent Variables *P. Dhillon, J. Rodu, M. Collins, D. Foster, and L. Ungar* .. **p. 163**	Multiple Aspect Summarization Using Integer Linear Programming *K. Woodsend and M. Lapata* ..... **p. 163**
2:30pm – 3:00pm	A Weakly Supervised Model for Sentence-Level Semantic Orientation Analysis with Multiple Experts *L. Qu, R. Gemulla, and G. Weikum* ..... **p. 161**	Local and Global Context for Supervised and Unsupervised Metonymy Resolution *V. Nastase, A. Judea, K. Markert, and M. Strube* ..... **p. 162**	A Phrase-Discovering Topic Model Using Hierarchical Pitman-Yor Processes *R. Lindsey, W. Headden, and M. Stipicevic* ..... **p. 163**	Minimal Dependency Length in Realization Ranking *M. White and R. Rajkumar* **p. 164**
3:00pm – 3:30pm	Collocation Polarity Disambiguation Using Web-based Pseudo Contexts *Y. Zhao, B. Qin, and T. Liu* **p. 162**	Learning Verb Inference Rules from Linguistically-Motivated Evidence *H. Weisman, J. Berant, I. Szpektor, and I. Dagan* ..... **p. 162**	A Bayesian Model for Learning SCFGs with Discontiguous Rules *A. Levenberg, C. Dyer, and P. Blunsom* ..... **p. 163**	Framework of Automatic Text Summarization Using Reinforcement Learning *S. Ryang and T. Abekawa* . **p. 164**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
4:00pm – 5:30pm	**Machine Translation** **Chair:** *Xiong Deyi*	**Dependency Parsing** **Chair:** *Emily Pitler*	**Phonemes, Words and Speech** **Chair:** *Pascale Fung*	**Question Answering** **Chair:** *Alessandro Moschitti*
4:00pm – 4:30pm	Large Scale Decipherment for Out-of-Domain Machine Translation *Q. Dou and K. Knight* .... **p. 164**	Exploiting Reducibility in Unsupervised Dependency Parsing *D. Mareček and Z. Žabokrtský* ..... **p. 165**	Universal Grapheme-to-Phoneme Prediction Over Latin Alphabets *Y.-B. Kim and B. Snyder* .. **p. 166**	Why Question Answering using Sentiment Analysis and Word Classes *J.-H. Oh, K. Torisawa, C. Hashimoto, T. Kawada, S. De Saeger, J. Kazama, and Y. Wang* ..... **p. 166**
4:30pm – 5:00pm	N-gram-based Tense Models for Statistical Machine Translation *Z. Gong, M. Zhang, C. L. Tan, and G. Zhou* ..... **p. 164**	Improving Transition-Based Dependency Parsing with Buffer Transitions *D. Fernández-González and C. Gómez-Rodríguez* ..... **p. 165**	Name Phylogeny: A Generative Model of String Variation *N. Andrews, J. Eisner, and M. Dredze* ..... **p. 166**	Natural Language Questions for the Web of Data *M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum* .. **p. 167**

5:00pm – 5:30pm	Source Language Adaptation for Resource-Poor Machine Translation *P. Wang, P. Nakov, and H. T. Ng* ..... **p. 165**	Generalized Higher-Order Dependency Parsing with Cube Pruning *H. Zhang and R. McDonald* ..... **p. 165**	Syntactic Surprisal Affects Spoken Word Duration in Conversational Contexts *V. Demberg, A. Sayeed, P. Gorinski, and N. Engonopoulos* ..... **p. 166**	Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews *J. Yu, Z.-J. Zha, and T.-S. Chua* ..... **p. 167**
6:00pm – 10:00pm	**— 3F Lobby (ICC) —** **Poster Session and Reception**			

Day 2 (July 13, 2012)

9:15am – 10:30pm	**Invited Talk: Patrick Pantel**	**(Youngju A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Machine Translation: Role of Syntax**	**(Youngju A (ICC))**
	Information Extraction: Temporally-Aware Extraction	**(203 (ICC))**
	Discourse and Generation	**(202 A (ICC))**
	CoNLL Shared Task: Oral Presentations	**(202 B (ICC))**
12:30pm – 1:45pm	**Lunch**	
1:45pm – 2:30pm	**SIGDAT and SIGNLL Business Meetings**	**(Youngju A (ICC))**
2:30pm – 3:30pm	**Semantics: Words and Topics**	**(Youngju A (ICC))**
	Machine Translation: Pruning	**(203 (ICC))**
	Evaluation	**(202 A (ICC))**
	CoNLL Shared Task: Poster Presentations	**(202 B (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:30pm	**Information Extraction: Relation and Event Extraction**	**(Youngju A (ICC))**
	Parsing Models and Evaluation	**(203 (ICC))**
	Large-Scale NLP Algorithms	**(202 A (ICC))**
	Machine Learning: Inference	**(202 B (ICC))**

Schedule (July 13, 2012)

7:30am – 5:00pm	Registration (3F Lobby (ICC))

9:15am – 10:30am	Invited Talk (Youngju A (ICC)) *Patrick Pantel:* **"The Appification of the Web and the Renaissance of Conversational User Interfaces"** **Chair:** *Marius Pasca*			
10:30am – 11:00am	**Coffee Break (3F Lobby (ICC))**			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
11:00am – 12:30pm	**Machine Translation: Role of Syntax** **Chair:** *Qun Liu*	**Information Extraction: Temporally-Aware Extraction** **Chair:** *Jong-Hoon Oh*	**Discourse and Generation** **Chair:** *Anette Frank*	**CoNLL Shared Task: Oral Presentations** **Chair:** *Sameer Pradhan*
11:00am – 11:30am	Inducing a Discriminative Parser to Optimize Machine Translation Reordering *G. Neubig, T. Watanabe, and S. Mori* ..... **p. 179**	Learning Constraints for Consistent Timeline Extraction *D. McClosky and C. D. Manning* ..... **p. 179**	A Novel Discriminative Framework for Sentence-Level Discourse Analysis *S. Joty, G. Carenini, and R. Ng* ..... **p. 180**	Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution *E. Fernandes, C. dos Santos, and R. Milidiú* ..... **p. 196**
11:30am – 12:00pm	Re-training Monolingual Parser Bilingually for Syntactic SMT *S. Liu, C.-H. Li, M. Li, and M. Zhou* ..... **p. 179**	Identifying Constant and Unique Relations by using Time-Series Text *Y. Takaku, N. Kaji, N. Yoshinaga, and M. Toyoda* ..... **p. 180**	Using Discourse Information for Paraphrase Extraction *M. Regneri and R. Wang* .. **p. 180**	Data-driven Multilingual Coreference Resolution using Resolver Stacking *A. Björkelund and R. Farkas* ..... **p. 196**
12:00pm – 12:30pm	Transforming Trees to Improve Syntactic Convergence *D. Burkett and D. Klein* .. **p. 179**	No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities *T. Lin, . Mausam, and O. Etzioni* ..... **p. 180**	Generating Non-Projective Word Order in Statistical Linearization *B. Bohnet, A. Björkelund, J. Kuhn, W. Seeker, and S. Zarriess* **p. 180**	Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution *C. Chen and V. Ng* ..... **p. 196** Using Syntactic Dependencies to Solve Coreferences *M. Stamborg, D. Medved, P. Exner, and P. Nugues* ..... **p. 196**
12:30pm – 1:45pm	**Lunch Break**			

1:45pm – 2:30pm	SIGDAT and SIGNLL Business Meetings (Youngju A (ICC))			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
2:30pm – 3:30pm	**Semantics: Words and Topics** **Chair:** *Roi Reichart*	**Machine Translation: Pruning** **Chair:** *Kevin Knight*	**Evaluation** **Chair:** *Billy Tak-Ming Wong*	**CoNLL Shared Task: Poster Presentations** **Chair:** *Alessandro Moschitti*
2:30pm – 3:00pm	Learning Syntactic Categories Using Paradigmatic Representations of Word Context *M. A. Yatbaz, E. Sert, and D. Yuret* ..... **p. 181**	Entropy-based Pruning for Phrase-based Machine Translation *W. Ling, J. a. Graça, I. Trancoso, and A. Black* ..... **p. 181**	Probabilistic Finite State Machines for Regression-based MT Evaluation *M. Wang and C. D. Manning* ..... **p. 182**	*See p. 157 for details*
3:00pm – 3:30pm	Exploring Topic Coherence over Many Models and Many Topics *K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler* ..... **p. 181**	A Systematic Comparison of Phrase Table Pruning Techniques *R. Zens, D. Stanton, and P. Xu* ..... **p. 181**	An Empirical Investigation of Statistical Significance in NLP *T. Berg-Kirkpatrick, D. Burkett, and D. Klein* ..... **p. 182**	
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
4:00pm – 5:30pm	**Information Extraction: Relation and Event Extraction** **Chair:** *Mausam*	**Parsing Models and Evaluation** **Chair:** *Ivan Titov*	**Large-Scale NLP Algorithms** **Chair:** *Benjamin van Durme*	**Machine Learning: Inference** **Chair:** *Jennifer Gillenwater*
4:00pm – 4:30pm	Employing Compositional Semantics and Discourse Consistency in Chinese Event Extraction *P. Li, G. Zhou, Q. Zhu, and L. Hou* ..... **p. 182**	Forest Reranking through Subtree Ranking *R. Farkas and H. Schmid* . **p. 183**	Fast Large-Scale Approximate Graph Construction for NLP *A. Goyal, H. Daume III, and R. Guerra* ..... **p. 184**	Monte Carlo MCMC: Efficient Inference by Approximate Sampling *S. Singh, M. Wick, and A. McCallum* ..... **p. 184**
4:30pm – 5:00pm	Reading The Web with Learned Syntactic-Semantic Inference Rules *N. Lao, A. Subramanya, F. Pereira, and W. W. Cohen* ..... **p. 182**	Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output *J. K. Kummerfeld, D. Hall, J. R. Curran, and D. Klein* **p. 183**	Building a Lightweight Semantic Model for Unsupervised Information Extraction on Short Listings *D. S. Kim, K. Verma, and P. Yeh* ..... **p. 184**	On Amortizing Inference Cost for Structured Prediction *V. Srikumar, G. Kundu, and D. Roth* ..... **p. 184**

5:00pm – 5:30pm	Ensemble Semantics for Large-scale Unsupervised Relation Extraction *B. Min, S. Shi, R. Grishman, and C.-Y. Lin* . . . . . **p. 182**	Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level *B. T. M. Wong and C. Kit* . **p. 183**	Sketch Algorithms for Estimating Point Queries in NLP *A. Goyal, H. Daume III, and G. Cormode* . . . . . **p. 184**	Exact Sampling and Decoding in High-Order Hidden Markov Models *S. Carter, M. Dymetman, and G. Bouchard* . . . . . **p. 185**

Day 3 (July 14, 2012)

9:00am – 10:30pm	**Plenary Session**	**(Youngju A (ICC))**
10:30am – 11:00am	**Coffee Break**	**(3F Lobby (ICC))**
11:00am – 12:30pm	**Machine Translation: Decoding**	**(Youngju A (ICC))**
	Distributional and Compositional Semantics	**(203 (ICC))**
	Discourse: Coreference Resolution	**(202 A (ICC))**
	Information Retrieval	**(202 B (ICC))**
12:30pm – 2:00pm	**Lunch**	
2:00pm – 3:30pm	**Machine Learning: Transfer and Biases**	**(Youngju A (ICC))**
	Opinion Mining: Discovering Opinion Expressions	**(203 (ICC))**
	Part of Speech Tagging	**(202 A (ICC))**
	Word Sense Disambiguation	**(202 B (ICC))**
3:30pm – 4:00pm	**Coffee Break**	**(3F Lobby (ICC))**
4:00pm – 5:30pm	**Syntax and Parsing: Joint Parsing Models**	**(Youngju A (ICC))**
	Social Media	**(203 (ICC))**
	NLP Applications	**(202 A (ICC))**
5:30pm – 5:45pm	**Closing Remarks**	**(202 B (ICC))**

Schedule (July 14, 2012)

7:30am – 5:00pm	Registration (3F Lobby (ICC))

Parallel Sessions	Youngju A (ICC)
9:00am – 10:30pm	Plenary Session
	Chair: *Jun'ichi Tsujii*
9:00am – 9:30am	PATTY: A Taxonomy of Relational Patterns with Semantic Types *N. Nakashole, G. Weikum, and F. Suchanek* ..... **p. 186**
9:30am – 10:00am	Training Factored PCFGs with Expectation Propagation *D. Hall and D. Klein* ..... **p. 186**
10:00am – 10:30am	A Coherence Model Based on Syntactic Patterns *A. Louis and A. Nenkova* ..... **p. 186**

Parallel Sessions	Youngju A (ICC)	203 (ICC)	202 A (ICC)	202 B (ICC)
11:00am – 12:30pm	**Machine Translation: Decoding** **Chair:** *Preslav Nakov*	**Distributional and Compositional Semantics** **Chair:** *Bo Pang*	**Discourse: Coreference Resolution** **Chair:** *Mihai Surdeanu*	**Information Retrieval** **Chair:** *Jianfeng Gao*
11:00am – 11:30am	Language Model Rest Costs and Space-Efficient Storage *K. Heafield, P. Koehn, and A. Lavie* ..... **p. 186**	Semantic Compositionality through Recursive Matrix-Vector Spaces *R. Socher, B. Huval, C. D. Manning, and A. Y. Ng* ..... **p. 187**	Learning-based Multi-Sieve Co-reference Resolution with Knowledge *L. Ratinov and D. Roth* ... **p. 188**	Entity based Q&A Retrieval *A. Singh* ..... **p. 189**
11:30am – 12:00pm	Document-Wide Decoding for Phrase-Based Statistical Machine Translation *C. Hardmeier, J. Nivre, and J. Tiedemann* ..... **p. 187**	Polarity Inducing Latent Semantic Analysis *W.-t. Yih, G. Zweig, and J. Platt* ..... **p. 187**	Joint Learning for Coreference Resolution with Markov Logic *Y. Song, J. Jiang, W. X. Zhao, S. Li, and H. Wang* ..... **p. 188**	Constructing Task-Specific Taxonomies for Document Collection Browsing *H. Yang* ..... **p. 189**

12:00pm – 12:30pm	Left-to-Right Tree-to-String Decoding with Prediction *Y. Feng, Y. Liu, Q. Liu, and T. Cohn* ..... **p. 187**	First Order vs. Higher Order Modification in Distributional Semantics *G. Boleda, E. M. Vecchi, M. Cornudella, and L. McNally* ..... **p. 188**	Resolving "This-issue" Anaphora *V. Kolhatkar and G. Hirst* . **p. 188**	Besting the Quiz Master: Crowdsourcing Incremental Classification Games *J. Boyd-Graber, B. Satinoff, H. He, and H. Daume III* ..... **p. 189**
12:30pm – 2:00pm	**Lunch Break**			
Parallel Sessions	**Youngju A (ICC)**	**203 (ICC)**	**202 A (ICC)**	**202 B (ICC)**
2:00pm – 3:30pm	**Machine Learning: Transfer and Biases** **Chair:** *Benjamin Snyder*	**Opinion Mining: Discovering Opinion Expressions** **Chair:** *Yejin Choi*	**Part of Speech Tagging** **Chair:** *Slav Petrov*	**Word Sense Disambiguation** **Chair:** *Marc Dymetman*
2:00pm – 2:30pm	Multi-Domain Learning: When Do Domains Matter? *M. Joshi, M. Dredze, W. W. Cohen, and C. Rose* ..... **p. 189**	Extracting Opinion Expressions with semi-Markov Conditional Random Fields *B. Yang and C. Cardie* .... **p. 190**	Learning to Map into a Universal POS Tagset *Y. Zhang, R. Reichart, R. Barzilay, and A. Globerson* ..... **p. 191**	Joining Forces Pays Off: Multilingual Joint Word Sense Disambiguation *R. Navigli and S. P. Ponzetto* ..... **p. 192**
2:30pm – 3:00pm	Biased Representation Learning for Domain Adaptation *F. Huang and A. Yates* .... **p. 190**	Opinion Target Extraction Using Word-Based Translation Model *K. Liu, L. Xu, and J. Zhao* . **p. 191**	Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features *J. Zhao, X. Qiu, S. Zhang, F. Ji, and X. Huang* ..... **p. 191**	A New Minimally-Supervised Framework for Domain Word Sense Disambiguation *S. Faralli and R. Navigli* .. **p. 192**
3:00pm – 3:30pm	Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars *K. Tu and V. Honavar* .... **p. 190**	Word Salad: Relating Food Prices and Descriptions *V. Chahuneau, K. Gimpel, B. R. Routledge, L. Scherlis, and N. A. Smith* ..... **p. 191**	Wiki-ly Supervised Part-of-Speech Tagging *S. Li, J. a. Graça, and B. Taskar* ..... **p. 192**	Grounded Models of Semantic Representation *C. Silberer and M. Lapata* **p. 192**
3:30pm – 4:00pm	**Coffee Break (3F Lobby (ICC))**			

Parallel Sessions	Youngju A (ICC)	203 (ICC)	202 A (ICC)
4:00pm – 5:30pm	**Syntax and Parsing: Joint Parsing Models** **Chair:** *Jason Eisner*	**Social Media** **Chair:** *Alan Ritter*	**NLP Applications** **Chair:** *Chikara Hashimoto*

4:00pm – 4:30pm	Improved Parsing and POS Tagging Using Inter-Sentence Consistency Constraints *A. Rush, R. Reichart, M. Collins, and A. Globerson* ..... **p. 192**	Identifying Event-related Bursts via Social Media Activities *X. Zhao, B. Shu, J. Jiang, Y. Song, H. Yan, and X. Li* ..... **p. 193**	Supervised Text-based Geolocation Using Language Models on an Adaptive Grid *S. Roller, M. Speriosu, S. Rallapalli, B. Wing, and J. Baldridge* ..... **p. 194**
4:30pm – 5:00pm	Unified Dependency Parsing of Chinese Morphological and Syntactic Structures *Z. Li and G. Zhou* ..... **p. 193**	User Demographics and Language in an Implicit Social Network *K. Filippova* ..... **p. 193**	A Discriminative Model for Query Spelling Correction with Latent Structural SVM *H. Duan, Y. Li, C. Zhai, and D. Roth* .. **p. 194**
5:00pm – 5:30pm	A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing *B. Bohnet and J. Nivre* ..... **p. 193**	Revisiting the Predictability of Language: Response Completion in Social Media *B. Pang and S. Ravi* ..... **p. 194**	Characterizing Stylistic Elements in Syntactic Structure *S. Feng, R. Banerjee, and Y. Choi* .... **p. 195**
5:30pm – 5:45pm	**— Youngju A (ICC) —** **Closing Remarks**		

Invited Talks

"On Learning Sparse Structured Input-Output Models"

Eric Xing, Carnegie Mellon University

In many modern problems across areas such as natural language processing, computer vision, and social media inference, one is often interested in learning a Sparse Structured Input-Output Model (SIOM), in which the input variables of the model such as lexicons in a document bear rich structures due to the syntactic and semantic dependences between them in the text; and the output variables such as the elements in a multi-way classification, a parse, or a topic representation are also structured because of their interrelatedness. A SIOM can nicely capture rich structural properties in the data and in the problem, but it also raises severe computational and theoretical challenge on sparse, consistent, and tractable model identification and inference.

In this talk, I will present models, algorithms, and theories that learn Sparse SIOMs of various kinds in very high dimensional input/output space, with fast and highly scalable optimization procedures, and strong statistical guarantees. I will demonstrate application of our approach to problems in large-scale text classification, topic modeling, and dependency parsing.

"The Appification of the Web and the Renaissance of Conversational User Interfaces"

Patrick Pantel, Microsoft Research

The appification of the Web is triggering a fundamental shift in how users access information. We are moving from centralized access points, such as search engines, towards highly specialized, and yet fragmented, functionalities in disconnected apps. This talk explores an entity-centric conversational interface as a mechanism to overcome this fragmentation, highlighting the numerous associated NLP challenges and opportunities that lie ahead.

Consider mobile scenarios, where the traditional search engine paradigm is being cannibalized by search and browse functionalities built directly into specialized apps. For example, while users can search for restaurants and products using their mobile browser, they are increasingly turning directly to applications such as Yelp, Urbanspoon and Amazon. However, interoperability between applications and lacking generalized interfaces to their functionalities pose serious scalability challenges. In this talk, we argue for an entity-centric conversational interface in which natural user interactions with entities are paired with actions that can be performed on the entities, thus enabling the brokering of web pages and applications that can satisfy the intended action. In this vision, the broker is aware of all entities and actions of interest to its users, understands the intent of the user, and provides direct actionable results through APIs with external providers satisfying the intent. The user saves clicks and time to accomplish her intended action and can discover related actions. New revenue streams open up from paid action placement and lead generation opportunities. At the forefront of this direction are a number of NLP challenges in the areas of entity recognition, entity linking, knowledge extraction, intent recognition, and dialog modeling, to name a few.

