

ALTA 2022

**Proceedings of the 20th Workshop of the  
Australasian Language Technology Association**



14–16 December 2022  
Flinders University  
Adelaide, Australia

## Sponsors

Platinum



**Australian Government**

**Defence**

Gold

# Google

Event Sponsors



**Flinders**  
UNIVERSITY

**MELBOURNE  
DATA ANALYTICS  
PLATFORM**



**CROSSING**  
French Australian Laboratory for Humans-Autonomous Agents Teaming

## Introduction

Welcome to The 20th Annual Workshop of the Australasian Language Technology Association and Flinders University.

ALTA is interested in all aspects of language technology, from speech to text, hearing to reading, phonology to morphology, syntax to semantics. Our program reflects this, as does our ongoing coordination with the Australian Document Computing Symposium, which will have a separate stream as well as joint sessions. Participants are welcome to join sessions across the two co-located events.

ALTA is interested in both language and technology, and these days learning is an important aspect of the technology and our keynote speakers will also dig deeper into the human side of language, learning and logic, the tradeoffs between neural learning and symbolic reasoning, as well as the ontological grounding of syntax, semantics and thought. Ed Hovy, from CMU and Melbourne University, will start the ball rolling with a discussion of the limitations of neural NLP and need for reasoning. Stephane Dufau, from CNRS and currently on sabbatical at the University of Queensland, will end the day with a look at how to understanding the process of reading will round off Thursday with a look at how computational models can help us understand the human processes involved in reading, while on Friday a double billing of Thora Tenbrink from Bangor and Barbara Tversky from Stanford will explore how language and thought and behaviour are situated in space and time, and the importance for joint spatial awareness in for properly grounding our understanding of language in humans and robots.

Turning to submitted papers, we had 40 submitted papers (short and long) for formal publication and presentation, as well as allowing for submission of shorter abstracts including work in progress, recently published work or half-baked ideas, for more informal presentation (without publication in the proceedings, and a special session after lunch on Friday). In addition, we once again offered a shared task (with papers from successful entrants published in the proceedings, with a special session before lunch on Friday). We accepted 10 (25%) of the submitted papers for oral presentation and a further 16 (44%) for poster presentation. We will also be presenting a best paper award and a best student paper award in the closing session on Friday afternoon.

After two years of online workshops due to COVID we have made a real effort to allow everyone to come together again in person - and expect close to a hundred in-person attendees across the combined ALTA/ADCS events. In addition, we are allowing on-line participation and will be operating in a hybrid mode with some speakers presenting remotely. For those attending in person, all meals and refreshments are provided and we have a special mentoring lunch for students and mentors as part of our doctoral symposium, mentoring program and tutorial day on Wednesday at our Victoria Square city campus, while the main ALTA workshop and ADCS symposium sessions will be at our Tonsley campus (a bus or train ride away) and our conference dinner will be held at the Tonsley Hotel on Thursday evening.

We would like to thank all the referees, committee members, local organizers and student helpers who have helped this event come together, and in particular we'd like to thank our Platinum Sponsor, the Defence Science and Technology Group, our Gold Sponsor Google, as well as the University of Melbourne, Flinders University, CNRS International Research Lab and CROSSING for their support of the event and its keynote speakers.

Welcome to Flinders, to Adelaide, to South Australia and Australia - our submissions have come from all over the world and we look forward to a rich and rewarding time together. We hope that this hybrid event will be a worthwhile and enjoyable experience for everyone.

David Powers, Jennifer Biggs and Pradeesh Parameswaran

Adelaide and Dunedin

December 2022

**Organisers:**

*Program Co-chairs:* David Powers, Jennifer Biggs and Pradeesh Parameswaran

*Shared Task Chair:* Diego Mollá

*ALTA Execs:* Maria Kim, Meladel Mistica, Sarvnaz Karimi, Massimo Piccardi, Afshin Rahimi, Diego Mollá

*Local Chairs:* Richard Leibbrandt, Paulo Santos, Mehwish Nasim, Tina Du, Joel Mackenzie, Maciek Rybinski

*Session Chairs:* Gabriela Ferraro, Jonathan Kummerfeld

**Program Committee:**

Abeed Sarkar, Afshin Rahimi, Antonio Jimeno, David Powers, Diego Molla, Fajri Koto, Hamed Hassanzadeh, Guido Zuccon, Hiyori Yoshikawa, Jennifer Biggs, Karin Verspoor, Massimo Piccardi, Mel Mistica, Maria Kim, Nitin Indurkha, Paulo Santos, Pradeesh Parameswaran, Richard Leibbrandt, Sarvnaz Karimi, Sunghwan Mac Kim, Timothy Baldwin, Veronica Liesaputra, Xiang Dai.

