ELEVENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION

With the Support of the Japanese Ministry of Education, Culture, Sports, Science and Technology

MAY 7 – 12, 2018

CONFERENCE PROGRAMME

For more information about ELRA:
ELRA-ELDA
9, rue des Cordelières
75013 Paris – FRANCE
Tel.: +33 1 43 13 33 33
Fax: + 33 1 43 13 33 30
info@elda.org

The LREC 2018 Proceedings are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
## LREC 2018 Committees

### Conference Programme Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicoletta Calzolari</td>
<td>ILC/CNR, Pisa, Italy (Conference chair)</td>
</tr>
<tr>
<td>Khalid Choukri</td>
<td>ELRA, Paris, France</td>
</tr>
<tr>
<td>Christopher Cieri</td>
<td>LDC, Philadelphia, USA</td>
</tr>
<tr>
<td>Thierry Declerck</td>
<td>DFKI GmbH, Saarbrücken, Germany</td>
</tr>
<tr>
<td>Koiti Hasida</td>
<td>The University of Tokyo, Tokyo, Japan</td>
</tr>
<tr>
<td>Hitoshi Isahara</td>
<td>Toyohashi University of Technology, Toyohashi, Japan</td>
</tr>
<tr>
<td>Bente Maegaard</td>
<td>CST, University of Copenhagen, Denmark</td>
</tr>
<tr>
<td>Joseph Mariani</td>
<td>LIMSI-CNRS, Orsay, France</td>
</tr>
<tr>
<td>Jan Odijk</td>
<td>UIL-OTS, Utrecht, The Netherlands</td>
</tr>
<tr>
<td>Asuncion Moreno</td>
<td>Universitat Politècnica de Catalunya, Barcelona, Spain</td>
</tr>
<tr>
<td>Stelios Piperidis</td>
<td>ILSP, Athens, Greece</td>
</tr>
<tr>
<td>Takenobu Tokunaga</td>
<td>Tokyo Institute of Technology, Tokyo, Japan</td>
</tr>
</tbody>
</table>

### Advisory Board

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shyam S. Agrawal</td>
<td>KIIT, Gurgaon (India)</td>
</tr>
<tr>
<td>Hiroya Fujisaki</td>
<td>University of Tokyo (Japan)</td>
</tr>
<tr>
<td>Eva Hajičová</td>
<td>UFAL, Charles University, Prague (Czech Republic)</td>
</tr>
<tr>
<td>Yuming Li</td>
<td>Beijing Language and Culture University (PRC)</td>
</tr>
<tr>
<td>Mark Liberman</td>
<td>Linguistic Data Consortium, Philadelphia (USA)</td>
</tr>
<tr>
<td>Makoto Nagao (Chair)</td>
<td>Professor Emeritus University of Kyoto (Japan)</td>
</tr>
<tr>
<td>Jun’ichi Tsujii</td>
<td>Artificial Intelligence Research Center, Tokyo (Japan)</td>
</tr>
</tbody>
</table>

### Local Liaison Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key-Sun Choi</td>
<td>KAIST (Korea)</td>
</tr>
<tr>
<td>Chu-Ren Huang</td>
<td>The Hong Kong Polytechnic University (Hong Kong SAR - PRC)</td>
</tr>
<tr>
<td>Toru Ishida</td>
<td>Department of Social Informatics, Kyoto University (Japan)</td>
</tr>
<tr>
<td>Haizhou Li</td>
<td>National University of Singapore (Singapore)</td>
</tr>
<tr>
<td>Satoshi Nakamura</td>
<td>Nara Institute of Science and Technology (Japan)</td>
</tr>
<tr>
<td>Byong-Rae Ryu</td>
<td>Chungnam National University (Korea)</td>
</tr>
</tbody>
</table>
Virach Sornlertlamvanich  
*Sirindhorn International Institute of Technology, Thammasat University (Thailand)*

Le Sun  
*Chinese Academy of Sciences, Beijing (PRC)*

Kam-Fai Wong  
*The Chinese University of Hong Kong (Hong Kong SAR - PRC)*

Chengqing Zong  
*Chinese Academy of Sciences, Beijing (PRC)*

**Scientific Committee**

The Programme Committee is very grateful to Scientific Committee members who reviewed the submissions and contributed to designing the conference programme. The list of the members of Scientific Committee is published on the LREC 2018 web site.

**Local Committee**

*Hitoshi Isahara (Chair)*  
*Toyohashi University of Technology, Toyohashi, Japan*

*Kyoko Kanzaki*  
*Toyohashi University of Technology, Toyohashi, Japan*

**Conference Editorial Committee**

*Sara Goggi*  
*ILC/CNR, Pisa, Italy*

*Hélène Mazo*  
*ELDA/ELRA, Paris, France*

**Organising Committee**

*Roberto Bartolini*  
*ILC/CNR, Pisa, Italy*

*Damien Bihel*  
*ELDA/ELRA, Paris, France*

*Irene De Felice*  
*University of Pisa, Italy*

*Riccardo Del Gratta*  
*ILC/CNR, Pisa, Italy*

*Sara Goggi*  
*ILC/CNR, Pisa, Italy (Co-chair)*

*Valérie Mapelli*  
*ELDA/ELRA, Paris, France*

*Hélène Mazo*  
*ELDA/ELRA, Paris, France (Co-chair)*

*Monica Monachini*  
*ILC/CNR, Pisa, Italy*

*Vincenzo Parrinelli*  
*ILC/CNR, Pisa, Italy*

*Vladimír Popescu*  
*ELDA/ELRA, Paris, France*
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valeria Quochi</td>
<td>ILC/CNR, Pisa, Italy</td>
</tr>
<tr>
<td>Caroline Rannaud</td>
<td>ELDA/ELRA, Paris, France</td>
</tr>
<tr>
<td>Alexandre Sicard</td>
<td>ELDA/ELRA, Paris, France</td>
</tr>
<tr>
<td><strong>Sponsorship Committee</strong></td>
<td></td>
</tr>
<tr>
<td>Nicoletta Calzolari</td>
<td>ILC/CNR, Pisa, Italy</td>
</tr>
<tr>
<td>Khalid Choukri</td>
<td>ELRA, Paris, France</td>
</tr>
<tr>
<td>Tatjana Gornostaja</td>
<td>Tilde, Riga, Latvia</td>
</tr>
<tr>
<td>Hitoshi Isahara</td>
<td>Toyohashi University of Technology, Toyohashi, Japan</td>
</tr>
<tr>
<td>Kyoko Kanzaki</td>
<td>Toyohashi University of Technology, Toyohashi, Japan</td>
</tr>
<tr>
<td>Jimmy Kunzmann</td>
<td>EML GmbH, Heidelberg, Germany</td>
</tr>
<tr>
<td>Joseph Mariani</td>
<td>LIMSI-CNRS &amp; IMMI, Orsay, France</td>
</tr>
<tr>
<td>Satoshi Sekine</td>
<td>New York University, New York City, USA</td>
</tr>
</tbody>
</table>
Acknowledgements

The European Language Resources Association, ELRA, and the LREC Committees acknowledge with gratitude the support and sponsoring of the following institutions.

Sponsors and supporters

- Google *(Platinum)*
- AIP, Center for Advanced Intelligent Project *(Bronze)*
- Amazon AWS *(Bronze)*
- Arcadia Computing Innovation *(Bronze)*
- EML European Media Laboratory *(Bronze)*
- Yahoo Research Japan *(Bronze)*
- GSK Language Resources Association *(Supporter)*
- Hituzi Syobo *(Publisher)*
- Multilingual *(Media Sponsor)*

Supporting Institutions

- Evaluations and Language resources Distribution Agency (ELDA), Paris (France)
- Istituto di Linguistica Computazionale (ILC) of the Italian National Research Council (CNR), Pisa (Italy)
- Toyohashi University of Technology, Japan
- Miyazaki Convention and Visitors Bureau
PROGRAMME AT A GLANCE
# Pre-conference Workshops and Tutorials

## Monday, May 7

<table>
<thead>
<tr>
<th>Tengyoku</th>
<th>Tenyo</th>
<th>Tenju</th>
<th>Tenran</th>
<th>Amber</th>
<th>Ivory</th>
<th>Marble</th>
<th>Crystal</th>
<th>Orchard N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oriental COCOSDA Conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W29</td>
<td>W2</td>
<td>W17</td>
<td>W24</td>
<td>W12</td>
<td>W7</td>
<td>T4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13th Workshop on Asian Language Resources (AsianLang 2018)</td>
<td>ParlaCLARIN</td>
<td>6th Workshop on Challenges in the Management of Large Corpora (CMUC-6)</td>
<td>7th International Workshop on Mining Scientific Publications (WOSP 2018)</td>
<td>Games &amp; Gamification for Natural Language Processing (Games4NLP)</td>
<td>Annotation Recognition &amp; Evaluation of Actions (AREA)</td>
<td>Distributional compositional semantics in the age of word embeddings: tasks, resources &amp; methodology</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Afternoon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oriental COCOSDA Conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W29</td>
<td>W2</td>
<td>W27</td>
<td>W24</td>
<td>W14</td>
<td>W20</td>
<td>T3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Tuesday, May 8

<table>
<thead>
<tr>
<th>Tengyoku</th>
<th>Tenyo</th>
<th>Tenju</th>
<th>Tenran</th>
<th>Amber</th>
<th>Ivory</th>
<th>Marble</th>
<th>Crystal</th>
<th>Orchard N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oriental COCOSDA Conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W34</td>
<td>W8</td>
<td>W33</td>
<td>W6</td>
<td>W30</td>
<td>W9</td>
<td>T5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Afternoon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oriental COCOSDA Conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W34</td>
<td>W8</td>
<td>W33</td>
<td>W3</td>
<td>W21</td>
<td>W30</td>
<td>W31</td>
<td>T8</td>
<td></td>
</tr>
<tr>
<td>Post-conference Workshops and Tutorials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturday, May 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tengyoku</th>
<th>Tenyo</th>
<th>Tenju</th>
<th>Tenran</th>
<th>Amber</th>
<th>Ivory</th>
<th>Marble</th>
<th>Crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>W1</strong> 8th Workshop on the Representation &amp; Processing of Sign Languages: Involving the Language Community (SignLang2018)</td>
<td><strong>W26</strong> Collaboration &amp; Computing for Under-Resourced Languages Sustaining knowledge diversity in the digital age (CCURL 2018)</td>
<td><strong>W19</strong> MLP-MomenT 2018</td>
<td><strong>W23</strong> 6th Workshop on Linked Data in Linguistics: Towards Linguistic Data Science (LDL-2018)</td>
<td><strong>W13</strong> Natural Language Meets Journalism III (NLP4 Journalism)</td>
<td><strong>W22</strong> 1st Workshop on Language Resources &amp; Technologies for the Legal Knowledge Graph</td>
<td><strong>W32</strong> 2nd Workshop on Text Analytics for Cybersecurity &amp; Online Safety (TA-COS 2018)</td>
</tr>
<tr>
<td>Afternoon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Conference Programme at a Glance - Morning

### 9 May 2018

<table>
<thead>
<tr>
<th>09:00-09:40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keynote Speech</strong></td>
</tr>
<tr>
<td>Charles Yang: How Children Overcome the Sparsity Problem</td>
</tr>
</tbody>
</table>

### 9:30-10:45

- **Opening Ceremony**  
  Tenzui

### 10:50-11:10

- **Introductory Session**  
  Gail Kent, European Commission  
  Language Resources for a Multilingual Europe  
  Tenzui

### 11:35-13:15

- **4 Oral and 7 Poster Parallel Sessions**  
  - Machine Translation & Evaluation  
    Tenran  
  - Semantics & Lexicon (1)  
    Tengyoku  
  - Corpus Annotation & Tagging  
    Tenju  
  - Dialogue  
    Tenyo

### 11:45-13:05

- **Posters in Poster Area 1**  
  - Bibliometrics, Scientometrics, Infometrics  
  - Discourse Annotation, Representation and Processing (1)  
  - Evaluation Methodologies  
  - Information Extraction, Information Retrieval, Text Analytics (2)  
  - Multimodality  
  - Parsing, Syntax, Treebank (1)

### 13:10-13:30

- **Japanese Invited Talk**  
  Yukinori Takubo, How many languages are there in Japan?  
  Tenzui

### 10 May 2018

<table>
<thead>
<tr>
<th>09:00-09:40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keynote Speech</strong></td>
</tr>
<tr>
<td>Pascale Fung: Empathetic Dialog Systems</td>
</tr>
</tbody>
</table>

### 9:45-11:25

- **Industry Track – Industrial Systems**  
  Tenzui
- **4 Oral and 6 Poster Parallel Sessions**  
  - Paraphrase & Semantics  
    Tenran  
  - Emotion & Sentiment (2)  
    Tengyoku  
  - Semantics & Lexicon (2)  
    Tenju  
  - Bilingual Speech Corpora & Code-Switching  
    Tenyo

### 10:50-11:10

- **Posters in Poster Area 1**  
  - Dialects  
  - Document Classification, Text Categorisation (2)  
  - Information Extraction, Information Retrieval, Text Analytics (3)  
  - Machine Translation, SpeechToSpeech Translation (2)  
  - Morphology (2)  
  - Multilinguality  
  - Part-of-Speech Tagging

### 11:45-13:05

- **4 Oral and 6 Poster Parallel Sessions**  
  - Evaluation Methodologies  
    Tenran  
  - Semantics  
    Tengyoku  
  - Information Extraction & Neural Networks  
    Tenju  
  - Dialogue, Emotion, Multimodality  
    Tenyo

### 11:45-13:25

- **Posters in Poster Area 1**  
  - Anaphora, Coreference  
  - Collaborative Resource Construction & Crowdsourcing  
  - Information Extraction, Information Retrieval, Text Analytics (1)  
  - Infrastructural Issues/Large Projects (1)  
  - Knowledge Discovery/Representation  
  - Opinion Mining / Sentiment Analysis (1)  
  - Social Media Processing (1)

### 11 May 2018

<table>
<thead>
<tr>
<th>09:00-09:40</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keynote Speech</strong></td>
</tr>
</tbody>
</table>

### 11:45-13:05

- **4 Oral and 7 Poster Parallel Sessions**  
  - Lexicon  
    Tenran  
  - Knowledge Discovery  
    Tengyoku  
  - Multilingual Corpora & Machine Translation  
    Tenju  
  - Corpus Creation, Use & Evaluation (2)  
    Tenyo

### 11:45-13:25

- **Posters in Poster Area 1**  
  - Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (3)  
  - Discourse Annotation, Representation and Processing (2)  
  - Language Acquisition & CALL (2)  
  - Opinion Mining / Sentiment Analysis (3)  
  - Sign Language  
  - Speech Resource/Database (2)

---
## Conference Programme at a Glance - Afternoon

<table>
<thead>
<tr>
<th>9 May 2018</th>
<th>10 May 2018</th>
<th>11 May 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>14:35-16:15</strong></td>
<td><strong>14:50-16:30</strong></td>
<td><strong>14:45-16:05</strong></td>
</tr>
</tbody>
</table>
| 4 Oral and 7 Poster Parallel Sessions  
  - Language Resource Policies & Management **Tenran**  
  - Emotion & Sentiment (1) **Tengyoku**  
  - Knowledge Discovery & Evaluation (1) **Tenju**  
    - Corpus Creation, Use & Evaluation (1) **Tenyo** | 4 Oral and 7 Poster Parallel Sessions  
  - Discourse & Argumentation **Tenran**  
  - Less-Resourced & Ancient Languages **Tengyoku**  
  - Semantics & Evaluation **Tenju**  
  - Multimodal & Written Corpora **Tenyo** | 4 Oral & 5 Poster Parallel Sessions  
  - Anaphora & Coreference **Tenran**  
  - Corpus for Document Classification **Tengyoku**  
  - Knowledge Discovery & Evaluation (2) **Tenju**  
  - Multimodal & Written Corpora & Tools **Tenyo** |
| **Posters in Poster Area 2**  
  - Character Recognition and Annotation  
  - Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (1)  
  - Digital Humanities  
  - Lexicon (1)  
  - Machine Translation, SpeechToSpeech Translation (1)  
  - Semantics (1)  
  - Word Sense Disambiguation | **Posters in Poster Area 2**  
  - Document Classification, Text Categorisation (1)  
  - Morphology  
  - Opinion Mining/Sentiment Analysis (2)  
  - Phonetic Databases, Phonology  
  - Question Answering and Machine Reading  
    - Social Media Processing (2)  
    - Speech Resource/Database (1) | **Posters in Poster Area 2**  
  - Corpus Creation, Annotation, Use (2)  
  - Lexicon (3)  
  - Named Entity Recognition  
  - Parsing, Syntax, Treebank (2)  
  - Wordnets and Ontologies |
| **16:35-17:55**            | **16:50-18:00**             | **16:25-17:25**             |
| 4 Oral & 5 Poster Parallel Sessions  
  - Bio-medical Corpora **Tenran**  
  - MultiWord Expressions **Tengyoku**  
  - Time & Space **Tenju**  
  - Computer Assisted Language Learning **Tenyo** | 4 Oral and 7 Poster Parallel Sessions  
  - Social Media & Evaluation **Tenran**  
  - Standards, Validation, Workflows **Tengyoku**  
  - Treebanks & Parsing **Tenju**  
  - Morphology & Lexicons **Tenyo** | Antonio Zampolli Prize Talk **Tenzui** |
| **Posters in Poster Area 1**  
  - Annotation Methods and Tools  
  - Corpus Creation, Annotation, Use (1)  
  - Emotion Recognition/Generation  
  - Ethics and Legal Issues  
  - LR Infrastructures and Architectures | **Posters in Poster Area 1**  
  - Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (2)  
  - Language Modelling  
  - Natural Language Generation  
    - Semantics (2)  
    - Speech Processing  
    - Summarisation  
    - Textual Entailment and Paraphrasing |  |
| **18:00-18:45**            | **18:00-18:45**             | **17:25-18:15**             |
| ELRA Individual Members Assembly **Tenzui** | ELRA Individual Members Assembly **Tenzui** | ELRA Individual Members Assembly **Tenzui** |
| **19:30: Welcome Reception** | **19:30: Welcome Reception** | **19:30: Gala Dinner** |
Detailed Conference Programme

9 May 2018

9.30-10.45  
Opening Ceremony

10.50-11.10  
Introductory Session: Gail Kent (European Commission), Language Resources for a Multilingual Europe

11.15-11.35  
Coffee Break

01 ● Machine Translation & Evaluation
Chair: Bente Maegaard

Room: Tenran

11.35-11.55  
Augmenting Librispeech with French Translations: A Multimodal Corpus for Direct Speech Translation Evaluation
Ali Can Kocabiyikoglu, Laurent Besacier, Olivier Kraif

11.55-12.15  
Evaluating Domain Adaptation for Machine Translation Across Scenarios
Thierry Etchegoyhen, Anna Fernández Torné, Andoni Azpeitia, Eva Martínez García, Anna Matamala

12.15-12.35  
Upping the Ante: Towards a Better Benchmark for Chinese-to-English Machine Translation
Christian Hadiwinoto, Hwee Tou Ng