We end by proposing one particular technique for learning and mapping user intents in a search interface. In an annotation study conducted over a traffic sample of web usage logs, we found that a large proportion of user queries involve actions on entities, calling for an automatic approach to identifying relevant actions for entity-bearing queries. We

pose the problem of finding actions that can be performed on entities as the problem of doing probabilistic inference in a graphical model that captures how entity-bearing information requests are generated. Given a large collection of real-world queries and clicks from a commercial search engine, the models are learned efficiently through maximum likelihood estimation using an EM algorithm. Given a new query, inference enables the recommendation of a set of pertinent actions and providers. We propose an evaluation methodology for measuring the relevance of our recommended actions, and show empirical evidence of the quality and the diversity of the discovered actions.

CoNLL Shared Task: Oral Presentations

CoNLL Shared Task: Poster Presentations

EMNLP-CoNLL Paper Abstracts

Oral Presentations: Thursday, July 12, 2012

Syntactic Transfer Using a Bilingual Lexicon
Greg Durrett, Adam Pauls, and Dan Klein
Thursday 11:00am–11:30am — Youngju A (ICC)

We consider the problem of using a bilingual dictionary to transfer lexico-syntactic information from a resource-rich source language to a resource-poor target language. In contrast to past work that used bitexts to transfer analyses of specific sentences at the token level, we instead use features to transfer the behavior of words at a type level. In a discriminative dependency parsing framework, our approach produces gains across a range of target languages, using two different low-resource training methodologies (one weakly supervised and one indirectly supervised) and two different dictionary sources (one manually constructed and one automatically constructed).

Regularized Interlingual Projections: Evaluation on Multilingual Transliteration
Jagadeesh Jagarlamudi and Hal Daume III
Thursday 11:30am–12:00pm — Youngju A (ICC)

In this paper, we address the problem of building a multilingual transliteration system using an interlingual representation. Our approach uses international phonetic alphabet (IPA) to learn the interlingual representation and thus allows us to use any word and its IPA representation as a training example. Thus, our approach requires only monolingual resources: a phoneme dictionary that lists words and their IPA representations.1 By adding a phoneme dictionary of a new language, we can readily build a transliteration system into any of the existing previous languages, without the expense of all-pairs data or computation. We also propose a regularization framework for learning the interlingual representation, which accounts for language specific phonemic variability, and thus it can find better mappings between languages. Experimental results on the name transliteration task in five diverse languages show a maximum improvement of 29% accuracy and an average improvement of 17% accuracy compared to a state-of-the-art baseline system.

Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation
Akihiro Tamura, Taro Watanabe, and Eiichiro Sumita
Thursday 12:00pm–12:30pm — Youngju A (ICC)

This paper proposes a novel method for lexicon extraction that extracts translation pairs from comparable corpora by using graph-based label propagation. In previous work, it was established that performance drastically decreases when the coverage of a seed lexicon is small. We resolve this problem by utilizing indirect relations with the bilingual seeds together with direct relations, in which each word is represented by a distribution of translated seeds. The seed distributions are propagated over a graph representing relations among words, and translation pairs are extracted by identifying word pairs with a high similarity in the seed distributions. We propose two types of the graphs: a co-occurrence graph, representing co-occurrence relations between words, and a similarity graph, representing context similarities between words. Evaluations using English and Japanese patent comparable corpora show that our proposed graph propagation method outperforms conventional methods. Further, the similarity graph achieved improved performance by clustering synonyms into the same translation.

Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls
Kai Hong, Christian G. Kohler, Mary E. March, Amber A. Parker, and Ani Nenkova
Thursday 11:00am–11:30am — 203 (ICC)

We present a system for automatic identification of schizophrenic patients and healthy controls

based on narratives the subjects recounted about emotional experiences in their own life. The focus of the study is to identify the lexical features that distinguish the two populations. We report the results of feature selection experiments that demonstrate that the classifier can achieve accuracy on patient level prediction as high as 76.9% with only a small set of features. We provide an in-depth discussion of the lexical features that distinguish the two groups and the unexpected relationship between emotion types of the narratives and the accuracy of patient status prediction.

Streaming Analysis of Discourse Participants

Benjamin Van Durme

Thursday 11:30am–12:00pm — 203 (ICC)

Inferring attributes of discourse participants has been treated as a batch-processing task: data such as all tweets from a given author are gathered in bulk, processed, analyzed for a particular feature, then reported as a result of academic interest. Given the sources and scale of material used in these efforts, along with potential use cases of such analytic tools, discourse analysis should be reconsidered as a streaming challenge. We show that under certain common formulations, the batch-processing analytic framework can be decomposed into a sequential series of updates, using as an example the task of gender classification. Once in a streaming framework, and motivated by large data sets generated by social media services, we present novel results in approximate counting, showing its applicability to space efficient streaming classification.

Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev

Thursday 12:00pm–12:30pm — 203 (ICC)

A mixture of positive (friendly) and negative (antagonistic) relations exist among users in most social media applications. However, many such applications do not allow users to explicitly express the polarity of their interactions. As a result most research has either ignored negative links or was limited to the few domains where such relations are explicitly expressed (e.g. Epinions trust/distrust). We study text exchanged between users in online communities. We find that the polarity of the links between users can be predicted with high accuracy given the text they exchange. This allows us to build a signed network representation of discussions; where every edge has a sign: positive to denote a friendly relation, or negative to denote an antagonistic relation. We also connect our analysis to social psychology theories of balance. We show that the automatically predicted networks are consistent with those theories. Inspired by that, we present a technique for identifying subgroups in discussions by partitioning singed networks representing them.

Generative Goal-Driven User Simulation for Dialog Management

Aciel Eshky, Ben Allison, and Mark Steedman

Thursday 11:00am–11:30am — 202 A (ICC)

User simulation is frequently used to train statistical dialog managers for task-oriented domains. At present, goal-driven simulators (those that have a persistent notion of what they wish to achieve in the dialog) require some task-specific engineering, making them impossible to evaluate intrinsically. Instead, they have been evaluated extrinsically by means of the dialog managers they are intended to train, leading to circularity of argument. In this paper, we propose the first fully generative goal-driven simulator that is fully induced from data, without hand-crafting or goal annotation. Our goals are latent, and take the form of topics in a topic model, clustering together semantically equivalent and phonetically confusable strings, implicitly modelling synonymy and speech recognition noise. We evaluate on two standard dialog resources, the Communicator and Let's Go datasets, and demonstrate that our model has substantially better fit to held out data than competing approaches. We also show that features derived from our model allow significantly greater improvement over a baseline at distinguishing real from randomly

permuted dialogs.

Optimising Incremental Dialogue Decisions Using Information Density for Interactive Systems

Nina Dethlefs, Helen Hastie, Verena Rieser, and Oliver Lemon

Thursday 11:30am–12:00pm — 202 A (ICC)

Incremental processing allows system designers to address several discourse phenomena that have previously been somewhat neglected in interactive systems, such as backchannels or barge-ins, but that can enhance the responsiveness and naturalness of systems. Unfortunately, prior work has focused largely on deterministic incremental decision making, rendering system behaviour less flexible and adaptive than is desirable. We present a novel approach to incremental decision making that is based on Hierarchical Reinforcement Learning to achieve an interactive optimisation of Information Presentation (IP) strategies, allowing the system to generate and comprehend backchannels and barge-ins, by employing the recent psycholinguistic hypothesis of information density (ID) (Jaeger, 2010). Results in terms of average rewards and a human rating study show that our learnt strategy outperforms several baselines that are not sensitive to ID by more than 23%.

Mixed Membership Markov Models for Unsupervised Conversation Modeling

Michael J. Paul

Thursday 12:00pm–12:30pm — 202 A (ICC)

Recent work has explored the use of hidden Markov models for unsupervised discourse and conversation modeling, where each segment of text such as a message in a conversation is associated with a hidden state in a sequence. We extend this approach to allow each segment of text to be a mixture of multiple classes. Under our model, the probability of a class in a text segment is a log-linear function of the classes in the previous segment. We show that this model performs well at predictive tasks on two conversation data sets, improving thread reconstruction accuracy by up to 15 percentage points over a standard HMM. Additionally, we show quantitatively that the induced word clusters correspond to speech acts more closely than baseline models.

An Entity-Topic Model for Entity Linking

Xianpei Han and Le Sun

Thursday 11:00am–11:30am — 202 B (ICC)

Entity Linking (EL) has received considerable attention in recent years. Given many name mentions in a document, the goal of EL is to predict their referent entities in a knowledge base. Traditionally, there have been two distinct directions of EL research: one focusing on the effects of mention's context compatibility, assuming that "the referent entity of a mention is reflected by its context"; the other dealing with the effects of document's topic coherence, assuming that "a mention's referent entity should be coherent with the document's main topics". In this paper, we propose a generative model — called entity-topic model, to effectively join the above two complementary directions together. By jointly modeling and exploiting the context compatibility, the topic coherence and the correlation between them, our model can accurately link all mentions in a document using both the local information (including the words and the mentions in a document) and the global knowledge (including the topic knowledge, the entity context knowledge and the entity name knowledge). Experimental results demonstrate the effectiveness of the proposed model.

Linking Named Entities to Any Database

Avirup Sil, Ernest Cronin, Penghai Nie, Yinfei Yang, Ana-Maria Popescu, and Alexander Yates

Thursday 11:30am–12:00pm — 202 B (ICC)

Existing techniques for disambiguating named entities in text mostly focus on Wikipedia as a target catalog of entities. Yet for many types of entities, such as restaurants and cult movies,

relational databases exist that contain far more extensive information than Wikipedia. This paper introduces a new task, called Open-Database Named-Entity Disambiguation (Open-DB NED), in which a system must be able to resolve named entities to symbols in an arbitrary database, without requiring labeled data for each new database. We introduce two techniques for Open-DB NED, one based on distant supervision and the other based on domain adaptation. In experiments on two domains, one with poor coverage by Wikipedia and the other with near-perfect coverage, our Open-DB NED strategies outperform a state-of-the-art Wikipedia NED system by over 25% in accuracy.

Towards Efficient Named-Entity Rule Induction for Customizability
Ajay Nagesh, Ganesh Ramakrishnan, Laura Chiticariu, Rajasekar Krishnamurthy, Ankush Dharkar, and Pushpak Bhattacharyya
Thursday 12:00pm–12:30pm — 202 B (ICC)

Generic rule-based systems for Information Extraction (IE) have been shown to work reasonably well out-of-the-box, and achieve state-of-the-art accuracy with further domain customization. However, it is generally rec- ognized that manually building and customiz- ing rules is a complex and labor intensive pro- cess. In this paper, we discuss an approach that facilitates the process of building cus- tomizable rules for Named-Entity Recognition (NER) tasks via rule induction, in the Annota- tion Query Language (AQL). Given a set of basic features and an annotated document col- lection, our goal is to generate an initial set of rules with reasonable accuracy, that are in- terpretable and thus can be easily refined by a human developer. We present an efficient rule induction process, modeled on a four- stage manual rule development process and present initial promising results with our sys- tem. We also propose a simple notion of ex- tractor complexity as a first step to quantify the interpretability of an extractor, and study the effect of induction bias and customization of basic features on the accuracy and complex- ity of induced rules. We demonstrate through experiments that the induced rules have good accuracy and low complexity according to our complexity measure.

Active Learning for Imbalanced Sentiment Classification
Shoushan Li, Shengfeng Ju, Guodong Zhou, and Xiaojun Li
Thursday 2:00pm–2:30pm — Youngju A (ICC)

Active learning is a promising way for sentiment classification to reduce the annotation cost. In this paper, we focus on the imbalanced class distribution scenario for sentiment classification, wherein the number of positive samples is quite different from that of negative samples. This scenario posits new challenges to active learning. To address these challenges, we propose a novel active learning approach, named co-selecting, by taking both the imbalanced class distribution issue and uncertainty into account. Specifically, our co-selecting approach employs two feature subspace classifiers to collectively select most informative minority-class samples for manual annotation by leveraging a certainty measurement and an uncertainty measurement, and in the meanwhile, automatically label most informative majority-class samples, to reduce human-annotation efforts. Extensive experiments across four domains demonstrate great potential and effectiveness of our proposed co-selecting approach to active learning for imbalanced sentiment classification.

A Weakly Supervised Model for Sentence-Level Semantic Orientation Analysis with Multiple Experts
Lizhen Qu, Rainer Gemulla, and Gerhard Weikum
Thursday 2:30pm–3:00pm — Youngju A (ICC)

We propose the weakly supervised Multi-Experts Model (MEM) for analyzing the semantic orientation of opinions expressed in natural language reviews. In contrast to most prior work, MEM predicts both opinion polarity and opinion strength at the level of individual sentences; such fine-grained analysis helps to understand better why users like or dislike the entity under review. A key challenge in this setting is that it is hard to obtain sentence-level training data for both po-

larity and strength. For this reason, MEM is weakly supervised: It starts with potentially noisy indicators obtained from coarse-grained training data (i.e., document-level ratings), a small set of diverse base predictors, and, if available, small amounts of fine-grained training data. We integrate these noisy indicators into a unified probabilistic framework using ideas from ensemble learning and graph-based semi-supervised learning. Our experiments indicate that MEM outperforms state-of-the-art methods by a significant margin.

Collocation Polarity Disambiguation Using Web-based Pseudo Contexts
Yanyan Zhao, Bing Qin, and Ting Liu
Thursday 3:00pm–3:30pm — Youngju A (ICC)

This paper focuses on the task of collocation polarity disambiguation. The collocation refers to a binary tuple of a polarity word and a target (such as <long, battery life> or <long, startup>), in which the sentiment orientation of the polarity word ("long") changes along with different targets ("battery life" or "startup"). To disambiguate a collocation's polarity, previous work always turned to investigate the polarities of its surrounding contexts, and then assigned the majority polarity to the collocation. However, these contexts are limited, thus the resulting polarity is insufficient to be reliable. We therefore propose an unsupervised three-component framework to expand some pseudo contexts from web, to help disambiguate a collocation's polarity.Without using any additional labeled data, experiments show that our method is effective.

Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering
Michael Roth and Anette Frank
Thursday 2:00pm–2:30pm — 203 (ICC)

Generating coherent discourse is an important aspect in natural language generation. Our aim is to learn factors that constitute coherent discourse from data, with a focus on how to realize predicate-argument structures in a model that exceeds the sentence level. We present an important subtask for this overall goal, in which we align predicates across comparable texts, admitting partial argument structure correspondences. The contribution of this work is two-fold: We first construct a large corpus resource of comparable texts, including an evaluation set with manual predicate alignments. Secondly, we present a novel approach for aligning predicates across comparable texts using graph-based clustering with Mincuts. Our method outperforms other alignment techniques when applied to this novel alignment task, by a margin of at least 6.5 percentage points in F1-score.

Local and Global Context for Supervised and Unsupervised Metonymy Resolution
Vivi Nastase, Alex Judea, Katja Markert, and Michael Strube
Thursday 2:30pm–3:00pm — 203 (ICC)

Computational approaches to metonymy resolution have focused almost exclusively on the local context, especially the constraints placed on a potentially metonymic word by its grammatical collocates. We expand such approaches by taking into account the larger context. Our algorithm is tested on the data from the metonymy resolution task (Task 8) at SemEval 2007. The results show that incorporation of the global context can improve over the use of the local context alone, depending on the types of metonymies addressed. As a second contribution, we move towards unsupervised resolution of metonymies, made feasible by considering ontological relations as possible readings. We show that such an unsupervised approach delivers promising results: it beats the supervised most frequent sense baseline and performs close to a supervised approach using only standard lexico-syntactic features.

Learning Verb Inference Rules from Linguistically-Motivated Evidence
Hila Weisman, Jonathan Berant, Idan Szpektor, and Ido Dagan
Thursday 3:00pm–3:30pm — 203 (ICC)

Learning inference relations between verbs is at the heart of many semantic applications. How-

ever, most prior work on learning such rules focused on a rather narrow set of information sources: mainly distributional similarity, and to a lesser extent manually constructed verb co-occurrence patterns. In this paper, we present a rich and novel set of linguistically motivated cues for detecting entailment between verbs, which exploits information at varying textual scopes: verb co-occurrence within a sentence, verb co-occurrence within a document, as well as overall corpus statistics. These cues are combined as features in a supervised classification framework. We empirically demonstrate that our model significantly outperforms previous methods and that information from each textual scope contributes to the verb entailment learning task.

Spectral Dependency Parsing with Latent Variables
Paramveer Dhillon, Jordan Rodu, Michael Collins, Dean Foster, and Lyle Ungar
Thursday 2:00pm–2:30pm — 202 A (ICC)

Recently there has been substantial interest in using spectral methods to learn generative sequence models like HMMs. Spectral methods are attractive as they provide globally consistent estimates of the model parameters and are very fast and scalable, unlike EM methods, which can get stuck in local minima. In this paper, we present a novel extension of this class of spectral methods to learn dependency tree structures. We propose a simple yet powerful latent variable generative model for dependency parsing, and a spectral learning method to efficiently estimate it. As a pilot experimental evaluation, we use the spectral tree probabilities estimated by our model to re-rank the outputs of a near state-of-the-art parser. Our approach gives us a moderate reduction in error of up to 4.6% over the baseline re-ranker.

A Phrase-Discovering Topic Model Using Hierarchical Pitman-Yor Processes
Robert Lindsey, William Headden, and Michael Stipicevic
Thursday 2:30pm–3:00pm — 202 A (ICC)

Topic models traditionally rely on the bag-of-words assumption. In data mining applications, this often results in end-users being presented with inscrutable lists of topical unigrams, single words inferred as representative of their topics. In this article, we present a hierarchical generative probabilistic model of topical phrases. The model simultaneously infers the location, length, and topic of phrases within a corpus and relaxes the bag-of-words assumption within phrases by using a hierarchy of Pitman-Yor processes. We use Markov chain Monte Carlo techniques for approximate inference in the model and perform slice sampling to learn its hyperparameters. We show via an experiment on human subjects that our model finds substantially better, more interpretable topical phrases than do competing models.

A Bayesian Model for Learning SCFGs with Discontiguous Rules
Abby Levenberg, Chris Dyer, and Phil Blunsom
Thursday 3:00pm–3:30pm — 202 A (ICC)

We describe a nonparametric model and corresponding inference algorithm for learning Synchronous Context Free Grammars (SCFGs) directly from parallel text. The model employs a Pitman-Yor process prior which uses a novel base distribution over SCFG rules. Through both synthetic grammar induction and statistical machine translation experiments, we demonstrate that our model learns complex translational correspondences— including discontinuous, many-to-many alignments—and produces competitive translation results. Further, inference is efficient and we present results on significantly larger corpora than prior work.

Multiple Aspect Summarization Using Integer Linear Programming
Kristian Woodsend and Mirella Lapata
Thursday 2:00pm–2:30pm — 202 B (ICC)

Multi-document summarization involves many aspects of content selection and surface realization. The summaries must be informative, succinct, grammatical, and obey stylistic writing conventions. We present a method where such individual aspects are learned separately from

data (without any hand-engineering) but optimized jointly using an integer linear programme. The ILP framework allows us to combine the decisions of the expert learners and to select and rewrite source content through a mixture of objective setting, soft and hard constraints. Experimental results on the TAC-08 data set show that our model achieves state-of-the-art performance using ROUGE and significantly improves the informativeness of the summaries.

Minimal Dependency Length in Realization Ranking
Michael White and Rajakrishnan Rajkumar
Thursday 2:30pm–3:00pm — 202 B (ICC)

Comprehension and corpus studies have found that the tendency to minimize dependency length has a strong influence on constituent ordering choices. In this paper, we investigate dependency length minimization in the context of discriminative realization ranking, focusing on its potential to eliminate egregious ordering errors as well as better match the distributional characteristics of sentence orderings in news text. We find that with a stateof- the-art, comprehensive realization ranking model, dependency length minimization yields statistically significant improvements in BLEU scores and significantly reduces the number of heavy/light ordering errors. Through distributional analyses, we also show that with simpler ranking models, dependency length minimization can go overboard, too often sacrificing canonical word order to shorten dependencies, while richer models manage to better counterbalance the dependency length minimization preference against (sometimes) competing canonical word order preferences.

Framework of Automatic Text Summarization Using Reinforcement Learning
Seonggi Ryang and Takeshi Abekawa
Thursday 3:00pm–3:30pm — 202 B (ICC)

We present a new approach to the problem of automatic text summarization called Automatic Summarization using Reinforcement Learning (ASRL) in this paper, which models the process of constructing a summary within the framework of reinforcement learning and attempts to optimize the given score function with the given feature representation of a summary. We demonstrate that the method of reinforcement learning can be adapted to automatic summarization problems naturally and simply, and other summarizing techniques, such as sentence compression, can be easily adapted as actions of the framework.

