



LREC 2026

**The 15th Workshop on Cognitive Modeling and
Computational Linguistics (CMCL) @ LREC 2026**

Workshop Proceedings

Editors

**Byung-Doh Oh, Tatsuki Kuribayashi, Giulia Rambelli,
Ece Takmaz, Philipp Wicke, Jixing Li, Ryo Yoshida**

16 May 2026

Proceedings of the 15th Workshop on Cognitive Modeling and Computational Linguistics
(CMCL) @ LREC 2026

This workshop has been supported by the National Institute for Japanese Language and
Linguistics (NINJAL), Japan.



National Institute for Japanese Language and Linguistics

Advanced Language Science
(E3P) Research Center

©ELRA Language Resources Association (ELRA), 2026

These proceedings are licensed under a Creative Commons Attribution-
NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-55-5

Preface

Welcome to the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2026)!

Now in its 15th edition, CMCL continues to be one of the leading workshops at the intersection of computational (psycho-)linguistics, cognitive modeling, and natural language processing. Building on the success of last year's event, we are thrilled to gather once again this year to explore the latest advancements in this exciting, interdisciplinary field.

This year, we received 48 submissions in total, including 38 regular archival submissions and 10 non-archival submissions. We accepted 26 archival papers, resulting in an acceptance rate of $26/38=68.4\%$, which is higher than that of last year (62.9%). As you will see, we believe this is largely due to the higher quality of submissions in this edition. We also accepted 7 non-archival papers for presentation. The program features a wide array of topics, spanning topics such as real-time sentence processing, language acquisition, and neuroscience-inspired evaluation of language models. We also see an increased focus on bilingual processing and pragmatic reasoning compared to last year's edition. This diversity reflects the growing engagement in our field and the increasing importance of understanding human language through computational methods.

We thank the Program Committee members for their timely and constructive reviews, which were crucial for shaping the program. We also extend our gratitude to our invited speakers, Dr. Claire Stevenson and Dr. Cory Shain, for enriching this year's program with their inspiring talks and expertise. We would also like to thank the LREC Workshop Chairs for selecting CMCL as one of the co-located workshops, as well as for helping us organize the workshop.

Last but not least, we are immensely grateful to our sponsor, the National Institute for Japanese Language and Linguistics (NINJAL). Their generous support allowed us to host the invited speakers for this year's event.

Thank you all for being a part of CMCL 2026. We look forward to an engaging and stimulating exchange of ideas!

The CMCL 2026 Organizing Committee

Organizing Committee

Byung-Doh Oh, Nanyang Technological University
Tatsuki Kuribayashi, Mohammed bin Zayed University of Artificial Intelligence
Giulia Rambelli, Aix-Marseille University
Ece Takmaz, Utrecht University
Philipp Wicke, AURYAL
Jixing Li, City University of Hong Kong
Ryo Yoshida, The University of Tokyo

Invited Speakers

Claire Stevenson, University of Amsterdam
Cory Shain, Stanford University

Program Committee

Giorgia Albertin, University of Bologna
Suhas Arehalli, Macalester College
Masayuki Asahara, National Institute for Japanese Language and Linguistics
Anna Bavaresco, University of Amsterdam
Philippe Blache, LPL CNRS
Emmanuele Chersoni, Hong Kong Polytechnic University
Rendi Chevi, Mohammed bin Zayed University of Artificial Intelligence
Marianne de Heer Kloots, University of Amsterdam
Aniello De Santo, University of Utah
Cui Ding, University of Zurich
Luca Dini, Istituto di Linguistica Computazionale "Antonio Zampolli" - CNR Pisa
Jakub Dotlacil, Utrecht University
Nadine El-Naggar, Mohammed bin Zayed University of Artificial Intelligence
Micha Elsner, The Ohio State University
Diego Frassinelli, LMU Munich
Richard Futrell, University of California, Irvine
Hai Hu, City University of Hong Kong
Shinnosuke Isono, National Institute for Japanese Language and Linguistics
Cassandra L. Jacobs, University at Buffalo
Kohei Kajikawa, Georgetown University
Shunsuke Kando, The University of Tokyo
Jordan Kodner, Stony Brook University
Junlin Li, Mohammed bin Zayed University of Artificial Intelligence
Tal Linzen, New York University
Agnese Lombardi, University of Pisa
Veronica Mangiaterra, University School for Advanced Studies IUSS Pavia, Italy
Sathvik Nair, University of Maryland, College Park
Miyu Oba, Nara Institute of Science and Technology
Yohei Oseki, The University of Tokyo
Francesca Padovani, University of Groningen
Ludovica Pannitto, University of Bologna
Laurent Prévot, University of Aix-Marseille