12.35-12.55  
ESCAPE: a Large-scale Synthetic Corpus for Automatic Post-Editing
Matteo Negri, Marco Turchi, Rajen Chatterjee, Nicola Bertoldi

12.55-13.15  
Evaluating Machine Translation Performance on Chinese Idioms with a Blacklist Method
Yutong Shao, Rico Sennrich, Bonnie Webber, Federico Fancellu

02 ● Semantics & Lexicon (1)
Chair: Yoshihiko Hayashi

Room: Tengyoku

11.35-11.55  
Network Features Based Co-hyponymy Detection
Abhik Jana, Pawan Goyal

11.55-12.15  
Cross-Lingual Generation and Evaluation of a Wide-Coverage Lexical Semantic Resource
Attila Novák, Borbála Novák

12.15-12.35  
Advances in Pre-Training Distributed Word Representations
Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, Armand Joulin
12.35-12.55 Integrating Generative Lexicon Event Structures into VerbNet
Susan Windisch Brown, James Pustejovsky, Annie Zaenen, Martha Palmer

12.55-13.15 FontLex: A Typographical Lexicon based on Affective Associations
Tugba Kulahcioglu, Gerard De Melo

O3 ● Corpus Annotation & Tagging
Chair: Tomaž Erjavec

Room: Tenju

11.35-11.55 Multi-layer Annotation of the Rigveda
Oliver Hellwig, Heinrich Hettrich, Ashutosh Modi, Manfred Pinkal

11.55-12.15 The Natural Stories Corpus
Richard Futrell, Edward Gibson, Harry J. Tily, Idan Blank, Anastasia Vishnevetsy, Steven Piantadosi, Evelina Fedorenko

12.15-12.35 Semi-automatic Korean FrameNet Annotation over KAIST Treebank
Younggyun Hahm, Jiseong Kim, Sunggoo Kwon, Key-Sun Choi

12.35-12.55 Handling Normalization Issues for Part-of-Speech Tagging of Online Conversational Text
Géraldine Damnati, Jérémy Auguste, Alexis Nasr, Delphine Charlet, Johannes Heinecke, Frédéric Béchet

12.55-13.15 Multi-Dialect Arabic POS Tagging: A CRF Approach
Kareem Darwish, Hamdy Mubarak, Ahmed Abdelali, Mohamed Eldesouki, Younes Samih, Randah Alharbi, Mohammed Attia, Walid Magdy, Laura Kallmeyer

O4 ● Dialogue
Chair: Anna Rumshinsky

Room: Tenyo

11.35-11.55 A Corpus for Modeling Word Importance in Spoken Dialogue Transcripts
Sushant Kafle, Matt Huenerfauth

11.55-12.15 Dialogue Structure Annotation for Multi-Floor Interaction
David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervits, Kimberly Pollard, Claire Bonial, Su Lei, Clare Voss, Matthew Marge, Cory Hayes, Susan Hill

12.15-12.35 Effects of Gender Stereotypes on Trust and Likability in Spoken Human-Robot Interaction
Matthias Kraus, Johannes Kraus, Martin Baumann, Wolfgang Minker

12.35-12.55 A Multimodal Corpus for Mutual Gaze and Joint Attention in Multiparty Situated Interaction
Dimosthenis Kontogiorgos, Vanya Avramova, Simon Alexandersson, Patrik Jonell, Catharine Oertel, Jonas Beskow, Gabriel Skantze, Joakim Gustafson

12.55-13.15 Improving Dialogue Act Classification for Spontaneous Arabic Speech and Instant Messages at Utterance Level
AbdelRahim Elmadany, Sherif Abdou, Mervat Gheith
**Poster Sessions – Poster Area 1**

**P1 ● Anaphora, Coreference**  
Chair: Scott Piao

Coreference Resolution in FreeLing 4.0  
*Montserrat Marimon, Lluís Padró, Jordi Turmo*

BASHI: A Corpus of Wall Street Journal Articles Annotated with Bridging Links  
*Ina Roesiger*

SACR: A Drag-and-Drop Based Tool for Coreference Annotation  
*Bruno Oberle*

Deep Neural Networks for Coreference Resolution for Polish  
*Bartłomiej Nitoń, Paweł Morawiecki, Maciej Ogrodniczuk*

SzegedKoref: A Hungarian Coreference Corpus  
*Veronika Vincze, Klára Hegedűs, Alex Sliz-Nagy, Richárd Farkas*

A Corpus to Learn Refer-to-as Relations for Nominals  
*Wasi Ahmad, Kai-Wei Chang*

Sanaphor++: Combining Deep Neural Networks with Semantics for Coreference Resolution  
*Julien Plu, Roman Prokofyev, Alberto Tonon, Philippe Cudré-Mauroux, Djellel Eddine Difallah, Raphael Troncy, Giuseppe Rizzo*

ANCOR-AS: Enriching the ANCOR Corpus with Syntactic Annotations  
*Loïc Grobol, Isabelle Tellier, Eric De La Clergerie, Marco Dinarelli, Frédéric Landragin*

ParCorFull: a Parallel Corpus Annotated with Full Coreference  
*Ekaterina Lapshinova-Koltunski, Christian Hardmeier, Pauline Krielke*

**P2 ● Collaborative Resource Construction & Crowdsourcing**  
Chair: Asad Sayeed

An Application for Building a Polish Telephone Speech Corpus  
*Bartosz Ziółko, Piotr Żelasko, Ireneusz Gawlik, Tomasz Pędzimąż, Tomasz Jadczyk*

CPJD Corpus: Crowdsourced Parallel Speech Corpus of Japanese Dialects  
*Shinnosuke Takamichi, Hiroshi Saruwatari*

Korean L2 Vocabulary Prediction: Can a Large Annotated Corpus be Used to Train Better Models for Predicting Unknown Words?  
*Kevin Yancey, Yves Lepage*

Crowdsourcing-based Annotation of the Accounting Registers of the Italian Comedy  
*Adeline Granet, Benjamin Hervy, Geoffrey Roman-Jimenez, Marouane Hachicha, Emmanuel Morin, Harold Mouchère, Solen Quiniou, Guillaume Raschia, Françoise Rubellin, Christian Viard-Gaudin*

FEIDEGGER: A Multi-modal Corpus of Fashion Images and Descriptions in German  
*Leonidas Lefakis, Alan Akbik, Roland Vollgraf*
Toward a Lightweight Solution for Less-resourced Languages: Creating a POS Tagger for Alsatian Using Voluntary Crowdsourcing
Alice Millour, Karën Fort

Crowdsourced Corpus of Sentence Simplification with Core Vocabulary
Akihiro Katsuta, Kazuhide Yamamoto

A Multilingual Wikified Data Set of Educational Material
Iris Hendrickx, Eirini Takoulidou, Thanasis Naskos, Katia Lida Kermanidis, Vilemmini Sosoni, Hugo De Vos, Maria Stasimioti, Menno Van Zaanen, Panayota Georgakopoulou, Valia Kordoni, Maja Popovic, Markus Egg, Antal Van den Bosch

Using Crowd Agreement for Wordnet Localization
Amarsanaa Ganbold, Altangerel Chagnaa, Gábor Bella

Translation Crowdsourcing: Creating a Multilingual Corpus of Online Educational Content
Vilemmini Sosoni, Katia Lida Kermanidis, Maria Stasimioti, Thanasis Naskos, Eirini Takoulidou, Menno Van Zaanen, Sheila Castilho, Panayota Georgakopoulou, Valia Kordoni, Markus Egg

Building an English Vocabulary Knowledge Dataset of Japanese English-as-a-Second-Language Learners Using Crowdsourcing
Yo Ehara

P3 • Information Extraction, Information Retrieval, Text Analytics (1)
Chair: Hikaru Yokono

Chinese Relation Classification using Long Short Term Memory Networks
Linrui Zhang, Dan Moldovan

The UIR Uncertainty Corpus for Chinese: Annotating Chinese Microblog Corpus for Uncertainty Identification from Social Media
Binyang Li, Jun Xiang, Le Chen, Xu Han, Xiaoyan Yu, Ruifeng Xu, Tengjiao Wang, Kam-Fai Wong

EventWiki: A Knowledge Base of Major Events
Tao Ge, Lei Cui, Baobao Chang, Zhifang Sui, Furu Wei, Ming Zhou

Annotating Spin in Biomedical Scientific Publications: the case of Random Controlled Trials (RCTs)
Anna Koroleva, Patrick Paroubek

Visualization of the occurrence trend of infectious diseases using Twitter
Ryusei Matsumoto, Minoru Yoshida, Kazuyuki Matsumoto, Hironobu Matsuda, Kenji Kita

Reusable workflows for gender prediction
Matej Martinc, Senja Pollak

Knowing the Author by the Company His Words Keep
Armin Hoenen, Niko Schenk
Towards a Gold Standard Corpus for Variable Detection and Linking in Social Science Publications

Andrea Zielinski, Peter Mutschke

KRAUTS: A German Temporally Annotated News Corpus

Jannik Strötgen, Anne-Lyse Minard, Lukas Lange, Manuela Speranza, Bernardo Magnini

P4 ● Infrastructural Issues/Large Projects (1)

Chair: Denise Di Persio

CogCompNLP: Your Swiss Army Knife for NLP

Daniel Khashabi, Mark Sammons, Ben Zhou, Tom Redman, Christos Christodoulopoulos, Vivek Srikumar, Nickolas Rizzolo, Lev Ratinov, Guanheng Luo, Quang Do, Chen-Tse Tsai, Subhro Roy, Stephen Mayhew, Zhili Feng, John Wieting, Xiaodong Yu, Yangqiu Song, Shashank Gupta, Shyam Upadhyay, Naveen Arivazhagan, Qiang Ning, Shaoshi Ling, Dan Roth

A Framework for the Needs of Different Types of Users in Multilingual Semantic Enrichment

Jan Nehring, Felix Sasaki

The LREC Workshops Map

Roberto Bartolini, Sara Goggi, Monica Monachini, Gabriella Pardelli

Preserving Workflow Reproducibility: The RePlay-DH Client as a Tool for Process Documentation

Markus Gärtner, Uli Hahn, Sibylle Hermann

The ACoLi CoNLL Libraries: Beyond Tab-Separated Values

Christian Chiarcos, Niko Schenk

What's Wrong, Python? -- A Visual Differ and Graph Library for NLP in Python

Balázs Indig, András Simonyi, Noémi Ligeti-Nagy

P5 ● Knowledge Discovery/Representation

Chair: Dan Tufiș

ScholarGraph: a Chinese Knowledge Graph of Chinese Scholars

Shuo Wang, Zehui Hao, Xiaofeng Meng, Qiuyue Wang

Enriching Frame Representations with Distributionally Induced Senses

Stefano Faralli, Alexander Panchenko, Chris Biemann, Simone Paolo Ponzetto

An Integrated Formal Representation for Terminological and Lexical Data included in Classification Schemes

Thierry Declerck, Kseniya Egorova, Eileen Schnur

One event, many representations. Mapping action concepts through visual features.

Alessandro Panunzi, Lorenzo Gregori, Andrea Amelio Ravelli

Tel(s)-Telle(s)-Signs: Highly Accurate Automatic Crosslingual Hypernym Discovery

Ada Wan
P6 ● Opinion Mining / Sentiment Analysis (1)
Chair: Cristina Bosco

Disambiguation of Verbal Shifters
*Michael Wiegand, Sylvette Loda, Josef Ruppenhofer*

Bootstrapping Polar-Opposite Emotion Dimensions from Online Reviews
*Luwen Huangfu, Mihai Surdeanu*

*Pavithra Rajendran, Danushka Bollegala, Simon Parsons*

Resource Creation Towards Automated Sentiment Analysis in Telugu (a low resource language) and Integrating Multiple Domain Sources to Enhance Sentiment Prediction
*Rama Rohit Reddy Gangula, Radhika Mamidi*

Multilingual Multi-class Sentiment Classification Using Convolutional Neural Networks
*Mohammed Attia, Younes Samih, Ali Elkahky, Laura Kallmeyer*

A Large Self-Annotated Corpus for Sarcasm
*Mikhail Khodak, Nikunj Saunshi, Kiran Vodrahalli*

HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments
*Akari Asai, Sara Evensen, Behzad Golshan, Alon Halevy, Vivian Li, Andrei Lopatenko, Daniela Stepanov, Yoshihiko Suhara, Wang-Chiew Tan, Yinzhan Xu*

MultiBooked: A Corpus of Basque and Catalan Hotel Reviews Annotated for Aspect-level Sentiment Classification
*Jeremy Barnes, Toni Badia, Patrik Lambert*

P7 ● Social Media Processing (1)
Chair: Paul Cook

BlogSet-BR: A Brazilian Portuguese Blog Corpus
*Henrique Santos, Vinicius Woloszyn, Renata Vieira*

SoMeWeTa: A Part-of-Speech Tagger for German Social Media and Web Texts
*Thomas Proisl*

Collecting Code-Switched Data from Social Media
*Gideon Mendels, Victor Soto, Aaron Jaech, Julia Hirschberg*

Classifying the Informative Behaviour of Emoji in Microblogs
*Giulia Donato, Patrizia Paggio*

A Taxonomy for In-depth Evaluation of Normalization for User Generated Content
*Rob Van der Goot, Rik Van Noord, Gertjan Van Noord*

Gaining and Losing Influence in Online Conversation
*Arun Sharma, Tomek Strzalkowski*

Arap-Tweet: A Large Multi-Dialect Twitter Corpus for Gender, Age and Language Variety Identification
*Wajdi Zaghouani, Anis Charfi*

13.15-14.35 Lunch Break
O5: Language Resource Policies & Management
Chair: Stelios Piperidis

Room: Tenran

14.35-14.55  Data Management Plan (DMP) for Language Data under the New General Data Protection Regulation (GDPR)
            Pawel Kamocki, Khalid Choukri, Valérie Mapelli

14.55-15.15  We Are Depleting Our Research Subject as We Are Investigating It: In Language Technology, more Replication and Diversity Are Needed
            António Branco

15.15-15.35  Lessons Learned: On the Challenges of Migrating a Research Data Repository from a Research Institution to a University Library.
            Thorsten Trippel, Claus Zinn

15.35-15.55  Introducing NIEUW: Novel Incentives and Workflows for Eliciting Linguistic Data
            Christopher Cieri, James Fiumara, Mark Liberman, Chris Callison-Burch, Jonathan Wright

15.55-16.15  Three Dimensions of Reproducibility in Natural Language Processing
            K. Bretonnel Cohen, Jingbo Xia, Pierre Zweigenbaum, Tiffany Callahan, Orin Hargraves, Foster Goss, Nancy Ide, Aurélie Névélol, Cyril Grouin, Lawrence E. Hunter

O6 ● Emotion & Sentiment (1)
Chair: Pushpak Bhattacharyya

Room: Tengyoku

14.35-14.55  Content-Based Conflict of Interest Detection on Wikipedia
            Udochukwu Orizu, Yulan He

14.55-15.15  Word Affect Intensities
            Saif Mohammad

15.15-15.35  Representation Mapping: A Novel Approach to Generate High-Quality Multi-Lingual Emotion Lexicons
            Sven Buechel, Udo Hahn

15.35-15.55  Unfolding the External Behavior and Inner Affective State of Teammates through Ensemble Learning: Experimental Evidence from a Dyadic Team Corpus
            Aggeliki Vlachostergiou, Mark Dennison, Catherine Neubauer, Stefan Scherer, Peter Khooshabeh, Andre Harrison

15.55-16.15  Understanding Emotions: A Dataset of Tweets to Study Interactions between Affect Categories
            Saif Mohammad, Svetlana Kiritchenko
O7 ● Knowledge Discovery & Evaluation (1)
Chair: Andrejs Vasiljevs

Room: Tenju

14.35-14.55 When ACE met KBP: End-to-End Evaluation of Knowledge Base Population with Component-level Annotation
Bonan Min, Marjorie Freedman, Roger Bock, Ralph Weischedel

14.55-15.15 Simple Large-scale Relation Extraction from Unstructured Text
Christos Christodouloupolos, Arpit Mittal

15.15-15.35 Joint Learning of Sense and Word Embeddings
Mohammed Alsuitoibani, Danushka Bollegala

15.35-15.55 Comparing Pretrained Multilingual Word Embeddings on an Ontology Alignment Task
Dagmar Gromann, Thierry Declerck

15.55-16.15 A Large Resource of Patterns for Verbal Paraphrases
Octavian Popescu, Ngoc Phuoc An Vo, Vadim Sheinin

O8 ● Corpus Creation, Use & Evaluation (1)
Chair: Patrizia Paggio

Room: Tenyo

14.35-14.55 Building Parallel Monolingual Gan Chinese Dialects Corpus
Fan Xu, Mingwen Wang, Maoxi Li

14.55-15.15 A Recorded Debating Dataset
Shachar Mirkin, Michal Jacovi, Tamar Lavee, Hong-Kwang Kuo, Samuel Thomas, Leslie Sager, Lili Kotlerman, Elad Venezian, Noam Slonim

15.15-15.35 Building a Corpus from Handwritten Picture Postcards: Transcription, Annotation and Part-of-Speech Tagging
Kyoko Sugisaki, Nicolas Wiedmer, Heiko Hausendorf

15.35-15.55 A Lexical Tool for Academic Writing in Spanish based on Expert and Novice Corpora
Marcos Garcia Salido, Marcos Garcia, Milka Villayandre-Llamazares, Margarita Alonso-Ramos

15.55-16.15 Framing Named Entity Linking Error Types
Adrian Brasoveanu, Giuseppe Rizzo, Philipp Kuntschick, Albert Weichselbraun, Lyndon J.B. Nixon

S-O1 ● Special Speech Session: Speech Resources Collection in Real-World Situations
Chair: Yuichi Ishimoto, Kikuo Maekawa

Room: Tenzui

14.35-14.55 Spontaneous Speech Resources in Japan
Yuichi Ishimoto, Tomoko Ohsuga

14.55-15.15 Challenges on Authentic Emotional Speech Corpus from Spontaneous Japanese Dialog
Yoshiko Arimoto
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speakers/Authors</th>
</tr>
</thead>
</table>

**Poster Sessions – Poster Area 2**

**P8 ● Character Recognition and Annotation**

Chair: Jordi Turmo

Transc&Anno: A Graphical Tool for the Transcription and On-the-Fly Annotation of Handwritten Documents  
Nadezda Okinina, Lionel Nicolas, Verena Lyding

Correction of OCR Word Segmentation Errors in Articles from the ACL Collection through Neural Machine Translation Methods  
Vivi Nastase, Julian Hitschler

From Manuscripts to Archetypes through Iterative Clustering  
Armin Hoenen

Building A Handwritten Cuneiform Character Imageset  
Kenji Yamauchi, Hajime Yamamoto, Wakaha Mori

PDF-to-Text Reanalysis for Linguistic Data Mining  
Michael Wayne Goodman, Ryan Georgi, Fei Xia

**P9 ● Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (1)**

Chair: Leo Wanner

Crowdsourced Multimodal Corpora Collection Tool  
Patrik Jonell, Catharine Oertel, Dimosthenis Kontogiorgos, Jonas Beskow, Joakim Gustafson