**Invited Speakers:**

Barbara Tversky, Stanford University

Eduard Hovy, University of Melbourne and Carnegie Mellon University

Thora Tenbrink, Bangor University

Stephane Dufau, Laboratoire de psychologie cognitive (CNRS)

## Invited Talks

### **Barbara Tversky: Mind in Motion: How Action Shapes Thought**

I will present a case that actions in space and with the things in it are the foundation of thought, not the entire edifice, but the foundation. To this end, I will bring evidence from neuroscience, from behavior, from language, and from gesture.

### **Eduard Hovy: On the complementarity of neural and symbolic approaches, and on how to transfer between them**

Today's neural NLP can do amazing things, leading some people to expect human-level performance soon. But it also fails spectacularly, in ways we find hard to predict and explain. Is perfection just a matter of doing additional neural architecture engineering and more-advanced training to overcome these problems, or are there deeper reasons for the failures? I argue that trying to understand the nature and reason for failures by couching the necessary operations in terms of symbolic reasoning is a good way to discover what neural networks will remain unable to do despite additional architecture engineering and training.

### **Thora Tenbrink: Beyond physical robots: How to achieve joint spatial reference with a smart environment**

Interacting with a smart environment involves joint understanding of where things and people are or where they should be. Face-to-face interaction between humans, or between humans and robots, implies clearly identifiable perspectives on the environment that can be used to establish such a joint understanding. A smart environment, in contrast, is ubiquitous and thus perspective-independent. In this talk I will review the implications of this situation in terms of the challenges for establishing joint spatial reference between humans and smart systems, and present a somewhat unconventional solution as an opportunity.

### **Stephane Dufau: How a reading brain works: insights from experimental studies and modelling**

Understanding how a human brain processes language in its written form has been at the heart of numerous research efforts over the last century, from the experimental works carried on in the first psychology labs to the use of modern computational models. In my talk, I will briefly review the research domain in an historical perspective and discuss the current concepts that help frame our understanding of our ability to read. I will argue that, in order to deeply represent the interaction between the core reading processes found in perception, attention, and language functions, reading is better investigated with a set of simple models rather than modelled with fully integrated neural networks. Whether computational or not, such simple models are built on basic principles like delta rule and random walks and are constrained by patterns of experimental results from both psycho- and neuro-linguistics. A series of research showcasing the method will be presented, with applications related to Natural Language Processing. More specifically, I will illustrate how text simplification has helped children with reading difficulties read better.

## PROGRAMME

### **14th December (Wednesday) Tutorial, Day 1**

---

- 09:00 REGISTRATION
- 09:30 - 11:00 Doctoral Symposium
- Improving Care in Clinical Dentistry with Natural Language Processing of Electronic Dental Records*  
Hanna Pethani
- Misinformation Detection via Transfer Learning*  
Lin Tian
- A Discourse-Analytic Approach to the Study of Information Disorders: How Online Communities Legitimate Social Bonds When Communing Around Misinformation and Disinformation*  
Olivia Inwood
- Learning to Adapt Neural Models with Limited Human Supervision in Natural Language Processing*  
Thuy-Trang Vu
- 11:00 - 11.30 MORNING TEA
- 11:30 - 12.30 Mentoring Session
- 12:30 - 14:00 LUNCH (AND MENTORING SESSION)
- 14:00 - 14:30 Tutorial (Part 1: Misinformation and Human Perception)  
Xiuzhen (Jenny) Zhang
- 14:30 - 15:30 Tutorial (Part 2: Misinformation and Detection)  
Jey Han Lau
- 15:30 - 16:00 AFTERNOON TEA
- 16:00 - 17:00 Tutorial (Part 3: Misinformation Mitigation)  
Xiuzhen (Jenny) Zhang
- 17:00 Informal Meet-n-Greet

