



LREC 2026

**Fourth Workshop on Language Technologies for
Historical and Ancient Languages (LT4HALA 2026) @
LREC 2026**

Workshop Proceedings

Editors

Marco Passarotti and Rachele Sprugnoli

11 May 2026

Proceedings of the Fourth Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2026) @ LREC 2026

©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 978-2-493814-58-6

Preface

These proceedings include the papers accepted for presentation at the Fourth Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2026).¹ The workshop was held on May 11 2026 in Palma, Mallorca (Spain), co-located with LREC 2026.²

The workshop wants to provide a venue to discuss research works on a wide range of topics concerning the building, analysis, exploitation and distribution of collections of digitized texts written in historical and ancient languages as well as of their lexica, with a specific focus on the development and application of Language Technologies (LTs) for such purposes. The topics of the workshop are strictly bound to the peculiar characteristics of textual and lexical data for historical and ancient languages, which set them apart from modern languages, with a significant impact on the use and development of LTs for their processing and study. The workshop welcomed contributions addressing a broad range of topics. These included the handling of spelling variation and the detection and correction of OCR errors, as well as the creation and annotation of linguistic resources, both lexical and textual. Further areas of interest were deciphering, and the morphological, syntactic, and semantic analysis of textual data, alongside the adaptation of existing tools to account for diachronic, diatopic, and diastratic variation. The workshop also encouraged work on the use of LTs for teaching ancient languages, NLP-driven theoretical studies in historical linguistics, and the computational analysis of ancient literary texts. Additional relevant themes included the evaluation of NLP tools specifically designed for historical and ancient languages, the application of Large Language Models to the automatic analysis of ancient texts, and the role of digital infrastructures (such as CLARIN) in supporting research on language resources for historical and ancient languages.

Overall, the languages tackled in the papers published in these proceedings are almost 30. This is not an exhaustive list but it gives an idea of the linguistic richness covered: Ancient Chinese, Ancient Greek, Ancient Hebrew, Arabic, Coptic, Czech, Dutch, Estonian, French, German, Icelandic, Italian, Japanese, Judeo-Arabic, Latin, Maltese, Medieval Greek, Old English, Old High German, Old Norse, Old Saxon, Phoenician, Rapa Nui, Serbian, Swedish.

Like for LREC 2026, the submission process was double-blind. Each paper was reviewed but three independent reviewers from a program committee made of 40 scholars from 17 countries. In total, we received 39 submissions (against the 32 of the previous edition). After the reviewing process, we accepted 29 submissions, leading to an acceptance rate of 74.3%.

LT4HALA 2026 was also the venue of the fourth edition of EvaLatin, the campaign devoted to the evaluation of NLP tools for Latin.³ EvaLatin was started in 2020 (co-located with the first edition of LT4HALA) considering the important role played by textual data and linguistic metadata in the study of historical and ancient languages, with a special focus on Latin due to its prominence among such languages, both for the size and for the degree of diversity of its texts. Running evaluation campaigns in such a scenario is essential to understand the level of accuracy of the NLP tools used to build and analyze resources featuring texts that show those peculiar characteristics mentioned above. The fourth edition of EvaLatin focused on two shared tasks, i.e. Dependency Parsing, and Named Entities Recognition and Classification. Organizers and participants of EvaLatin submitted a report to be published in this proceedings thus this volume includes two overviews (one for each EvaLatin task) and 6 reports from participants.

Besides EvaLatin, LT4HALA 2024 hosted also a new edition of EvaHan, the evaluation campaign of NLP tools for the Ancient Chinese language, organized by a team of scholars

¹<https://circse.github.io/LT4HALA/2026/>

²<https://lrec2026.info>

³<https://circse.github.io/LT4HALA/2026/EvaLatin>

directed by Bin Li (Nanjing Normal University).⁴ The new edition of EvaHan focused on OCR tasks for multimodal large language models in ancient Chinese. Like for EvaLatin, the participants of EvaHan were required to submit a short technical report which received a light review by the organizers. Overall, these proceedings include an overview of the EvaHan campaign (authored by the organizers) and 12 technical reports, corresponding to as many participants. We wish to thank the organizers of EvaHan, whose contribution broadened the scope of historical and ancient languages represented at the LT4HALA 2026 workshop and highlighted how certain NLP challenges are inherent to ancient and historical languages as such, regardless of their typological differences.