Expert Evaluation of a Spoken Dialogue System in a Clinical Operating Room  
Juliana Miehle, Nadine Gerstenlauer, Daniel Ostler, Hubertus Feußner, Wolfgang Minker, Stefan Ultes

JAIST Annotated Corpus of Free Conversation  
Kiyoaki Shirai, Tomotaka Fukuoka

The Metalogue Debate Trainee Corpus: Data Collection and Annotations  
Volha Petukhova, Andrei Malchanau, Youssef Oualil, Dietrich Klakow, Saturnino Luz, Fasih Haider, Nick Campbell, Dimitris Koryzis, Dimitris Spiliotopoulos, Pierre Albert, Nicklas Linz, Jan Alexandersson
Towards Continuous Dialogue Corpus Creation: writing to corpus and generating from it
*Andrei Malchanau, Volha Petukhova, Harry Bunt*

MYCanCor: A Video Corpus of spoken Malaysian Cantonese
*Andreas Liesenfeld*

KTH Tangrams: A Dataset for Research on Alignment and Conceptual Pacts in Task-Oriented Dialogue
*Todd Shore, Theofronia Androulakaki, Gabriel Skantze*

On the Vector Representation of Utterances in Dialogue Context
*Louisa Pragst, Niklas Rach, Wolfgang Minker, Stefan Ultes*

ES-Port: a Spontaneous Spoken Human-Human Technical Support Corpus for Dialogue Research in Spanish
*Laura García-Sardiña, Manex Serras, Arantza Del Pozo*

From analysis to modeling of engagement as sequences of multimodal behaviors
*Soumia Dermouche, Catherine Pelachaud*

**P10 ● Digital Humanities**

Chair: Giorgio Maria Di Nunzio

A corpus of German political speeches from the 21st century
*Adrien Barbaresi*

Building Literary Corpora for Computational Literary Analysis - A Prototype to Bridge the Gap between CL and DH
*Andrew Frank, Christine Ivanovic*

Towards faithfully visualizing global linguistic diversity
*Garland McNew, Curdin Derungs, Steven Moran*

The GermaParl Corpus of Parliamentary Protocols
*Andreas Blätte, Andre Blessing*

Identifying Speakers and Addressees in Dialogues Extracted from Literary Fiction
*Adam Ek, Mats Wirén, Robert Östling, Kristina Nilsson Björkenstam, Gintare Grigonyte, Sofia Gustafson Capková*

**P11 ● Lexicon (1)**

Chair: Francesca Frontini

Word Embedding Evaluation Datasets and Wikipedia Title Embedding for Chinese
*Chi-Yen Chen, Wei-Yun Ma*

An Automatic Learning of an Algerian Dialect Lexicon by using Multilingual Word Embeddings
*ABIDI Karima, Kamel Smaili*

Candidate Ranking for Maintenance of an Online Dictionary
*Claire Broad, Helen Langone, David Guy Brizan*

Language adaptation experiments via cross-lingual embeddings for related languages
*Serge Sharoff*
Tools for Building an Interlinked Synonym Lexicon Network
Zdenka Uresova, Eva Fucikova, Eva Hajicova, Jan Hajic

Very Large-Scale Lexical Resources to Enhance Chinese and Japanese Machine Translation
Jack Halpern

Combining Concepts and Their Translations from Structured Dictionaries of Uralic Minority Languages
Mika Hämäläinen, Liisa Lotta Tarvainen, Jack Rueter

Transfer of Frames from English FrameNet to Construct Chinese FrameNet: A Bilingual Corpus-Based Approach
Tsung-Han Yang, Hen-Hsen Huang, An-Zi Yen, Hsin-Hsi Chen

EFLlex: A Graded Lexical Resource for Learners of English as a Foreign Language
Luise Dürlich, Thomas Francois

P12 ● Machine Translation, SpeechToSpeech Translation (1)
Chair: Laurent Besacier

English-Basque Statistical and Neural Machine Translation
Inigo Jauregi Unanue, Lierni Garmendia Arratibel, Ehsan Zare Borzeshi, Massimo Piccardi

TQ-AutoTest – An Automated Test Suite for (Machine) Translation Quality
Vivien Macketanz, Renlong Ai, Aljoscha Burchardt, Hans Uszkoreit

Exploiting Pre-Ordering for Neural Machine Translation
Yang Zhao, Jiajun Zhang, Chengqing Zong

Improving a Multi-Source Neural Machine Translation Model with Corpus Extension for Low-Resource Languages
Gyu Hyeon Choi, Jong Hun Shin, Young Kil Kim

Dynamic Oracle for Neural Machine Translation in Decoding Phase
Zi-Yi Dou, Hao Zhou, Shu-Jian Huang, Xin-Yu Dai, Jia-Jun Chen

One Sentence One Model for Neural Machine Translation
Xiaoqing Li, Jiajun Zhang, Chengqing Zong

A Parallel Corpus of Arabic-Japanese News Articles
Go Inoue, Nizar Habash, Yuji Matsumoto, Hiroyuki Aoyama

Examining the Tip of the Iceberg: A Data Set for Idiom Translation
Marzieh Fadaee, Arianna Bisazza, Christof Monz

Automatic Enrichment of Terminological Resources: the IATE RDF Example
Mihael Arcan, Elena Montiel-Ponsoda, John Philip McCrae, Paul Buitelaar

A Comparative Study of Extremely Low-Resource Transliteration of the World’s Languages
Winston Wu, David Yarowsky

Translating Web Search Queries into Natural Language Questions
Adarsh Kumar, Sandipan Dandapat, Sushil Chordia
P13 ● Semantics (1)
Chair: Kyoko Kanzaki

Construction of a Japanese Word Similarity Dataset
Yuuya Sakaizawa, Mamoru Komachi

Acquiring Verb Classes Through Bottom-Up Semantic Verb Clustering
Olga Majewska, Diana McCarthy, Ivan Vulić, Anna Korhonen

Constructing High Quality Sense-specific Corpus and Word Embedding via Unsupervised Elimination of Pseudo Multi-sense
Haoyue Shi, Xihao Wang, Yuqi Sun, Junfeng Hu

Urdu Word Embeddings
Samar Haider

Social Image Tags as a Source of Word Embeddings: A Task-oriented Evaluation
Mika Hasegawa, Tetsunori Kobayashi, Yoshihiko Hayashi

Towards AMR-BR: A SemBank for Brazilian Portuguese Language
Rafael Anchiêta, Thiago Pardo

Towards a Welsh Semantic Annotation System
Scott Piao, Paul Rayson, Dawn Knight, Gareth Watkins

Semantic Frame Parsing for Information Extraction: the CALOR corpus
Gabriel Marzinotto, Jeremy Auguste, Frederic Bechet, Géraldine Damnati, Alexis Nasr

Using a Corpus of English and Chinese Political Speeches for Metaphor Analysis
Kathleen Ahrens, Huiheng Zeng, Shun-han Rebekah Wong

A Multi-versus a Single-classifier Approach for the Identification of Modality in the Portuguese Language
João Sequeira, Teresa Gonçalves, Paulo Quaresma, Amália Mendes, Iris Hendrickx

P14 ● Word Sense Disambiguation
Chair: Maite Melero

All-words Word Sense Disambiguation Using Concept Embeddings
Rui Suzuki, Kanako Komiya, Masayuki Asahara, Minoru Sasaki, Hiroyuki Shinnou

Enhancing Modern Supervised Word Sense Disambiguation Models by Semantic Lexical Resources
Stefano Melacci, Achille Globo, Leonardo Rigutini

An Unsupervised Word Sense Disambiguation System for Under-Resourced Languages
Dmitry Ustalov, Denis Teslenko, Alexander Panchenko, Mikhail Chersnoskutov, Chris Biemann, Simone Paolo Ponzetto

Unsupervised Korean Word Sense Disambiguation using CoreNet
Kijong Han, Sangha Nam, Jiseong Kim, Younggyun Hahm, Key-Sun Choi

UFSAC: Unification of Sense Annotated Corpora and Tools
Loïc Vial, Benjamin Lecouteux, Didier Schwab
Retrofitting Word Representations for Unsupervised Sense Aware Word Similarities
*Steffen Remus, Chris Biemann*

FastSense: An Efficient Word Sense Disambiguation Classifier
*Tolga Uslu, Alexander Mehler, Daniel Baumartz, Alexander Henlein, Wahed Hemati*

**16.15-16.35**  
**Coffee Break**

**O9 ● Bio-medical Corpora**
Chair: Paul Rayson
**Room: Tenran**

**16.35-16.55**  
A FrameNet for Cancer Information in Clinical Narratives: Schema and Annotation
*Kirk Roberts, Yuqi Si, Anshul Gandhi, Elmer Bernstam*

**16.55-17.15**  
A New Corpus to Support Text Mining for the Curation of Metabolites in the ChEBI Database
*Matthew Shardlow, Nhung Nguyen, Gareth Owen, Claire O’Donovan, Andrew Leach, John McNaught, Steve Turner, Sophia Ananiadou*

**17.15-17.35**  
Parallel Corpora for the Biomedical Domain
*Aurélie Névéol, Antonio Jimeno Yepes, Mariana Neves, Karin Verspoor*

**17.35-17.55**  
Medical Entity Corpus with PICO elements and Sentiment Analysis
*Markus Zlabinger, Linda Andersson, Allan Hanbury, Michael Andersson, Vanessa Quasnik, Jon Brassey*

**O10 ● MultiWord Expressions**
Chair: Francis Bond
**Room: Tengyoku**

**16.35-16.55**  
Word Embedding Approach for Synonym Extraction of Multi-Word Terms
*Amir Hazem, Béatrice Daille*

**16.55-17.15**  
A Large Automatically-Acquired All-Words List of Multiword Expressions Scored for Compositionality
*Will Roberts, Markus Egg*

**17.15-17.35**  
A Hybrid Approach for Automatic Extraction of Bilingual Multiword Expressions from Parallel Corpora
*Nasredine Semmar*

**17.35-17.55**  
No more beating about the bush: A Step towards Idiom Handling for Indian Language NLP
*Ruchit Agrawal, Vighnesh Chenthil Kumar, Vigneshwaran Muralidaran, Dipti Sharma*

**O11 ● Time & Space**
Chair: Kyoko Ohara
**Room: Tenju**

**16.35-16.55**  
Sentence Level Temporality Detection using an Implicit Time-sensed Resource
*Sabyasachi Kamila, Asif Ekbal, Pushpak Bhattacharyya*
16.55-17.15  Comprehensive Annotation of Various Types of Temporal Information on the Time Axis  
Tomohiro Sakaguchi, Daisuke Kawahara, Sadao Kurohashi

17.15-17.35  Systems’ Agreements and Disagreements in Temporal Processing: An Extensive Error Analysis of the TempEval-3 Task  
Tommaso Caselli, Roser Morante

17.35-17.55  Annotating Temporally-Anchored Spatial Knowledge by Leveraging Syntactic Dependencies  
Alakananda Vempala, Eduardo Blanco

**O12 ● Computer Assisted Language Learning**  
Chair: Zygmunt Vetulani  
Room: Tenyo

16.35-16.55  Contextualized Usage-Based Material Selection  
Dirk De Hertog, Piet Desmet

16.55-17.15  CBFC: a parallel L2 speech corpus for Korean and French learners  
Hyon Yoo, Inyoung Kim

17.15-17.35  SW4ALL: a CEFR Classified and Aligned Corpus for Language Learning  
Rodrigo Wilkens, Leonardo Zilio, Cédrick Fairon

17.35-17.55  Towards a Diagnosis of Textual Difficulties for Children with Dyslexia  
Solen Quiniou, Béatrice Daille

**S-O2 ● Special Speech Session - Speech Resources Collection in Real-World Situations**  
Chair: Yuichi Ishimoto, Kikuo Maekawa  
Room: Tenzui

16.35-16.55  Miraikan SC Corpus: A Trial of Data Collection in Semi-Opened and Semi-Controlled Environment  
Mayumi Bono, Rui Sakaida, Ryosaku Makino, Ayami Joh

16.55-17.15  A Multimodal Multiparty Human-Robot Dialogue Corpus for Real World Interaction  
Kotaro Funakoshi

17.15-17.35  Speech and Language Resources for the Development of Dialogue Systems and Problems Arising from Their Deployment  
Ryuichiro Higashinaka, Ryo Ishii, Narimune Matsumura, Tadashi Nunobiki, Atsushi Itoh, Ryuichi Inagawa, Junji Tomita

17.35-17.55  General discussion – Moderator: Kikuo Maekawa

16:35 - 17:55  
**Poster Sessions – Poster Area 1**

**P15 ● Annotation Methods and Tools**  
Chair: Ron Artstein

Text Annotation Graphs: Annotating Complex Natural Language Phenomena  
Angus Forbes, Kristine Lee, Gus Hahn-Powell, Marco A. Valenzuela-Escarcega, Mihai Surdeanu
Manzanilla: An Image Annotation Tool for TKB Building
Arianne Reimerink, Pilar León-Araúz

Tools for The Production of Analogical Grids and a Resource of N-gram Analogical Grids in 11 Languages
Rashel Fam, Yves Lepage

The Automatic Annotation of the Semiotic Type of Hand Gestures in Obama’s Humorous Speeches
Costanza Navarretta

WASA: A Web Application for Sequence Annotation
Fahad AlGhamdi, Mona Diab

Annotation and Quantitative Analysis of Speaker Information in Novel Conversation Sentences in Japanese
Makoto Yamazaki, Yumi Miyazaki, Wakako Kashino

PDFAnno: a Web-based Linguistic Annotation Tool for PDF Documents
Hiroyuki Shindo, Yohei Munesada, Yuji Matsumoto

A Lightweight Modeling Middleware for Corpus Processing
Markus Gärtner, Jonas Kuhn

An Annotation Language for Semantic Search of Legal Sources
Adeline Nazarenko, Francois Levy, Adam Wyner

Resource Interoperability for Sustainable Benchmarking: The Case of Events
Chantal Van Son, Oana Inel, Roser Morante, Lora Aroyo, Piek Vossen

Parsivar: A Language Processing Toolkit for Persian
Salar Mohtaj, Behnam Roshanfekr, Atefeh Zafarian, Habibollah Asghari

Multilingual Word Segmentation: Training Many Language-Specific Tokenizers Smoothly Thanks to the Universal Dependencies Corpus
Erwan Moreau, Carl Vogel

Build Fast and Accurate Lemmatization for Arabic
Hamdy Mubarak

P16 ● Corpus Creation, Annotation, Use (1)
Chair: Prokopis Prokopidis

JESC: Japanese-English Subtitle Corpus
Reid Pryzant, Youngjoo Chung, Dan Jurafsky, Denny Britz

Building a Corpus for Personality-dependent Natural Language Understanding and Generation
Ricelli Ramos, Georges Neto, Barbara Silva, Danielle Monteiro, Ivandré Paraboni, Rafael Dias

Linguistic and Sociolinguistic Annotation of 17th Century Dutch Letters
Marijn Schraagen, Feike Dietz, Marjo Van Koppen

Simplified Corpus with Core Vocabulary
Takumi Maruyama, Kazuhide Yamamoto

A Pragmatic Approach for Classical Chinese Word Segmentation
Shilei Huang, Jiangqin Wu
ASAP++: Enriching the ASAP Automated Essay Grading Dataset with Essay Attribute Scores  
*Sandeep Mathias, Pushpak Bhattacharyya*

MirasText: An Automatically Generated Text Corpus for Persian  
*Behnam Sabeti, Hossein Abedi Firouzaee, Ali Janalizadeh Choobbasti, Seyed Hani Elamahdi Mortazavi Najafabadi, Amir Vaheb*

The Reference Corpus of the Contemporary Romanian Language (CoRoLa)  
*Verginica Barbu Mititelu, Dan Tufiș, Elena Irimia*

A Corpus of Drug Usage Guidelines Annotated with Type of Advice  
*Sarah Masud Preum, Md. Rizwan Parvez, Kai-Wei Chang, John Stankovic*

BioRo: The Biomedical Corpus for the Romanian Language  
*Maria Mitrofan, Dan Tufiș*

**P17 ● Emotion Recognition/Generation**  
Chair: Lluís Padró

A Comparison Of Emotion Annotation Schemes And A New Annotated Data Set  
*Ian Wood, John Philip McCrae, Vladimir Andryushechkin, Paul Buitelaar*

Humor Detection in English-Hindi Code-Mixed Social Media Content: Corpus and Baseline System  
*Ankush Khandelwal, Sahil Swami, Syed Sarfaraz Akhtar, Manish Shrivastava*

Dialogue Scenario Collection of Persuasive Dialogue with Emotional Expressions via Crowdsourcing  
*Koichiro Yoshino, Yoko Ishikawa, Masahiro Mizukami, Yu Suzuki, Sakriani Sakti, Satoshi Nakamura*

SentiArabic: A Sentiment Analyzer for Standard Arabic  
*Ramy Eskander*

Contextual Dependencies in Time-Continuous Multidimensional Affect Recognition  
*Dmitrii Fedotov, Denis Ivanko, Maxim Sidorov, Wolfgang Minker*

WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art  
*Saif Mohammad, Svetlana Kiritchenko*

Arabic Data Science Toolkit: An API for Arabic Language Feature Extraction  
*Paul Rodrigues, Valerie Novak, C. Anton Rytting, Julie Yelle, Jennifer Bouitza*

Sentence and Clause Level Emotion Annotation, Detection, and Classification in a Multi-Genre Corpus  
*Shabnam Tafreshi, Mona Diab*

**P18 ● Ethics and Legal Issues**  
Chair: Karën Fort

A Swedish Cookie-Theft Corpus  
*Dimitrios Kokkinakis, Kristina Lundholm Fors, Kathleen Fraser, Arto Nordlund*

Sharing Copies of Synthetic Clinical Corpora without Physical Distribution — A Case Study to Get Around IPRs and Privacy Constraints Featuring the German JSYNCC Corpus  
*Christina Lohr, Sven Buechel, Udo Hahn*
A Legal Perspective on Training Models for Natural Language Processing
Richard Eckart de Castilho, Giulia Dore, Thomas Margoni, Penny Labropoulou, Iryna Gurevych