The experimental results indicated ASRL was superior to the best performing method in DUC2004 and comparable to the state of the art ILP-style method, in terms of ROUGE scores. The results also revealed ASRL can search for sub-optimal solutions efficiently under conditions for effectively selecting features and the score function.

Large Scale Decipherment for Out-of-Domain Machine Translation
Qing Dou and Kevin Knight
Thursday 4:00pm–4:30pm — Youngju A (ICC)

We apply slice sampling to Bayesian decipherment and use our new decipherment framework to improve out-of-domain machine translation. Compared with the state of the art algorithm, our approach is highly scalable and produces better results, which allows us to decipher ciphertext with billions of tokens and hundreds of thousands of word types with high accuracy. We decipher a large amount of monolingual data to improve out of domain translation and achieve significant gains of up to 3.8 BLEU points.

N-gram-based Tense Models for Statistical Machine Translation
Zhengxian Gong, Min Zhang, Chew Lim Tan, and Guodong Zhou
Thursday 4:30pm–5:0pm — Youngju A (ICC)

Tense is a small element to a sentence, however, error tense can raise odd grammars and result in misunderstanding. Recently, tense has drawn attention in many natural language processing applications. However,most of current Statistical Machine Translation (SMT) systems mainly

utilize translation model and language model. They never consider and make full use of such tense infomation. In this paper, we propose n-gram-based tense models for SMT and successfully integrate such models into a state-of-the-art phrase-based SMT system via two additional features. Experimental results on the NIST Chinese-English translation task show that our proposed tense model is very effective, contributing performance improvement by 0.62 BLUE points over a strong baseline.

Source Language Adaptation for Resource-Poor Machine Translation

Pidong Wang, Preslav Nakov, and Hwee Tou Ng

Thursday 5:00pm–5:30pm — Youngju A (ICC)

We propose a novel, language-independent approach for improving machine translation from a resource-poor language to X by adapting a large bi-text for a related resource-rich language and X (the same target language). We assume a small bi-text for the resource-poor language to X pair, which we use to learn word-level and phrase-level paraphrases and cross-lingual morphological variants between the resource-rich and the resource-poor language; we then adapt the former to get closer to the latter. Our experiments for Indonesian/Malay—English translation show that using the large adapted resource-rich bi-text yields 6.7 BLEU points of improvement over the unadapted one and 2.6 BLEU points over the original small bi-text. Moreover, combining the small bi-text with the adapted bi-text outperforms the corresponding combinations with the unadapted bi-text by 1.5-3 BLEU points. We also demonstrate applicability to other languages and domains.

Exploiting Reducibility in Unsupervised Dependency Parsing

David Mareček and Zdeněk Žabokrtský

Thursday 4:00pm–4:30pm — 203 (ICC)

The possibility of deleting a word from a sentence without violating its syntactic correctness belongs to traditionally known manifestations of syntactic dependency. We introduce a novel unsupervised parsing approach that is based on a new n-gram reducibility measure. We perform experiments across 18 languages available in CoNLL data and we show that our approach achieves better accuracy for the majority of the languages then previously reported results. d

Improving Transition-Based Dependency Parsing with Buffer Transitions

Daniel Fernández-González and Carlos Gómez-Rodríguez

Thursday 4:30pm–5:00pm — 203 (ICC)

In this paper, we show that significant improvements in the accuracy of well-known transition-based parsers can be obtained, without sacrificing efficiency, by enriching the parsers with simple transitions that act on buffer nodes.

First, we show how adding a specific transition to create either a left or right arc of length one between the first two buffer nodes produces improvements in the accuracy of Nivre's arc-eager projective parser on a number of datasets from the CoNLL-X shared task. Then, we show that accuracy can also be improved by adding transitions involving the topmost stack node and the second buffer node (allowing a limited form of non-projectivity).

None of these transitions has a negative impact on the computational complexity of the algorithm. Although the experiments in this paper use the arc-eager parser, the approach is generic enough to be applicable to any stack-based dependency parser.

Generalized Higher-Order Dependency Parsing with Cube Pruning

Hao Zhang and Ryan McDonald

Thursday 5:00pm–5:30pm — 203 (ICC)

State-of-the-art graph-based parsers use features over higher-order dependencies that rely on decoding algorithms that are slow and difficult to generalize. On the other hand, transition-based dependency parsers can easily utilize such features without increasing the linear complexity of the shift-reduce system beyond a constant. In this paper, we attempt to address this imbalance for

graph-based parsing by generalizing the Eisner (1996) algorithm to handle arbitrary features over higher-order dependencies. The generalization is at the cost of asymptotic efficiency. To account for this, cube pruning for decoding is utilized (Chiang, 2007). For the first time, label tuple and structural features such as valencies can be scored efficiently with third-order features in a graph-based parser. Our parser achieves the state-of-art unlabeled accuracy of 93.06% and labeled accuracy of 91.86% on the standard test set for English, at a faster speed than a reimplementation of the third-order model of Koo et al. (2010).

Universal Grapheme-to-Phoneme Prediction Over Latin Alphabets

Young-Bum Kim and Benjamin Snyder

Thursday 4:00pm–4:30pm — 203 (ICC)

We consider the problem of inducing grapheme-to-phoneme mappings for unknown languages written in a Latin alphabet. First, we collect a data-set of 107 languages with known grapheme-phoneme relationships, along with a short text in each language. We then cast our task in the framework of supervised learning, where each known language serves as a training example, and predictions are made on unknown languages. We induce an undirected graphical model that learns phonotactic regularities, thus relating textual patterns to plausible phonemic interpretations across the entire range of languages. Our model correctly predicts grapheme-phoneme pairs with over 88% F1-measure.

Name Phylogeny: A Generative Model of String Variation

Nicholas Andrews, Jason Eisner, and Mark Dredze

Thursday 4:30pm–5:00pm — 203 (ICC)

Many linguistic and textual processes involve transduction of strings. We show how to learn a stochastic transducer from an unorganized collection of strings (rather than string pairs). The role of the transducer is to organize the collection. Our generative model explains similarities among the strings by supposing that some strings in the collection were not generated ab initio, but were instead derived by transduction from other, "similar" strings in the collection. Our variational EM learn- ing algorithm alternately reestimates this phy- logeny and the transducer parameters. The fi- nal learned transducer can quickly link any test name into the final phylogeny, thereby locat- ing variants of the test name. We find that our method can effectively find name variants in a corpus of web strings used to refer to per- sons in Wikipedia, improving over standard untrained distances such as Jaro-Winkler and Levenshtein distance.

Syntactic Surprisal Affects Spoken Word Duration in Conversational Contexts

Vera Demberg, Asad Sayeed, Philip Gorinski, and Nikolaos Engonopoulos

Thursday 5:00pm–5:30pm — 203 (ICC)

We present results of a novel experiment to investigate speech production in conversational data that links speech rate to information density. We provide the first evidence for an association between syntactic surprisal and word duration in recorded speech. Using the AMI corpus which contains transcriptions of focus group meetings with precise word durations, we show that word durations correlate with syntactic surprisal estimated from the incremental Roark parser over and above simpler measures, such as word duration estimated from a state-of-the-art text-to-speech system and word frequencies, and that the syntactic surprisal estimates are better predictors of word durations than a simpler version of surprisal based on trigram probabilities. This result supports the uniform information density (UID) hypothesis and points a way to more realistic artificial speech generation.

Why Question Answering using Sentiment Analysis and Word Classes

Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun'ichi Kazama, and Yiou Wang

Thursday 4:00pm–4:30pm — 202 B (ICC)

In this paper we explore the utility of sentiment analysis and semantic word classes for improving

why-question answering on a large-scale web corpus. Our work is motivated by the observation that a why-question and its answer often follow the pattern that "if something undesirable happens, the reason is also often something undesirable," and "if something desirable happens, the reason is also often something desirable." To the best of our knowledge, this is the first work that introduces sentiment analysis to non-factoid question answering. We combine this simple idea with semantic word classes for ranking answers to why-questions and show that on a set of 850 why-questions our method gains 15.2% improvement in precision at the top-1 answer over a baseline state-of-the-art QA system that achieved the best performance in a shared task of Japanese non-factoid QA in NTCIR-6.

Natural Language Questions for the Web of Data

Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp, and Gerhard Weikum

Thursday 4:30pm–5:00pm — 202 B (ICC)

The Linked Data initiative comprises structured databases in the Semantic-Web data model RDF. Exploring this heterogeneous data by structured query languages is tedious and error-prone even for skilled users. To ease the task, this paper presents a methodology for translating natural language questions into structured SPARQL queries over linked-data sources.

Our method is based on an integer linear program to solve several disambiguation tasks jointly: the segmentation of questions into phrases; the mapping of phrases to semantic entities, classes, and relations; and the construction of SPARQL triple patterns. Our solution harnesses the rich type system provided by knowledge bases in the web of linked data, to constrain our semantic-coherence objective function. We present experiments on both the question translation and the resulting query answering.

Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews

Jianxing Yu, Zheng-Jun Zha, and Tat-Seng Chua

Thursday 5:00pm–5:30pm — 202 B (ICC)

This paper proposes to generate appropriate answers for opinion questions about products by exploiting the hierarchical organization of consumer reviews. The hierarchy organizes product aspects as nodes following their parent-child relations. For each aspect, the reviews and corresponding opinions on this aspect are stored. We develop a new framework for opinion Questions Answering, which enables accurate question analysis and effective answer generation by making use the hierarchy. In particular, we first identify the (explicit/implicit) product aspects asked in the questions and their sub-aspects by referring to the hierarchy. We then retrieve the corresponding review fragments relevant to the aspects from the hierarchy. In order to generate appropriate answers from the review fragments, we develop a multi-criteria optimization approach for answer generation by simultaneously taking into account review salience, coherence, diversity, and parent-child relations among the aspects. We conduct evaluations on 11 popular products in four domains. The evaluated corpus contains 70,359 consumer reviews and 220 questions on these products. Experimental results demonstrate the effectiveness of our approach.

Poster Presentations: Thursday, July 12, 2012

Locally Training the Log-Linear Model for SMT

Lemao Liu, Hailong Cao, Taro Watanabe, Tiejun Zhao, Mo Yu, and Conghui Zhu

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In statistical machine translation, minimum error rate training (MERT) is a standard method for tuning a single weight with regard to a given development data. However, due to the diversity and uneven distribution of source sentences, there are two problems suffered by this method. First, its performance is highly dependent on the choice of a development set, which may lead to an unstable performance for testing. Second, translations become inconsistent at the sentence level since tuning is performed globally on a document level. In this paper, we propose a novel local training method to address these two problems. Unlike a global training method, such as MERT, in which a single weight is learned and used for all the input sentences, we perform training and testing in one step by learning a sentencewise weight for each input sentence. We propose efficient incremental training methods to put the local training into practice. In NIST Chinese-to-English translation tasks, our local training method significantly outperforms MERT with the maximal improvements up to 2.0 BLEU points, meanwhile its efficiency is comparable to that of the global method.

Iterative Annotation Transformation with Predict-Self Reestimation for Chinese Word Segmentation

Wenbin Jiang, Fandong Meng, Qun Liu, and Yajuan Lü

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper we first describe the technology of automatic annotation transformation, which is based on the annotation adaptation algorithm. It can automatically transform a human-annotated corpus from one annotation guideline to another. We then propose two optimization strategies, iterative training and predict-self reestimation, to further improve the accuracy of annotation guideline transformation. Experiments on Chinese word segmentation show that, the iterative training strategy together with predict-self reestimation brings significant improvement over the simple annotation transformation baseline, and leads to classifiers with significantly higher accuracy and several times faster processing than annotation adaptation does. On the Penn Chinese Treebank 5.0, it achieves an F-measure of 98.43%, significantly outperforms previous works although using a single classifier with only local features.

Automatically Constructing a Normalisation Dictionary for Microblogs

Bo Han, Paul Cook, and Timothy Baldwin

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Microblog normalisation methods often utilise complex models and struggle to differenti- ate between correctly-spelled unknown words and lexical variants of known words. In this paper, we propose a method for construct- ing a dictionary of lexical variants of known words that facilitates lexical normalisation via simple string substitution (e.g. tomorrow for tmrw). We use context information to generate possible variant and normalisation pairs and then rank these by string similarity. Highly- ranked pairs are selected to populate the dic- tionary. We show the combined dictionary- based approach achieves state-of-the-art per- formance for both F-score and word error rate. Compared with other methods, it offers a fast, lightweight and easy-to-use solution, and is thus suitable for high-volume microblog pre- processing.

Unsupervised PCFG Induction for Grounded Language Learning with Highly Ambiguous Supervision

Joohyun Kim and Raymond Mooney

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

”Grounded” language learning employs training data in the form of sentences paired with relevant but ambiguous perceptual contexts. Borschinger et al. (2011) introduced an approach to

grounded language learning based on unsupervised PCFG induction. Their approach works well when each sentence potentially refers to one of a small set of possible meanings, such as in the sportscasting task on which it was evaluated. However, it does not scale to problems with a large set of potential meanings for each sentence, such as the navigation instruction following task studied by Chen and Mooney (2011). This paper presents an enhancement of the PCFG approach that scales to such problems with highly-ambiguous supervision. Experimental results on the navigation task demonstrates the effectiveness of our approach.

Forced Derivation Tree based Model Training to Statistical Machine Translation
Nan Duan, Mu Li, and Ming Zhou
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

A forced derivation tree (FDT) of a sentence pair f,e denotes a derivation tree that can translate f into its accurate target translation e. In this paper, we present an approach that leverages structured knowledge contained in FDTs to train component models for statistical machine translation (SMT) systems. We first describe how to generate different FDTs for each sentence pair in training corpus, and then present how to infer the optimal FDTs based on their derivation and alignment qualities. As the first step in this line of research, we verify the effectiveness of our approach in a BTG-based phrasal system, and propose four FDT-based component models. Experiments are carried out on large scale English-to-Japanese and Chinese-to-English translation tasks, and significant improvements are reported on both translation quality and alignment quality.

Multi-instance Multi-label Learning for Relation Extraction
Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Distant supervision for relation extraction (RE) - gathering training data by aligning a database of facts with text - is an efficient approach to scale RE to thousands of different relations. However, this introduces a challenging learning scenario where the relation expressed by a pair of entities found in a sentence is unknown. For example, a sentence containing Balzac and France may express BornIn or Died, an unknown relation, or no relation at all. Because of this, traditional supervised learning, which assumes that each example is explicitly mapped to a label, is not appropriate. We propose a novel approach to multi-instance multi-label learning for RE, which jointly models all the instances of a pair of entities in text and all their labels using a graphical model with latent variables. Our model performs competitively on two difficult domains.

An "AI readability" Formula for French as a Foreign Language
Thomas François and Cédrick Fairon
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper present a new readability formula for French as a foreign language (FFL), which relies on 46 textual features representative of the lexical, syntactic, and semantic levels as well as some of the specificities of the FFL context. We report comparisons between several techniques for feature selection and various learning algorithms. Our best model, based on support vector machines (SVM), significantly outperforms previous FFL formulas. We also found that semantic features behave poorly in our case, in contrast with some previous readability studies on English as a first language.

Dynamic Programming for Higher Order Parsing of Gap-Minding Trees
Emily Pitler, Sampath Kannan, and Mitchell Marcus
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We introduce gap inheritance, a new structural property on trees, which provides a way to quantify the degree to which intervals of descendants can be nested. Based on this property, two new classes of trees are derived that provide a closer approximation to the set of plausible natural

language dependency trees than some alternative classes of trees: unlike projective trees, a word can have descendants in more than one interval; unlike spanning trees, these intervals cannot be nested in arbitrary ways. The 1-Inherit class of trees has exactly the same empirical coverage of natural language sentences as the class of mildly non-projective trees, yet the optimal scoring tree can be found in an order of magnitude less time. Gap-minding trees (the second class) have the property that all edges into an interval of descendants come from the same node, and thus an algorithm which uses only single intervals can produce trees in which a node has descendants in multiple intervals.

Joint Entity and Event Coreference Resolution across Documents
Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We introduce a novel coreference resolution system that models entities and events jointly. Our iterative method cautiously constructs clusters of entity and event mentions using lin- ear regression to model cluster merge opera- tions. As clusters are built, information flows between entity and event clusters through fea- tures that model semantic role dependencies. Our system handles nominal and verbal events as well as entities, and our joint formulation allows information from event coreference to help entity coreference, and vice versa. In a cross-document domain with comparable doc- uments, joint coreference resolution performs significantly better (over 3 CoNLL F1 points) than two strong baselines that resolve entities and events separately.

Joint Chinese Word Segmentation, POS Tagging and Parsing
Xian Qian and Yang Liu
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper, we propose a novel decoding algorithm for discriminative joint Chinese word segmentation, part-of-speech (POS) tagging, and parsing. Previous work often used a pipeline method – Chinese word segmentation followed by POS tagging and parsing, which suffers from error propagation and is unable to leverage information in later modules for earlier components. In our approach, we train the three individual models separately during training, and incorporate them together in a unified framework during decoding. We extend the CYK parsing algorithm so that it can deal with word segmentation and POS tagging features. As far as we know, this is the first work on joint Chinese word segmentation, POS tagging and parsing. Our experimental results on Chinese Tree Bank 5 corpus show that our approach outperforms the state-of-the-art pipeline system.

Translation Model Based Cross-Lingual Language Model Adaptation: from Word Models to Phrase Models
Shixiang Lu, Wei Wei, Xiaoyin Fu, and Bo Xu
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper, we propose a novel translation model (TM) based cross-lingual data selection model for language model (LM) adaptation in statistical machine translation (SMT), from word models to phrase models. Given a source sentence in the translation task, this model directly estimates the probability that a sentence in the target LM training corpus is similar. Compared with the traditional approaches which utilize the first pass translation hypotheses, cross-lingual data selection model avoids the problem of noisy proliferation. Furthermore, phrase TM based cross-lingual data selection model is more effective than the traditional approaches based on bag-of-words models and word-based TM, because it captures contextual information in modeling the selection of phrase as a whole. Experiments conducted on large-scale data sets demonstrate that our approach significantly outperforms the state-of-the-art approaches on both LM perplexity and SMT performance.

Open Language Learning for Information Extraction
Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Open Information Extraction (IE) systems extract relational tuples from text, without requiring a pre-specified vocabulary, by identifying relation phrases and associated arguments in arbitrary sentences. However, state-of-the-art Open IE systems such as REVERB and WOE share two important weaknesses — (1) they extract only relations that are mediated by verbs, and (2) they ignore context, thus extracting tuples that are not asserted as factual. This paper presents OLLIE, a substantially improved Open IE system that addresses both these limitations. First, OLLIE achieves high yield by extracting relations mediated by nouns, adjectives, and more. Second, a context-analysis step increases precision by including contextual information from the sentence in the extractions. OLLIE obtains 2.7 times the area under precision-yield curve (AUC) compared to REVERB and 1.9 times the AUC of WOE-parse.

Modelling Sequential Text with an Adaptive Topic Model
Lan Du, Wray Buntine, and Huidong Jin
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Topic models are increasingly being used for text analysis tasks, often times replacing earlier semantic techniques such as latent semantic analysis. In this paper, we develop a novel adaptive topic model with the ability to adapt topics from both the previous segment and the parent document. For this proposed model, a Gibbs sampler is developed for doing posterior inference. Experimental results show that with topic adaptation, our model significantly improves over existing approaches in terms of perplexity, and is able to uncover clear sequential structure on, for example, Herman Melville's book "Moby Dick".

A Comparison of Vector-based Representations for Semantic Composition
William Blacoe and Mirella Lapata
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper we address the problem of modeling compositional meaning for phrases and sentences using distributional methods. We experiment with several possible combinations of representation and composition, exhibiting varying degrees of sophistication. Some are shallow while others operate over syntactic structure, rely on parameter learning, or require access to very large corpora. We find that shallow approaches are as good as more computationally intensive alternatives with regards to two particular tests: (1) phrase similarity and (2) paraphrase detection. The sizes of the involved training corpora and the generated vectors are not as important as the fit between the meaning representation and compositional method.

Exploiting Chunk-level Features to Improve Phrase Chunking
Junsheng Zhou, Weiguang Qu, and Fen Zhang
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Most existing systems solved the phrase chunking task with the sequence labeling approaches, in which the chunk candidates cannot be treated as a whole during parsing process so that the chunk-level features cannot be exploited in a natural way. In this paper, we formulate phrase chunking as a joint segmentation and labeling task. We propose an efficient dynamic programming algorithm with pruning for decoding, which allows the direct use of the features describing the internal characteristics of chunk and the features capturing the correlations between adjacent chunks. A relaxed, online maximum margin training algorithm is used for learning. Within this framework, we explored a variety of effective feature representations for Chinese phrase chunking. The experimental results show that the use of chunk-level features can lead to significant performance improvement, and that our approach achieves state-of-the-art performance. In particular, our approach is much better at recognizing long and complicated phrases.