William Schuler, The Ohio State University
Alice Suozzi, Ca' Foscari University of Venice
Weijie Xu, University of California, Irvine
He Zhou, Hong Kong Polytechnic University

Table of Contents

<i>Transformer Attention as a Unified Model of Encoding and Retrieval in Human Sentence Processing</i> Dan Parker	1
<i>ChineseDevBench: A Chinese Developmental Benchmark for Language Development</i> Shaonan Wang, Yiwen Wu, Na Li, Zesheng Chen, Gan Wang, Shuchen Zhang, Xin Sun, Luan Li and Yaran Chen	10
<i>Predicting Sentence Acceptability Judgments in Multimodal Contexts</i> Hyewon Jang, Nikolai Ilinykh, Sharid Loaiciga, Jey Han Lau and Shalom Lappin	25
<i>ToM in LLM is not ToM, but a Pragmatic Effect</i> Agnese Lombardi and Alessandro Lenci	35
<i>What Kind of Language is Easy to Language-Model Under Curriculum Learning?</i> Nadine El-Naggar, Tatsuki Kuribayashi and Ted Briscoe	48
<i>Modeling semantic association in self-paced reading with language model embeddings</i> Sara Møller Østergaard, Kenneth Enevoldsen, Afra Alishahi and Bruno Nicenboim	60
<i>Character-aware Transformers Learn an Irregular Morphological Pattern Yet None Generalize Like Humans</i> Akhilesh Kakolu Ramarao, Kevin Tang and Dinah Baer-Henney	74
<i>Comparing Transformer Model Interpretability with Human Cognition: A Dual Analysis of Attention and Attribution</i> Lingchen Kong, Jinnie Shin and Pavlo Antonenko	86
<i>Correlating Language Model Surprisal With Cloze and Plausibility: Getting the Best of Both Measures</i> Kate Rebecca Belcher and Matthew Crocker	99
<i>Is Cross-Lingual Transfer in Bilingual Models Human-Like? A Study with Overlapping Word Forms in Dutch and English</i> Iza Škrjanec, Irene Elisabeth Winther, Vera Demberg and Stefan L. Frank	110
<i>Brain-to-Text Decoding with Brain Atlases and Brain Foundation Models</i> Haruka Akama, Ryo Yoshida, Max Müller-Eberstein and Yohei Oseki	124
<i>Floating or Suggesting Ideas? A Large-Scale Contrastive Analysis of Metaphorical and Literal Verb–Object Constructions</i> Prisca Piccirilli, Alexander Fraser and Sabine Schulte im Walde	145
<i>Social Meaning in Large Language Models: Structure, Magnitude, and Pragmatic Prompting</i> Roland Mühlenbernd	162
<i>Headlines You Won’t Forget: Can Pronoun Insertion Increase Memorability?</i> Selina Meyer, Magdalena Abel and Michael Roth	172