P19 ● LR Infrastructures and Architectures
Chair: Dieter van Uytvanck
LREMap, a Song of Resources and Evaluation
Riccardo Del Gratta, Sara Goggi, Gabriella Pardelli, Nicoletta Calzolari
Metadata Collection Records for Language Resources
Henk Van den Heuvel, Erwin Komen, Nelleke Oostdijk
Managing Public Sector Data for Multilingual Applications Development
Stelios Piperidis, Penny Labropoulou, Miltos Deligiannis, Maria Giagkou
Bridging the LAPPs Grid and CLARIN
Erhard Hinrichs, Nancy Ide, James Pustejovsky, Jan Hajic, Marie Hinrichs, Mohammad Fazleh Elahi, Keith Suderman, Marc Verhagen, Kyeongmin Rim, Pavel Stranak, Jozef Misutka
Fluid Annotation: A Granularity-aware Annotation Tool for Chinese Word Fluidity
Shu-Kai Hsieh, Yu-Hsiang Tseng, Chi-Yao Lee, Chiung-Yu Chiang
E-magyar -- A Digital Language Processing System
Tamás Várádi, Eszter Simon, Bálint Sass, Iván Mittelholcz, Attila Novák, Balázs Indig, Richárd Farkas, Veronika Vincze
ILCM - A Virtual Research Infrastructure for Large-Scale Qualitative Data
Andreas Niekler, Arnim Bleier, Christian Kahmann, Lisa Posch, Gregor Wiedemann, Kenan Erdogan, Gerhard Heyer, Markus Strohmaier
CLARIN’s Key Resource Families
Darja Fišer, Jakob Lenardič, Tomaž Erjavec
Indra: A Word Embedding and Semantic Relatedness Server
Juliano Efson Sales, Leonardo Souza, Siamak Barzegar, Brian Davis, André Freitas, Siegfried Handschuh
A UIMA Database Interface for Managing NLP-related Text Annotations
Giuseppe Abrami, Alexander Mehler
European Language Resource Coordination: Collecting Language Resources for Public Sector Multilingual Information Management
Andrea Lösch, Valérie Mapelli, Stelios Piperidis, Andrejs Vasiļjevs, Lilli Smal, Thierry Declerck, Eileen Schnur, Khalid Choukri, Josef Van Genabith

18.00-18.45
ELRA Individual Members Assembly
19.30
Welcome Reception
Detailed Conference Programme

10 May 2018

9.00-9.40  
**Keynote Speech**

Charles Yang  
How Children Overcome the Sparsity Problem  
Chaiperson: Jan Odijk  

*Room: Tenzui*

9.40-9.45  
**5-minute break**

9.45-10.05  
**I-O1 ● Industry Track ● Industrial systems**

Chair: Martin Jansche  

*Room: Tenzui*

9.45-10.05  
Tilde MT Platform for Developing Client Specific MT Solutions  
Mārcis Pinnis, Andrejs Vasiljevs, Rihards Kalniņš, Roberts Rozis, Raivis Skadiņš, Valters Šics

10.05-10.25  
Improving homograph disambiguation with supervised machine learning  
Kyle Gorman, Gleb Mazovetskiy, Vitaly Nikolaev

10.25-10.45  
Text Normalization Infrastructure that Scales to Hundreds of Language Varieties  
Mason Chua, Daan Van Esch, Noah Coccaro, Eunjoon Cho, Sujeet Bhandari, Libin Jia

10.45-11.05  
**O13 ● Paraphrase & Semantics**

Chair: Udo Kruschwitz

*Room: Tenran*

9.45-10.05  
DeModify: A Dataset for Analyzing Contextual Constraints on Modifier Deletion  
Vivi Nastase, Devon Fritz, Anette Frank

10.05-10.25  
Open Subtitles Paraphrase Corpus for Six Languages  
Mathias Creutz

10.25-10.45  
Fine-grained Semantic Textual Similarity for Serbian  
Vuk Batanović, Miloš Cvetanović, Boško Nikolić

10.45-11.05  
SPADE: Evaluation Dataset for Monolingual Phrase Alignment  
Yuki Arase, Jun’ichi Tsujii

11.05-11.25  
ETPC - A Paraphrase Identification Corpus Annotated with Extended Paraphrase Typology and Negation  
Venelin Kovatchev, Toni Marti, Maria Salamo
O14 ● Emotion & Sentiment (2)
Chair: Min Zhang
Room: Tengyoku

9.45-10.05     Introducing a Lexicon of Verbal Polarity Shifters for English
Marc Schulder, Michael Wiegand, Josef Ruppenhofer, Stephanie Köser

10.05-10.25    JFCKB: Japanese Feature Change Knowledge Base
Tetsuaki Nakamura, Daisuke Kawahara

10.25-10.45    Quantifying Qualitative Data for Understanding Controversial Issues
Michael Wojatzki, Saif Mohammad, Torsten Zesch, Svetlana Kiritchenko

10.45-11.05    Distribution of Emotional Reactions to News Articles in Twitter
Omar Juárez Gambino, Hiram Calvo, Consuelo-Varinia García-Mendoza

11.05-11.25    Aggression-annotated Corpus of Hindi-English Code-mixed Data
Ritesh Kumar, Aishwarya N. Reganti, Akshit Bhatia, Tushar Maheshwari

O15 ● Semantics & Lexicon (2)
Chair: Reinhard Rapp
Room: Tenju

9.45-10.05     Creating a Verb Synonym Lexicon Based on a Parallel Corpus
Zdenka Uresova, Eva Fucikova, Eva Hajicova, Jan Hajic

10.05-10.25    Evaluation of Domain-specific Word Embeddings using Knowledge Resources
Farhad Nooralahzadeh, Lilja Øvrelid, Jan Tore Lønning

10.25-10.45    Automatic Thesaurus Construction for Modern Hebrew
Chaya Liebeskind, Ido Dagan, Jonathan Schler

10.45-11.05    Automatic Wordnet Mapping: from CoreNet to Princeton WordNet
Jiseong Kim, Younggyun Hahm, Sunggoo Kwon, KEY-SUN CHOI

11.05-11.25    The New Propbank: Aligning Propbank with AMR through POS Unification
Tim O’Gorman, Sameer Pradhan, Martha Palmer, Julia Bonn, Kathryn Conger, James Gung

O16 ● Bilingual Speech Corpora & Code-Switching
Chair: Christopher Cieri
Room: Tenyo

9.45-10.05     The Boarnsterhim Corpus: A Bilingual Frisian-Dutch Panel and Trend Study
Marjoleine Sloos, Eduard Drenth, Wilbert Heeringa

10.05-10.25    The French-Algerian Code-Switching Triggered audio corpus (FACST)
Amazouz Djegdjiga, Martine Adda-Decker, Lori Lamel

10.25-10.45    Strategies and Challenges for Crowdsourcing Regional Dialect Perception Data for
Swiss German and Swiss French
Jean-Philippe Goldman, Simon Clematide, Mathieu Avanzi, Raphaël Tandler
10.45-11.05  Phonetically Balanced Code-Mixed Speech Corpus for Hindi-English Automatic Speech Recognition
Ayushi Pandey, Brij Mohan Lal Srivastava, Rohit Kumar, Bhanu Teja Nellore, Kasi Sai Teja, Suryakanth V Gangashetty

11.05-11.25  Chinese-Portuguese Machine Translation: A Study on Building Parallel Corpora from Comparable Texts
Siyou Liu, Longyue Wang, Chao-Hong Liu

9:45 - 11:25  Poster Sessions – Poster Area 2

P20 ● Bibliometrics, Scientometrics, Infometrics
Chair: Richard Eckart de Castilho
A High-Quality Gold Standard for Citation-based Tasks
Michael Färber, Alexander Thiemann, Adam Jatowt
Measuring Innovation in Speech and Language Processing Publications
Joseph Mariani, Gil Francopoulo, Patrick Paroubek
PDFdigest: an Adaptable Layout-Aware PDF-to-XML Textual Content Extractor for Scientific Articles
Daniel Ferrés, Horacio Saggion, Francesco Ronzano, Alex Bravo
Automatic Identification of Research Fields in Scientific Papers
Eric Kergosien, Amin Farvardin, Maguelonne Teisseire, Marie-Noelle BESSAGNET, Joachim Schöpfel, Stéphane Chaudiron, Bernard Jacquemin, Annig Lacayrelle, Mathieu Roche, Christian Sallaberry, Jean-Philippe Tonneau

P21 ● Discourse Annotation, Representation and Processing (1)
Chair: Silvia Pareti
A «Portrait» Approach to Multichannel Discourse
Andrej Kibrik, Olga Fedorova
Multilingual Extension of PDTB-Style Annotation: The Case of TED Multilingual Discourse Bank
Deniz Zeyrek, Amália Mendes, Murathan Kurfalı
Building a Macro Chinese Discourse Treebank
Xiaomin Chu, Feng Jiang, Sheng Xu, Qiaoming Zhu
Enhancing the AI2 Diagrams Dataset Using Rhetorical Structure Theory
Tuomo Hiippala, Serafina Orekhova
QUD-Based Annotation of Discourse Structure and Information Structure: Tool and Evaluation
Kordula De Kuthy, Nils Reiter, Arndt Riester
The Spot the Difference corpus: a multi-modal corpus of spontaneous task oriented spoken interactions
José Lopes, Nils Hemmingsson, Oliver Åstrand
Attention for Implicit Discourse Relation Recognition
Andre Cianflone, Leila Kosseim
A Context-based Approach for Dialogue Act Recognition using Simple Recurrent Neural Networks
Chandrakant Bothe, Cornelius Weber, Sven Magg, Stefan Wermter

TreeAnnotator: Versatile Visual Annotation of Hierarchical Text Relations
Philipp Helfrich, Elias Rieb, Giuseppe Abrami, Andy Lücking, Alexander Mehler

Chats and Chunks: Annotation and Analysis of Multiparty Long Casual Conversations
Emer Gilmartin, Carl Vogel, Nick Campbell

P22 ● Evaluation Methodologies
Chair: Edouard Geoffrois

Extending the gold standard for a lexical substitution task: is it worth it?
Ludovic Tanguy, Cécile Fabre, Laura Rivière

Lexical and Semantic Features for Cross-lingual Text Reuse Classification: an Experiment in English and Latin Paraphrases
Maria Moritz, David Steding

Investigating the Influence of Bilingual MWU on Trainee Translation Quality
Yu Yuan, Serge Sharoff

Evaluation of Dictionary Creating Methods for Finno-Ugric Minority Languages
Zsanett Ferenczi, Iván Mittelholcz, Eszter Simon, Tamás Váradi

Dysarthric speech evaluation: automatic and perceptual approaches
Imed Laaridh, Christine Meunier, Corinne Fredouille

Towards an Automatic Assessment of Crowdsourced Data for NLU
Patricia Braunger, Wolfgang Maier, Jan Wessling, Maria Schmidt

Visual Choice of Plausible Alternatives: An Evaluation of Image-based Commonsense Causal Reasoning
Jinyoung Yeo, Gyeongbok Lee, Gengyu Wang, Seungtaek Choi, Hyunsouk Cho, Reinald Kim Amplayo, Seung-won Hwang

Is it worth it? Budget-related evaluation metrics for model selection
Filip Klubička, Giancarlo D. Salton, John D. Kelleher

Automated Evaluation of Out-of-Context Errors
Patrick Huber, Jan Niehues, Alex Waibel

Matics Software Suite: New Tools for Evaluation and Data Exploration
Olivier Galibert, Guillaume Bernard, Agnes Delaborde, Sabrina Lecadre, Juliette Kahn

MGAD: Multilingual Generation of Analogy Datasets
Mostafa Abdou, Artur Kulmizev, Vinit Ravishankar

P23 ● Information Extraction, Information Retrieval, Text Analytics (2)
Chair: Pierre Zweigenbaum

MIsA: Multilingual "IsA" Extraction from Corpora
Stefano Faralli, Els Lefever, Simone Paolo Ponzetto

Biomedical term normalization of EHRs with UMLS
Naiara Perez, Montse Cuadros, German Rigau
Revisiting the Task of Scoring Open IE Relations  
William Lechelle, Phillippe Langlais

A supervised approach to taxonomy extraction using word embeddings  
Rajdeep Sarkar, John Philip McCrae, Paul Buitelaar

A Chinese Dataset with Negative Full Forms for General Abbreviation Prediction  
Yi Zhang, Sun Xu

Korean TimeBank Including Relative Temporal Information  
Chae-Gyun Lim, Young-Seob Jeong, Ho-Jin Choi

Mining Biomedical Publications With The LAPPs Grid  
Nancy Ide, Keith Suderman, Jin-Dong Kim

An Initial Test Collection for Ranked Retrieval of SMS Conversations  
Rashmi Sankepally, Douglas W. Oard

FrNewsLink : a corpus linking TV Broadcast News Segments and Press Articles  
Nathalie Camelin, Géraldine Damnati, Abdessalam Boucheikif, Anais Landeau, Delphine Charlet, Yannick Esteve

PyRATA, Python Rule-based feAture sTructure Analysis  
Nicolas Hernandez, Amir Hazem

**P24 • Multimodality**  
Chair: Martin Braschler

Towards Processing of the Oral History Interviews and Related Printed Documents  
Zbynek Zajic, Lucie Skorkovska, Petr Neduchal, Pavel Ircing, Josef V. Psutka, Marek Hruz, Ales Prazak, Daniel Soutner, Jan Švec, Lukas Bures, Luked Muller

Multi Modal Distance - An Approach to Stemma Generation With Weighting  
Armin Hoenen

A Corpus of Natural Multimodal Spatial Scene Descriptions  
Ting Han, David Schlangen

The Effects of Unimodal Representation Choices on Multimodal Learning  
Fernando T. Ito, Helena de Medeiros Caseli, Jander Moreira

An Evaluation Framework for Multimodal Interaction  
Nikhil Krishnaswamy, James Pustejovsky

The WAW Corpus: The First Corpus of Interpreted Speeches and their Translations for English and Arabic  
Ahmed Abdelali, Irina Temnikova, Samy Hedaya, Stephan Vogel

Polish Corpus of Annotated Descriptions of Images  
Alina Wróblewska

Action Verb Corpus  
Stephanie Gross, Matthias Hirschmanner, Brigitte Krenn, Friedrich Neubarth, Michael Zillich

EMO&LY (EMOtion and AnomaLY): A new corpus for anomaly detection in an audiovisual stream with emotional context.  
Cédric Fayet, Arnaud Delhay, Damien Lolive, Pierre-françois Marteau
Development of an Annotated Multimodal Dataset for the Investigation of Classification and Summarisation of Presentations using High-Level Paralinguistic Features

Keith Curtis, Nick Campbell, Gareth Jones

**P25 ● Parsing, Syntax, Treebank (1)**

Chair: Simonetta Montemagni

BKTtreebank: Building a Vietnamese Dependency Treebank

Kiem-Hieu Nguyen

GeCoTagger: Annotation of German Verb Complements with Conditional Random Fields

Roman Schneider, Monica Fürbacher

AET: Web-based Adjective Exploration Tool for German

Tatiana Bladier, Esther Seyffarth, Oliver Hellwig, Wiebke Petersen

ZAP: An Open-Source Multilingual Annotation Projection Framework

Alan Akbij, Roland Vollgraf

Palmyra: A Platform Independent Dependency Annotation Tool for Morphologically Rich Languages

Talha Javed, Nizar Habash, Dima Taji

A Web-based System for Crowd-in-the-Loop Dependency Treebanking

Stephen Tratz, Nhien Phan

Building Universal Dependency Treebanks in Korean

Jayeol Chun, Na-Rae Han, Jena D. Hwang, Jinho D. Choi

Moving TIGER beyond Sentence-Level

Agnieszka Falenska, Kerstin Eckart, Jonas Kuhn

Spanish HPSG Treebank based on the AnCora Corpus

Luis Chiruzzo, Dina Wonsever

Universal Dependencies for Amharic

Binyam Ephrem Seyoum, Yusuke Miyao, Baye Yimam Mekonnen

A Parser for LTAG and Frame Semantics

David Arps, Simon Petitjean

Multilingual Dependency Parsing for Low-Resource Languages: Case Studies on North Saami and Komi-Zyrian

KyungTae Lim, Niko Partanen, Thierry Poibeau

11.25-11.45 **Coffee Break**

**O17 ● Evaluation Methodologies**

Chair: Maciej Piasecki

Room: Tenran
Evaluating the WordsEye Text-to-Scene System: Imaginative and Realistic Sentences  
Morgan Ulinski, Bob Coyne, Julia Hirschberg

Computer-assisted Speaker Diarization: How to Evaluate Human Corrections  
Pierre-Alexandre Broux, David Doukhan, Simon Petitrenaud, Sylvain Meignier, Jean Carrive

Performance Impact Caused by Hidden Bias of Training Data for Recognizing Textual Entailment  
Masatoshi Tsuchiya

Evaluation of Croatian Word Embeddings  
Lukas Svoboda, Slobodan Beliga

O18 ● Semantics  
Chair: James Pustejovsky  

Room: Tengyoku

C-HTS: A Concept-based Hierarchical Text Segmentation approach  
Mostafa Bayomi, Seamus Lawless

Semantic Supersenses for English Possessives  
Austin Blodgett, Nathan Schneider

A Corpus of Metaphor Novelty Scores for Syntactically-Related Word Pairs  
Natalie Parde, Rodney Nielsen

Improving Hypernymy Extraction with Distributional Semantic Classes  
Alexander Panchenko, Dmitry Ustalov, Stefano Faralli, Simone Paolo Ponzetto, Chris Biemann

O19 ● Information Extraction & Neural Networks  
Chair: Bernardo Magnini  

Room: Tenju

Laying the Groundwork for Knowledge Base Population: Nine Years of Linguistic Resources for TAC KBP  
Jeremy Getman, Joe Ellis, Stephanie Strassel, Zhiyi Song, Jennifer Tracey

A Dataset for Inter-Sentence Relation Extraction using Distant Supervision  
Angrosh Mandya, Danushka Bollegala, Frans Coenen, Katie Atkinson

Diacritics Restoration Using Neural Networks  
Jakub Náplava, Milan Straka, Pavel Strahák, Jan Hajic

Ensemble Romanian Dependency Parsing with Neural Networks  
Radu Ion, Elena Irimia, Virginica Barbu Mititelu

O20 ● Dialogue, Emotion, Multimodality  
Chair: Nick Campbell  

Room: Tenyo
11:45-12:05 Classifying Sluice Occurrences in Dialogue
*Austin Baird, Anissa Hamza, Daniel Hardt*

12:05-12:25 Collection of Multimodal Dialog Data and Analysis of the Result of Annotation of Users' Interest Level
*Masahiro Araki, Sayaka Tomimasu, Mikio Nakano, Kazunori Komatani, Shogo Okada, Shinya Fujie, Hiroaki Sugiyama*

12:25-12:45 Recognizing Behavioral Factors while Driving: A Real-World Multimodal Corpus to Monitor the Driver’s Affective State
*Alicia Lotz, Klas Ihme, Audrey Charnoz, Pantelis Maroudis, Ivan Dmitriev, Andreas Wendemuth*

12:45-13:05 EmotionLines: An Emotion Corpus of Multi-Party Conversations
*Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao Huang, Lun-Wei Ku*

11:45 - 13:05

**Poster Sessions – Poster Area 1**

I-P1 ● Industry Track - Industrial Systems

Chair: Linne Ha

**FonBund: A Library for Combining Cross-lingual Phonological Segment Data**
*Alexander Gutkin, Martin Jansche, Tatiana Merkulova*

**Voice Builder: A Tool for Building Text-To-Speech Voices**
*Pasindu De Silva, Theeraphol Wattanavekin, Tang Hao, Knot Pipatsrisawat*

**Sudachi: a Japanese Tokenizer for Business**
*Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, Yuji Matsumoto*

**Chemical Compounds Knowledge Visualization with Natural Language Processing and Linked Data**
*Kazunari Tanaka, Tomoya Iwakura, Yusuke Koyanagi, Noriko Ikeda, Hiroyuki Shindo, Yuji Matsumoto*