A Beam-Search Decoder for Grammatical Error Correction

Daniel Dahlmeier and Hwee Tou Ng

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We present a novel beam-search decoder for grammatical error correction. The decoder iteratively generates new hypothesis corrections from current hypotheses and scores them based on features of grammatical correctness and fluency. These features include scores from discriminative classifiers for specific error categories, such as articles and prepositions. Unlike all previous approaches, our method is able to perform correction of whole sentences with multiple and interacting errors while still taking advantage of powerful existing classifier approaches. Our decoder achieves an F1 correction score significantly higher than all previous published scores on the Helping Our Own (HOO) shared task data set.

A Statistical Relational Learning Approach to Identifying Evidence Based Medicine Categories

Mathias Verbeke, Vincent Van Asch, Roser Morante, Paolo Frasconi, Walter Daelemans, and Luc De Raedt

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Evidence-based medicine is an approach whereby clinical decisions are supported by the best available findings gained from scientific research. This requires efficient access to such evidence. To this end, abstracts in evidence-based medicine can be labeled using a set of predefined medical categories, the so-called PICO criteria. This paper presents an approach to automatically annotate sentences in medical abstracts with these labels. Since both structural and sequential information are important for this classification task, we use kLog, a new language for statistical relational learning with kernels. Our results show a clear improvement with respect to state-of-the-art systems.

Lyrics, Music, and Emotions

Rada Mihalcea and Carlo Strapparava

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper, we explore the classification of emotions in songs, using the music and the lyrics representation of the songs. We introduce a novel corpus of music and lyrics, consisting of 100 songs annotated for emotions. We show that textual and musical features can both be successfully used for emotion recognition in songs. Moreover, through comparative experiments, we show that the joint use of lyrics and music brings significant improvements over each of the individual textual and musical classifiers, with error rate reductions of up to 31%.

Assessment of ESL Learners' Syntactic Competence Based on Similarity Measures

Su-Youn Yoon and Suma Bhat

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This study presents a novel method that measures English language learners' syntactic competence towards improving automated speech scoring systems. In contrast to most previous studies which focus on the length of production units such as the mean length of clauses, we focused on capturing the differences in the distribution of morpho-syntactic features or grammatical expressions across proficiency. We estimated the syntactic competence through the use of corpus-based NLP techniques. Assuming that the range and sophistication of grammatical expressions can be captured by the distribution of Part-of-Speech (POS) tags, vector space models of POS tags were constructed. We use a large corpus of English learners' responses that are classified into four proficiency levels by human raters. Our proposed feature measures the similarity of a given response with the most proficient group and is then estimates the learner's syntactic competence level.

Widely outperforming the state-of-the-art measures of syntactic complexity, our method attained a significant correlation with human-rated scores. The correlation between human-rated scores and features based on manual transcription was 0.43 and the same based on ASR-hypothesis was slightly lower, 0.42. An important advantage of our method is its robustness

against speech recognition errors not to mention the simplicity of feature generation that captures a reasonable set of learner-specific syntactic errors.

A Unified Approach to Transliteration-based Text Input with Online Spelling Correction
Hisami Suzuki and Jianfeng Gao
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper presents an integrated, end-to-end approach to online spelling correction for text input. Online spelling correction refers to the spelling correction as you type, as opposed to post-editing. The online scenario is particularly important for languages that routinely use transliteration-based text input methods, such as Chinese and Japanese, because the desired target characters cannot be input at all unless they are in the list of candidates provided by an input method, and spelling errors prevent them from appearing in the list. For example, a user might type suesheng by mistake to mean xuesheng 生/‘ (s) “t”udent’ in Chinese; existing input methods fail to convert this misspelled input to the desired target Chinese characters. In this paper, we propose a unified approach to the problem of spelling correction and transliteration-based character conversion using an approach inspired by the phrase-based statistical machine translation framework. At the phrase (substring) level, k most probable pinyin (Romanized Chinese) corrections are generated using a monotone decoder; at the sentence level, input pinyin strings are directly transliterated into target Chinese characters by a decoder using a log-linear model that refer to the features of both levels. A new method of automatically deriving parallel training data from user keystroke logs is also presented. Experiments on Chinese pinyin conversion show that our integrated method reduces the character error rate by 20% (from 8.9% to 7.12%) over the previous state-of-the art based on a noisy channel model.

Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh, and Jun’ichi Kazama
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We propose a new semantic orientation, Excitation, and its automatic acquisition method. Excitation is a semantic property of predicates that classifies them into excitatory, inhibitory and neutral. We show that Excitation is useful for extracting contradiction pairs (e.g., destroy cancer vs. develop cancer) and causality pairs (e.g., increase in crime -> heighten anxiety). Our experiments show that with automatically acquired Excitation knowledge we can extract one million contradiction pairs and 500,000 causality pairs with about 70% precision from a 600 million page Web corpus. Furthermore, by combining these extracted causality and contradiction pairs, we can generate one million plausible causality hypotheses that are not written in any single sentence in our corpus with reasonable precision.

Enlarging Paraphrase Collections through Generalization and Instantiation
Atsushi Fujita, Pierre Isabelle, and Roland Kuhn
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper presents a paraphrase acquisition method that uncovers and exploits generalities underlying paraphrases: paraphrase patterns are first induced and then used to collect novel instances. Unlike existing methods, ours uses both bilingual parallel and monolingual corpora. While the former are regarded as a source of high-quality seed paraphrases, the latter are searched for paraphrases that match patterns learned from the seed paraphrases. We show how one can use monolingual corpora, which are far more numerous and larger than bilingual corpora, to obtain paraphrases that rival in quality those derived directly from bilingual corpora. In our experiments, the number of paraphrase pairs obtained in this way from monolingual corpora was a large multiple of the number of seed paraphrases. Human evaluation through a paraphrase substitution test demonstrated that the newly acquired paraphrase pairs are of reasonable quality. Remaining noise can be further reduced by filtering seed paraphrases.

Concurrent Acquisition of Word Meaning and Lexical Categories
Afra Alishahi and Grzegorz Chrupala
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Learning the meaning of words from ambiguous and noisy context is a challenging task for language learners. It has been suggested that children draw on syntactic cues such as lexical categories of words to constrain potential referents of words in a complex scene. Although the acquisition of lexical categories should be interleaved with learning word meanings, it has not previously been modeled in that fashion. In this paper, we investigate the interplay of word learning and category induction by integrating an LDA-based word class learning module with a probabilistic word learning model. Our results show that the incrementally induced word classes significantly improve word learning, and their contribution is comparable to that of manually assigned part of speech categories.

Do Neighbours Help? An Exploration of Graph-based Algorithms for Cross-domain Sentiment Classification
Natalia Ponomareva and Mike Thelwall
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper presents a comparative study of graph-based approaches for cross-domain sentiment classification. In particular, the paper analyses two existing methods: an optimisation problem and a ranking algorithm. We compare these graph-based methods with each other and with the other state-of-the-art approaches and conclude that graph domain representations offer a competitive solution to the domain adaptation problem. Analysis of the best parameters for graph-based algorithms reveals that there are no optimal values valid for all domain pairs and that these values are dependent on the characteristics of corresponding domains.

Learning Lexicon Models from Search Logs for Query Expansion
Jianfeng Gao, Shasha Xie, Xiaodong He, and Alnur Ali
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper explores log-based query expansion (QE) models for Web search. Three lexicon models are proposed to bridge the lexical gap between Web documents and user queries. These models are trained on pairs of user queries and titles of clicked documents. Evaluations on a real world data set show that the lexicon models, integrated into a ranker-based QE system, not only significantly improve the document retrieval performance but also outperform two state-of-the-art log-based QE methods.

Joint Inference for Event Timeline Construction
Quang Do, Wei Lu, and Dan Roth
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper addresses the task of constructing a timeline of events mentioned in a given text. To accomplish that, we present a novel representation of the temporal structure of a news article based on time intervals. We then present an algorithmic approach that jointly optimizes the temporal structure by coupling local classifiers that predict associations and temporal relations between pairs of temporal entities with global constraints. Moreover, we present ways to leverage knowledge provided by event coreference to further improve the system performance. Overall, our experiments show that the joint inference model significantly outperformed the local classifiers by 9.2% of relative improvement in F_1. The experiments also suggest that good event coreference could make remarkable contribution to a robust event timeline construction system.

Three Dependency-and-Boundary Models for Grammar Induction
Valentin I. Spitkovsky, Hiyan Alshawi, and Daniel Jurafsky
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We present a new family of models for unsupervised parsing, Dependency and Boundary models, that use cues at constituent boundaries to inform head-outward dependency tree generation.

We build on three intuitions that are explicit in phrase-structure grammars but only implicit in standard dependency formulations: (i) Distributions of words that occur at sentence boundaries – such as English determiners – resemble constituent edges. (ii) Punctuation at sentence boundaries further helps distinguish full sentences from fragments like headlines and titles, allowing us to model grammatical differences between complete and incomplete sentences. (iii) Sentence-internal punctuation boundaries help with longer-distance dependencies, since punctuation correlates with constituent edges. Our models induce state-of-the-art dependency grammars for many languages without special knowledge of optimal input sentence lengths or biased, manually-tuned initializers.

Exploring Adaptor Grammars for Native Language Identification

Sze-Meng Jojo Wong, Mark Dras, and Mark Johnson

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

The task of inferring the native language of an author based on texts written in a second language has generally been tackled as a classification problem, typically using as features a mix of n-grams over characters and part of speech tags (for small and fixed n) and unigram function words. To capture arbitrarily long n-grams that syntax-based approaches have suggested are useful, adaptor grammars have some promise. In this work we investigate their extension to identifying n-gram collocations of arbitrary length over a mix of PoS tags and words, using both maxent and induced syntactic language model approaches to classification. After presenting a new, simple baseline, we show that learned collocations used as features in a maxent model perform better still, but that the story is more mixed for the syntactic language model.

Discovering Diverse and Salient Threads in Document Collections

Jennifer Gillenwater, Alex Kulesza, and Ben Taskar

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We propose a novel probabilistic technique for modeling and extracting salient structure from large document collections. As in clustering and topic modeling, our goal is to provide an organizing perspective into otherwise overwhelming amounts of information. We are particularly interested in revealing and exploiting relationships between documents. To this end, we focus on extracting diverse sets of threads—singly-linked, coherent chains of important documents. To illustrate, we extract research threads from citation graphs and construct timelines from news articles. Our method is highly scalable, running on a corpus of over 30 million words in about four minutes, more than 75 times faster than a dynamic topic model. Finally, the results from our model more closely resemble human news summaries according to several metrics and are also preferred by human judges.

Generalizing Sub-sentential Paraphrase Acquisition across Original Signal Type of Text Pairs

Aurélien Max, Houda Bouamor, and Anne Vilnat

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper describes a study on the impact of the original signal (text, speech, visual scene, event) of a text pair on the task of both man- ual and automatic sub-sentential paraphrase acquisition. A corpus of 2,500 annotated sen- tences in English and French is described, and performance on this corpus is reported for an efficient system combination exploiting a large set of features for paraphrase recogni- tion. A detailed quantified typology of sub- sentential paraphrases found in our corpus types is given.

Parse, Price and Cut—Delayed Column and Row Generation for Graph Based Parsers

Sebastian Riedel, David Smith, and Andrew McCallum

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Graph-based dependency parsers suffer from the sheer number of higher order edges they need

to (a) score and (b) consider during optimization. Here we show that when working with LP relaxations, large fractions of these edges can be pruned before they are fully scored—without any loss of optimality guarantees and, hence, accuracy. This is achieved by iteratively parsing with a subset of higher-order edges, adding higher-order edges that may improve the score of the current solution, and adding higher-order edges that are implied by the current best first order edges. This amounts to delayed column and row generation in the LP relaxation and is guaranteed to provide the optimal LP solution. For second order grandparent models, our method considers, or scores, no more than 6–13% of the second order edges of the full model. This yields up to an eightfold parsing speedup, while providing the same empirical accuracy and certificates of optimality as working with the full LP relaxation. We also provide a tighter LP formulation for grandparent models that leads to a smaller integrality gap and higher speed.

Domain Adaptation for Coreference Resolution: An Adaptive Ensemble Approach
Jian Bo Yang, Qi Mao, Qiao Liang Xiang, Ivor Wai-Hung Tsang, Kian Ming Adam Chai, and Hai Leong Chieu
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We propose an adaptive ensemble method to adapt coreference resolution across domains. This method has three features: (1) it can optimize for any user-specified objective measure; (2) it can make document-specific prediction rather than rely on a fixed base model or a fixed set of base models; (3) it can automatically adjust the active ensemble members during prediction. With simplification, this ethod can be used in the traditional within-domain case, while still retaining the above features. To the best of our knowledge, this work is the first to both (i) develop a domain adaptation algorithm for the coreference resolution problem and (ii) have the above features as an ensemble method. Empirically, we show the benefits of (i) on the six domains of the ACE 2005 data set in domain adaptation setting, and of (ii) on both the MUC-6 and the ACE 2005 data sets in within-domain setting.

Weakly Supervised Training of Semantic Parsers
Jayant Krishnamurthy and Tom Mitchell
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We present a method for training a semantic parser using only a knowledge base and an unlabeled text corpus, without any individually annotated sentences. Our key observation is that multiple forms of weak supervision can be combined to train an accurate semantic parser: semantic supervision from a knowledge base, and syntactic supervision from dependency-parsed sentences. We apply our approach to train a semantic parser that uses 77 relations from Freebase in its knowledge representation. This semantic parser extracts instances of binary relations with state-of-the-art accuracy, while simultaneously recovering much richer semantic structures, such as conjunctions of multiple relations with partially shared arguments. We demonstrate recovery of this richer structure by extracting logical forms from natural language queries against Freebase. On this task, the trained semantic parser achieves 80% precision and 56% recall, despite never having seen an annotated logical form.

Cross-Lingual Language Modeling with Syntactic Reordering for Low-Resource Speech Recognition
Ping Xu and Pascale Fung
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

This paper proposes cross-lingual language modeling for transcribing source resource-poor languages and translating them into target resource-rich languages if necessary. Our focus is to improve the speech recognition performance of low-resource languages by leveraging the language model statistics from resource-rich languages. The most challenging work of cross-lingual language modeling is to solve the syntactic discrepancies between the source and target languages. We therefore propose syntactic reordering for cross-lingual language modeling, and present a first result that compares inversion transduction grammar (ITG) reordering constraints to IBM

and local constraints in an integrated speech transcription and translation system. Evaluations on resource-poor Cantonese speech transcription and Cantonese to resource-rich Mandarin translation tasks show that our proposed approach improves the system performance significantly, up to 3.4% relative WER reduction in Cantonese transcription and 13.3% relative bilingual evaluation understudy (BLEU) score improvement in Mandarin transcription compared with the system without reordering.

Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge
Altaf Rahman and Vincent Ng
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

We examine the resolution of complex cases of definite pronouns, specifically those for which traditional syntactic and semantic constraints on coreference (e.g., Binding Constraints, agreement in gender, number and semantic class) as well as commonly-used resolution heuristics (e.g., string-matching facilities, syntactic salience) are not useful. We hypothesized that coreference resolvers that achieved state-of-the-art performance on standard evaluation corpora would not work well for these difficult pronouns, and investigate a variety of knowledge sources for their resolution. Experimental results on our dataset indicate that our resolver outperforms state-of-the-art resolvers by nearly 18 points in accuracy.

A Sequence Labelling Approach to Quote Attribution
Timothy O'Keefe, Silvia Pareti, James R. Curran, Irena Koprinska, and Matthew Honnibal
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Quote extraction and attribution is the task of automatically extracting quotes from text and attributing each quote to its correct speaker. The present state-of-the-art system uses gold standard information from previous decisions in its features, which, when removed, results in a large drop in performance. We treat the problem as a sequence labelling task, which allows us to incorporate sequence features without using gold standard information. We present results on two new corpora and an augmented version of a third, achieving a new state-of-the-art for systems using only realistic features.

SSHLDA: A Semi-Supervised Hierarchical Topic Model
Xian-Ling Mao, Zhao-Yan Ming, Tat-Seng Chua, Si Li, Hongfei Yan, and Xiaoming Li
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Supervised hierarchical topic modeling and unsupervised hierarchical topic modeling are usually used to obtain hierarchical topics, such as hLLDA and hLDA. Supervised hierarchical topic modeling makes heavy use of the information from observed hierarchical labels, but cannot explore new topics; while unsupervised hierarchical topic modeling is able to detect automatically new topics in the data space, but does not make use of any information from hierarchical labels. In this paper, we propose a semi-supervised hierarchical topic model which aims to explore new topics automatically in the data space while incorporating the information from observed hierarchical labels into the modeling process, called Semi-Supervised Hierarchical Latent Dirichlet Allocation (SSHLDA). We also prove that hLDA and hLLDA are the special cases of SSHLDA. We conduct experiments on Yahoo! Answers and ODP datasets, and assess the performance in terms of perplexity and clustering. The experimental results show that predictive ability of SSHLDA is better than that of baselines, and SSHLDA can also achieve significant improvement over baselines for clustering on the FScore measure.

Improving NLP through Marginalization of Hidden Syntactic Structure
Jason Naradowsky, Sebastian Riedel, and David Smith
Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Many NLP tasks make predictions that are inherently coupled to syntactic relations, but for many languages the resources required to provide such syntactic annotations are unavailable. For

others it is unclear exactly how much of the syntactic annotations can be effectively leveraged with current models, and what structures in the syntactic trees are most relevant to the current task.

We propose a novel method which avoids the need for any syntactically annotated data when predicting a related NLP task. Our method couples latent syntactic representations, constrained to form valid dependency graphs or constituency parses, with the prediction task via specialized factors in a Markov random field. At both training and test time we marginalize over this hidden structure, learning the optimal latent representations for the problem. Results show that this approach provides significant gains over a syntactically uninformed baseline, outperforming models that observe syntax on an English relation extraction task, and performing comparably to them in semantic role labeling.

Type-Supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries

Dan Garrette and Jason Baldridge

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

Past work on learning part-of-speech taggers from tag dictionaries and raw data has reported good results, but the assumptions made about those dictionaries are often unrealistic: due to historical precedents, they assume access to information about labels in the raw and test sets. Here, we demonstrate ways to learn hidden Markov model taggers from incomplete tag dictionaries. Taking the MIN-GREEDY algorithm (Ravi et al., 2010) as a starting point, we improve it with several intuitive heuristics. We also define a simple HMM emission initialization that takes advantage of the tag dictionary and raw data to capture both the openness of a given tag and its estimated prevalence in the raw data. Altogether, our augmentations produce improvements to performance over the original MIN-GREEDY algorithm for both English and Italian data.

Explore Person Specific Evidence in Web Person Name Disambiguation

Liwei Chen, Yansong Feng, Lei Zou, and Dongyan Zhao

Thursday 6:00pm–10:00pm — 3F Lobby (ICC)

In this paper, we investigate different usages of feature representations in the web person name disambiguation task which has been suffering from the mismatch of vocabulary and lack of clues in web environments. In literature, the latter receives less attention and remains more challenging. We explore the feature space in this task and argue that collecting person specific evidences from a corpus level can provide a more reasonable and robust estimation for evaluating a feature's importance in a given web page. This can alleviate the *lack of clues* where discriminative features can be reasonably weighted by taking their corpus level importance into account, not just relying on the current local context. We therefore propose a topic-based model to exploit the person specific global importance and embed it into the person name similarity. The experimental results show that the corpus level topic information provides more stable evidences for discriminative features and our method outperforms the state-of-the-art systems on three WePS datasets.

Oral Presentations: Friday, July 13, 2012

Inducing a Discriminative Parser to Optimize Machine Translation Reordering
Graham Neubig, Taro Watanabe, and Shinsuke Mori
Friday 11:00am–11:30am — Youngju A (ICC)

This paper proposes a method for learning a discriminative parser for machine translation reordering using only aligned parallel text. This is done by treating the parser's derivation tree as a latent variable in a discriminative model that is trained to maximize reordering accuracy. We demonstrate that efficient large-margin training is possible by showing that two measures of reordering accuracy can be factored over the parse tree. Using this model in the pre-ordering framework results in significant gains in translation accuracy over standard phrase-based SMT and previously proposed unsupervised syntax induction methods.

Re-training Monolingual Parser Bilingually for Syntactic SMT
Shujie Liu, Chi-Ho Li, Mu Li, and Ming Zhou
Friday 11:30am–12:00pm — Youngju A (ICC)

The training of most syntactic SMT approaches involves two essential components, word alignment and monolingual parser. In the current state of the art these two components are mutually independent, thus causing problems like lack of rule generalization, and violation of syntactic correspondence in translation rules. In this paper, we propose two ways of re-training monolingual parser with the target of maximizing the consistency between parse trees and alignment matrices. One is targeted self-training with a simple evaluation function; the other is based on training data selection from forced alignment of bilingual data. We also propose an auxiliary method for boosting alignment quality, by symmetrizing alignment matrices with respect to parse trees. The best combination of these novel methods achieves 3 Bleu point gain in an IWSLT task and more than 1 Bleu point gain in NIST tasks.