LT4HALA 2026 also included EvaCun, an evaluation campaign on cuneiform LTs organized by Adam Anderson and Farhan Afzal Abdulla (UC Berkeley, USA).⁵ The proposed shared tasks focused on transliteration normalization, morphological analysis and lemmatization, Named Entity Recognition. Unfortunately, this edition of EvaCun attracted no participants, and the proceedings therefore contain no contributions on this topic. We nonetheless remain hopeful: the computational study of cuneiform languages represents one of the most fascinating and open challenges in the field of ancient language processing, and we trust that future editions will succeed in bringing together the researchers and resources needed to rise to it.

Now in its fourth edition, LT4HALA continues to grow steadily, both in the number of participants and in the range and diversity of languages and topics addressed by scholarly contributions. We welcome this expansion as an encouraging sign for the field, though perhaps not a surprising one, since the study of ancient and historical languages has always been deeply rooted in the analysis of textual evidence. Processing the large digital collections of such texts through the most advanced LTs, with the aim of bringing computational analysis to bear on them, holds genuine promise for advancing a century-long scholarly tradition. LT4HALA aspires to be a venue that supports and accompanies this computational turn.

Marco Passarotti and Rachele Sprugnoli

⁴<https://circse.github.io/LT4HALA/2026/EvaHan>

⁵<https://circse.github.io/LT4HALA/2026/EvaCun>

Organizing Committee

Workshop Organizers:

- Marco Passarotti, Università Cattolica del Sacro Cuore di Milano (Italy)
- Rachele Sprugnoli, Università Cattolica del Sacro Cuore di Milano (Italy)

Program Committee:

- Alaa Mamdouh Akef, Peking University, China
- Adam Anderson, UC Berkeley, USA
- Masayuki Asahara, National Institute for Japanese Language and Linguistics, Japan
- Monica Berti, University of Leipzig, Germany
- Federico Boschetti, Università Ca' Foscari, Istituto di Linguistica Computazionale - CNR & CLARIN Knowledge Centre for Digital and Public Textual Scholarship (DiPTeX-KC), Italy
- Luca Brigada Villa, University of Bergamo and University of Pavia, Italy
- Patrick J. Burns, Institute for the Study of the Ancient World / NYU, USA
- Flavio Massimiliano Cecchini, KU Leuven, Belgium
- Thibault Cl rice, Inria Paris, France
- Claudia Corbetta, Universit  di Bergamo and Universit  di Pavia, Italy
- Angelo Mario Del Grosso, Istituto di Linguistica Computazionale - CNR & CLARIN Knowledge Centre for Digital and Public Textual Scholarship (DiPTeX-KC), Italy
- Mathieu Dehouck, Laboratoire Lattice - CNRS, France
- Margherita Fantoli, KU Leuven, Belgium
- Andrea Farina, King's College London, UK
- Federica Gamba, Charles University, Czech Republic
- Shai Gordin, Ariel University, Israel
- Marisa Hudspeth, University of Massachusetts Amherst, USA
- Alek Keersmaekers, KU Leuven, Belgium
- Federica Iurescia, Universit  Cattolica del Sacro Cuore di Milano, Italy
- Bin Li, Nanjing Normal University, China
- Eleonora Litta, Universit  Cattolica del Sacro Cuore di Milano, Italy
- Yudong Liu, Western Washington University, USA
- Francesco Mambrini, Universit  Cattolica del Sacro Cuore di Milano, Italy
- Barbara McGillivray, King's College London, United Kingdom
- Domenico Giuseppe Muscianisi, Universit  di Parma, Italy
- Vivi Nastase, University of Geneva/Idiap Research Institute, Martigny, Switzerland
- Carlos Javier Nusch, Universidad Nacional de La Plata, Argentina
- Teresa Paccosi, Royal Netherlands Academy of Arts & Sciences (KNAW), The Netherlands
- Chiara Palladino, Durham University, United Kingdom

- Alessio Palmero Aprosio, Università di Trento, Italy
- Matteo Pellegrini, University of Surrey, UK
- Eva Pettersson, Uppsala University & CLARIN Knowledge Centre on Diachronic Language Resources (DiaRes), Sweden
- Sophie Prévost, Laboratoire Lattice - CNRS, France
- Matteo Romanello, Swiss Art Research Infrastructure (SARI), University of Zurich, Switzerland
- Colin Swaelens, Ghent University, Belgium
- Muhammad Shoaib Tahir, COMSATS University Islamabad, Lahore Campus, Pakistan
- James K. Tauber, FAU Erlangen, Germany
- Tatiana Tommasi, Università Ca' Foscari, Venezia, Italy
- Tariq Yousef, University of Southern Denmark, Denmark
- Chiara Zanchi, Università di Pavia, Italy