P26 ● Language Acquisition & CALL (1)

Chair: Donghui Lin

**Using Discourse Information for Education with a Spanish-Chinese Parallel Corpus**
*Shuyuan Cao, Harritxu Gete*

**A 2nd Longitudinal Corpus for Children's Writing with Enhanced Output for Specific Spelling Patterns**
*Kay Berkling*

**Development of a Mobile Observation Support System for Students: FishWatchr Mini**
*Masaya Yamaguchi, Masanori Kitamura, Naomi Yanagida*

**The AnnCor CHILDES Treebank**
*Jan Odijk, Alexis Dimitriadis, Martijn Van der Klis, Marjo Van Koppen, Meie Otten, Remco Van der Veen*
BabyCloud, a Technological Platform for Parents and Researchers
Xuan-Nga Cao, Cyrille Dakhlia, Patricia Del Carmen, Mohamed-Amine Jaouani, Malik Ould-Arbi, Emmanuel Dupoux

Infant Word Comprehension-to-Production Index Applied to Investigation of Noun Learning Predominance Using Cross-lingual CDI database
Yasuhiro Minami, Tessei Kobayashi, Yuko Okumura

Building a TOCFL Learner Corpus for Chinese Grammatical Error Diagnosis
Lung-Hao Lee, Yuen-Hsien Tseng, Liping Chang

MIAPARLE: Online training for the discrimination of stress contrasts
Jean-Philippe Goldman, Sandra Schwab

ESCRITO - An NLP-Enhanced Educational Scoring Toolkit
Torsten Zesch, Andrea Horbach

A Leveled Reading Corpus of Modern Standard Arabic
Muhammed Al Khalil, H índ Saddiki, Nizar Habash, Latifa Al Falasi

P27 ● Less-Resourced/Endangered Languages (1)
Chair: Valérie Mapelli

Developing New Linguistic Resources and Tools for the Galician Language
Rodrigo Agerri, Xavier Gómez Guinovart, German Rigau, Miguel Anxo Solla Portela

Modeling Northern Haida Verb Morphology
Jordan Lachler, Lene Antonsen, Trond Trosterud, Sjur Moshagen, Antti Arppe

Low-resource Post Processing of Noisy OCR Output for Historical Corpus Digitisation
Caitlin Richter, Matthew Wickes, Deniz Beser, Mitchell Marcus

Introducing the CLARIN Knowledge Centre for Linguistic Diversity and Language Documentation
Hanna Hedeland, Timm Lehmberg, Felix Rau, Sophie Salffner, Mandana Seyfeddinipur, Andreas Witt

Low Resource Methods for Medieval Document Sections Analysis
Petra Galuscakova, Lucie Neuzilova

SB-CH: A Swiss German Corpus with Sentiment Annotations
Ralf Grubenmann, Don Tuggener, Pius Von Däniken, Jan Deriu, Mark Cieliebak

Universal Dependencies for Ainu
Hajime Senuma, Akiko Aizawa

P28 ● Lexicon (2)
Chair: Lionel Nicolas

Signbank: Software to Support Web Based Dictionaries of Sign Language
Steve Cassidy, Onno Crasborn, Henri Nieminen, Wessel Stoop, Micha Hulsbosch, Susan Even, Erwin Komen, Trevor Johnson

J-MeDic: A Japanese Disease Name Dictionary based on Real Clinical Usage
Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Shoko Wakamiya, Tomohide Iwao, Eiji Aramaki
Building a List of Synonymous Words and Phrases of Japanese Compound Verbs
Kyoko Kanzaki, Hitoshi Isahara

Evaluating EcoLexiCAT: a Terminology-Enhanced CAT Tool
Pilar Leόn-Araúz, Arianne Reimerink

A Danish FrameNet Lexicon and an Annotated Corpus Used for Training and Evaluating a Semantic Frame Classifier
Bolette Pedersen, Sanni Nimb, Anders Søgaard, Mareike Hartmann, Sussi Olsen

SLIDE - a Sentiment Lexicon of Common Idioms
Charles Jochim, Francesca Bonin, Roy Bar-Haim, Noam Slonim

PronouncUR: An Urdu Pronunciation Lexicon Generator
Harris Bin Zia, Agha Ali Raza, Anis Athar

SimLex-999 for Polish
Agnieszka Mykowiecka, Malgorzata Marciniak, Piotr Rychlik

Finely Tuned, 2 Billion Token Based Word Embeddings for Portuguese
João Rodrigues, António Branco

P29 ● Linked Data
Chair: Monica Monachini

Teanga: A Linked Data based platform for Natural Language Processing
Housam Ziad, John Philip McCrae, Paul Buitleaar

Automatic and Manual Web Annotations in an Infrastructure to handle Fake News and other Online Media Phenomena
Georg Rehm, Julian Moreno-Schneider, Peter Bourgonje

The LODeXporter: Flexible Generation of Linked Open Data Triples from NLP Frameworks for Automatic Knowledge Base Construction
René Witte, Bahar Sateli

LiDo RDF: From a Relational Database to a Linked Data Graph of Linguistic Terms and Bibliographic Data
Bettina Klimek, Robert Schädlich, Dustin Kröger, Edwin Knese, Benedikt Elßmann

Towards a Linked Open Data Edition of Sumerian Corpora
Christian Chiarcos, Émilie Pagé-Perron, Ilya Khait, Niko Schenk, Lucas Reckling

P30 ● Infrastructural Issues/Large Projects (2)
Chair: Krister Lindén

A Bird’s-eye View of Language Processing Projects at the Romanian Academy
Dan Tufiş, Cristea Dan

PMKI: an European Commission action for the interoperability, maintainability and sustainability of Language Resources
Peter Schmitz, Enrico Francesconi, Najeh Hajlaoui, Brahim Batouche

The Abkhaz National Corpus
Paul Meurer
Collecting Language Resources from Public Administrations in the Nordic and Baltic Countries
Andrejs Vasiļjevs, Rihards Kalniņš, Roberts Rozis, Aivars Bērziņš

P31 ● MultiWord Expressions & Collocations
Chair: Brigitte Krenn

LIdioms: A Multilingual Linked Idioms Data Set
Diego Moussallem, Mohamed Ahmed Sherif, Diego Esteves, Marcos Zampieri, Axel-Cyrille Ngonga Ngomo

Annotating Modality Expressions and Event Factuality for a Japanese Chess Commentary Corpus
Suguru Matsuyoshi, Hirotaka Kameko, Yugo Murawaki, Shinsuke Morů

Annotating Chinese Light Verb Constructions according to PARSEME guidelines
Menghan Jiang, Natalia Klyueva, Hongzhi Xu, Chu-Ren Huang

Using English Baits to Catch Serbian Multi-Word Terminology
Cvetana Krstev, Branislava Šandrih, Ranka Stankovic, Miljana Mladenović

Construction of Large-scale English Verbal Multiword Expression Annotated Corpus
Akihiko Kato, Hiroyuki Shindo, Yuji Matsumoto

Konbitzul: an MWE-specific database for Spanish-Basque
Uxoa Iñurrieta, Itzizar Aduriz, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola

A Multilingual Test Collection for the Semantic Search of Entity Categories
Juliano Efson Sales, Siham Barzegar, Wellington Franco, Bernhard Bermeitinger, Tiago Cunha, Brian Davis, André Freitas, Siegfried Handschuh

Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study
Melania Cabezas-García, Pilar León-Araúz

Generation of a Spanish Artificial Collocation Error Corpus
Sara Rodríguez-Fernández, Roberto Carlini, Leo Wanner

Improving a Neural-based Tagger for Multiword Expressions Identification
Dušan Variš, Natalia Klyueva

Designing a Russian Idiom-Annotated Corpus
Katsiaryna Aharodnik, Anna Feldman, Jing Peng

13.05-13.10 5-minute break
13.10-13.30 Japanese Invited Talk

Yukinori Takubo
How many languages are there in Japan?
Chair: Koiti Hasida
Room: Tenzui

13.30-14.50 Lunch Break

I-O2 ● Industry Track ● Human computation in industry
Chair: Kyle Gorman
Room: Tenzui
14.50-15.10  Academic-Industrial Perspective on the Development and Deployment of a Moderation System for a Newspaper Website
Dieter Schabus, Marcin Skowron

15.10-15.30  Community-Driven Crowdsourcing: Data Collection with Local Developers
Christina Funk, Michael Tseng, Ravindran Rajakumar, Linne Ha

15.30-15.50  Building Open Javanese and Sundanese Corpora for Multilingual Text-to-Speech
Jaka Aris Eko Wibawa, Supheakmungkol Sarin, Chenfeng Li, Knot Pipatsrisawat, Keshan Sodimana, Oddur Kjartansson, Alexander Gutkin, Martin Jansche, Linne Ha

16.30-16.50  Coffee Break

16.50-18.30  Linguists and Linguistics in Industry: Panel discussion and Q&A - Moderator: Linne Ha

**O21 ● Discourse & Argumentation**

Chair: Costanza Navarretta

*Room: Tenran*

14.50-15.10  An Integrated Representation of Linguistic and Social Functions of Code-Switching
Silvana Hartmann, Monojit Choudhury, Kalika Bali

15.10-15.30  A Corpus of eRulemaking User Comments for Measuring Evaluability of Arguments
Joonsuk Park, Claire Cardie

15.30-15.50  A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations
Elena Musi, Manfred Stede, Leonard Kriese, Smaranda Muresan, Andrea Rocci

15.50-16.10  Discourse Coherence Through the Lens of an Annotated Text Corpus: A Case Study
Eva Hajicova, Jiří Mirovský

16.10-16.30  Automatic Prediction of Discourse Connectives
Eric Malmi, Daniele Pighin, Sebastian Krause, Mikhail Kozhevnikov

**O22 ● Less-Resourced & Ancient Languages**

Chair: Mark Liberman

*Room: Tengyoku*

14.50-15.10  Handling Rare Word Problem using Synthetic Training Data for Sinhala and Tamil Neural Machine Translation
Pasindu Tennage, Prabath Sandaruwan, Malith Thilakarathne, Achini Herath, Surangika Ranathunga

15.10-15.30  BDPROTO: A Database of Phonological Inventories from Ancient and Reconstructed Languages
Egidio Marsico, Sebastien Flavier, Annemarie Verkerk, Steven Moran

15.30-15.50  Creating a Translation Matrix of the Bible’s Names Across 591 Languages
Winston Wu, Nidhi Vyas, David Yarowsky

15.50-16.10  Building a Word Segmenter for Sanskrit Overnight
Vikas Reddy, Amrith Krishna, Vishnu Sharma, Prateek Gupta, Vineeth M R, Pawan Goyal
16.10-16.30 Simple Semantic Annotation and Situation Frames: Two Approaches to Basic Text Understanding in LORELEI
*Kira Griffitt, Jennifer Tracey, Ann Bies, Stephanie Strassel*

**O23 ● Semantics & Evaluation**

Chair: Gerard de Melo

***Room: Tenju***

14.50-15.10 Abstract Meaning Representation of Constructions: The More We Include, the Better the Representation
*Claire Bonial, Bianca Badarau, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Tim O’Gorman, Martha Palmer, Nathan Schneider*

15.10-15.30 Evaluating Scoped Meaning Representations
*Rik Van Noord, Lasha Abzianidze, Hessel Haagsma, Johan Bos*

15.30-15.50 Huge Automatically Extracted Training-Sets for Multilingual Word Sense Disambiguation
*Tommaso Pasini, Francesco Elia, RobertoNavigli*

15.50-16.10 SentEval: An Evaluation Toolkit for Universal Sentence Representations
*Alexis Conneau, Douwe Kiela*

*Steven Neale*

**O24 ● Multimodal & Written Corpora**

Chair: Erhard Hinrichs

***Room: Tenyo***

14.50-15.10 Linguistically-driven Framework for Computationally Efficient and Scalable Sign Recognition
*Dimitri Metaxas, Mark Dilsizian, Carol Neidle*

15.10-15.30 CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control
*Lei Chen, Sylvie Gibet, Camille Marteau*

15.30-15.50 Neural Caption Generation for News Images
*Vishwash Batra, Yulan He, George Vogiatzis*

15.50-16.10 MPST: A Corpus of Movie Plot Synopses with Tags
*Sudipta Kar, Suraj Maharjan, Adrian Pastor López Monroy, Thamar Solorio*

16.10-16.30 OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora
*Pierre Lison, Jörg Tiedemann, Milen Kouylekov*

14:50 - 16:30

Poster Sessions – Poster Area 2

**P32 ● Document Classification, Text Categorisation (1)**

Chair: Monserrat Marimon
DeepTC – An Extension of DKPro Text Classification for Fostering Reproducibility of Deep Learning Experiments
Tobias Horsmann, Torsten Zesch

Improving Hate Speech Detection with Deep Learning Ensembles
Steven Zimmerman, Udo Kruschwitz, Chris Fox

Distributional Term Set Expansion
Amaru Cuba Gyllensten, Magnus Sahlgren

Can Domain Adaptation be Handled as Analogies?
Núria Bel, Joel Pocostales

Author Profiling from Facebook Corpora
Fernando Hsieh, Rafael Dias, Ivandré Paraboni

Semantic Relatedness of Wikipedia Concepts -- Benchmark Data and a Working Solution
Liat Ein Dor, Alon Halfon, Yoav Kantor, Ran Levy, Yosi Mass, Ruty Rinott, Eyal Shnarch, Noam Slonim

Experiments with Convolutional Neural Networks for Multi-Label Authorship Attribution
Dainis Boumber, Yifan Zhang, Arjun Mukherjee

P33 ● Morphology (1)
Chair: Piotr Banski

A Fast and Accurate Vietnamese Word Segmenter
Dat Quoc Nguyen, Dai Quoc Nguyen, Thanh Vu, Mark Dras, Mark Johnson

Finite-state morphological analysis for Gagauz
Francis Tyers, Sevilay Bayatli, Güllü Karanfil, Memduh Gokirmak

Albanian Part-of-Speech Tagging: Gold Standard and Evaluation
Besim Kabashi, Thomas Proisl

Morphology Injection for English-Malayalam Statistical Machine Translation
Sreelekha S, Pushpak Bhattacharyya

The Morpho-syntactic Annotation of Animacy for a Dependency Parser
Mohammed Attia, Vitaly Nikolaev, Ali Elkahky

MADARI: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction
Ossama Obeid, Salam Khalifa, Nizar Habash, Houda Bouamor, Wajdi Zaghouani, Kemal Oflazer

A Morphological Analyzer for St. Lawrence Island / Central Siberian Yupik
Emily Chen, Lane Schwartz

Universal Morphologies for the Caucasus region
Christian Chiarcos, Kathrin Donandt, Maxim Ionov, Monika Rind-Pawlowski, Hasmik Sargsian, Jesse Wichers Schreur, Frank Abromeit, Christian Fäth

P34 ● Opinion Mining / Sentiment Analysis (2)
Chair: Patrick Paroubek

EMTC: Multilabel Corpus in Movie Domain for Emotion Analysis in Conversational Text
Phan Duc-Anh, Yuji Matsumoto
Complex and Precise Movie and Book Annotations in French Language for Aspect Based Sentiment Analysis
Stefania Pecore, Jeanne Villaneau

Lingmotif-lex: a Wide-coverage, State-of-the-art Lexicon for Sentiment Analysis
Antonio Moreno-Ortiz, Chantal Pérez-Hernández

A Japanese Corpus for Analyzing Customer Loyalty Information
Yiou Wang, Takuji Tahara

FooTweets: A Bilingual Parallel Corpus of World Cup Tweets
Henny Sluyter-Gäthje, Pintu Lohar, Haithem Afli, Andy Way

The SSIX Corpora: Three Gold Standard Corpora for Sentiment Analysis in English, Spanish and German Financial Microblogs
Thomas Gaillat, Manel Zarrouk, André Freitas, Brian Davis

Sarcasm Target Identification: Dataset and An Introductory Approach
Aditya Joshi, Pranav Goel, Pushpak Bhattacharyya, Mark Carman

Annotating Opinions and Opinion Targets in Student Course Feedback
Janaka Chathuranga, Shanika Ediriweera, Ravindu Hasantha, Pranidhith Munasinghe, Surangika Ranathunga

Generating a Gold Standard for a Swedish Sentiment Lexicon
Jacobo Rouces, Nina Tahmasebi, Lars Borin, Stian Rødven Eide

P35 ● : Phonetic Databases, Phonology
Chair: Martine Adda-Decker

WordKit: a Python Package for Orthographic and Phonological Featurization
Stephan Tulkens, Dominiek Sandra, Walter Daelemans

Pronunciation Variants and ASR of Colloquial Speech: A Case Study on Czech
David Lukeš, Marie Kopřivová, Zuzana Komrsková, Petra Klimešová

Epitran: Precision G2P for Many Languages
David R. Mortensen, Siddharth Dalmia, Patrick Littell

P36 ● Question Answering and Machine Reading
Chair: António Branco

A Multilingual Approach to Question Classification

Dataset for the First Evaluation on Chinese Machine Reading Comprehension
Yiming Cui, Ting Liu, Zhipeng Chen, Wentao Ma, Shijin Wang, Guoping Hu

A Multi-Domain Framework for Textual Similarity. A Case Study on Question-to-Question and Question-Answering Similarity Tasks
Amir Hazem, Basma El Amel Boussaha, Nicolas Hernandez

WorldTree: A Corpus of Explanation Graphs for Elementary Science Questions supporting Multi-hop Inference
Peter Jansen, Elizabeth Wainwright, Steven Marmorstein, Clayton Morrison
Analysis of Implicit Conditions in Database Search Dialogues
Shun-ya Fukunaga, Hitoshi Nishikawa, Takenobu Tokunaga, Hikaru Yokono, Tetsuro Takahashi

An Information-Providing Closed-Domain Human-Agent Interaction Corpus
Jelte Van Waterschoot, Guillaume Dubuisson Duplessis, Lorenzo Gatti, Merijn Brui jnes, Dirk Heylen

Augmenting Image Question Answering Dataset by Exploiting Image Captions
Masashi Yokota, Hideki Nakayama

Semi-supervised Training Data Generation for Multilingual Question Answering
Kyungjae Lee, Kyoungho Yoon, Sunghyun Park, Seung-won Hwang

PhotoshopQuiA: A Corpus of Non-Factoid Questions and Answers for Why-Question Answering
Andrei Dulceanu, Thang Le Dinh, Walter Chang, Trung Bui, Doo Soon Kim, Manh Chien Vu, Seokhwan Kim

BioRead: A New Dataset for Biomedical Reading Comprehension
Dimitris Pappas, Ion Androutsopoulos, Haris Papageorgiou

MMQA: A Multi-domain Multi-lingual Question-Answering Framework for English and Hindi
Deepak Gupta, Surabhi Kumari, Asif Ekbal, Pushpak Bhattacharyya

P37 ● Social Media Processing (2)
Chair: Tetsuro Takahashi

The First 100 Days: A Corpus Of Political Agendas on Twitter
Nathan Green, Septina Larasati

Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System
Shweta Yadav, Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya

An Italian Twitter Corpus of Hate Speech against Immigrants
Manuela Sanguinetti, Fabio Poletto, Cristina Bosco, Viviana Patti, Marco Stranisci

A Large Multilingual and Multi-domain Dataset for Recommender Systems
Giorgia Di Tommaso, Stefano Faralli, Paola Velardi

RtGender: A Corpus for Studying Differential Responses to Gender
Rob Voigt, David Jurgens, Vinodkumar Prabhakaran, Dan Jurafsky, Yulia Tsvetkov

A Neural Network Model for Part-Of-Speech Tagging of Social Media Texts
Sara Meftah, Nasredine Semmar

Utilizing Large Twitter Corpora to Create Sentiment Lexica
Valerij Fredriksen, Brage Jahren, Björn Gambäck

P38 ● Speech Resource/Database (1)
Chair: Gilles Adda

The Nautilus Speaker Characterization Corpus: Speech Recordings and Labels of Speaker Characteristics and Voice Descriptions
Laura Fernández Gallardo, Benjamin Weiss
Evaluation of Automatic Formant Trackers
Florian Schiel, Thomas Zitzelsberger

Design and Development of Speech Corpora for Air Traffic Control Training
Luboš Šmídl, Jan Švec, Daniel Tihelka, Jindrich Matousek, Jan Romportl, Pavel Irčing

A First South African Corpus of Multilingual Code-switched Soap Opera Speech
Ewald Van der westhuizen, Thomas Niesler

A Web Service for Pre-segmenting Very Long Transcribed Speech Recordings
Nina Poerner, Florian Schiel

A Real-life, French-accented Corpus of Air Traffic Control Communications
Estelle Delpech, Marion Laignelet, Christophe Pimm, Céline Raynal, Michal Trzos, Alexandre Arnold, Dominique Pronto

Creating Lithuanian and Latvian Speech Corpora from Inaccurately Annotated Web Data
Askars Salimbajevs

Discovering Canonical Indian English Accents: A Crowdsourcing-based Approach
Sunayana Sitaram, Varun Manjunath, Varun Bharadwaj, Monojit Choudhury, Kalika Bali, Michael Tjalve

Extending Search System based on Interactive Visualization for Speech Corpora
Tomoko Ohsuga, Yuichi Ishimoto, Tomoko Kajiyama, Shunsuke Kozawa, Kiyotaka Uchimoto, Shuichi Itahashi

German Radio Interviews: The GRAIN Release of the SFB732 Silver Standard Collection
Katrin Schweitzer, Kerstin Eckart, Markus Gärtner, Agnieszka Falenska, Arndt Riester, Ina Roesiger, Antje Schweitzer, Sabrina Stehwien, Jonas Kuhn

Preparing Data from Psychotherapy for Natural Language Processing
Margot Mieskes, Andreas Stiegelmayr

MirasVoice: A bilingual (English-Persian) speech corpus
Amir Vaheb, Ali Janalizadeh Choobbasti, Mahdi Mortazavi, Saeid Safavi, Behnam Sabeti

16.30-16.50 Coffee Break

O25 ● Social Media & Evaluation
Chair: Nasredine Semmar

Room: Tenran

16.50-17.10 Building an Ellipsis-aware Chinese Dependency Treebank for Web Text
Xuancheng Ren, Xu SUN, Ji Wen, Bingzhen Wei, Weidong Zhan, Zhiyuan Zhang

17.10-17.30 EuroGames16: Evaluating Change Detection in Online Conversation
Cyril Goutte, Yunli Wang, FangMing Liao, Zachary Zanussi, Samuel Larkin, Yuri Grinberg

17.30-17.50 A Deep Neural Network based Approach for Entity Extraction in Code-Mixed Indian Social Media Text
Deepak Gupta, Asif Ekbal, Pushpak Bhattacharyya
17.50-18.10  PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies
   Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, Oronzo
   Antonelli, Fabio Tamburini

18.10-18.30  Annotating If the Authors of a Tweet are Located at the Locations They Tweet About
   Vivek Reddy Doudagiri, Alakananda Vempala, Eduardo Blanco

O26 ● Standards, Validation, Workflows
   Chair: Yohei Murakami
   Room: Tengyoku

16.50-17.10  MOCCA: Measure of Confidence for Corpus Analysis - Automatic Reliability Check of
   Transcript and Automatic Segmentation
   Thomas Kisler, Florian Schiel

17.10-17.30  Towards an ISO Standard for the Annotation of Quantification
   Harry Bunt, James Pustejovsky, Kiyong Lee

17.30-17.50  Lightweight Grammatical Annotation in the TEI: New Perspectives
   Piotr Banski, Susanne Haaf, Martin Mueller

17.50-18.10  A Gold Standard for Multilingual Automatic Term Extraction from Comparable
   Corpora: Term Structure and Translation Equivalents
   Ayla Rigouts Terryn, Veronique Hoste, Els Lefever

18.10-18.30  Handling Big Data and Sensitive Data Using EUDAT's Generic Execution Framework
   and the WebLicht Workflow Engine.
   Claus Zinn, Wei Qui, Marie Hinrichs, Emanuel Dima, Alexandr Chernov

O27 ● Treebanks & Parsing
   Chair: Wenliang Chen
   Room: Tenju

16.50-17.10  Building a Web-Scale Dependency-Parsed Corpus from CommonCrawl
   Alexander Panchenko, Eugen Ruppert, Stefano Faralli, Simone Paolo Ponzetto, Chris
   Biemann

17.10-17.30  Universal Dependencies Version 2 for Japanese
   Masayuki Asahara, Hiroshi Kanayama, Takaaki Tanaka, Yusuke Miyao, Sumire
   Uematsu, Shinsuke Mori, Yuji Matsumoto, Mai Omura, Yugo Murawaki

17.30-17.50  Developing the Bangla RST Discourse Treebank
   Debopam Das, Manfred Stede

17.50-18.10  A New Version of the Składnica Treebank of Polish Harmonised with the Walenty
   Valency Dictionary
   Marcin Woliński, Elżbieta Hajnicz, Tomasz Bartosiak

18.10-18.30  Parse Me if You Can: Artificial Treebanks for Parsing Experiments on Elliptical Constructions
   Kira Droganova, Daniel Zeman, Jenna Kanerva, Filip Ginter
O28 ● Morphology & Lexicons
Chair: Tamás Varádi

Room: Tenyo

16.50-17.10  Semi-Automatic Construction of Word-Formation Networks (for Polish and Spanish)
Mateusz Lango, Magda Sevcikova, Zdeněk Žabokrtský

17.10-17.30  A multilingual collection of CoNLL-U-compatible morphological lexicons
Benoît Sagot

17.30-17.50  UniMorph 2.0: Universal Morphology
Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sebastian J. Mielke, Arya McCarthy, Sandra Kübler, David Yarowsky, Jason Eisner, Mans Hulden

17.50-18.10  A Computational Architecture for the Morphology of Upper Tanana
Olga Lovick, Christopher Cox, Miikka Silfverberg, Antti Arppe, Mans Hulden

18.10-18.30  Expanding Abbreviations in a Strongly Inflected Language: Are Morphosyntactic Tags Sufficient?
Piotr Żelasko

16:50 - 18:30  Poster Sessions – Poster Area 1

P39 ● Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (2)
Chair: Johannes Kraus

Silvia Pareti, Tatiana Lando

JDCFC: A Japanese Dialogue Corpus with Feature Changes
Tetsuaki Nakamura, Daisuke Kawahara

Koichiro Yoshino, Hiroki Tanaka, Kyoshiro Sugiyama, Makoto Kondo, Satoshi Nakamura

The Niki and Julie Corpus: Collaborative Multimodal Dialogues between Humans, Robots, and Virtual Agents
Ron Artstein, Jill Boberg, Alesia Gainer, Jonathan Gratch, Emmanuel Johnson, Anton Leuski, Gale Lucas, David Traum

Constructing a Chinese Medical Conversation Corpus Annotated with Conversational Structures and Actions
Nan Wang, Yan Song, Fei Xia

Predicting Nods by using Dialogue Acts in Dialogue
Ryo Ishii, Ryuichiro Higashinaka, Junji Tomita
Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus
Maria Koutsombogera, Carl Vogel

Magalie Ochs, Philippe Blache, Grégoire De Montcheuil, Jean-Marie Pergandi, Jorane Saubesty, Daniel Francon, Daniel Mestre

Construction of English-French Multimodal Affective Conversational Corpus from TV Dramas
Sashi Novitasari, Quoc Truong Do, Sakriani Sakti, Dessi Lestari, Satoshi Nakamura

QUEST: A Natural Language Interface to Relational Databases
Vadim Sheinin, Elahe Khorasani, Hangu Yeo, Kun Xu, Ngoc Phuoc An Vo, Octavian Popescu

P40 ● Language Modelling
Chair: Bolette Pedersen

TF-LM: TensorFlow-based Language Modeling Toolkit
Lyan Verwimp, Hugo Van hamme, Patrick Wambacq

Grapheme-level Awareness in Word Embeddings for Morphologically Rich Languages
Suzi Park, Hyopil Shin

Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments
Katherine Schmirler, Antti Arppe, Trond Trosterud, Lene Antonsen

BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages
Benjamin Heinzerling, Michael Strube

P41 ● Natural Language Generation
Chair: Ineke Schuurman

Reference production in human-computer interaction: Issues for Corpus-based Referring Expression Generation
Danilo Rocha, Ivandré Paraboni

Definite Description Lexical Choice: taking Speaker's Personality into account
Alex Lan, Ivandré Paraboni

Referring Expression Generation in time-constrained communication
André Mariotti, Ivandré Paraboni

Incorporating Semantic Attention in Video Description Generation
Natsuda Laokulrat, Naoaki Okazaki, Hideki Nakayama

GenDR: A Generic Deep Realizer with Complex Lexicalization
François Lareau, Florie Lambrey, Ieva Dubinskaite, Daniel Galarreta-Piquette, Maryam Nejat

A Detailed Evaluation of Neural Sequence-to-Sequence Models for In-domain and Cross-domain Text Simplification
Sanja Štajner, Sergiu Nisioi
Don't Annotate, but Validate: a Data-to-Text Method for Capturing Event Data
Piek Vossen, Filip Ilievski, Marten Postma, Roxane Segers

RDF2PT: Generating Brazilian Portuguese Texts from RDF Data
Diego Moussalem, Thiago Ferreira, Marcos Zampieri, Maria Cláudia Cavalcanti, Geraldo Xexéo, Mariana Neves, Axel-Cyrille Ngonga Ngomo

Towards a music-language mapping
Michele Berlingerio, Francesca Bonin

Up-cycling Data for Natural Language Generation
Amy Isard, Jon Oberlander, Claire Grover

P42 ● Semantics (2)
Chair: Cécile Fabre

Neural Models of Selectional Preferences for Implicit Semantic Role Labeling
Minh Le, Antske Fokkens

A database of German definitory contexts from selected web sources
Adrien Barbaresi, Lothar Lemnitzer, Alexander Geyken

Annotating Abstract Meaning Representations for Spanish
Noelia Migueles-Abaira, Rodrigo Agerri, Arantza Diaz de Ibarra

Browsing the Terminological Structure of a Specialized Domain: A Method Based on Lexical Functions and their Classification
Marie-Claude L’Homme, Benoît Robichaud, Nathalie Prévil

Rollenwechsel-English: a large-scale semantic role corpus
Asad Sayeed, Pavel Shkadzko, Vera Demberg

Towards a Standardized Dataset for Noun Compound Interpretation
Girishkumar Ponkiya, Kevin Patel, Pushpak Bhattacharyya, Girish K. Palshikar

Structured Interpretation of Temporal Relations
Yuchen Zhang, Nianwen Xue

NL2Bash: A Corpus and Semantic Parser for Natural Language Interface to the Linux Operating System
Xi Victoria Lin, Chenglong Wang, Luke Zettlemoyer, Michael D. Ernst

World Knowledge for Abstract Meaning Representation Parsing
Charles Welch, Jonathan K. Kummerfeld, Song Feng, Rada Mihalcea

P43 ● Speech Processing
Chair: Sebastian Stüker

Improved Transcription and Indexing of Oral History Interviews for Digital Humanities Research
Michael Gref, Joachim Köhler, Almut Leh

Sound Signal Processing with Seq2Tree Network
Weicheng Ma, Kai Cao, Zhaocheng Ni, Peter Chin, Xiang Li

Open ASR for Icelandic: Resources and a Baseline System
Anna Björk Nikulásdóttir, Inga Rún Helgadóttir, Matthias Pétursson, Jón Guðnason
Towards Neural Speaker Modeling in Multi-Party Conversation: The Task, Dataset, and Models
Zhao Meng, Lil Mou, Zhi Jin

Discriminating between Similar Languages on Imbalanced Conversational Texts
Junqing He, Xian Huang, Xuemin Zhao, Yan Zhang, Yonghong Yan

Data-Driven Pronunciation Modeling of Swiss German Dialectal Speech for Automatic Speech Recognition
Michael Stadtschnitzer, Christoph Schmidt

Simulating ASR errors for training SLU systems
Edwin Simonnet, Sahar Ghannay, Nathalie Camelin, Yannick Estève

Evaluation of Feature-Space Speaker Adaptation for End-to-End Acoustic Models
Natalia Tomashenko, Yannick Estève

Creating New Language and Voice Components for the Updated MaryTTS Text-to-Speech Synthesis Platform
Ingmar Steiner, Sébastien Le Maguer

Speech Rate Calculations with Short Utterances: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task
Akira Hayakawa, Carl Vogel, Saturnino Luz, Nick Campbell

**P44 ● Summarisation**
Chair: Senja Pollak

Beyond Generic Summarization: A Multi-faceted Hierarchical Summarization Corpus of Large Heterogeneous Data
Christopher Tauchmann, Thomas Arnold, Andreas Hanselowski, Christian M. Meyer, Margot Mieskes

A New Annotated Portuguese/Spanish Corpus for the Multi-Sentence Compression Task
Elvys Linhares Pontes, Juan-Manuel Torres-Moreno, Stéphane Huet, Andréa carneiro Linhares

Live Blog Corpus for Summarization
Avinesh PVS, Maxime Peyrard, Christian M. Meyer

TSix: A Human-involved-creation Dataset for Tweet Summarization
Minh-Tien Nguyen, Dac Viet Lai, Huy-Tien Nguyen, Minh-Le Nguyen

A Workbench for Rapid Generation of Cross-Lingual Summaries
Nisarg Jhaveri, Manish Gupta, Vasudeva Varma

Annotation and Analysis of Extractive Summaries for the Kyutech Corpus
Takashi Yamamura, Kazutaka Shimada

A Repository of Corpora for Summarization
Franck Dernoncourt, Mohammad Ghassemi, Walter Chang

Auto-hMDS: Automatic Construction of a Large Heterogeneous Multilingual Multi-Document Summarization Corpus
Markus Zopf

PyrEval: An Automated Method for Summary Content Analysis
Yanjun Gao, Andrew Warner, Rebecca Passonneau
Mapping Texts to Scripts: An Entailment Study
Simon Ostermann, Hannah Seitz, Stefan Thater, Manfred Pinkal

Semantic Equivalence Detection: Are Interrogatives Harder than Declaratives?
João Rodrigues, Chakaveh Saedi, António Branco, João Silva

CEFR-based Lexical Simplification Dataset
Satoru Uchida, Shohei Takada, Yuki Arase
Detailed Conference Programme

11 May 2018

9.00-9.40  
**Keynote Speech**

Pascale Fung  
Empathetic Dialog Systems  
Chairperson: Joseph Mariani  
*Room: Tenzui*

9.40-9.45  
5-minute break

O29 • Language Resource Infrastructures

Chair: Takenobu Tokunaga  
*Room: Tenran*

9.45-10.05  
CLARIN: Towards FAIR and Responsible Data Science Using Language Resources  
Franciska De Jong, Bente Maegaard, Koenraad De Smedt, Darja Fišer, Dieter Van Uytvanck

10.05-10.25  
From ‘Solved Problems’ to New Challenges: A Report on LDC Activities  
Christopher Cieri, Mark Liberman, Stephanie Strassel, Denise DiPersio, Jonathan Wright, Andrea Mazzucchli

10.25-10.45  
New directions in ELRA activities  
Valérie Mapelli, Victoria Arranz, Hélène Mazo, Pawel Kamocki, Vladimir Popescu

10.45-11.05  
A Framework for Multi-Language Service Design with the Language Grid  
Donghui Lin, Yohei Murakami, Toru Ishida

11.05-11.25  
Language Technology for Multilingual Europe: An Analysis of a Large-Scale Survey regarding Challenges, Demands, Gaps and Needs  
Georg Rehm, Stefanie Hegele

O30 • Digital Humanities & Text Analytics

Chair: Thierry Declerck  
*Room: Tengyoku*

9.45-10.05  
Annotating High-Level Structures of Short Stories and Personal Anecdotes  
Boyang Li, Beth Cardier, Tong Wang, Florian Metze

10.05-10.25  
Discovering the Language of Wine Reviews: A Text Mining Account  
Els Lefever, Iris Hendrickx, Ilja Croijmans, Antal Van den Bosch, Asifa Majid
10.25-10.45  Toward An Epic Epigraph Graph  
*Francis Bond, Graham Matthews*

10.45-11.05  Delta vs. N-Gram Tracing: Evaluating the Robustness of Authorship Attribution Methods  
*Thomas Proisl, Stefan Evert, Fotis Jannidis, Christof Schöch, Leonard Konle, Steffen Pielström*

11.05-11.25  An Attribution Relations Corpus for Political News  
*Edward Newell, Drew Margolin, Derek Ruths*

**O31 ● Crowdsourcing & Collaborative Resource Construction**  
**Chair: Steve Cassidy**  
**Room: Tenju**

9.45-10.05  Face2Text: Collecting an Annotated Image Description Corpus for the Generation of Rich Face Descriptions  
*Albert Gatt, Marc Tanti, Adrian Muscat, Patrizia Paggio, Reuben A Farrugia, Claudia Borg, Kenneth Camilleri, Mike Rosner, Lonneke Van der Plas*

10.05-10.25  Adapting Serious Game for Fallacious Argumentation to German: Pitfalls, Insights, and Best Practices  
*Ivan Habernal, Patrick Pauli, Iryna Gurevych*

10.25-10.45  Crowdsourcing Regional Variation Data and Automatic Geolocalisation of Speakers of European French  
*Jean-Philippe Goldman, Yves Scherrer, Julie Glikman, Mathieu Avanzi, Christophe Benzitoun, Philippe Boula de Mareüil*

10.45-11.05  Improving Machine Translation of Educational Content via Crowdsourcing  
*Maximiliana Behnke, Antonio Valerio Miceli Barone, Rico Sennrich, Vilelmini Sosoni, Thanasis Naskos, Eirini Takoulidou, Maria Stasimioti, Menno Van Zaanen, Sheila Castilho, Federico Gaspari, Panayota Georgakopoulou, Valia Kordoni, Markus Egg, Katia Lida Kermanidis*