### **15th December (Thursday) Day 2**

---

- 8:15 - 9:00 REGISTRATION
- 9:00 - 9:30 ALTA/ADCS OPENING

- 9:30 - 10:30 Keynote 1: Ed Hovy  
*On the complementarity of neural and symbolic approaches, and on how to transfer between them*
- 10:30 - 11:00 MORNING TEA
- 11:00 - 12:30 Session A – (Session Chair: Jonathan Kummerfeld)  
Presentations are 20 minutes.
- Using public domain resources and off-the-shelf tools to produce high-quality multimedia texts*  
Manny Rayner, Belinda Chiera and Cathy Chua  
*TCG-Event: Effective Task Conditioning for Generation-based Event Extraction*  
Fatemeh Shiri, Tongtong Wu, Yuanfang Li and Gholamreza Haffari  
*Complex Reading Comprehension Through Question Decomposition*  
Xiao-Yu Guo, Yuan-Fang Li and Gholamreza Haffari  
*Using Aspect-Based Sentiment Analysis to Classify Attitude-bearing Words*  
Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers
- 12:30 - 13:30 LUNCH
- 11:00 - 12:30 Session B (ADCS/ALTA) – (Session Chair: Sarvnaz Karimi)  
ADCS presentations are 20 minutes and ALTA presentations are 10 minutes.
- Investigating Language Use by Polarised Groups on Twitter: A Case Study of the Bushfires (ADCS)*  
Mehwish Nasim, Naeha Sharif, Pranav Bhandari, Derek Weber, Martin Wood, Lucia Falzon, and Yoshihisa Kashima  
*Robustness of Neural Rankers to Typos: A Comparative Study (ADCS)*  
Shengyao Zhuang, Xinyu Mao, and Guido Zuccon  
*Automatic Explanation Generation For Climate Science Claims (ALTA)*  
Rui Xing, Shraey Bhatia, Timothy Baldwin and Jey Han Lau  
*Improving Text-based Early Prediction by Distillation from Privileged Time-Series Text*  
Jinghui Liu, Daniel Capurro, Anthony Nguyen and Karin Verspoor
- 14:30 - 15:30 Poster Session
- 15:30 - 16:00 AFTERNOON TEA
- 16:00 - 17:00 Session C – (Session Chair: Jonathan Kummerfeld)  
ADCS presentations are 20 minutes and ALTA presentations are 10 minutes.
- Fine-tuning a Subtle Parsing Distinction Using a Probabilistic Decision Tree: the Case of Postnominal "that" in Noun Complement Clauses vs. Relative Clauses*  
Zineddine Tighidet and Nicolas Ballier  
*Robustness of Hybrid Models in Cross-domain Readability Assessment*



Ho Hung Lim, Tianyuan Cai, John S. Y. Lee and Meichun Liu

- 17:00 - 17:30 Keynote 2: Stephane Dufau  
*How a reading brain works: insights from experimental studies and modelling*
- 17:30 - 18:30 Poster Session
- 18:30 - late ALTA/ADCS DINNER @ TONSLEY HOTEL RESTAURANT

### 16th December (Friday) Day 3

---

- 09:00 - 10:00 Keynote 3: Thora Tenbrink  
*Beyond physical robots: How to achieve joint spatial reference with a smart environment*
- 10:00 - 10:30 Keynote 4: Barbara Tversky  
*Mind in Motion: How Action Shapes Thought*
- 10:30 - 11:00 Rolling Discussion
- 11:00 - 11:30 Morning Tea
- 11:30 - 12:15 Session D (Shared Task Challenge) – (Session Chair: Diego Molla)  
Presentations are 10 minutes
- Overview of the 2022 ALTA Shared task: PIBOSO sentence classification, 10 years later*  
Diego Molla
- Context-Aware Sentence Classification in Evidence-Based Medicine*  
Biaoyan Fang and Fajri Koto
- Enhancing the DeBERTa Transformers Model for Classifying Sentences from Biomedical Abstracts*  
Abdul Aziz, Md. Akram Hossain, and Abu Nowshed Chy
- Automatic Classification of Evidence Based Medicine Using Transformers*  
Necva Bolucu, Pinar Uskaner Hepsag
- 12:15 - 12:30 ALTA AGM
- 12:30 - 13:30 LUNCH
- 13:30 - 15:30 Session E (Abstracts) – (Session Chair: Gabriela Ferraro)  
Presentations are 20 minutes
- Verifying Urarina Language Phonemes With TensorFlow*  
Michael Dorin and Judith Dorin
- A Multi-Faceted Reward for Adversarial Attacks on Text Classifiers*  
Tom Roth, Inigo Jauregi Unanue, Alsharif Abuadbbba and Massimo Piccardi

*Probing of Quantitative Values in Abstractive Summarization Models*

Nathan White

*Zero-shot Slot Filling with Slot-Prefix Prompting and Attention Relationship Descriptor*

Qiaoyang Luo and Lingqiao Liu

*Writing Progress in Australian Schools: An Experimental Proof-of-concept Application*

Charbel El-Khaissi

15:30 - 16:00 AFTERNOON TEA

16:00 - 16:50 Session F (Best Papers) – (Session Chair: Jennifer Biggs)

Presentations are 20 minutes

*The Corpus of Australian and New Zealand Spoken English: A new resource of naturalistic speech transcripts*

Steven Coats

*The Role of Context in Vaccine Stance Prediction for Twitter Users*

Aleney Khoo, Maciej Rybinski, Sarvnaz Karimi and Adam Dunn.