Transforming Trees to Improve Syntactic Convergence
David Burkett and Dan Klein
Friday 12:00pm–12:30pm — Youngju A (ICC)

We describe a transformation-based learning method for learning a sequence of monolingual tree transformations that improve the agreement between constituent trees and word alignments in bilingual corpora. Using the manually annotated English Chinese Translation Treebank, we show how our method automatically discovers transformations that accommodate differences in English and Chinese syntax. Furthermore, when transformations are learned on automatically generated trees and alignments from the same domain as the training data for a syntactic MT system, the transformed trees achieve a 0.9 BLEU improvement over baseline trees.

Learning Constraints for Consistent Timeline Extraction
David McClosky and Christopher D. Manning
Friday 11:00am–11:30am — 203 (ICC)

We present a distantly supervised system for extracting the temporal bounds of fluents (relations which only hold during certain times, such as attends school). Unlike previous pipelined approaches, our model does not assume independence between each fluent or even between named entities with known connections (parent, spouse, employer, etc.). Instead, we model what makes timelines of fluents consistent by learning cross-fluent constraints, potentially spanning entities as well. For example, our model learns that someone is unlikely to start a job at age two or to marry someone who hasn't been born yet. Our system achieves a 36% error reduction over a pipelined baseline.

Identifying Constant and Unique Relations by using Time-Series Text
Yohei Takaku, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda
Friday 11:30am–12:00pm — 203 (ICC)

Because the real world evolves over time, numerous relations between entities written in presently available texts are already obsolete or will potentially evolve in the future. This study aims at resolving the intricacy in consistently compiling relations extracted from text, and presents a method for identifying constancy and uniqueness of the relations in the context of supervised learning. We exploit massive time-series web texts to induce features on the basis of time-series frequency and linguistic cues. Experimental results confirmed that the time-series frequency distributions contributed much to the recall of constancy identification and the precision of the uniqueness identification.

No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities
Thomas Lin, Mausam, and Oren Etzioni
Friday 12:00pm–12:30pm — 203 (ICC)

Entity linking systems link noun-phrase mentions in text to their corresponding Wikipedia articles. However, NLP applications would gain from the ability to detect and type all entities mentioned in text, including the long tail of entities not prominent enough to have their own Wikipedia articles. In this paper we show that once the Wikipedia entities mentioned in a corpus of textual assertions are linked, this can further enable the detection and fine-grained typing of the unlinkable entities. Our proposed method for detecting unlinkable entities achieves 24% greater accuracy than a Named Entity Recognition baseline, and our method for fine-grained typing is able to propagate over 1,000 types from linked Wikipedia entities to unlinkable entities. Detection and typing of unlinkable entities can increase yield for NLP applications such as typed question answering.

A Novel Discriminative Framework for Sentence-Level Discourse Analysis
Shafiq Joty, Giuseppe Carenini, and Raymond Ng
Friday 11:00am–11:30am — 202 A (ICC)

We propose a complete probabilistic discriminative framework for performing sentence-level discourse analysis. Our framework comprises a discourse segmenter, based on a binary classifier, and a discourse parser, which applies an optimal CKY-like parsing algorithm to probabilities inferred from a Dynamic Conditional Random Field. We show on two corpora that our approach outperforms the state-of-the-art, often by a wide margin.

Using Discourse Information for Paraphrase Extraction
Michaela Regneri and Rui Wang
Friday 11:30am–12:00pm — 202 A (ICC)

Previous work on paraphrase extraction using parallel or comparable corpora has generally not considered the documents' discourse structure as a useful information source. We propose a novel method for collecting paraphrases relying on the sequential event order in the discourse, using multiple sequence alignment with a semantic similarity measure. We show that adding discourse information boosts the performance of sentence-level paraphrase acquisition, which consequently gives a tremendous advantage for extracting phrase-level paraphrase fragments from matched sentences. Our system beats an informed baseline by a margin of 50%.

Generating Non-Projective Word Order in Statistical Linearization
Bernd Bohnet, Anders Björkelund, Jonas Kuhn, Wolfgang Seeker, and Sina Zarriess
Friday 12:00pm–12:30pm — 202 A (ICC)

We propose a technique to generate non-projective word orders in an efficient statistical linearization system. Our approach predicts liftings of edges in an unordered syntactic tree by means of a classifier, and uses a projective algorithm for tree linearization. We obtain statistically significant improvements on six typologically different languages: English, German, Dutch, Danish,

Hungarian, and Czech.

Learning Syntactic Categories Using Paradigmatic Representations of Word Context
Mehmet Ali Yatbaz, Enis Sert, and Deniz Yuret
Friday 2:30pm–3:00pm — Youngju A (ICC)

We investigate paradigmatic representations of word context in the domain of unsupervised syntactic category acquisition. Paradigmatic representations of word context are based on potential substitutes of a word in contrast to syntagmatic representations based on properties of neighboring words. We compare a bigram based baseline model with several paradigmatic models and demonstrate significant gains in accuracy. Our best model based on Euclidean co-occurence embedding combines paradigmatic features with morphological and orthographic features and achieves 80% many-to-one accuracy on a 45-tag million-word test corpus.

Exploring Topic Coherence over Many Models and Many Topics
Keith Stevens, Philip Kegelmeyer, David Andrzejewski, and David Buttler
Friday 3:00pm–3:30pm — Youngju A (ICC)

We apply two new automated semantic evaluations to three distinct latent topic models. Both metrics have been shown to align with human evaluations and provide a balance between internal measures of information gain and comparisons to human ratings of coherent topics. We improve upon the measures by introducing new aggregate measures that allows for comparing complete topic models. We further compare the automated measures to other metrics for topic models, comparison to manually crafted semantic tests and document classification. Our experiments reveal that LDA and LSA each have different strengths; LDA best learns descriptive topics while LSA is best at creating a compact semantic representation of documents and words in a corpus.

Entropy-based Pruning for Phrase-based Machine Translation
Wang Ling, João Graça, Isabel Trancoso, and Alan Black
Friday 2:30pm–3:00pm — 203 (ICC)

Phrase-based machine translation models have shown to yield better translations than Word-based models, since phrase pairs encode the contextual information that is needed for a more accurate translation. However, many phrase pairs do not encode any relevant context, which means that the translation event encoded in that phrase pair is led by smaller translation events that are independent from each other, and can be found on smaller phrase pairs, with little or no loss in translation accuracy. In this work, we propose a relative entropy model for translation models, that measures how likely a phrase pair encodes a translation event that is derivable using smaller translation events with similar probabilities. This model is then applied to phrase table pruning. Tests show that considerable amounts of phrase pairs can be excluded, without much impact on the translation quality. In fact, we show that better translations can be obtained using our pruned models, due to the compression of the search space during decoding.

A Systematic Comparison of Phrase Table Pruning Techniques
Richard Zens, Daisy Stanton, and Peng Xu
Friday 3:00pm–3:30pm — 203 (ICC)

When trained on very large parallel corpora, the phrase table component of a machine translation system grows to consume vast computational resources. In this paper, we introduce a novel pruning criterion that places phrase table pruning on a sound theoretical foundation. Systematic experiments on four language pairs under various data conditions show that our principled approach is superior to existing ad hoc pruning methods.

Probabilistic Finite State Machines for Regression-based MT Evaluation
Mengqiu Wang and Christopher D. Manning
Friday 2:30pm–3:00pm — 202 A (ICC)

Accurate and robust metrics for automatic evaluation are key to the development of statistical machine translation (MT) systems. We first introduce a new regression model that uses a probabilistic finite state machine (pFSM) to compute weighted edit distance as predictions of translation quality. We also propose a novel pushdown automaton extension of the pFSM model for modeling word swapping and cross alignments that cannot be captured by standard edit distance models. Our models can easily incorporate a rich set of linguistic features, and automatically learn their weights, eliminating the need for ad-hoc parameter tuning. Our methods achieve state-of-the-art correlation with human judgments on two different prediction tasks across a diverse set of standard evaluations (NIST OpenMT06,08; WMT06- 08).

An Empirical Investigation of Statistical Significance in NLP
Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein
Friday 3:00pm–3:30pm — 202 A (ICC)

We investigate two aspects of the empirical behavior of paired significance tests for NLP systems. First, when one system appears to outperform another, how does significance level relate in practice to the magnitude of the gain, to the size of the test set, to the similarity of the systems, and so on? Is it true that for each task there is a gain which roughly implies significance? We explore these issues across a range of NLP tasks using both large collections of past systems' outputs and variants of single systems. Next, once significance levels are computed, how well does the standard i.i.d. notion of significance hold up in practical settings where future distributions are neither independent nor identically distributed, such as across domains? We explore this question using a range of test set variations for constituency parsing.

Employing Compositional Semantics and Discourse Consistency in Chinese Event Extraction
Peifeng Li, Guodong Zhou, Qiaoming Zhu, and Libin Hou
Friday 4:00pm–4:30pm — Youngju A (ICC)

Current Chinese event extraction systems suffer much from two problems in trigger identification: unknown triggers and word segmentation errors to known triggers. To resolve these problems, this paper proposes two novel inference mechanisms to explore special characteristics in Chinese via compositional semantics inside Chinese triggers and discourse consistency between Chinese trigger mentions. Evaluation on the ACE 2005 Chinese corpus justifies the effectiveness of our approach over a strong baseline.

Reading The Web with Learned Syntactic-Semantic Inference Rules
Ni Lao, Amarnag Subramanya, Fernando Pereira, and William W. Cohen
Friday 4:30pm–5:00pm — Youngju A (ICC)

We study how to extend a large knowledge base (Freebase) by reading relational information from a large Web text corpus. Previous studies on extracting relational knowledge from text show the potential of syntactic patterns for extraction, but they do not exploit background knowledge of other relations in the knowledge base. We describe a distributed, Web-scale implementation of a path-constrained random walk model that learns syntactic-semantic inference rules for binary relations from a graph representation of the parsed text and the knowledge base. Experiments show significant accuracy improvements in binary relation prediction over methods that consider only text, or only the existing knowledge base.

Ensemble Semantics for Large-scale Unsupervised Relation Extraction
Bonan Min, Shuming Shi, Ralph Grishman, and Chin-Yew Lin
Friday 5:00pm–5:30pm — Youngju A (ICC)

Discovering significant types of relations from the web is challenging because of its open nature.

Unsupervised algorithms are developed to extract relations from a corpus without knowing the relations in advance, but most of them rely on tagging arguments of predefined types. Recently, a new algorithm was proposed to jointly extract relations and their argument semantic classes, taking a set of relation instances extracted by an open IE algorithm as input. However, it cannot handle polysemy of relation phrases and fails to group many similar ("synonymous") relation instances because of the sparseness of features. In this paper, we present a novel unsupervised algorithm that provides a more general treatment of the polysemy and synonymy problems. The algorithm incorporates various knowledge sources which we will show to be very effective for unsupervised extraction. Moreover, it explicitly disambiguates polysemous relation phrases and groups synonymous ones. While maintaining approximately the same precision, the algorithm achieves significant improvement on recall compared to the previous method. It is also very efficient. Experiments on a real-world dataset show that it can handle 14.7 million relation instances and extract a very large set of relations from the web.

Forest Reranking through Subtree Ranking

Richard Farkas and Helmut Schmid

Friday 4:00pm–4:30pm — 203 (ICC)

We propose the subtree ranking approach to parse forest reranking which is a generalization of current perceptron-based reranking methods. For the training of the reranker, we extract competing local subtrees, hence the training instances (candidate subtree sets) are very similar to those used during beam-search parsing. This leads to better parameter optimization. Another chief advantage of the framework is that arbitrary learning to rank methods can be applied. We evaluated our reranking approach on German and English phrase structure parsing tasks and compared it to various state-of-the-art reranking approaches such as the perceptron-based forest reranker. The subtree ranking approach with a Maximum Entropy model significantly outperformed the other approaches.

Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output

Jonathan K. Kummerfeld, David Hall, James R. Curran, and Dan Klein

Friday 4:30pm–5:00pm — 203 (ICC)

Constituency parser performance is primarily interpreted through a single metric, F-score on WSJ section 23, that conveys no linguistic information regarding the remaining errors. We classify errors within a set of linguistically meaningful types using tree transformations that repair groups of errors together. We use this analysis to answer a range of questions about parser behaviour, including what linguistic constructions are difficult for state-of-the-art parsers, what types of errors are being resolved by rerankers, and what types are introduced when parsing out-of-domain text.

Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level

Billy T. M. Wong and Chunyu Kit

Friday 5:00pm–5:30pm — 203 (ICC)

This paper proposes the utilization of lexical cohesion to facilitate evaluation of machine translation at the document level. As a linguistic means to achieve text coherence, lexical cohesion ties sentences together into a meaningfully interwoven structure through words with the same or related meaning. A comparison between machine and human translation is conducted to illustrate one of their critical distinctions that human translators tend to use more cohesion devices than machine. Various ways to apply this feature to evaluate machine-translated documents are presented, including one without reliance on reference translation. Experimental results show that incorporating this feature into sentence-level evaluation metrics can enhance their correlation with human judgements.

Fast Large-Scale Approximate Graph Construction for NLP

Amit Goyal, Hal Daume III, and Raul Guerra

Friday 4:00pm–4:30pm — 202 A (ICC)

Many natural language processing problems involve constructing large nearest-neighbor graphs. We propose a system called FLAG to construct such graphs approximately from large data sets. To handle the large amount of data, our algorithm maintains approximate counts based on sketching algorithms. To find the approximate nearest neighbors, our algorithm pairs a new distributed online-PMI algorithm with novel fast approximate nearest neighbor search algorithms (variants of PLEB). These algorithms return the approximate nearest neighbors quickly. We show our system's efficiency in both intrinsic and extrinsic experiments. We further evaluate our fast search algorithms both quantitatively and qualitatively on two NLP applications.

Building a Lightweight Semantic Model for Unsupervised Information Extraction on Short Listings

Doo Soon Kim, Kunal Verma, and Peter Yeh

Friday 4:30pm–5:00pm — 202 A (ICC)

Short listings such as classified ads or product listings abound on the web. If a computer can reliably extract information from them, it will greatly benefit a variety of applications. Short listings are, however, challenging to process due to their informal styles. In this paper, we present an unsupervised information extraction system for short listings. Given a corpus of listings, the system builds a semantic model that represents typical objects and their attributes in the domain of the corpus, and then uses the model to extract information. Two key features in the system are a semantic parser that extracts objects and their attributes and a listing-focused clustering module that helps group together the extracted tokens of same type. Our evaluation shows that the semantic model learned by these two modules is effective across multiple domains.

Sketch Algorithms for Estimating Point Queries in NLP

Amit Goyal, Hal Daume III, and Graham Cormode

Friday 5:00pm–5:30pm — 202 A (ICC)

Many NLP tasks rely on accurate statistics from large corpora. Tracking complete statistics is memory intensive, so recent work has proposed using compact approximate sketches of frequency distributions. We describe 10 sketch methods, including existing and novel variants. We compare and study the errors (over-estimation and under-estimation) made by the sketches. We evaluate several sketches on three important NLP problems. Our experiments show that one sketch performs best for all the three tasks.

Monte Carlo MCMC: Efficient Inference by Approximate Sampling

Sameer Singh, Michael Wick, and Andrew McCallum

Friday 4:00pm–4:30pm — 202 B (ICC)

Conditional random fields and other graphical models have achieved state of the art results in a variety of tasks such as coreference, relation extraction, data integration, and parsing. Increasingly, practitioners are using models with more complex structure—higher tree-width, larger fan-out, more features, and more data—rendering even approximate inference methods such as MCMC inefficient. In this paper we propose an alternative MCMC sampling scheme in which transition probabilities are approximated by sampling from the set of relevant factors. We demonstrate that our method converges more quickly than a traditional MCMC sampler for both marginal and MAP inference. In an author coreference task with over 5 million mentions, we achieve a 13 times speedup over regular MCMC inference.

On Amortizing Inference Cost for Structured Prediction

Vivek Srikumar, Gourab Kundu, and Dan Roth

Friday 4:30pm–5:00pm — 202 B (ICC)

This paper deals with the problem of predicting structures in the context of NLP. Typically, in

structure prediction, an inference procedure is applied to each example independently of the others. In this paper, we seek to optimize the inference time complexity over entire datasets, rather than individual examples. By considering the general inference representation provided by integer linear programs, we propose three exact inference theorems which allows us to reuse earlier solutions for certain examples, thereby completely avoiding possibly expensive calls to an ILP solver. We also identify several approximation schemes which can provide further speedup. We instantiate these ideas to the structure prediction task of semantic role labeling and show that we can achieve a speedup of over 2.5 using our approach while retaining the guarantees of exactness and a further speedup of over 3 using an approximation that does not degrade performance.

Exact Sampling and Decoding in High-Order Hidden Markov Models

Simon Carter, Marc Dymetman, and Guillaume Bouchard

Friday 5:00pm–5:30pm — 202 B (ICC)

We present a method for exact optimization and sampling from high order Hidden Markov Models (HMMs), which are generally handled by approximation techniques. Motivated by adaptive rejection sampling and heuristic search, we propose a strategy based on sequentially refining a lower-order language model that is an upper bound on the true model we wish to decode and sample from. This allows us to build tractable variable-order HMMs. The ARPA format for language models is extended to enable an efficient use of the *max-backoff* quantities required to compute the upper bound. We evaluate our approach on two problems: a SMS-retrieval task and a POS tagging experiment using 5-gram models. Results show that the same approach can be used for exact optimization and sampling, while explicitly constructing only a fraction of the huge implicit state-space.

Oral Presentations: Saturday, July 14, 2012

PATTY: A Taxonomy of Relational Patterns with Semantic Types
Ndapandula Nakashole, Gerhard Weikum, and Fabian Suchanek
Saturday 9:00am–9:30am — Youngju A (ICC)

This paper presents PATTY: a large resource for textual patterns that denote binary relations between entities. The patterns are semantically typed and organized into a subsumption taxonomy. The PATTY system is based on efficient algorithms for frequent itemset mining and can process Web-scale corpora. It harnesses the rich type system and entity population of large knowledge bases. The PATTY taxonomy comprises 350,569 pattern synsets. Random-sampling-based evaluation shows a pattern accuracy of 84.7%. PATTY has 8,162 subsumptions, with a random-sampling-based precision of 75%

Training Factored PCFGs with Expectation Propagation
David Hall and Dan Klein
Saturday 9:30am–10:00am — Youngju A (ICC)

PCFGs can grow exponentially as additional annotations are added to an initially simple base grammar. We present an approach where multiple annotations coexist, but in a factored manner that avoids this combinatorial explosion. Our method works with linguistically-motivated annotations, induced latent structure, lexicalization, or any mix of the three. We use a structured expectation propagation algorithm that makes use of the factored structure in two ways. First, by partitioning the factors, it speeds up parsing exponentially over the unfactored approach. Second, it minimizes the redundancy of the factors during training, improving accuracy over an independent approach. Using purely latent variable annotations, we can efficiently train and parse with up to 8 latent bits per symbol, achieving F1 scores up to 88.4 on the Penn Treebank while using two orders of magnitudes fewer parameters compared to the naive approach. Combining latent, lexicalized, and unlexicalized annotations, our best parser gets 89.4 F1 on all sentences from section 23 of the Penn Treebank.

A Coherence Model Based on Syntactic Patterns
Annie Louis and Ani Nenkova
Saturday 10:00am–10:30am — Youngju A (ICC)

We introduce a model of coherence which captures the intentional discourse structure in text. Our work is based on the hypothesis that syntax provides a proxy for the communicative goal of a sentence and therefore the sequence of sentences in a coherent discourse should exhibit detectable structural patterns. Results show that our method has high discriminating power for separating out coherent and incoherent news articles reaching accuracies of up to 90%. We also show that our syntactic patterns are correlated with manual annotations of intentional structure for academic conference articles and can successfully predict the coherence of abstract, introduction and related work sections of these articles.

Language Model Rest Costs and Space-Efficient Storage
Kenneth Heafield, Philipp Koehn, and Alon Lavie
Saturday 11:00am–11:30am — Youngju A (ICC)

Approximate search algorithms, such as cube pruning in syntactic machine translation, rely on the language model to estimate probabilities of sentence fragments. We contribute two changes that trade between accuracy of these estimates and memory, holding sentence-level scores constant. Common practice uses lower-order entries in an N-gram model to score the first few words of a fragment; this violates assumptions made by common smoothing strategies, including Kneser-Ney. Instead, we use a unigram model to score the first word, a bigram for the second, etc. This improves search at the expense of memory. Conversely, we show how to save memory by collapsing probability and backoff into a single value without changing sentence-level scores, at the expense of less accurate estimates for sentence fragments. These changes can be stacked, achieving better estimates with unchanged memory usage. In order to interpret changes in search

accuracy, we adjust the pop limit so that accuracy is unchanged and report the change in CPU time. In a German-English Moses system with target-side syntax, improved estimates yielded a 63% reduction in CPU time; for a Hiero-style version, the reduction is 21%. The compressed language model uses 26% less RAM while equivalent search quality takes 27% more CPU. Source code is released as part of KenLM.