11.05-11.25  Grounding Gradable Adjectives through Crowdsourcing  
*Rebecca Sharp, Mithun Paul, Ajay Nagesh, Dane Bell, Mihai Surdeanu*

**O32 ● Less-Resourced Languages Speech & Multimodal Corpora**  
**Chair: Shyam Agrawal**  
**Room: Tenyo**

9.45-10.05  Evaluation Phonemic Transcription of Low-Resource Tonal Languages for Language Documentation  
*Oliver Adams, Trevor Cohn, Graham Neubig, Hilaria Cruz, Steven Bird, Alexis MICHAUD*

10.05-10.25  A Very Low Resource Language Speech Corpus for Computational Language Documentation Experiments  
*Pierre Godard, Gilles Adda, Martine Adda-Decker, Juan Benjumea, Laurent Besacier, Jamison Cooper-Leavitt, Guy-Noel Kouarata, Lori Lamel, Hélène Maynard, Markus Mueller, Annie Rialland, Sebastian Stueker, François Yvon, Marcely Zanon Boito*
10.25-10.45 Chahta Anumpa: A multimodal corpus of the Choctaw Language
Jacqueline Brixey, Eli Pincus, Ron Artstein

10.45-11.05 BULBasaa: A Bilingual Basaa-French Speech Corpus for the Evaluation of Language Documentation Tools
Fatima Hamlaoui, Emmanuel-Moselly Makasso, Markus Müller, Jonas Engelmann, Gilles Adda, Alex Waibel, Sebastian Stüker

11.05-11.25 Researching Less-Resourced Languages – the DigiSami Corpus
Kristiina Jokinen

9:45 - 11:25 Poster Sessions – Poster Area 2

P46 ● Dialects
Chair: Claudia Soria

Creating dialect sub-corpora by clustering: a case in Japanese for an adaptive method
Yo Sato, Kevin Heffernan

A Fast and Flexible Webinterface for Dialect Research in the Low Countries
Roel Van Hout, Nicoline Van der Sijs, Erwin Komen, Henk Van den Heuvel

Arabic Dialect Identification in the Context of Bivalency and Code-Switching
Mahmoud El-Haj, Paul Rayson, Mariam Aboelezz

Unified Guidelines and Resources for Arabic Dialect Orthography
Nizar Habash, Fadhil Eryani, Salam Khalifa, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghouani, Houda Bouamor, Nasser Zalmout, Sara Hassan, Faisal Al shargi, Sakhar Alkhereyf, Basma Abdulkareem, Ramy Eskander, Mohammad Salameh, Hind Saddiki

Automatic Identification of Maghreb Dialects Using a Dictionary-Based Approach
Houda SAADANE, Hosni Seffih, Christian Fluhr, Khalid Choukri, Nasredine SEMMAR

Shami: A Corpus of Levantine Arabic Dialects
Chatrine Qwaider, Motaz Saad, Stergos Chatzikyriakidis, Simon Dobnik

You Tweet What You Speak: A City-Level Dataset of Arabic Dialects
Muhammad Abdul-Mageed, Hassan Alhuzali, Mohamed Elaraby

Visualizing the "Dictionary of Regionalisms of France" (DRF)
Ada Wan

DART: A Large Dataset of Dialectal Arabic Tweets
Israa Alsarsour, Esraa Mohamed, Reem Suwaileh, Tamer Elsayed

Classification of Closely Related Sub-dialects of Arabic Using Support-Vector Machines
Samantha Wray

P47 ● Document Classification, Text Categorisation (2)
Chair: Piotr Pęzik

Page Stream Segmentation with Convolutional Neural Nets Combining Textual and Visual Features
Gregor Wiedemann, Gerhard Heyer
Automating Document Discovery in the Systematic Review Process: How to Use Chaff to Extract Wheat  
*Christopher Norman, Mariska Leeflang, Pierre Zweigenbaum, Aurélie Névéol*

Two Multilingual Corpora Extracted from the Tenders Electronic Daily for Machine Learning and Machine Translation Applications.  
*Oussama Ahmia, Nicolas Béchet, Pierre-François Marteau*

Using Adversarial Examples in Natural Language Processing  
*Petr Bělohlávek, Ondřej Plátek, Zdeněk Žabokrtský, Milan Straka*

Modeling Trolling in Social Media Conversations  
*Luis Gerardo Mojica de la Vega, Vincent Ng*

**P48 ● Information Extraction, Information Retrieval, Text Analytics (3)**

Chair: Patrick Drouin

Automatic Annotation of Semantic Term Types in the Complete ACL Anthology Reference Corpus  
*Anne-Kathrin Schumann, Héctor Martínez Alonso*

Annotated Corpus of Scientific Conference's Homepages for Information Extraction  
*Piotr Andruszkiewicz, Rafal Hazan*

Improving Unsupervised Keyphrase Extraction using Background Knowledge  
*Yang Yu, Vincent Ng*

WikiDragon: A Java Framework For Diachronic Content And Network Analysis Of MediaWikis  
*Rüdiger Gleim, Alexander Mehler, Sung Y. Song*

Studying Muslim Stereotyping through Microportrait Extraction  
*Antske Fokkens, Nel Ruigrok, Camiel Beukeboom, Gagstein Sarah, Wouter Van Attveldt*

Analyzing the Quality of Counseling Conversations: the Tell-Tale Signs of High-quality Counseling  
*Verónica Pérez-Rosas, Xuetong Sun, Christy Li, Yuchen Wang, Kenneth Resnicow, Rada Mihalcea*

Interpersonal Relationship Labels for the CALLHOME Corpus  
*Denys Katerenchuk, David Guy Brizan, Andrew Rosenberg*

Text Mining for History: first steps on building a large dataset  
*Suemi Higuchi, Cláudia Freitas, Bruno Cuconato, Alexandre Rademaker*

Building Evaluation Datasets for Cultural Microblog Retrieval  
*Lorraine Goeuriot, Josiane Mothe, Philippe Mulhem, Eric SanJuan*

**P49 ● Machine Translation, SpeechToSpeech Translation (2)**

Chair: Mona Diab

Training and Adapting Multilingual NMT for Less-resourced and Morphologically Rich Languages  
*Matīss Rikters, Mārcis Pinnis, Rihards Krišlauks*

Cross-lingual Terminology Extraction for Translation Quality Estimation  
*Yu Yuan, Yuze Gao, Yue Zhang, Serge Sharoff*
Machine Translation of Low-Resource Spoken Dialects: Strategies for Normalizing Swiss German
Pierre-Edouard Honnet, Andrei Popescu-Belis, Claudiu Musat, Michael Baeriswyl
Improving domain-specific SMT for low-resourced languages using data from different domains
Fathima Farhath, Pranavan Theivendiram, Surangika Ranathunga, Sanath Jayasena, Gihan Dias
Discovering Parallel Language Resources for Training MT Engines
Vassilis Papavassiliou, Prokopis Prokopidis, Stelios Piperidis
A fine-grained error analysis of NMT, SMT and RBMT output for English-to-Dutch
Laura Van Brussel, Arda Tezcan, Lieve Macken
Collection and Analysis of Code-switch Egyptian Arabic-English Speech Corpus
Injy Hamed, Mohamed Elmahdy, Slim Abdennadher
Multimodal Lexical Translation
Chiraag Lala, Lucia Specia
Literality and cognitive effort: Japanese and Spanish
Isabel Lacruz, Michael Carl, Masaru Yamada
Evaluation of Machine Translation Performance Across Multiple Genres and Languages
Marlies Van der Wees, Arianna Bisazza, Christof Monz
A Multilingual Dataset for Evaluating Parallel Sentence Extraction from Comparable Corpora
Pierre Zweigenbaum, Serge Sharoff, Reinhard Rapp
Manual vs Automatic Bitext Extraction
Aibek Makazhanov, Bagdat Myrzakhmetov, Zhenisbek Assylbekov

P50 ● Morphology (2)
Chair: Amália Mendes
A Morphologically Annotated Corpus of Emirati Arabic
Salam Khalifa, Nizar Habash, Fadhl Eryani, Ossama Obeid, Dana Abdulrahim, Meera Al Kaabi
CoNLL-UL: Universal Morphological Lattices for Universal Dependency Parsing
Amir More, Özlem Çetinoğlu, Çağrı Çöltekin, Nizar Habash, Benoît Sagot, Djamé Seddah, Dima Taji, Reut Tsarfaty
Manually Annotated Corpus of Polish Texts Published between 1830 and 1918
Witold Kieraś, Marcin Woliński
Evaluating Inflectional Complexity Crosslinguistically: a Processing Perspective
Claudia Marzi, Marcello Ferro, Ouafae Nahli, Patrizia Belik, Stavros Bompolas, Vito Pirrelli
Parser combinators for Tigrinya and Oromo morphology
Patrick Littell, Tom McCoy, Na-Rae Han, Shruti Rijhwani, Zaid Sheikh, David R. Mortensen, Teruko Mitamura, Lori Levin
Massively Translingual Compound Analysis and Translation Discovery
Winston Wu, David Yarowsky
Building a Morphological Treebank for German from a Linguistic Database
Petra Steiner, Josef Ruppenhofer
P51 ● Multilinguality
Chair: Pavel Straňák
Baselines and Test Data for Cross-Lingual Inference
Željko Agić, Natalie Schluter
CATS: A Tool for Customized Alignment of Text Simplification Corpora
Sanja Štajner, Marc Franco-Salvador, Paolo Rosso, Simone Paolo Ponzetto
KIT-Multi: A Translation-Oriented Multilingual Embedding Corpus
Thanh-Le Ha, Jan Niehues, Matthias Sperber, Ngoc Quan Pham, Alexander Waibel
Multi-lingual Argumentative Corpora in English, Turkish, Greek, Albanian, Croatian, Serbian, Macedonian, Bulgarian, Romanian and Arabic
Alfred Sliwa, Yuan Man, Ruishen Liu, Niravkumar Borad, Seyyedeh Ziyaei, Mina Ghobadi, Firas Sabbah, Ahmet Aker
SemR-11: A Multi-Lingual Gold-Standard for Semantic Similarity and Relatedness for Eleven Languages
Siamak Barzegar, Brian Davis, Manel Zarrouk, Siegfried Handschuh, André Freitas

P52 ● Part-of-Speech Tagging
Chair: Andreas Witt
Corpora with Part-of-Speech Annotations for Three Regional Languages of France: Alsatian, Occitan and Picard
Delphine Bernhard, Anne-Laure Ligazat, Fanny Martin, Myriam Bras, Pierre Magistry, Marianne Vergez-Couret, Lucie Steiblé, Pascale Erhart, Nabil Hathout, Dominique Huck, Christophe Rey, Philippe Reynés, Sophie Rosset, Jean Sibille, Thomas Lavergne
Part-of-Speech Tagging for Arabic Gulf Dialect Using Bi-LSTM
Randah Alharbi, Walid Magdy, Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak
Web-based Annotation Tool for Inflectional Language Resources
Abdulrahman Alosaimy, Eric Atwell
HiNTS: A Tagset for Middle Low German
Fabian Barteld, Sarah Ihden, Katharina Dreessen, Ingrid Schröder
Leveraging Lexical Resources and Constraint Grammar for Rule-Based Part-of-Speech Tagging in Welsh
Steven Neale, Kevin Donnelly, Gareth Watkins, Dawn Knight
Graph Based Semi-Supervised Learning Approach for Tamil POS tagging
Mukanarangan Thayaparan, Surangika Ranathunga, Uthayasanker Thayasivam

11.25-11.45 Coffee Break

O33 ● Lexicon
Chair: Simon Krek
Room: Tenran

11.45-12.05 The MADAR Arabic Dialect Corpus and Lexicon
Houda Bouamor, Nizar Habash, Mohammad Salameh, Wajdi Zaghouani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhil Eryani, Alexander Erdmann, Kemal Oflazer
<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.05-12.25</td>
<td>Designing a Collaborative Process to Create Bilingual Dictionaries of Indonesian Ethnic Languages</td>
<td>Arbi Haza Nasution, Yohei Murakami, Toru Ishida</td>
</tr>
<tr>
<td>12.25-12.45</td>
<td>Constructing a Lexicon of Relational Nouns</td>
<td>Edward Newell, Jackie Chi Kit Cheung</td>
</tr>
<tr>
<td>12.45-13.05</td>
<td>Creating Large-Scale Multilingual Cognate Tables</td>
<td>Winston Wu, David Yarowsky</td>
</tr>
<tr>
<td>13.05-13.25</td>
<td>Lexical Profiling of Environmental Corpora</td>
<td>Patrick Drouin, Marie-Claude L'Homme, Benoît Robichaud</td>
</tr>
</tbody>
</table>

**O34 ● Knowledge Discovery**

Chair: German Rigau

**Room: Tengyoku**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.45-12.05</td>
<td>Linking, Searching, and Visualizing Entities in Wikipedia</td>
<td>Marcus Klang, Pierre Nugues</td>
</tr>
<tr>
<td>12.05-12.25</td>
<td>Learning to Map Natural Language Statements into Knowledge Base Representations for Knowledge Base Construction</td>
<td>Chin-Ho Lin, Hen-Hsen Huang, Hsin-Hsi Chen</td>
</tr>
<tr>
<td>12.25-12.45</td>
<td>Building a Knowledge Graph from Natural Language Definitions for Interpretable Text Entailment Recognition</td>
<td>Vivian Silva, André Freitas, Siegfried Handschuh</td>
</tr>
<tr>
<td>12.45-13.05</td>
<td>Combining rule-based and embedding-based approaches to normalize textual entities with an ontology</td>
<td>Arnaud Ferré, Louise Deléger, Pierre Zweigenbaum, Claire Nédellec</td>
</tr>
<tr>
<td>13.05-13.25</td>
<td>T-REx: A Large Scale Alignment of Natural Language with Knowledge Base Triples</td>
<td>Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, Elena Simperl</td>
</tr>
</tbody>
</table>

**O35 ● Multilingual Corpora & Machine Translation**

Chair: Eva Hajičová

**Room: Tenju**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.45-12.05</td>
<td>Multilingual Parallel Corpus for Global Communication Plan</td>
<td>Kenji Imamura, Eiichiro Sumita</td>
</tr>
<tr>
<td>12.05-12.25</td>
<td>A Large Parallel Corpus of Full-Text Scientific Articles</td>
<td>Felipe Soares, Viviane Moreira, Karin Becker</td>
</tr>
<tr>
<td>12.25-12.45</td>
<td>NegPar: A parallel corpus annotated for negation</td>
<td>Qianchu Liu, Federico Fancellu, Bonnie Webber</td>
</tr>
<tr>
<td>12.45-13.05</td>
<td>The IIT Bombay English-Hindi Parallel Corpus</td>
<td>Anoop Kunchukuttan, Pratik Mehta, Pushpak Bhattacharyya</td>
</tr>
<tr>
<td>13.05-13.25</td>
<td>Extracting an English-Persian Parallel Corpus from Comparable Corpora</td>
<td>Akbar Karimi, Ebrahim Ansari, Bahram Sadeghi Bigham</td>
</tr>
</tbody>
</table>
O36 ● Corpus Creation, Use & Evaluation (2)

Chair: Satoshi Nakamura

Room: Tenyo

11.45-12.05 Learning Word Vectors for 157 Languages
   *Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, Tomas Mikolov*

12.05-12.25 SumeCzech: Large Czech News-Based Summarization Dataset
   *Milan Straka, Nikita Mediankin, Tom Kocmi, Zdeněk Žabokrtský, Vojtěch Hudeček, Jan Hajic*

12.25-12.45 A Diachronic Corpus for Literary Style Analysis
   *Carmen Klaussner, Carl Vogel*

12.45-13.05 Text Simplification from Professionally Produced Corpora
   *Carolina Scarton, Gustavo Paetzold, Lucia Specia*

13.05-13.25 Intertextual Correspondence for Integrating Corpora
   *Jacky Visser, Rory Duthie, John Lawrence, Chris Reed*

11:45 - 13:25 Poster Sessions – Poster Area 1

P53 ● Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction (3)

Chair: Kalika Bali

What Causes the Differences in Communication Styles? A Multicultural Study on Directness and Elaborateness
   *Juliana Miehle, Wolfgang Minker, Stefan Ultes*

FARMI: A FrAmework for Recording Multi-Modal Interactions
   *Patrik Jonell, Mattias Bystedt, Per Fallgren, Dimosthenis Kontogiorgos, José Lopes, Zofia Malisz, Samuel Mascarenhas, Catharine Oertel, Eran Raveh, Todd Shore*

Creating Large-Scale Argumentation Structures for Dialogue Systems
   *Kazuki Sakai, Akari Inago, Ryuichiro Higashinaka, Yuichiro Yoshikawa, Hiroshi Ishiguro, Junji Tomita*

Exploring Conversational Language Generation for Rich Content about Hotels
   *Marilyn Walker, Albry Smither, Shereen Oraby, Vrindavan Harrison, Hadar Shemtov*

Identification of Personal Information Shared in Chat-Oriented Dialogue
   *Sarah Fillwock, David Traum*

A Vietnamese Dialog Act Corpus Based on ISO 24617-2 Standard
   *Thi Lan Ngo, Pham Khac Linh, Takeda Hideaki*

Annotating Reflections for Health Behavior Change Therapy
   *Nishitha Guntakandla, Nielsen Rodney*

P54 ● Discourse Annotation, Representation and Processing (2)

Chair: Bruno Cartoni

Annotating Attribution Relations in Arabic
   *Amal Alsaif, Tasniem Alyahya, Madawi Alotaibi, Huda Almuzaini, Abeer Algahtani*
The ADELE Corpus of Dyadic Social Text Conversations: Dialog Act Annotation with ISO 24617-2

Emer Gilmartin, Christian Saam, Brendan Spillane, Maria O'Reilly, Ketong Su, Arturo Calvo, Loredana Cerrato, Killian Levacher, Nick Campbell, Vincent Wade

An Assessment of Explicit Inter- and Intra-sentential Discourse Connectives in Turkish Discourse Bank

Deniz Zeyrek, Murathan Kurfalı

Compilation of Corpora for the Study of the Information Structure—Prosody Interface

Alicia Burga, Monica Domínguez, Mireia Farrús, Leo Wanner

Preliminary Analysis of Embodied Interactions between Science Communicators and Visitors Based on a Multimodal Corpus of Japanese Conversations in a Science Museum

Rui Sakaida, Ryosaku Makino, Mayumi Bono

Improving Crowdsourcing-Based Annotation of Japanese Discourse Relations

Yudai Kishimoto, Shinnosuke Sawada, Yugo Murawaki, Daisuke Kawahara, Sadao Kurohashi

Persian Discourse Treebank and Coreference Corpus

Azadeh Mirzaei, Pegah Safari

Automatic Labeling of Problem-Solving Dialogues for Computational Microgenetic Learning Analytics

Yuanliang Meng, Anna Rumshisky, Florence Sullivan

Increasing Argument Annotation Reproducibility by Using Inter-annotator Agreement to Improve Guidelines

Milagro Teruel, Cristian Cardellino, Fernando Cardellino, Laura Alonso Alemany, Serena Villata