16:50 - 17:30 Best Paper Award/Shared Task Award & Closing Remarks

## Table of Contents

<b>The Corpus of Australian and New Zealand Spoken English: A new resource of naturalistic speech transcripts</b> . . . . .	1
<i>Steven Coats</i>	
<b>Using public domain resources and off-the-shelf tools to produce high-quality multimedia texts</b> . .	6
<i>Manny Rayner, Belinda Chiera and Cathy Chua</i>	
<b>The Role of Context in Vaccine Stance Prediction for Twitter Users</b> . . . . .	16
<i>Aleney Khoo, Maciej Rybinski, Sarvnaz Karimi and Adam Dunn</i>	
<b>TCG-Event: Effective Task Conditioning for Generation-based Event Extraction</b> . . . . .	22
<i>Fatemeh Shiri, Tongtong Wu, Yuanfang Li and Gholamreza Haffari</i>	
<b>Complex Reading Comprehension Through Question Decomposition</b> . . . . .	31
<i>Xiao-Yu Guo, Yuan-Fang Li and Gholamreza Haffari</i>	
<b>Using Aspect-Based Sentiment Analysis to Classify Attitude-bearing Words</b> . . . . .	41
<i>Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers</i>	
<b>Fine-tuning a Subtle Parsing Distinction Using a Probabilistic Decision Tree: the Case of Postnominal "that" in Noun Complement Clauses vs. Relative Clauses</b> . . . . .	52
<i>Zineddine Tighidet and Nicolas Ballier</i>	
<b>Robustness of Hybrid Models in Cross-domain Readability Assessment</b> . . . . .	62
<i>Ho Hung Lim, Tianyuan Cai, John S. Y. Lee and Meichun Liu</i>	
<b>Specifying Optimisation Problems for Declarative Programs in Precise Natural Language</b> . . . . .	68
<i>Rolf Schwitter</i>	
<b>Improving Text-based Early Prediction by Distillation from Privileged Time-Series Text</b> . . . . .	73
<i>Jinghui Liu, Daniel Capurro, Anthony Nguyen and Karin Verspoor</i>	
<b>A DistilBERTopic Model for Short Text Documents</b> . . . . .	84
<i>Junaid Rashid, Jungeun Kim, Usman Naseem and Amir Hussain</i>	
<b>Generating Code-Switched Text from Monolingual Text with Dependency Tree</b> . . . . .	90
<i>Bryan Gregorius and Takeshi Okadome</i>	
<b>Stability of Forensic Text Comparison System</b> . . . . .	98
<i>Susan Brown and Shunichi Ishihara</i>	
<b>Academic Curriculum Generation using Wikipedia for External Knowledge</b> . . . . .	107
<i>Anurag Reddy Muthyala and Vikram Pudi</i>	
<b>Interactive Rationale Extraction for Text Classification</b> . . . . .	115
<i>Jiayi Dai, Mi-Young Kim and Randy Goebel</i>	
<b>Automatic Explanation Generation For Climate Science Claims</b> . . . . .	122
<i>Rui Xing, Shraey Bhatia, Timothy Baldwin and Jey Han Lau</i>	
<b>Zhangzhou Implosives and Their Variations</b> . . . . .	123
<i>Yishan Huang and Gwendolyn Hyslop</i>	
<b>Evaluating the Examiner: The Perils of Pearson Correlation for Validating Text Similarity Metrics</b>	131
<i>Gisela Vallejo, Timothy Baldwin and Lea Frermann</i>	

<b>Can Language Models Help in System Security? Investigating Log Anomaly Detection using BERT</b>	140
<i>Crispin Almodovar, Fariza Sabrina, Sarvnaz Karimi and Salahuddin Azad</i>	
<b>A Semantics of Spatial Expressions for interacting with unmanned aerial vehicles</b>	149
<i>Lucas Domingos and Paulo Santos</i>	
<b>Enhancing the DeBERTa Transformers Model for Classifying Sentences from Biomedical Abstracts</b>	157
<i>Abdul Aziz, Md. Akram Hossain and Abu Nowshed Chy</i>	
<b>Textstar: a Fast and Lightweight Graph-Based Algorithm for Extractive Summarization and Keyphrase Extraction</b>	162
<i>David Brock, Ali Khan, Tam Doan, Alicia Lin, Yifan Guo and Paul Tarau</i>	
<b>Contrastive Visual and Language Learning for Visual Relationship Detection</b>	171
<i>Thanh Tran, Maelic Neau, Paulo Santos and David Powers</i>	
<b>Shared Task (Not Peer Reviewed)</b>	
<b>Overview of the 2022 ALTA Shared task: PIBOSO sentence classification, 10 years later</b>	172
<i>Diego Mollá</i>	
<b>Estimating the Strength of Authorship Evidence with a Deep-Learning-Based Approach</b>	177
<i>Shunichi Ishihara, Satoru Tsuge, Mitsuyuki Inaba and Wataru Zaito</i>	
<b>Automatic Classification of Evidence Based Medicine Using Transformers</b>	182
<i>Necva Bolucu and Pinar Uskaner Hepsag</i>	
<b>Context-Aware Sentence Classification in Evidence-Based Medicine</b>	187
<i>Biaoyan Fang and Fajri Koto</i>	