Document-Wide Decoding for Phrase-Based Statistical Machine Translation

Christian Hardmeier, Joakim Nivre, and Jörg Tiedemann

Saturday 11:30am–12:00pm — Youngju A (ICC)

Independence between sentences is an assumption deeply entrenched in the models and algorithms used for Statistical Machine Translation (SMT), particularly in the popular dynamic programming beam search decoding algorithm. This restriction is an obstacle to research on more sophisticated discourse-level models for SMT. We propose a stochastic local search decoding method for phrase-based SMT, which permits free document-wide dependencies in the models. We explore the stability and the search parameters of this method and demonstrate that it can be successfully used to optimise a document-level semantic language model.

Left-to-Right Tree-to-String Decoding with Prediction

Yang Feng, Yang Liu, Qun Liu, and Trevor Cohn

Saturday 12:00pm–12:30pm — Youngju A (ICC)

Decoding algorithms for syntax based machine translation suffer from high computational complexity, a consequence of intersecting a language model with a context free grammar. Left-to-right decoding, which generates the target string in order, can improve decoding efficiency by simplifying the language model evaluation. This paper presents a novel left to right decoding algorithm for tree-to-string translation, using a bottom-up parsing strategy and dynamic future cost estimation for each partial translation. Our method outperforms previously published tree-to-string decoders, including a competing left-to-right method.

Semantic Compositionality through Recursive Matrix-Vector Spaces

Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng

Saturday 11:00am–11:30am — 203 (ICC)

Single-word vector space models have been very successful at learning lexical information. However, they cannot capture the compositional meaning of longer phrases, preventing them from a deeper understanding of language. We introduce a recursive neural network (RNN) model that learns compositional vector representations for phrases and sentences of arbitrary syntactic type and length. Our model assigns a vector and a matrix to every node in a parse tree: the vector captures the inherent meaning of the constituent, while the matrix captures how it changes the meaning of neighboring words or phrases. This matrix-vector RNN can learn the meaning of operators in propositional logic and natural language. The model obtains state of the art performance on three different experiments: predicting fine-grained sentiment distributions of adverb-adjective pairs; classifying sentiment labels of movie reviews and classifying semantic relationships such as cause-effect or topic-message between nouns using the syntactic path between them.

Polarity Inducing Latent Semantic Analysis

Wen-tau Yih, Geoffrey Zweig, and John Platt

Saturday 11:30am–12:00pm — 203 (ICC)

Existing vector space models typically map synonyms and antonyms to similar word vectors, and thus fail to represent antonymy. We introduce a new vector space representation where antonyms lie on opposite sides of a sphere: in the word vector space, synonyms have cosine similarities close to one, while antonyms are close to minus one.

We derive this representation with the aid of a thesaurus and latent semantic analysis (LSA). Each entry in the thesaurus – a word sense along with its synonyms and antonyms - is treated as

a "document," and the resulting document collection is subjected to LSA. The key contribution of this work is to show how to assign signs to the entries in the co-occurrence matrix on which LSA operates, so as to induce a subspace with the desired property.

We evaluate this procedure with the Graduate Record Examination questions of (Mohammed et al., 2008) and find that the method improves on the results of that study. Further improvements result from refining the subspace representation with discriminative training, and augmenting the training data with general newspaper text. Altogether, we improve on the best previous results by 11 points absolute in F measure.

First Order vs. Higher Order Modification in Distributional Semantics

Gemma Boleda, Eva Maria Vecchi, Miquel Cornudella, and Louise McNally

Saturday 12:00pm–12:30pm — 203 (ICC)

Adjectival modification, particularly by expressions that have been treated as higher-order modifiers in the formal semantics tradition, raises interesting challenges for semantic composition in distributional semantic models. We contrast three types of adjectival modifiers —- intersectively used color terms (as in "white towel", clearly first-order), subsectively used color terms ("white wine", which have been modeled as both first- and higher-order), and intensional adjectives ("former bassist", clearly higher-order) —- and test the ability of different composition strategies to model their behavior. In addition to opening up a new empirical domain for research on distributional semantics, our observations concerning the attested vectors for the different types of adjectives, the nouns they modify, and the resulting noun phrases yield insights into modification that have been little evident in the formal semantics literature to date.

Learning-based Multi-Sieve Co-reference Resolution with Knowledge

Lev Ratinov and Dan Roth

Saturday 11:00am–11:30am — 202 A (ICC)

We explore the interplay of knowledge and structure in co-reference resolution. To inject knowledge, we use a state-of-the-art system which cross-links (or "grounds") expressions in free text to Wikipedia. We explore ways of using the resulting grounding to boost the performance of a state-of-the-art co-reference resolution system. To maximize the utility of the injected knowledge, we deploy a learning-based multi-sieve approach and develop novel entity-based features. Our end system outperforms the state-of-the-art baseline by 2 B^3 F1 points on non-transcript portion of the ACE 2004 dataset.

Joint Learning for Coreference Resolution with Markov Logic

Yang Song, Jing Jiang, Wayne Xin Zhao, Sujian Li, and Houfeng Wang

Saturday 11:30am–12:00pm — 202 A (ICC)

Pairwise coreference resolution models must merge pairwise coreference decisions to generate final outputs. Traditional merging methods adopt different strategies such as the best-first method and enforcing the transitivity constraint, but most of these methods are used independently of the pairwise learning methods as an isolated inference procedure at the end. We propose a joint learning model which combines pairwise classification and mention clustering with Markov logic. Experimental results show that our joint learning system outperforms independent learning systems. Our system gives a better performance than all the learning-based systems from the CoNLL-2011 shared task on the same dataset. Compared with the best system from CoNLL-2011, which employs a rule-based method, our system shows competitive performance.

Resolving "This-issue" Anaphora

Varada Kolhatkar and Graeme Hirst

Saturday 12:00pm–12:30pm — 202 A (ICC)

We annotate and resolve a particular case of abstract anaphora, namely, "this-issue" anaphora. We propose a candidate ranking model for "this-issue" anaphora resolution that explores different "issue"-specific and general abstract-anaphora features. The model is not restricted to nominal

or verbal antecedents; rather, it is able to identify antecedents that are arbitrary spans of text. Our results show that (a) the model outperforms the strong adjacent-sentence baseline; (b) general abstract-anaphora features, as distinguished from "issue"-specific features, play a crucial role in "this-issue" anaphora resolution, suggesting that our approach can be generalized for other NPs such as "this problem" and "this debate"; and (c) it is possible to reduce the search space in order to improve performance.

Entity based Q&A Retrieval

Amit Singh

Saturday 11:00am–11:30am — 202 B (ICC)

Bridging the lexical gap between the user's question and the question-answer pairs in the Q&A archives has been a major challenge for Q&A retrieval. State-of-the-art approaches address this issue by implicitly expanding the queries with additional words using statistical translation models. While useful, the effectiveness of these models is highly dependant on the availability of quality corpus in the absence of which they are troubled by noise issues. Moreover these models perform word based expansion in a context agnostic manner resulting in translation that might be mixed and fairly general. This results in degraded retrieval performance. In this work we address the above issues by extending the lexical word based translation model to incorporate semantic concepts (entities). We explore strategies to learn the translation probabilities between words and the concepts using the Q&A archives and a popular entity catalog. Experiments conducted on a large scale real data show that the proposed techniques are promising.

Constructing Task-Specific Taxonomies for Document Collection Browsing

Hui Yang

Saturday 11:30am–12:00pm — 202 B (ICC)

Taxonomies can serve as browsing tools for document collections. However, given an arbitrary collection, pre-constructed taxonomies could not easily adapt to the specific topic/task present in the collection. This paper explores techniques to quickly derive task-specific taxonomies supporting browsing in arbitrary document collections. The supervised approach directly learns semantic distances from users to propose meaningful task-specific taxonomies. The approach aims to produce globally optimized taxonomy structures by incorporating path consistency control and user-generated task specification into the general learning framework. A comparison to state-of-the-art systems and a user study jointly demonstrate that our techniques are highly effective.

Besting the Quiz Master: Crowdsourcing Incremental Classification Games

Jordan Boyd-Graber, Brianna Satinoff, He He, and Hal Daume III

Saturday 12:00pm–12:30pm — 202 B (ICC)

Cost-sensitive classification, where the *features* used in machine learning tasks have a cost, has been explored as a means of balancing knowledge and the expense of obtaining new features.

In this paper, we introduce a domain where humans engage in classification with incrementally revealed features: the collegiate trivia circuit. By providing the community with a web-based system to practice, we collected a dataset of tens of thousands of implicit word-by-word ratings of how useful features are for eliciting the correct answers. By observing humans' process of classification, we are able to improve the performance of a state-of-the art classifier. In addition, we also use the dataset to evaluate a system to compete in the incremental classification task through a reduction of reinforcement learning to classification. Our system learns **when** to answer a question, performing better than batch baselines and most human players.

Multi-Domain Learning: When Do Domains Matter?

Mahesh Joshi, Mark Dredze, William W. Cohen, and Carolyn Rose

Saturday 2:00pm–2:30pm — Youngju A (ICC)

We present a systematic analysis of existing multi-domain learning approaches with respect to

two questions. First, many multi-domain learning algorithms resemble ensemble learning algorithms. (1) Are multi-domain learning improvements the result of ensemble learning effects? Second, these algorithms are traditionally evaluated in a balanced class label setting, although in practice many multi-domain settings have domain-specific class label biases. When multi-domain learning is applied to these settings, (2) are multi-domain methods improving because they capture domain-specific class biases? An understanding of these two issues presents a clearer idea about where the field has had success in multi-domain learning, and it suggests some important open questions for improving beyond the current state of the art.

Biased Representation Learning for Domain Adaptation
Fei Huang and Alexander Yates
Saturday 2:30pm–3:00pm — Youngju A (ICC)

Representation learning is a promising technique for discovering features that allow supervised classifiers to generalize from a source domain dataset to arbitrary new domains. We present a novel, formal statement of the representation learning task. We argue that because the task is computationally intractable in general, it is important for a representation learner to be able to incorporate expert knowledge during its search for helpful features. Leveraging the Posterior Regularization framework, we develop an architecture for incorporating biases into representation learning. We investigate three types of biases, and experiments on two domain adaptation tasks show that our biased learners identify significantly better sets of features than unbiased learners, resulting in a relative reduction in error of more than 16% for both tasks, with respect to existing state-of-the-art representation learning techniques.

Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars
Kewei Tu and Vasant Honavar
Saturday 3:00pm–3:30pm — Youngju A (ICC)

We introduce a novel approach named unambiguity regularization for unsupervised learning of probabilistic natural language grammars. The approach is based on the observation that natural language is remarkably unambiguous in the sense that only a tiny portion of the large number of possible parses of a natural language sentence are syntactically valid. We incorporate an inductive bias into grammar learning in favor of grammars that lead to unambiguous parses on natural language sentences. The resulting family of algorithms includes the expectation-maximization algorithm (EM) and its variant, Viterbi EM, as well as a so-called softmax-EM algorithm. The softmax-EM algorithm can be implemented with a simple and computationally efficient extension to standard EM. In our experiments of unsupervised dependency grammar learning, we show that unambiguity regularization is beneficial to learning, and in combination with annealing (of the regularization strength) and sparsity priors it leads to improvement over the current state of the art.

Extracting Opinion Expressions with semi-Markov Conditional Random Fields
Bishan Yang and Claire Cardie
Saturday 2:00pm–2:30pm — 203 (ICC)

Extracting opinion expressions from text is usually formulated as a token-level sequence labeling task tackled using Conditional Random Fields (CRFs). CRFs, however, do not readily model potentially useful segment-level information like syntactic constituent structure. Thus, we propose a semi-CRF-based approach to the task that can perform sequence labeling at the segment level. It allows the modeling of arbitrarily long opinion expressions and accounts for likely syntactic structure when modeling segment boundaries. We evaluate performance on two opinion extraction tasks, and, in contrast to previous sequence labeling approaches to the task, explore the usefulness of segment-level syntactic parse features. Experimental results demonstrate that our approach outperforms state-of-the-art methods for both opinion expression tasks.

Opinion Target Extraction Using Word-Based Translation Model
Kang Liu, Liheng Xu, and Jun Zhao
Saturday 2:30pm–3:00pm — 203 (ICC)

This paper proposes a novel approach to extract opinion targets based on word-based translation model (WTM). At first, we apply WTM in a monolingual scenario to mine the associations between opinion targets and opinion words. Then, a graph-based algorithm is exploited to extract opinion targets, where candidate opinion relevance estimated from the mined associations, is incorporated with candidate importance to generate a global measure. By using WTM, our method can capture opinion relations more precisely, especially for long-span relations. In particular, compared with previous syntax-based methods, our method can effectively avoid noises from parsing errors when dealing with informal texts in large Web corpora. By using graph-based algorithm, opinion targets are extracted in a global process, which can effectively alleviate the problem of error propagation in traditional bootstrap-based methods, such as Double Propagation. The experimental results on three real world datasets in different sizes and languages show that our approach is more effective and robust than state-of-art methods.

Word Salad: Relating Food Prices and Descriptions
Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis, and Noah A. Smith
Saturday 3:00pm–3:30pm — 203 (ICC)

We investigate the use of language in food writing, specifically on restaurant menus and in customer reviews. Our approach is to build predictive models of concrete external variables, such as restaurant menu prices. We make use of a dataset of menus and customer reviews for thousands of restaurants in several U.S. cities. By focusing on prediction tasks and doing our analysis at scale, our methodology allows quantitative, objective measurements of the words and phrases used to describe food in restaurants. We also explore interactions in language use between menu prices and sentiment as expressed in user reviews.

Learning to Map into a Universal POS Tagset
Yuan Zhang, Roi Reichart, Regina Barzilay, and Amir Globerson
Saturday 2:00pm–2:30pm — 202 A (ICC)

We present an automatic method for mapping language-specific part-of-speech tags to a set of universal tags. This unified representation plays a crucial role in cross-lingual syntactic transfer of multilingual dependency parsers. Until now, however, such conversion schemes have been created manually. Our central hypothesis is that a valid mapping yields POS annotations with coherent linguistic properties which are consistent across source and target languages. We encode this intuition in an objective function that captures a range of distributional and typological characteristics of the derived mapping. Given the exponential size of the mapping space, we propose a novel method for optimizing over soft mappings, and use entropy regularization to drive those towards hard mappings. Our results demonstrate that automatically induced mappings rival the quality of their manually designed counterparts when evaluated in the context of multilingual parsing.

Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features
Jiayi Zhao, Xipeng Qiu, Shu Zhang, Feng Ji, and Xuanjing Huang
Saturday 2:30pm–3:00pm — 202 A (ICC)

In modern Chinese articles or conversations, it is very popular to involve a few English words, especially in emails and Internet literature. Therefore, it becomes an important and challenging topic to analyze Chinese-English mixed texts. The underlying problem is how to tag part-of-speech (POS) for the English words involved. Due to the lack of specially annotated corpus, most of the English words are tagged as the oversimplified type, "foreign words". In this paper, we present a method using dynamic features to tag POS of mixed texts. Experiments show that our method achieves higher performance than traditional sequence labeling methods. Meanwhile, our method also boosts the performance of POS tagging for pure Chinese texts.

Wiki-ly Supervised Part-of-Speech Tagging
Shen Li, João Graça, and Ben Taskar
Saturday 3:00pm–3:30pm — 202 A (ICC)

Despite significant recent work, unsupervised techniques for part-of-speech (POS) tagging have not achieved useful accuracies required by many language processing tasks. Use of parallel text between resource-rich and resource-poor languages is one source of weak supervision that significantly improves accuracy. However, parallel text is not always available and techniques for using it require multiple complex algorithmic steps. In this paper we show that we can build POS-taggers exceeding state-of-the-art bilingual methods by using simple hidden Markov models and freely available and naturally growing resource, Wiktionary. Across eight languages for which we have labeled data to evaluate results, we achieve accuracy that significantly exceeds best unsupervised and parallel text methods. We achieve highest accuracy reported for several languages and show that our approach yields better out-of-domain taggers than those trained using fully supervised Penn Treebank.

Joining Forces Pays Off: Multilingual Joint Word Sense Disambiguation
Roberto Navigli and Simone Paolo Ponzetto
Saturday 2:00pm–2:30pm — 202 B (ICC)

We present a multilingual joint approach to Word Sense Disambiguation (WSD). Our method exploits BabelNet, a very large multilingual knowledge base, to perform graph-based WSD across different languages, and brings together empirical evidence from these languages using ensemble methods. The results show that, thanks to complementing wide-coverage multilingual lexical knowledge with robust graph-based algorithms and combination methods, we are able to achieve the state of the art in both monolingual and multilingual WSD settings.

A New Minimally-Supervised Framework for Domain Word Sense Disambiguation
Stefano Faralli and Roberto Navigli
Saturday 2:30pm–3:00pm — 202 B (ICC)

We present a new minimally-supervised framework for performing domain-driven Word Sense Disambiguation (WSD). Glossaries for several domains are iteratively acquired from the Web by means of a bootstrapping technique. The acquired glosses are then used as the sense inventory for fully-unsupervised domain WSD. Our experiments, on new and gold-standard datasets, show that our wide-coverage framework enables high-performance results on dozens of domains at a coarse and fine-grained level.

Grounded Models of Semantic Representation
Carina Silberer and Mirella Lapata
Saturday 3:00pm–3:30pm — 202 B (ICC)

A popular tradition of studying semantic representation has been driven by the assumption that word meaning can be learned from the linguistic environment, despite ample evidence suggesting that language is grounded in perception and action. In this paper we present a comparative study of models that represent word meaning based on linguistic and perceptual data. Linguistic information is approximated by naturally occurring corpora and sensorimotor experience by feature norms (i.e., attributes native speakers consider important in describing the meaning of a word). The models differ in terms of the mechanisms by which they integrate the two modalities. Experimental results show that a closer correspondence to human data can be obtained by uncovering latent information shared among the textual and perceptual modalities rather than arriving at semantic knowledge by concatenating the two.

Improved Parsing and POS Tagging Using Inter-Sentence Consistency Constraints
Alexander Rush, Roi Reichart, Michael Collins, and Amir Globerson
Saturday 4:00pm–4:30pm — Youngju A (ICC)

State-of-the-art statistical parsers and POS taggers perform very well when trained with large

amounts of in-domain data. When training data is out-of-domain or limited, accuracy degrades. In this paper, we aim to compensate for the lack of available training data by exploiting similarities between test set sentences. We show how to augment sentence-level models for parsing and POS tagging with inter-sentence consistency constraints. To deal with the resulting global objective, we present an efficient and exact dual decomposition decoding algorithm. In experiments, we add consistency constraints to the MST parser and the Stanford part-of-speech tagger and demonstrate significant error reduction in the domain adaptation and the lightly supervised settings across five languages.

Unified Dependency Parsing of Chinese Morphological and Syntactic Structures
Zhongguo Li and Guodong Zhou
Saturday 4:30pm–5:00pm — Youngju A (ICC)

Most previous approaches to syntactic parsing of Chinese rely on a preprocessing step of word segmentation, thereby assuming there was a clearly defined boundary between morphology and syntax in Chinese. We show how this assumption can fail badly, leading to many out-of-vocabulary words and incompatible annotations. Hence in practice the strict separation of morphology and syntax in the Chinese language proves to be untenable. We present a unified dependency parsing approach for Chinese which takes unsegmented sentences as input and outputs both morphological and syntactic structures with a single model and algorithm. By removing the intermediate word segmentation, the unified parser no longer needs separate notions for words and phrases. Evaluation proves the effectiveness of the unified model and algorithm in parsing structures of words, phrases and sentences simultaneously.

A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing
Bernd Bohnet and Joakim Nivre
Saturday 5:00pm–5:30pm — Youngju A (ICC)

Most current dependency parsers presuppose that input words have been morphologically disambiguated using a part-of-speech tagger before parsing begins. We present a transition-based system for joint part-of-speech tagging and labeled dependency parsing with non-projective trees. Experimental evaluation on Chinese, Czech, English and German shows consistent improvements in both tagging and parsing accuracy when compared to a pipeline system, which lead to improved state-of-the-art results for all languages.

Identifying Event-related Bursts via Social Media Activities
Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan, and Xiaoming Li
Saturday 4:00pm–4:30pm — 203 (ICC)

Activities on social media increase at a dramatic rate. When an external event happens, there is a surge in the degree of activities related to the event. These activities may be temporally correlated with one another, but they may also capture different aspects of an event and therefore exhibit different bursty patterns. In this paper, we propose to identify event-related bursts via social media activities. We study how to correlate multiple types of activities to derive a global bursty pattern. To model smoothness of one state sequence, we propose a novel function which can capture the state context. The experiments on a large Twitter dataset shows our methods are very effective.