<i>Aggregated Transformer Attention Measures Predict Reading Times Beyond Surprisal</i>	
Lukas Mielczarek and Laura Kallmeyer	188
<i>'Layer su Layer': Identifying and Disambiguating the Italian NPN Construction in BERT's family</i>	
Greta Gorzoni, Ludovica Pannitto and Francesca Masini	203
<i>Teaching LLMs to unveil tendentious implicit contents of Italian political communication</i>	
Walter Paci, Lorenzo Gregori and Alessandro Panunzi	221
<i>Examining Algebraic Recombination for Compositional Generalisation</i>	
Joaquin Cardona Ruiz, Antske Fokkens and Lucia Donatelli	230
<i>Can LLMs Simulate Human Behavioral Variability? A Case Study in the Phonemic Fluency Task</i>	
Mengyang Qiu, Zoe Brisebois and Siena Sun	250
<i>Translation from the Information Bottleneck Perspective: an Efficiency Analysis of Spatial Prepositions in Bitexts</i>	
Antoine Taroni, Ludovic Moncla and Frederique Laforest	264
<i>Causal Inferences Are Driven by Noun Concept Specificity Evidence from Self-Paced Reading and Large Language Model Surprisal</i>	
Fabian Schlotterbeck and Raphael Barth	276
<i>Topic-context Dependency on Continuous Semantic Reconstruction of Language from fMRI Signals</i>	
Fermin Travi, Agustin Delmagro, Diego Fernandez Slezak, Bruno Bianchi and Juan E. Kamienkowski	286
<i>Disentanglement and Compositionality of Letter Identity and Letter Position in Variational Auto-Encoder Vision Models</i>	
Bruno Bianchi, Aakash Agrawal, Stanislas Dehaene, Emmanuel Chemla and Yair Lakretz	293
<i>Plausibility as Commonsense Reasoning: Humans Succeed, Large Language Models Do not</i>	
Sercan Karakas	309
<i>Benchmarking Source-Sensitive Reasoning in Turkish: Humans and LLMs under Evidential Trust Manipulation</i>	
Sercan Karakas and Yusuf Şimşek	329
<i>Toward Cognitive Alignment in Large Language Models: Integrating Linguistic Theory and Human Data</i>	
Wajdi Zaghouani	347

Non-archival papers

Quantifying the Pragmatics of Discourse Markers: A Surprisal-Based Analysis of Taiwanese "Ah"

Ri-Sheng Huang

Acoustic Representations Support Statistical Learning of Syllable Sequences: A Computational Study

Nika Jurov and Thomas Schatz

Simulating Cross-Linguistic Influence in Bilingual Reading - a Knowledge Distillation Approach

Irene E. Winther and Stefan L. Frank

Investigating Brain-LLM Alignment with Category-Deprived Text

Shantanu Nath, Marco Tettamanti and Roberto Zamparelli

Semantic Norms and Structural Relations in Chinese Semantic Composition: A Dataset and Analysis

He Zhou, Emmanuele Chersoni and Yu-Yin Hsu

Polar Questions in SPA–TTR: Linking Dialogue, Acquisition, and Neurosemantics

Jonathan Ginzburg, Shiyun Dong, Robin Cooper, Andy Luecking and Staffan Larsson

Generics as Useful Defeasible Rules of Inference

Justin Helms

Workshop Program

Saturday, 16 May 2026

09:10–09:20 Opening Remarks

09:20–10:20 Invited Talk: Claire Stevenson

Saturday, 16 May 2026 (continued)

10:30–11:40 Session 1: Poster Presentations

Transformer Attention as a Unified Model of Encoding and Retrieval in Human Sentence Processing

Dan Parker

ChineseDevBench: A Chinese Developmental Benchmark for Language Development

Shaonan Wang, Yiwen Wu, Na Li, Zesheng Chen, Gan Wang, Shuchen Zhang, Xin Sun, Luan Li and Yaran Chen

Predicting Sentence Acceptability Judgments in Multimodal Contexts

Hyewon Jang, Nikolai Ilinykh, Sharid Loaiciga, Jey Han Lau and Shalom Lappin

ToM in LLM is not ToM, but a Pragmatic Effect

Agnese Lombardi and Alessandro Lenci

What Kind of Language is Easy to Language-Model Under Curriculum Learning?

Nadine El-Naggar, Tatsuki Kuribayashi and Ted Briscoe

Modeling semantic association in self-paced reading with language model embeddings

Sara Møller Østergaard, Kenneth Enevoldsen, Afra Alishahi and Bruno Nicenboim

Quantifying the Pragmatics of Discourse Markers: A Surprisal-Based Analysis of Taiwanese "Ah"

Ri-Sheng Huang

Acoustic Representations Support Statistical Learning of Syllable Sequences: A Computational Study

Nika Jurov and Thomas Schatz

Simulating Cross-Linguistic Influence in Bilingual Reading - a Knowledge Distillation Approach

Irene E. Winther and Stefan L. Frank

Investigating Brain-LLM Alignment with Category-Deprived Text

Shantanu Nath, Marco Tettamanti and Roberto Zamparelli

Saturday, 16 May 2026 (continued)