P55 • Language Acquisition & CALL (2)

Chair: Serge Sharoff

Semi-Supervised Clustering for Short Answer Scoring

Andrea Horbach, Manfred Pinkal

Analyzing Vocabulary Commonality Index Using Large-scaled Database of Child Language Development

Yan Cao, Yasuhiro Minami, Yuko Okumura, Tessei Kobayashi

The ICOn Corpus of Academic Written Italian (L1 and L2)

Mirko Tavosanis, Federica Cominetti

Revita: a Language-learning Platform at the Intersection of ITS and CALL

Anisia Katinskaia, Javad Nouri, Roman Yangarber

The Distribution and Prosodic Realization of Verb Forms in German Infant-Directed Speech

Bettina Braun, Katharina Zahner

Cross-linguistically Small World Networks are Ubiquitous in Child-directed Speech

Steven Moran, Danica Pajović, Sabine Stoll

L1-L2 Parallel Treebank of Learner Chinese: Overused and Underused Syntactic Structures

Keying Li, John Lee
The Use of Text Alignment in Semi-Automatic Error Analysis: Use Case in the Development of the Corpus of the Latvian Language Learners
Roberts Darģis, Ilze Auziņa, Kristīne Levāne-Petrova

Error annotation in a Learner Corpus of Portuguese
Iria Del Río Gayo, Amália Mendes

An SLA Corpus Annotated with Pedagogically Relevant Grammatical Structures
Leonardo Zilio, Rodrigo Wilkens, Cédrick Fairon

**P56 ● Less-Resourced/Endangered Languages (2)**
Chair: Sonja Bosch

Portable Spelling Corrector for a Less-Resourced Language: Amharic
Andargachew Mekonnen Gezmu, Andreas Nürnberger, Binyam Ephrem Seyoum

A Speaking Atlas of the Regional Languages of France
Philippe Boula de Mareüil, Albert Rilliard, Frédéric Vernier

Towards Language Technology for Mi’kmaq
Anant Maheshwari, Leo Bousscarat, Paul Cook

Pronunciation Dictionaries for the Alsatian Dialects to Analyze Spelling and Phonetic Variation
Lucie Steiblé, Delphine Bernhard

ChAnot: An Intelligent Annotation Tool for Indigenous and Highly Agglutinative Languages in Peru
Rodolfo Mercado, José Pereira, Marco Antonio Sobrevilla Cabezudo, Arturo Oncevay

The DLDP Survey on Digital Use and Usability of EU Regional and Minority Languages
Claudia Soria, Valeria Quochi, Irene Russo

ASR for Documenting Acutely Under-Resourced Indigenous Languages
Robert Jimerson, Emily Prud’hommeaux

**P57 ● Opinion Mining / Sentiment Analysis (3)**
Chair: Rodrigo Agerri

Building a Sentiment Corpus of Tweets in Brazilian Portuguese
Henrico Brum, Maria das Graças Volpe Nunes

'Aye' or 'No'? Speech-level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts
Gavin Abercrombie, Riza Batista-Navarro

Scalable Visualisation of Sentiment and Stance
Jon Chamberlain, Udo Kruschwitz, Orland Hoeber

NoReC: The Norwegian Review Corpus
Erik Velldal, Lilja Øverlid, Eivind Alexander Bergem, Cathrine Stadsnes, Samia Touileb, Fredrik Jørgensen

SenSALDO: Creating a Sentiment Lexicon for Swedish
Jacobo Rouces, Nina Tahmasebi, Lars Borin, Stian Rødven Eide
Corpus Building and Evaluation of Aspect-based Opinion Summaries from Tweets in Spanish
Daniel Peñaloza, JuanJosé Tenorio, Rodrigo López, Héctor Gomez, Arturo Oncevay, Marco Antonio Sobrevilla Cabezudo

Application and Analysis of a Multi-layered Scheme for Irony on the Italian Twitter Corpus TWITTIRÒ
Alessandra Teresa Cignarella, Cristina Bosco, Viviana Patti, Mirko Lai

Classifier-based Polarity Propagation in a WordNet
Jan Kocor, Arkadiusz Janz, Maciej Piasecki

P58 ● Sign Language
Chair: Thomas Hanke

SMILE Swiss German Sign Language Dataset
Sarah Ebling, Necati Cihan Camgöz, Penny Boyes Braem, Katja Tissi, Sandra Sidler-Miserrez, Stephanie Stoll, Simon Hadfield, Tobias Haug, Richard Bowden, Sandrine Tornay, Marzieh Razavi, Mathew Magimai-Doss

IPSL: A Database of Iconicity Patterns in Sign Languages. Creation and Use
Vadim Kimmelman, Anna Klezovich, George Moroz

Sign Languages and the Online World Online Dictionaries & Lexicostatistics
Shi Yu, Carlo Geraci, Natasha Abner

Elicitation protocol and material for a corpus of long prepared monologues in Sign Language
Michael Filhol, Mohamed Nassime Hadjadj

Deep JSLC: A Multimodal Corpus Collection for Data-driven Generation of Japanese Sign Language Expressions
Heike Brock, Kazuhiro Nakadai

Modeling French Sign Language: a proposal for a semantically compositional system
Mohamed Nassime Hadjadj, Michael Filhol, Annelies Braffort

P59 ● Speech Resource/Database (2)
Chair: Christoph Draxler

Construction of the Corpus of Everyday Japanese Conversation: An Interim Report
Hanae Koiso, Yasuharu Den, Yuriko Iseki, Wakako Kashino, Yoshiko Kawabata, Ken’ya Nishikawa, Yayoi Tanaka, Yasuyuki Usuda

Carcinologic Speech Severity Index Project: A Database of Speech Disorder Productions to Assess Quality of Life Related to Speech After Cancer
Corine Astésano, Mathieu Balaguer, Jérôme Farinas, Corinne Fredouille, Pascal Gaillard, Alain Ghio, Imed Laaridh, Muriel Lalain, Benoît Lepage, Julie Mauclair, Olivier Nocaudie, Julien Pinquier, Oriol Pont, Gilles Pouchoulin, Michèle Puech, Danièle Robert, Etienne Sicard, Virginie Woïsard

Parallel Corpora in Mboshi (Bantu C25, Congo-Brazzaville)
Annie Rialland, Martine Adda-Decker, Guy-Noël Kouarata, Gilles Adda, Laurent Besacier, Lori Lamel, Elodie Gauthier, Pierre Godard, Jamison Cooper-Leavitt

A Multimodal Corpus of Expert Gaze and Behavior during Phonetic Segmentation Tasks
Arif Khan, Ingmar Steiner, Yusuke Sugano, Andreas Bulling, Ross Macdonald
Statistical Analysis of Missing Translation in Simultaneous Interpretation Using A Large-scale Bilingual Speech Corpus
Zhongxi Cai, Koichiro Ryu, Shigeki Matsubara

SynPaFlex-Corpus: An Expressive French Audiobooks Corpus dedicated to expressive speech synthesis.
Aghilas Sini, Damien Lolive, Gaëlle Vidal, Marie Tahon, Élisabeth Delais-Roussarie

Increasing the Accessibility of Time-Aligned Speech Corpora with Spokes Mix
Piotr Pęzik

The MonPaGe_HA Database for the Documentation of Spoken French Throughout Adulthood
Cécile Fougeron, Veronique Delvaux, Lucie Ménard, Marina Laganaro

Bringing Order to Chaos: A Non-Sequential Approach for Browsing Large Sets of Found Audio Data
Per Fallgren, Zofia Malisz, Jens Edlund

CoLoSS: Cognitive Load Corpus with Speech and Performance Data from a Symbol-Digit Dual-Task
Robert Herms, Maria Wirzberger, Maximilian Eibl, Günter Daniel Rey

VAST: A Corpus of Video Annotation for Speech Technologies
Jennifer Tracey, Stephanie Strassel

13.25-14.45 Lunch Break

O37 ● Anaphora & Coreference
Chair: Claire Bonial

Room: Tenran

14.45-15.05 A Gold Anaphora Annotation Layer on an Eye Movement Corpus
Olga Seminck, Pascal Amsili

15.05-15.25 Annotating Zero Anaphora for Question Answering
Yoshihiko Asao, Ryu Iida, Kentaro Torisawa

15.25-15.45 Building Named Entity Recognition Taggers via Parallel Corpora
Rodrigo Agerri, Yiling Chung, Itziar Aldabe, Nora Aranberri, Gorka Labaka, German Rigau

15.45-16.05 Cross-Document, Cross-Language Event Coreference Annotation Using Event Hoppers
Zhiyi Song, Ann Bies, Justin Mott, Xuansong Li, Stephanie Strassel, Christopher Caruso

O38 ● Corpus for Document Classification
Chair: Franciska de Jong

Room: Tengyoku

14.45-15.05 TAP-DLND 1.0 : A Corpus for Document Level Novelty Detection
Tirthankar Ghosal, Amitra Salam, Swati Tiwary, Asif Ekbal, Pushpak Bhattacharyya

15.05-15.25 A Corpus for Multilingual Document Classification in Eight Languages
Holger Schwenk, Xian Li
15.25-15.45 Analyzing Citation-Distance Networks for Evaluating Publication Impact  
*Drahomira Herrmannova, Petr Knoth, Robert Patton*

15.45-16.05 Annotating Educational Questions for Student Response Analysis  
*Andreea Godea, Rodney Nielsen*

**O39: Knowledge Discovery & Evaluation (2)**
Chair: Yuji Matsumoto

*Room: Tenju*

14.45-15.05 Incorporating Global Contexts into Sentence Embedding for Relational Extraction at the Paragraph Level with Distant Supervision  
*Eun-kyung Kim, Key-Sun Choi*

15.05-15.25 MCScript: A Novel Dataset for Assessing Machine Comprehension Using Script Knowledge  
*Simon Ostermann, Ashutosh Modi, Michael Roth, Stefan Thater, Manfred Pinkal*

15.25-15.45 A Neural Network Based Model for Loanword Identification in Uyghur  
*Chenggang Mi, Yating Yang, Lei Wang, Xi Zhou, Tonghai Jiang*

15.45-16.05 Revisiting Distant Supervision for Relation Extraction  
*Tingsong Jiang, Jing Liu, Chin-Yew Lin, Zhifang Sui*

**O40 ● Multimodal & Written Corpora & Tools**
Chair: Nancy Ide

*Room: Tenyo*

14.45-15.05 Incorporating Contextual Information for Language-Independent, Dynamic Disambiguation Tasks  
*Tobias Staron, Özge Alacam, Wolfgang Menzel*

15.05-15.25 Overcoming the Long Tail Problem: A Case Study on CO2-Footprint Estimation of Recipes using Information Retrieval  
*Melanie Geiger, Martin Braschler*

*Motoki Yatsu, Kenji Araki*

15.45-16.05 A vision-grounded dataset for predicting typical locations for verbs  
*Nelson Mukuze, Anna Rohrbach, Vera Demberg, Bernt Schiele*

14:45 - 16:05

**Poster Sessions – Poster Area 2**

**P60 ● Corpus Creation, Annotation, Use (2)**
Chair: Beatrice Daille

*Edit me: A Corpus and a Framework for Understanding Natural Language Image Editing  
*Ramesh Manuvinakurike, Jacqueline Brixey, Trung Bui, Walter Chang, Doo Soon Kim, Ron Artstein, Kallirroi Georgila*
Enriching a Lexicon of Discourse Connectives with Corpus-based Data
Anna Feltracco, Elisabetta Jezek, Bernardo Magnini

SimPA: A Sentence-Level Simplification Corpus for the Public Administration Domain
Carolina Scarton, Gustavo Paetzold, Lucia Specia

The brWaC Corpus: A New Open Resource for Brazilian Portuguese
Jorge Alberto Wagner Filho, Rodrigo Willems, Marco Idiart, Aline Villavicencio

Czech Text Document Corpus v 2.0
Pavel Kral, Ladislav Lenc

Corpora of Typical Sentences
Lydia Müller, Uwe Quasthoff, Maciej Sumalvico

The German Reference Corpus DeReKo: New Developments – New Opportunities
Marc Kupietz, Harald Lüngen, Pawel Kamocki, Andreas Witt

Risamálheild: A Very Large Icelandic Text Corpus
Steínþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson, Jon Gudnason

P61 • Lexicon (3)
Chair: John McCrae

TriMED: A Multilingual Terminological Database
Federica Vezzani, Giorgio Maria Di Nunzio, Geneviève Henrot

Preparation and Usage of Xhosa Lexicographical Data for a Multilingual, Federated Environment
Sonja Bosch, Thomas Eckart, Bettina Klimek, Dirk Goldhahn, Uwe Quasthoff

A Lexicon of Discourse Markers for Portuguese – LDM-PT
Amália Mendes, Iria Del Río Gayo, Manfred Stede, Felix Dombek

One Language to rule them all: modelling Morphological Patterns in a Large Scale Italian Lexicon with SWRL
Fahad Khan, Andrea Bellandi, Francesca Frontini, Monica Monachini

Metaphor Suggestions based on a Semantic Metaphor Repository
Gerard De Melo

The Linguistic Category Model in Polish (LCM-PL)
Aleksander Wawer, Justyna Sarzyńska

WordNet-Shp: Towards the Building of a Lexical Database for a Peruvian Minority Language
Diego Maguiña Valencia, Arturo Oncevay, Marco Antonio Sobrevilla Cabezudo

Retrieving Information from the French Lexical Network in RDF/OWL Format
Alexsandro Fonseca, Fatiha Sadat, François Lareau
P62 ● Named Entity Recognition
Chair: Gil Francopoulo

Transforming Wikipedia into a Large-Scale Fine-Grained Entity Type Corpus
Abbas Ghaddar, Philippe Langlais

Error Analysis of Uyghur Name Tagging: Language-specific Techniques and Remaining Challenges
Halidanmu Abudukelimu, Aduodukelimu Abulizi, Boliang Zhang, Xiaoman Pan, Di Lu, Heng Ji, Yang Liu

BiLSTM-CRF for Persian Named-Entity Recognition ArmanPersoNERCorpus: the First Entity-Annotated Persian Dataset
Hanieh Poostchi, Ehsan Zare Borzeshi, Massimo Piccardi

Data Anonymization for Requirements Quality Analysis: a Reproducible Automatic Error Detection Task
Juyeon Kang, Jungyeul Park

A German Corpus for Fine-Grained Named Entity Recognition and Relation Extraction of Traffic and Industry Events
Martin Schiersch, Veselina Mironova, Maximilian Schmitt, Philippe Thomas, Aleksandra Gabryszaq, Leonhard Hennig

A Corpus Study and Annotation Schema for Named Entity Recognition and Relation Extraction of Business Products
Saskia Schön, Veselina Mironova, Aleksandra Gabryszaq, Leonhard Hennig

Portuguese Named Entity Recognition using Conditional Random Fields and Local Grammars
Juliana Pirovani, Elias Oliveira

M-CNER: A Corpus for Chinese Named Entity Recognition in Multi-Domains
Qi Lu, YaoSheng Yang, Zhenghua Li, Wenliang Chen, Min Zhang

SlugNERDS: A Named Entity Recognition Tool for Open Domain Dialogue Systems
Kevin Bowden, Jiaqi Wu, Shereen Oraby, Amita Misra, Marilyn Walker

Transfer Learning for Named-Entity Recognition with Neural Networks
Ji Young Lee, Franck Dernoncourt, Peter Szolovits

P63 ● Parsing, Syntax, Treebank (2)
Chair: Jan Hajič

ForFun 1.0: Prague Database of Forms and Functions -- An Invaluable Resource for Linguistic Research
Marie Mikulová, Eduard Bejček

The LIA Treebank of Spoken Norwegian Dialects
Lilja Øvrelid, Andre Kåsen, Kristin Hagen, Anders Nøklestad, Per Erik Solberg, Janne Bondi Johannessen

Errator: a Tool to Help Detect Annotation Errors in the Universal Dependencies Project
Guillaume Wisniewski
SandhiKosh: A Benchmark Corpus for Evaluating Sanskrit Sandhi Tools  
*Shubham Bhardwaj, Neelamadhav Gantayat, Nikhil Chaturvedi, Rahul Garg, Sumeet Agarwal*

Czech Legal Text Treebank 2.0  
*Vincent Kríž, Barbora Hladka*

Creation of a Balanced State-of-the-Art Multilayer Corpus for NLU  
*Normunds Gružitis, Lauma Pretkalnina, Baiba Saulite, Laura Rituma, Gunta Nespore-Berzkalne, Arturs Znotins, Peteris Paikens*

Test Sets for Chinese Nonlocal Dependency Parsing  
*Manjuan Duan, William Schuler*

Adding Syntactic Annotations to Flickr30k Entities Corpus for Multimodal Ambiguous Prepositional-Phrase Attachment Resolution  
*Sebastien Delecraz, Alexis Nasr, FREDERIC BECHET, Benoit Favre*

Analyzing Middle High German Syntax with RDF and SPARQL  
*Christian Chiarcos, Benjamin Kosmehl, Christian Fäth, Maria Sukhareva*

Cheating a Parser to Death: Data-driven Cross-Treebank Annotation Transfer  
*Djamé Seddah, Eric De La Clergerie, Benoît Sagot, Héctor Martínez Alonso, Marie Candito*

Universal Dependencies and Quantitative Typological Trends. A Case Study on Word Order  
*Chiara Alzetta, Felice Dell’Orletta, Simonetta Montemagni, Giulia Venturi*

**P64 ● Wordnets and Ontologies**  
**Chair:** Elena Montiel-Ponsoda

Undersampling Improves Hyponymy Prototypicality Learning  
*Koki Washio, Tsuneaki Kato*

Interoperability of Language-related Information: Mapping the BLL Thesaurus to Lexvo and Glottolog  
*Vanya Dimitrova, Christian Fäth, Christian Chiarcos, Heike Renner-Westermann, Frank Abromeit*

Browsing and Supporting Pluricentric Global Wordnet, or just your Wordnet of Interest  
*António Branco, Ruben Branco, Chakaveh Saedi, João Silva*

Cross-checking WordNet and SUMO Using Meronymy  
*Javier Alvez, Itziar Gonzalez-Dios, German Rigau*

Extended HowNet 2.0 – An Entity-Relation Common-Sense Representation Model  
*Wei-Yun Ma, Yueh-Yin Shih*

The Circumstantial Event Ontology (CEO) and ECB+/CEO: an Ontology and Corpus for Implicit Causal Relations between Events  
*Roxane Segers, Tommaso Caselli, Piek Vossen*

Profiling Medical Journal Articles Using a Gene Ontology Semantic Tagger  
*Mahmoud El-Haj, Paul Rayson, Scott Piao, Jo Knight*
Towards a Conversation-Analytic Taxonomy of Speech Overlap
Felix Gervits, Matthias Scheutz

Indian Language Wordnets and their Linkages with Princeton WordNet
Diptesh Kanojia, Kevin Patel, Pushpak Bhattacharyya

16.05-16.25  **Coffee Break**

16.25-17.25  **Antonio Zampolli Prize Talk**

17.25-18.15  **Closing Session**

19.30  **Gala Dinner**