User Demographics and Language in an Implicit Social Network
Katja Filippova
Saturday 4:30pm–5:00pm — 203 (ICC)

We consider the task of predicting the gender of the YouTube1 users and contrast two information sources: the comments they leave and the social environment induced from the affiliation graph of users and videos. We propagate gender information through the videos and show that a user's

gender can be predicted from her social environment with the accuracy above 90%. We also show that the gender can be predicted from language alone (89%). A surprising result of our study is that the latter predictions correlate more strongly with the gender predominant in the user's environment than with the sex of the person as reported in the profile. We also investigate how the two views (linguistic and social) can be combined and analyse how prediction accuracy changes over different age groups.

Revisiting the Predictability of Language: Response Completion in Social Media
Bo Pang and Sujith Ravi
Saturday 5:00pm–5:30pm — 203 (ICC)

The question "how predictable is English?" has long fascinated researchers. While prior work has focused on formal English typically used in news articles, we turn to texts generated by users in online settings that are more informal in nature. We are motivated by a novel application scenario: given the difficulty of typing on mobile devices, can we help reduce typing effort with message completion, especially in conversational settings? We propose a method for automatic response completion. Our approach models both the language used in responses and the specific context provided by the original message. Our experimental results on a large-scale dataset show that both components help reduce typing effort. We also perform an information-theoretic study in this setting and examine the entropy of user-generated content, especially in conversational scenarios, to better understand predictability of user generated English.

Supervised Text-based Geolocation Using Language Models on an Adaptive Grid
Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge
Saturday 4:00pm–4:30pm — 202 A (ICC)

The geographical properties of words have recently begun to be exploited for geolocating documents based solely on their text, often in the context of social media and online content. One common approach for geolocating texts is rooted in information retrieval. Given training documents labeled with latitude/longitude coordinates, a grid is overlaid on the Earth and pseudo-documents constructed by concatenating the documents within a given grid cell; then a location for a test document is chosen based on the most similar pseudo-document. Uniform grids are normally used, but they are sensitive to the dispersion of documents over the earth. We define an alternative grid construction using k-d trees that more robustly adapts to data, especially with larger training sets. We also provide a better way of choosing the locations for pseudo-documents. We evaluate these strategies on existing Wikipedia and Twitter corpora, as well as a new, larger Twitter corpus. The adaptive grid achieves competitive results with a uniform grid on small training sets and outperforms it on the large Twitter corpus. The two grid constructions can also be combined to produce consistently strong results across all training sets.

A Discriminative Model for Query Spelling Correction with Latent Structural SVM
Huizhong Duan, Yanen Li, ChengXiang Zhai, and Dan Roth
Saturday 4:30pm–5:00pm — 202 A (ICC)

Discriminative training in query spelling correction is difficult due to the complex internal structures of the data. Recent work on query spelling correction suggests a two stage approach — a noisy channel model that is used to retrieve a number of candidate corrections, followed by discriminatively trained ranker applied to these candidates. The ranker, however, suffers from the fact the low recall of the first, suboptimal, search stage.

This paper proposes to directly optimize the search stage with a discriminative model based on latent structural SVM. In this model, we treat query spelling correction as a multi-class classification problem with structured input and output. The latent structural information is used to model the alignment of words in the spelling correction process. Experiment results show that as a standalone speller, our model outperforms all the baseline systems. It also attains a higher recall compared with the noisy channel model, and can therefore serve as a better filtering stage when combined with a ranker.

Characterizing Stylistic Elements in Syntactic Structure

Song Feng, Ritwik Banerjee, and Yejin Choi

Saturday 5:00pm–5:30pm — 202 A (ICC)

Much of the writing styles recognized in rhetorical and composition theories involve deep syntactic elements. However, most previous research for computational stylometric analysis has relied on shallow lexico-syntactic patterns. Some very recent work has shown that PCFG models can detect distributional difference in syntactic styles, but without offering much insights into exactly what constitute salient stylistic elements in sentence structure characterizing each authorship. In this paper, we present a comprehensive exploration of syntactic elements in writing styles, with particular emphasis on interpretable characterization of stylistic elements. We present analytic insights with respect to the authorship attribution task in two different domains.

CoNLL Shared Task Paper Abstracts

Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution

Eraldo Fernandes, Cícero dos Santos, and Ruy Milidiú

Friday 11:00am–11:20am — 202 B (ICC)

We describe a machine learning system based on large margin structure perceptron for unrestricted coreference resolution that introduces two key modeling techniques: latent coreference trees and entropy guided feature induction. The proposed latent tree modeling turns the learning problem computationally feasible. Additionally, using an automatic feature induction method, we are able to efficiently build nonlinear models and, hence, achieve high performances with a linear learning algorithm. Our system is evaluated on the CoNLL-2012 Shared Task closed track, which comprises three languages: Arabic, Chinese and English. We apply the same system to all languages, except for minor adaptations on some language dependent features, like static lists of pronouns. Our system achieves an official score of 59.96, the best one among all the competitors.

Data-driven Multilingual Coreference Resolution using Resolver Stacking

Anders Björkelund and Richárd Farkas

Friday 11:20am–11:40am — 202 B (ICC)

This paper describes our contribution to the CoNLL 2012 Shared Task. We present a novel decoding algorithm for coreference resolution which is combined with a standard pair-wise coreference resolver in a stacking approach. The stacked decoders are evaluated on the three languages of the Shared Task. We obtain an official overall score of 59.52 which is the second highest in the Shared Task.

Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution

Chen Chen and Vincent Ng

Friday 11:40am–12:00pm — 202 B (ICC)

We describe our system for the CoNLL-2012 shared task, which seeks to model coreference in OntoNotes for English, Chinese, and Arabic. We adopt a hybrid approach to coreference resolution, which combines the strengths of rule-based methods and learningbased methods. Our official combined score over all three languages is 58.73. In particular, our score on the Chinese test set is the best among the participating teams.

Using Syntactic Dependencies to Solve Coreferences

Marcus Stamborg, Dennis Medved, Peter Exner, and Pierre Nugues

Friday 12:00pm–12:20pm — 202 B (ICC)

This paper describes the structure of the LTH coreference solver used in the closed track of the CoNLL 2012 shared task (Pradhan et al., 2012). The solver core is a mention classifier that uses Soon et al. (2001)'s algorithm and features extracted from the dependency graphs of the sentences.

This system builds on Björkelund and Nugues (2011)'s solver that we extended so that it can be applied to the three languages of the task: English, Chinese, and Arabic. We designed a new mention detection module that removes pleonastic pronouns, prunes constituents, and recovers mentions when they do not match exactly a noun phrase. We carefully redesigned the features so that they reflect more complex linguistic phenomena as well as discourse properties. Finally, we introduced a minimal cluster model grounded in the first mention of an entity.

We optimized the feature sets for the three languages: We carried out an extensive evaluation of pairs of features and we complemented the single features with associations that improved the CoNLL score. We obtained the respective scores of 59.57, 56.62, and 48.25 on English, Chinese, and Arabic on the development set, 59.36, 60.13, and 49.43 on the test set, and the combined official score of 56.3.

ICT: System Description for CoNLL-2012

Hao Xiong and Qun Liu

Friday 2:00pm–3:30pm — 202 B (ICC)

In this paper, we present our system description for the CoNLL-2012 coreference resolution task on English, Chinese and Arabic. We investigate a projection-based model in which we first translate Chinese and Arabic into English, run a publicly available coreference system, and then use a new projection algorithm to map the coreferring entities back from English into mention candidates detected in the Chinese and Arabic source. We compare to a baseline that just runs the English coreference system on the supplied parses for Chinese and Arabic. Because our method does not beat the baseline system on the development set, we submit outputs generated by the baseline system as our final submission.

A Mixed Deterministic Model for Coreference Resolution

Bo Yuan, Qingcai Chen, Yang Xiang, Xiaolong Wang, Liping Ge, Zengjian Liu, Meng Liao, and Xianbo Si

Friday 2:00pm–3:30pm — 202 B (ICC)

This paper presents a mixed deterministic model for coreference resolution in CoNLL-2012 shared task. We separated the two main stages of our model, mention detection and coreference resolution, into several sub-tasks which would be solved by machine learning method and deterministic rules based on multi-filters, such as lexical, syntactic, semantic, gender and number information. We participated in the closed track of English and Chinese, and also submitted an open result of Chinese using tools to generate the required features. Finally we reached the average F1 scores 58.68, 66.46 and 66.38 on English closed task, Chinese closed and open tasks.

Simple Maximum Entropy Models for Multilingual Coreference Resolution

Xinxin Li, Xuan Wang, and Xingwei Liao

Friday 2:00pm–3:30pm — 202 B (ICC)

This paper describes our system participating in the CoNLL-2012 shared task: "Modeling Multilingual Unrestricted Coreference in Ontonotes". Maximum entropy models are used for our system as classifiers to determine the coreference relationship between every two mentions (usually noun phrases and pronouns) in each document. We exploit rich lexical, syntactic and semantic features for the system, and the final features are selected using a greedy forward and backward strategy from an initial feature set. Our system participated in the closed track for both English and Chinese languages.

UBIU for Multilingual Coreference Resolution in OntoNotes

Desislava Zhekova, Sandra Kübler, Joshua Bonner, Marwa Ragheb, and Yu-Yin Hsu

Friday 2:00pm–3:30pm — 202 B (ICC)

The current work presents the participation of UBIU (Zhekova and Kübler, 2010) in the CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes (Pradhan et al., 2012). Our system deals with all three languages: Arabic, Chinese and English. The system results show that UBIU works reliably across all three languages, reaching an average score of 40.57 for Arabic, 46.12 for Chinese, and 48.70 for English. For Arabic and Chinese, the system produces high precision, while for English, precision and recall are balanced, which leads to the highest results across languages.

Chinese Coreference Resolution via Ordered Filtering

Xiaotian Zhang, Chunyang Wu, and Hai Zhao

Friday 2:00pm–3:30pm — 202 B (ICC)

We in this paper present the model for our participation (BCMI) in the CoNLL-2012 Shared Task. This paper describes a pure rule-based method, which assembles different filters in a proper order. Different filters handle different situations and the filtering strategies are designed manually. These filters are assigned to different ordered tiers from general to special cases. We

participated in the Chinese and English closed tracks, scored 54.21 and 59.24 respectively.

A Multigraph Model for Coreference Resolution
Sebastian Martschat, Jie Cai, Samuel Broscheit, Éva Mújdricza-Maydt, and Michael Strube
Friday 2:00pm–3:30pm — 202 B (ICC)

This paper presents HITS' coreference resolution system that participated in the CoNLL-2012 shared task on multilingual unrestricted coreference resolution. Our system employs a simple multigraph representation of the relation between mentions in a document, where the nodes correspond to mentions and the edges correspond to relations between the mentions. Entities are obtained via greedy clustering. We participated in the closed tasks for English and Chinese. Our system ranked second in the English closed task.

Incorporating Rule-based and Statistic-based Techniques for Coreference Resolution
Ruifeng Xu, Jun Xu, Jie Liu, Chengxiang Liu, Chengtian Zou, Lin Gui, Yanzhen Zheng, and Peng Qu
Friday 2:00pm–3:30pm — 202 B (ICC)

This paper describes a coreference resolution system for CONLL 2012 shared task developed by HLT_HITSZ group, which incorporates rule-based and statistic-based techniques. The system performs coreference resolution through the mention pair classification and linking. For each detected mention pairs in the text, a Decision Tree (DT) based binary classifier is applied to determine whether they form a coreference. This classifier incorporates 51 and 61 selected features for English and Chinese, respectively. Meanwhile, a rule-based classifier is applied to recognize some specific types of coreference, especially the ones with long distances. The outputs of these two classifiers are merged. Next, the recognized coreferences are linked to generate the final coreference chain. This system is evaluated on English and Chinese sides (Closed Track), respectively. It achieves 0.5861 and 0.6003 F1 score on the development data of English and Chinese, respectively. As for the test dataset, the achieved F1 scores are 0.5749 and 0.6508, respectively. This encouraging performance shows the effectiveness of our proposed coreference resolution system.

Illinois-Coref: The UI System in the CoNLL-2012 Shared Task
Kai-Wei Chang, Rajhans Samdani, Alla Rozovskaya, Mark Sammons, and Dan Roth
Friday 2:00pm–3:30pm — 202 B (ICC)

The CoNLL-2012 shared task is an extension of the last year's coreference task. We participated in the closed track of the shared tasks in both years. In this paper, we present the improvements of Illinois-Coref system from last year. We focus on improving mention detection and pronoun coreference resolution, and present a new learning protocol. These new strategies boost the performance of the system by 5development set.

System paper for CoNLL-2012 shared task: Hybrid Rule-based Algorithm for Coreference Resolution.
Heming Shou and Hai Zhao
Friday 2:00pm–3:30pm — 202 B (ICC)

My submission for the system paper, still needed to be modified and updated before the Camera Ready deadline.

Update on June 7, 2012:

1. Corrected lots of grammatical errors. 2. Supplemented experiment results on development set and corrected the result since the submission for development set got the wrong version. 3. Add citation for this shared task. 4. Seriously consider the advice of reviewers and give some detailed explanation.

Update on June 13, 2012: Finish final submission with copyright information.

BART goes multilingual: The UniTN / Essex submission to the CoNLL-2012 Shared Task

Olga Uryupina, Alessandro Moschitti, and Massimo Poesio

Friday 2:00pm–3:30pm — 202 B (ICC)

This paper describes the UniTN/Essex submission to the CoNLL-2012 Shared Task on the Multilingual Coreference Resolution. We have extended our 2011 submission, based on BART, to cover two additional languages, Arabic and Chinese. This paper focuses on adapting BART to new languages, discussing the problems we have encountered and the solutions adopted. In particular, we propose a novel entity-mention detection algorithm that might help identify nominal mentions in an unknown language. We also discuss the impact of basic linguistic information on the overall performance level of our coreference resolution system.

Learning to Model Multilingual Unrestricted Coreference in OntoNotes

Baoli Li

Friday 2:00pm–3:30pm — 202 B (ICC)

Coreference resolution, which aims at correctly linking meaningful expressions in text, is a much challenging problem in Natural Language Processing community. This paper describes the multilingual coreference modeling system of Web Information Processing Group, Henan University of Technology, China, for the CoNLL-2012 shared task (closed track). The system takes a supervised learning strategy, and consists of two cascaded components: one for detecting mentions, and the other for clustering mentions. To make the system applicable for multiple languages, generic syntactic and semantic features are used to model coreference in text. The system obtained combined official score 42.32 over three languages (Arabic, Chinese, and English) and ranked 7th among the 15 systems in the closed track.

Sponsors

12

Workshops

Overview

2-day Workshops: July 12-13, 2012

1-day Workshops: July 12, 2012

1-day Workshops: July 13, 2012

Workshop 1:

SP-Sem-MRL2012: Parsing and Semantic Processing of Morphologically Rich Languages 2012

Thursday, July 12

Venue: 401 A (ICC)

Workshop Program

Thursday, July 12, 2012

(08:50-10:05) **Session 1: Opening Session**

08:50-09:05 Statistical Parsing and Semantic Processing of MRLs: Overview of the workshop by Reut Tsarfaty

09:05-10:05 Invited Talk (I) by Ivan Titov

(10:05-10:30) **Session 2: Syntactic Parsing of MRLs (I)**

10:05–10:30 Probabilistic Lexical Generalization for French Dependency Parsing
Enrique Henestroza Anguiano and Marie Candito

10:30-11:00 **Coffee Break**

(11:00-12:25) **Session 3: Semantic Processing of MRLs**

11:00–11:25 Supervised Learning of German Qualia Relations
Yannick Versley

11:25–11:40 Building an Arabic Multiword Expressions Repository
Abdelati Hawwari, Kfir Bar, and Mona Diab

11:40–11:55 Unsupervised frame based Semantic Role Induction: application to French and English
Alejandra Lorenzo and Christophe Cerisara

11:55–12:10 Using Synthetic Compounds for Word Sense Discrimination
Yannick Versley and Verena Henrich

12:10–12:25 Machine Learning of Syntactic Attachment from Morphosyntactic and Semantic Co-occurrence Statistics
Szymon Acedański, Adam Slaski, and Adam Przepiórkowski

12:30-14:00 **Lunch Break**

(14:00-15:30) **Session 4: Syntactic Parsing of MRLs (II)**

14:00-15:00 **Invited Talk (II) by Mark Steedman**

15:00–15:15 Combining Rule-Based and Statistical Syntactic Analyzers
Iakes Goenaga, Koldobika Gojenola, María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Kepa Bengoetxea

15:15–15:30 Statistical Parsing of Spanish and Data Driven Lemmatization
Joseph Le Roux, Benoit Sagot, and Djamél' Seddah

15:30-16:00 **Coffee Break**

(16:00-17:30) **Session 5: Syntactic Parsing of MRLs (III)**

16:00–16:25 Assigning Deep Lexical Types Using Structured Classifier Features for Grammatical Dependencies
João Silva and António Branco

16:25–16:40 Using an SVM Ensemble System for Improved Tamil Dependency Parsing
Nathan Green, Loganathan Ramasamy, and Zdeněk Žabokrtský

16:40–17:05 Korean Treebank Transformation for Parser Training
DongHyun Choi, Jungyeul Park, and Key-Sun Choi

17:05–17:30 Generative Constituent Parsing and Discriminative Dependency Reranking: Experiments on English and French
Joseph Le Roux, Benoit Favre, Alexis Nasr, and Seyed Abolghasem Mirroshandel

17:30-17:40 **Short Break**

(17:40-18:20) **Session 6: Closing Session**

17:40-18:10 **Panel: Disclosing the SPMRL 2013 Shared Task**

18:10-18:20 **Concluding Remarks by Reut Tsarfaty**

Workshop 2:
SMIAE 2012: Speech and Multimodal Interaction in Assistive Environments

Thursday, July 12

Venue: 401 B (ICC)

Workshop Program

Thursday, July 12, 2012

Session 1: Session

9:00–9:10 **Welcome and Introduction**

9:10–10:00 **Keynote: Robots that can learn new words and their grounded meanings through dialogues. Mikio Nakano**

Session 1a: Session

10:00–10:30 Multimodal Human-Machine Interaction for Service Robots in Home-Care Environments
Stefan Goetze, Sven Fischer, Niko Moritz, Jens-E. Appell, and Frank Wallhoff

10:30–11:00 **Coffee Break**

Session 1b: Session

11:00–11:30 Integration of Multimodal Interaction as Assistance in Virtual Environments
Kiran Pala, Ram Naresh, Sachin Joshi, and Suryakanth V Ganagshetty

11:30–12:00 Toward a Virtual Assistant for Vulnerable Users: Designing Careful Interaction
Ramin Yaghoubzadeh and Stefan Kopp

12:00–12:30 Speech and Gesture Interaction in an Ambient Assisted Living Lab
Dimitra Anastasiou, Cui Jian, and Desislava Zhekova

12:30–14:00 **Lunch Break**

Session 1c: Session

14:00–14:30 Reduction of Non-stationary Noise for a Robotic Living Assistant using Sparse Non-negative Matrix Factorization
Benjamin Cauchi, Stefan Goetze, and Simon Doclo

14:30–15:00 Towards a Self-Learning Assistive Vocal Interface: Vocabulary and Grammar Learning
Janneke van de Loo, Jort F. Gemmeke, Guy De Pauw, Joris Driesen, Hugo Van hamme, and Walter Daelemans

15:00–15:30 A Bengali Speech Synthesizer on Android OS
Sankar Mukherjee and Shyamal Kumar Das Mandal

15:30–16:00 **Coffee Break**

16:00–17:30 **Discussion and Conclusion**

Workshop 3:
The LAW VI: The 6th Linguistic Annotation Workshop

July 12–13

Venue: 301 (ICC)

Workshop Program

Thursday, July 12, 2012

8:45–9:00 **Opening Remarks**

Invited talk

9:00–9:35 The Role of Linguistic Models and Language Annotation in Feature Selection for Machine Learning
James Pustejovsky

Special Session: The LAW Challenge

9:35–9:40 **Presentation of LAW Challenge Award**

9:40–10:05 Who Did What to Whom? A Contrastive Study of Syntacto-Semantic Dependencies
Angelina Ivanova, Stephan Oepen, Lilja Øvrelid, and Dan Flickinger

10:05–10:30 Prague Markup Language Framework
Jirka Hana and Jan Štěpánek

10:30–11:00 **Morning coffee break**

Paper Session 1

11:00–11:25 Exploiting naive vs expert discourse annotations: an experiment using lexical cohesion to predict Elaboration / Entity-Elaboration confusions
Clémentine Adam and Marianne Vergez-Couret