11:40–13:00 Session 2: Oral Presentations

Character-aware Transformers Learn an Irregular Morphological Pattern Yet None Generalize Like Humans

Akhilesh Kakolu Ramarao, Kevin Tang and Dinah Baer-Henney

Comparing Transformer Model Interpretability with Human Cognition: A Dual Analysis of Attention and Attribution

Lingchen Kong, Jinnie Shin and Pavlo Antonenko

Correlating Language Model Surprisal With Cloze and Plausibility: Getting the Best of Both Measures

Kate Rebecca Belcher and Matthew Crocker

Is Cross-Lingual Transfer in Bilingual Models Human-Like? A Study with Overlapping Word Forms in Dutch and English

Iza Škrjanec, Irene Elisabeth Winther, Vera Demberg and Stefan L. Frank

13:00–14:00 Lunch Break

14:00–15:00 Invited Talk: Cory Shain

15:00–16:00 Session 3: Oral Presentations

Brain-to-Text Decoding with Brain Atlases and Brain Foundation Models

Haruka Akama, Ryo Yoshida, Max Müller-Eberstein and Yohei Oseki

Floating or Suggesting Ideas? A Large-Scale Contrastive Analysis of Metaphorical and Literal Verb–Object Constructions

Prisca Piccirilli, Alexander Fraser and Sabine Schulte im Walde

Social Meaning in Large Language Models: Structure, Magnitude, and Pragmatic Prompting

Roland Mühlenbernd

Semantic Norms and Structural Relations in Chinese Semantic Composition: A Dataset and Analysis

He Zhou, Emmanuele Chersoni and Yu-Yin Hsu

Saturday, 16 May 2026 (continued)

16:00–17:00 Session 4: Poster Presentations

Headlines You Won't Forget: Can Pronoun Insertion Increase Memorability?

Selina Meyer, Magdalena Abel and Michael Roth

Aggregated Transformer Attention Measures Predict Reading Times Beyond Surprisal

Lukas Mielczarek and Laura Kallmeyer

'Layer su Layer': Identifying and Disambiguating the Italian NPN Construction in BERT's family

Greta Gorzoni, Ludovica Pannitto and Francesca Masini

Teaching LLMs to unveil tendentious implicit contents of Italian political communication

Walter Paci, Lorenzo Gregori and Alessandro Panunzi

Examining Algebraic Recombination for Compositional Generalisation

Joaquin Cardona Ruiz, Antske Fokkens and Lucia Donatelli

Can LLMs Simulate Human Behavioral Variability? A Case Study in the Phonemic Fluency Task

Mengyang Qiu, Zoe Brisebois and Siena Sun

Translation from the Information Bottleneck Perspective: an Efficiency Analysis of Spatial Prepositions in Bitexts

Antoine Taroni, Ludovic Moncla and Frederique Laforest

Causal Inferences Are Driven by Noun Concept Specificity Evidence from Self-Paced Reading and Large Language Model Surprisal

Fabian Schlotterbeck and Raphael Barth

Polar Questions in SPA–TTR: Linking Dialogue, Acquisition, and Neurosemantics

Jonathan Ginzburg, Shiyun Dong, Robin Cooper, Andy Luecking and Staffan Larsson

Generics as Useful Defeasible Rules of Inference

Justin Helms

Saturday, 16 May 2026 (continued)

17:00–17:50 Session 5: Oral Presentations

Topic-context Dependency on Continuous Semantic Reconstruction of Language from fMRI Signals

Fermin Travi, Agustin Delmagro, Diego Fernandez Slezak, Bruno Bianchi and Juan E. Kamienkowski

Disentanglement and Compositionality of Letter Identity and Letter Position in Variational Auto-Encoder Vision Models

Bruno Bianchi, Aakash Agrawal, Stanislas Dehaene, Emmanuel Chemla and Yair Lakretz

Plausibility as Commonsense Reasoning: Humans Succeed, Large Language Models Do not

Sercan Karakas

Benchmarking Source-Sensitive Reasoning in Turkish: Humans and LLMs under Evidential Trust Manipulation

Sercan Karakas and Yusuf Şimşek

Toward Cognitive Alignment in Large Language Models: Integrating Linguistic Theory and Human Data

Wajdi Zaghouani

17:50–18:00 Closing Remarks