11:25–11:50 Pair Annotation: Adaption of Pair Programming to Corpus Annotation
Isin Demirsahin, Ihsan Yalcinkaya, and Deniz Zeyrek

11:50–12:15 Structured Named Entities in two distinct press corpora: Contemporary Broadcast News and Old Newspapers
Sophie Rosset, Cyril Grouin, Karën Fort, Olivier Galibert, Juliette Kahn, and Pierre Zweigenbaum

12:15–12:40 Intra-Chunk Dependency Annotation : Expanding Hindi Inter-Chunk Annotated Treebank
Prudhvi Kosaraju, Bharat Ram Ambati, Samar Husain, Dipti Misra Sharma, and Rajeev Sangal

12:40–14:15 **Lunch**

Paper Session 2

14:15–14:40 A Model for Linguistic Resource Description
Nancy Ide and Keith Suderman

14:40–15:05 A GrAF-compliant Indonesian Speech Recognition Web Service on the Language Grid for Transcription Crowdsourcing
Bayu Distiawan and Ruli Manurung

15:05–15:30 Towards Adaptation of Linguistic Annotations to Scholarly Annotation Formalisms on the Semantic Web
Karin Verspoor and Kevin Livingston

15:30–16:00 **Afternoon coffee break**

Paper Session 3

16:00–16:25 Intonosyntactic Data Structures: The Rhapsodie Treebank of Spoken French
Kim Gerdes, Sylvain Kahane, Anne Lacheret, Paola Pietandrea, and Arthur Truong

16:25–16:50 Annotation Schemes to Encode Domain Knowledge in Medical Narratives
Wilson McCoy, Cecilia Ovesdotter Alm, Cara Calvelli, Rui Li, Jeff B. Pelz, Pengcheng Shi, and Anne Haake

16:50–17:15 Usability Recommendations for Annotation Tools
Manuel Burghardt

17:15–17:30 **SIGANN business meeting**

Friday, July 13, 2012

(9:00-10:05am) **Poster Session**

- Search Result Diversification Methods to Assist Lexicographers
  Lars Borin, Markus Forsberg, Karin Friberg Heppin, Richard Johansson, and Annika Kjellandsson
- Simultaneous error detection at two levels of syntactic annotation
  Adam Przepiórkowski and MichałĆ Lenart
- Exploring Temporal Vagueness with Mechanical Turk
  Yuping Zhou and Nianwen Xue
- Developing Learner Corpus Annotation for Korean Particle Errors
  Sun-Hee Lee, Markus Dickinson, and Ross Israel
- Annotating Archaeological Texts: An Example of Domain-Specific Annotation in the Humanities
  Francesca Bonin, Fabio Cavulli, Aronne Noriller, Massimo Poesio, and Egon W. Stemle
- Annotating Preferences in Chats for Strategic Games
  Anais Cadilhac, Nicholas Asher, and Farah Benamara
- Morpheme Segmentation in the METU-Sabancı Turkish Treebank
  Ruket Cakici
- AlvisAE: a collaborative Web text annotation editor for knowledge acquisition
  Frédéric Papazian, Robert Bossy, and Claire Nédellec
- CSAF - a community-sourcing annotation framework
  Jin-Dong Kim and Yue Wang

Paper Session 4

10:05–10:30 Dependency Treebank of Urdu and its Evaluation
Riyaz Ahmad Bhat and Dr. Dipti Misra Sharma

10:30-11:30 **: Morning coffee break**

Paper Session 5

11:00–11:25 Annotating Coordination in the Penn Treebank
Wolfgang Maier, Sandra Kübler, Erhard Hinrichs, and Julia Kriwanek

11:25–11:50 Annotating Particle Realization and Ellipsis in Korean
Sun-Hee Lee and Jae-Young Song

11:50–12:15 Annotation of Adversarial and Collegial Social Actions in Discourse
David Bracewell, Marc Tomlinson, Mary Brunson, Jesse Plymale, Jiajun Bracewell, and Daniel Boerger

12:15–12:30 **Closing**

Workshop 4:

WASSA2012: 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis

Thursday, July 12

Venue: 302 (ICC)

Workshop Program

Thursday July 12, 2012

(8:30) **Opening Remarks**

(8:40) **Invited talk (I): Prof. Dr. Rada Mihalcea**

- Multimodal Sentiment Analysis
  Rada Mihalcea

(9:35) **Invited talk (II): Prof. Dr. Janyce Wiebe**

- Subjectivity Word Sense Disambiguation
  Janyce Wiebe

(10:30) **Break**

(11:00) **Session 1: Subjectivity and Sentiment Analysis in Social Media**

- Random Walk Weighting over SentiWordNet for Sentiment Polarity Detection on Twitter
  Arturo Montejo-Ráez, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña López
- Mining Sentiments from Tweets
  Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, and Vasudeva Varma
- SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media
  Muhammad Abdul-Mageed, Sandra Kuebler, and Mona Diab

(12:30) **Lunch Break**

(13:30) **Session 2: Affect Detection and Classification (I)**

- Opinum: statistical sentiment analysis for opinion classification
  Boyan Bonev, Gema Ramírez Sánchez, and Sergio Ortiz Rojas
- Sentimantics: Conceptual Spaces for Lexical Sentiment Polarity Representation with Contextuality
  Amitava Das and Gambäck Björn
- Analysis of Travel Review Data from Reader's Point of View
  Maya Ando and Shun Ishizaki
- Multilingual Sentiment Analysis using Machine Translation?
  Alexandra Balahur and Marco Turchi

(15:30) **Break**

(16:00) **Session 3: Affect Detection and Classification (II)**

- Unifying Local and Global Agreement and Disagreement Classification in Online Debates
  Jie Yin, Nalin Narang, Paul Thomas, and Cecile Paris
- Prior versus Contextual Emotion of a Word in a Sentence
  Diman Ghazi, Diana Inkpen, and Stan Szpakowicz
- Cross-discourse Development of Supervised Sentiment Analysis in the Clinical Domain
  Phillip Smith and Mark Lee
- POLITICAL-ADS: An annotated corpus for modeling event-level evaluativity
  Kevin Reschke and Pranav Anand

(17:30) **Session 4: Applications of Subjectivity and Sentiment Analysis**

- Automatically Annotating A Five-Billion-Word Corpus of Japanese Blogs for Affect and Sentiment Analysis
  Michal Ptaszynski, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi
- How to Evaluate Opinionated Keyphrase Extraction?
  Gábor Berend and Veronika Vincze
- Semantic frames as an anchor representation for sentiment analysis
  Josef Ruppenhofer and Ines Rehbein
- On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators
  Dasha Bogdanova, Paolo Rosso, and Thamar Solorio

Workshop 5:
ExProM: Extra-propositional aspects of meaning in computational linguistics

Friday, July 13

Venue: 402 A (ICC)

Workshop Program

Friday July 13, 2012

8:45 **Welcome**

Morning session

9:00 Disfluencies as Extra-Propositional Indicators of Cognitive Processing
Kathryn Womack, Wilson McCoy, Cecilia Ovesdotter Alm, Cara Calvelli, Jeff B. Pelz, Pengcheng Shi, and Anne Haake

9:30 How do Negation and Modality Impact on Opinions?
Farah Benamara, Baptiste Chardon, Yannick Mathieu, Vladimir Popescu, and Nicholas Asher

10:00 Linking Uncertainty in Physicians' Narratives to Diagnostic Correctness
Wilson McCoy, Cecilia Ovesdotter Alm, Cara Calvelli, Jeff B. Pelz, Pengcheng Shi, and Anne Haake

10:30 **Coffee break**

11:00 Factuality Detection on the Cheap: Inferring Factuality for Increased Precision in Detecting Negated Events
Erik Velldal and Jonathon Read

11:30 Improving Speculative Language Detection using Linguistic Knowledge
Guillermo Moncecchi, Jean-Luc Minel, and Dina Wonsever

12:00 Bridging the Gap Between Scope-based and Event-based Negation/Speculation Annotations: A Bridge Not Too Far
Pontus Stenetorp, Sampo Pyysalo, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii

12:30 **Lunch**

Afternoon session

14:00 **PASCAL2 Invited talk by Bonnie Webber: Alternatives and Extra-Propositional Meaning**

15:00 Statistical Modality Tagging from Rule-based Annotations and Crowdsourcing
Vinodkumar Prabhakaran, Michael Bloodgood, Mona Diab, Bonnie Dorr, Lori Levin, Christine D. Piatko, Owen Rambow, and Benjamin Van Durme

15:30 **Coffee break**

16:00 Annotating the Focus of Negation in terms of Questions Under Discussion
Pranav Anand and Craig Martell

16:30 Hedge Detection as a Lens on Framing in the GMO Debates: A Position Paper
Eunsol Choi, Chenhao Tan, Lillian Lee, Cristian Danescu-Niculescu-Mizil, and Jennifer Spindel

17:00 Recognizing Arguing Subjectivity and Argument Tags
Alexander Conrad, Janyce Wiebe, and Rebecca Hwa

17:30 **Discussion**

Workshop 6:
MM: Multilingual Modeling

Friday, July 13

Venue: 402 B (ICC)

Workshop Program

Friday, July 13, 2012

9:00 **Invited Talk by Reinhard Rapp**

9:40 Implementing a Language-Independent MT Methodology
Sokratis Sofianopoulos, Marina Vassiliou, and George Tambouratzis

10:05 **Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation**

Akihiro Tamura, Taro Watanabe and Eiichiro Sumita (Invited Paper)

10:30 **Coffee break**

11:00 **Invited Talk by Benjamin Snyder**

11:40 The Study of Effect of Length in Morphological Segmentation of Agglutinative Languages
Loganathan Ramasamy, Zdeněk Žabokrtský, and Sowmya Vajjala

12:05 **Unsupervised Structure Prediction with Non-Parallel Multilingual Guidance**

Shay B. Cohen, Dipanjan Das and Noah A. Smith (Invited Paper)

12:30 **Lunch break**

2:00 **Invited Talk by Slav Petrov**

2:40 Language Independent Named Entity Identification using Wikipedia
Mahathi Bhagavatula, Santosh GSK, and Vasudeva Varma

3:05 **Cross-Lingual Parse Disambiguation based on Semantic Correspondence**

Lea Frermann and Francis Bond (Invited Paper)

3:30 **Coffee break**

4:00 **Learning Discriminative Projections for Text Similarity Measures**

Wen-tau Yih, Kristina Toutanova, John Platt, and Chris Meek (Invited Paper)

4:25 **Untangling the Cross-Lingual Link Structure of Wikipedia**

Gerard de Melo and Gerhard Weikum (Invited Paper)

4:50 A Comparable Corpus Based on Aligned Multilingual Ontologies
Roger Granada, Lucelene Lopes, Carlos Ramisch, Cassia Trojahn, Renata Vieira, and Aline Villavicencio

5:15 **Discussion**

Workshop 7:
The People's Web meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP

Friday, July 13

Venue: 401 A (ICC)

Workshop Program

Friday July 13, 2012

09:15–09:30 **Opening remarks**

09:35–10:05 Sentiment Analysis Using a Novel Human Computation Game
Claudiu Cristian Musat, Alireza Ghasemi, and Boi Faltings

10:10–10:30 A Serious Game for Building a Portuguese Lexical-Semantic Network
Mathieu Mangeot and Carlos Ramisch

10:30–11:00 **Coffee break**

11:00–11:20 Collaboratively Building Language Resources while Localising the Web
Asanka Wasala, Reinhard Schäler, Ruvan Weerasinghe, and Chris Exton

11:25–12:30 **Invited talk: Phrase Detectives: The First Three Years by Massimo Poesio**

12:30–14:00 **Lunch break**

14:00–14:30 Resolving Task Specification and Path Inconsistency in Taxonomy Construction
Hui Yang

14:35–14:55 EAGER: Extending Automatically Gazetteers for Entity Recognition
Omer Farukhan Gunes, Tim Furche, Christian Schallhart, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo

15:00–15:30 Extracting Context-Rich Entailment Rules from Wikipedia Revision History
Elena Cabrio, Bernardo Magnini, and Angelina Ivanova

15:30–16:00 **Coffee break**

16:00–17:00 **Panel discussion: Collaboratively Looking Ahead: How to Make Sustainable Goods out of Collaboratively Constructed Semantic Resources?**

Workshop 8:
TextGraphs-7: Graph-based Methods for Natural Language Processing

Friday, July 13

Venue: 401 B (ICC)

Workshop Program

Friday July 13, 2012

8:45 **Opening Remarks**

9:00 **Session 1**

9:00 **Invited Talk by Rada Mihalcea and Dragomir Radev**

10:05 A New Parametric Estimation Method for Graph-based Clustering
Javid Ebrahimi and Mohammad Saniee Abadeh

10:30 **Coffee Break**

11:00 **Session 2**

11:00 Extracting Signed Social Networks from Text
Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev

11:25 Using Link Analysis to Discover Interesting Messages Spread Across Twitter
Min-Chul Yang, Jung-Tae Lee, and Hae-Chang Rim

11:50 Graph Based Similarity Measures for Synonym Extraction from Parsed Text
Einat Minkov and William Cohen

12:15 **Lunch Break**

14:00 Semantic Relatedness for Biomedical Word Sense Disambiguation
Kiem-Hieu Nguyen and Cheol-Young Ock

14:25 Identifying Untyped Relation Mentions in a Corpus given an Ontology
Gabor Melli

14:50 Cause-Effect Relation Learning
Zornitsa Kozareva

15:30 **Coffee Break**

16:00 **Session 3**

16:00 Bringing the Associative Ability to Social Tag Recommendation
Miao Fan, Yingnan Xiao, and Qiang Zhou

16:25 **Closing Session**

Workshop 9:
SSST-6: Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation

Thursday, July 12

Venue: 402 A (ICC)

Workshop Program

Thursday, July 12, 2012

8:45–9:00 **Opening remarks**

Session 1: Source language modeling

9:00–9:30 WSD for n-best reranking and local language modeling in SMT
Marianna Apidianaki, Guillaume Wisniewski, Artem Sokolov, Aurélien Max, and François Yvon

9:30–10:00 Linguistically-Enriched Models for Bulgarian-to-English Machine Translation
Rui Wang, Petya Osenova, and Kiril Simov

10:00–10:30 Enriching Parallel Corpora for Statistical Machine Translation with Semantic Negation Rephrasing
Dominikus Wetzel and Francis Bond

10:30–11:00 **Coffee break**

Session 2: MT output evaluation and processing

11:00–11:30 Towards a Predicate-Argument Evaluation for MT
Ondrej Bojar and Dekai Wu

11:30–12:00 Using Parallel Features in Parsing of Machine-Translated Sentences for Correction of Grammatical Errors
Rudolf Rosa, Ondřej Dušek, David Mareček, and Martin Popel

12:00–12:30 Unsupervised vs. supervised weight estimation for semantic MT evaluation metrics
Chi-kiu Lo and Dekai Wu

12:30–2:00 **Lunch**

Session 3: Semantic dependencies

2:00–2:30 Head Finalization Reordering for Chinese-to-Japanese Machine Translation
Dan Han, Katsuhito Sudoh, Xianchao Wu, Kevin Duh, Hajime Tsukada, and Masaaki Nagata

2:30–3:00 Extracting Semantic Transfer Rules from Parallel Corpora with SMT Phrase Aligners
Petter Haugereid and Francis Bond

3:00–3:30 Towards Probabilistic Acceptors and Transducers for Feature Structures
Daniel Quernheim and Kevin Knight

Session 4: Poster session at Coffee break

3:30–4:00 Using Domain-specific and Collaborative Resources for Term Translation
Mihael Arcan, Christian Federmann, and Paul Buitelaar

3:30–4:00 Improving Statistical Machine Translation through co-joining parts of verbal constructs in English-Hindi translation
Karunesh Kumar Arora and R. Mahesh K. Sinha

3:30–4:00 Application of Clause Alignment for Statistical Machine Translation
Svetla Koeva, Svetlozara Leseva, Ivelina Stoyanova, Rositsa Dekova, Angel Genov, Borislav Rizov, Tsvetana Dimitrova, Ekaterina Tarpomanova, and Hristina Kukova

3:30–4:00 Zero Pronoun Resolution can Improve the Quality of J-E Translation
Hirotoshi Taira, Katsuhito Sudoh, and Masaaki Nagata

4:00–5:00 **Panel**

Workshop 10:
DSSD: Detecting Structure in Scholarly Discourse

Thursday, July 12

Venue: 402 B (ICC)

Workshop Program

Thursday, July 12, 2012

Session 1: Exploiting Discourse Structure

9:00–9:45 Identifying Comparative Claim Sentences in Full-Text Scientific Articles
Dae Hoon Park and Catherine Blake

9:45–10:30 Identifying Claimed Knowledge Updates in Biomedical Research Articles
Ágnes Sándor and Anita de Waard

10:30–11:00 **Coffee Break**

Session 2: Detecting Discourse Elements

11:00–11:45 Detection of Implicit Citations for Sentiment Detection
Awais Athar and Simone Teufel

11:45–12:30 Open-domain Anatomical Entity Mention Detection
Tomoko Ohta, Sampo Pyysalo, Jun'ichi Tsujii, and Sophia Ananiadou

12:30–2:00 **Lunch**

Session 3: Taxonomies and Annotation

2:00–2:45 A three-way perspective on scientific discourse annotation for knowledge extraction
Maria Liakata, Paul Thompson, Anita de Waard, Raheel Nawaz, Henk Pander Maat, and Sophia Ananiadou

2:45–3:30 Epistemic Modality and Knowledge Attribution in Scientific Discourse: A Taxonomy of Types and Overview of Features
Anita de Waard and Henk Pander Maat

3:30–4:00 **Coffee Break**

4:00–5:00 **Discussion between the authors on detecting and using discourse structure for scholarly text**

5:00–5:30 **Wrapup and close**

Workshop 11:
NEWS2012: 4th Named Entities Workshop

Thursday, July 12

Venue: 400 (ICC)

Workshop Program

Thursday, July 12, 2012

9:00–9:10 **Opening Remarks by Min Zhang, Haizhou Li, A Kumaran and Ming Liu**

- Whitepaper of NEWS 2012 Shared Task on Machine Transliteration
  Min Zhang, Haizhou Li, A Kumaran, and Ming Liu
- Report of NEWS 2012 Machine Transliteration Shared Task
  Min Zhang, Haizhou Li, A Kumaran, and Ming Liu

9:10–10:30 **Session 1: Research Papers**

09:10–09:35 Accurate Unsupervised Joint Named-Entity Extraction from Unaligned Parallel Text
Robert Munro and Christopher D. Manning

09:35–10:00 Automatically generated NE tagged corpora for English and Hungarian
Eszter Simon and Dávid Márk Nemeskey

10:00–10:25 Latent Semantic Transliteration using Dirichlet Mixture
Masato Hagiwara and Satoshi Sekine

10:30–11:00 **Morning Break**

11:00–12:30 **Session 2: System Papers 1**

11:00–11:25 Rescoring a Phrase-based Machine Transliteration System with Recurrent Neural Network Language Models
Andrew Finch, Paul Dixon, and Eiichiro Sumita

11:25–11:50 Syllable-based Machine Transliteration with Extra Phrase Features
Chunyue Zhang, Tingting Li, and Tiejun Zhao

11:50–12:15 English-Korean Named Entity Transliteration Using Substring Alignment and Re-ranking Methods
Chun-Kai Wu, Yu-Chun Wang, and Richard Tzong-Han Tsai

12:15–13:45 **Lunch Break**

13:45–15:30 **Session 3: System Papers 2**

13:45–14:10 Applying mpaligner to Machine Transliteration with Japanese-Specific Heuristics
Yoh Okuno

14:10–14:35 Transliteration by Sequence Labeling with Lattice Encodings and Reranking
Waleed Ammar, Chris Dyer, and Noah Smith

14:35–15:00 Transliteration Experiments on Chinese and Arabic
Grzegorz Kondrak, Xingkai Li, and Mohammad Salameh

15:00–15:25 Cost-benefit Analysis of Two-Stage Conditional Random Fields based English-to-Chinese Machine Transliteration
Chan-Hung Kuo, Shih-Hung Liu, Mike Tian-Jian Jiang, Cheng-Wei Lee, and Wen-Lian Hsu

15:25–15:30 **Closing**

15:30–16:00 **Afternoon Break**

Index

We wish to thank our Sponsors

PLATINUM SPONSOR

GOLD SPONSORS

SILVER SPONSORS

Microsoft® Research

Daumsoft

MINING MINDS

BRONZE SPONSORS

Appen ButlerHill

Rakuten

ETS

Listening. Learning. Leading.®

SimSimi

LG Electronics

WISEnut

searchwise, search wiser

diquest

YesKiSTi

SUPPORTER

BAOBAB

mixi

Preferred Infrastructure.

STUDENT PAPER AWARD SPONSOR

IBM Research