EvaLatin 2026 Organizers:

Dependency Parsing:

- Federica Iurescia, Università Cattolica del Sacro Cuore di Milano, Italy
- Rachele Sprugnoli, Università Cattolica del Sacro Cuore di Milano, Italy
- Marco Passarotti, Università Cattolica del Sacro Cuore di Milano, Italy

Named Entity Recognition:

- Valeria Irene Boano, KU Leuven, Belgium
- Eleonora Litta, Università Cattolica del Sacro Cuore di Milano, Italy
- Matteo Romanello, Swiss Art Research Infrastructure (SARI), University of Zurich, Switzerland

EvaHan 2026 Organizers:

- Bin Li, Nanjing Normal University, P.R. China
- Bolin Chang, Nanjing Normal University, P.R. China
- Minxuan Feng, Nanjing Normal University, P.R. China
- Chao Xu, Nanjing Normal University, P.R. China
- Dongbo Wang, Nanjing Agricultural University, P.R. China

EvaCun 2026 Organizers:

- Adam Anderson, UC Berkeley, USA
- Farhan Afzal Abdulla, UC Berkeley, USA

Table of Contents

<i>Morphological Annotation of Old Serbian in Universal Dependencies</i> Vladimir Polomac and Silvie Cinkova	1
<i>Tracing Morph Origins in Czech: A Computational Approach to Morph-Level Etymology</i> Aleš Manuel Manuel Papáček and Zdeněk Žabokrtský	7
<i>Uncovering Work from Words: LLM-Based Information Extraction from Historical Petitions</i> Ellinor Lindqvist, Eva Pettersson and Joakim Nivre	19
<i>Extracting Volcanological Knowledge from Historical Texts: A Language-Technology Pipeline for Diachronic Geovisualization</i> Costanza Marini, Gianluca Casagrande, Alessio Palmero Aprosio and Claudia Principe	38
<i>When Lexicographic Quotations Become a Corpus: To Deduplicate or Not to Deduplicate?</i> Manuel Favaro, Elisa Guadagnini, Eva Sassolini, Marco Biffi and Simonetta Montemagni	49
<i>A New State-of-the-Art BERT Model for Judeo-Arabic</i> Elisha Rosensweig, Yitzchak Lindenbaum, Hillel Gershuni, Vered Raziel-Kretzmer, Daniel Caine and Avi Shmidman	58
<i>BEReshiT: an Ancient Hebrew Model based on DictaBERT</i> Iglia Nikolova-Stoupak, Maxime Amblard and Frédérique Rey	72
<i>Automatic Detection of Metaphorical Expressions in Classical Japanese Using WLSP-Enhanced BERT</i> Hang Zhu, Mitoki Ohara, Rei Kikuchi, Kanako Komiya, Masayuki Asahara and Sachi Kato	89
<i>Domain-Aware Error Correction for Citation NER in Medieval Hebrew Responsa</i> Shmuel Liebeskind, Maayan Zhitomirsky-Geffet, Binyamin Katzoff, Nati Ben-Gigi and Jonathan Schler	96
<i>From Lemmas to Links: A Lemma Bank for Ancient Greek</i> Colin Swaelens, Francesco Mambrini and Marco Passarotti	106
<i>Across Generations: A Comparative Analysis of NER for Latin Inscriptions from Classical Machine Learning to LLMs</i> Wenhui Cui and Phillip Benjamin Ströbel	112
<i>POS Tagging with Generative LLMs for Historical Germanic Low-Resource Languages: An Evaluation Against Fine-Tuned BERT</i> Irene Miani, Gregory Darwin and Sara Stymne	125
<i>From Manuscript to Model: Developing HTR for Medieval Greek</i> Nicklas Sindlev Andersen, Byron MacDougall, Tariq Yousef and Aglae Pizzone	139

<i>I, RE: Claudius 256: Towards Linking Classical Latin Person Mentions to a Domain-specific Knowledge Base</i>	
Marijke Beersmans, Evelien de Graaf, Julie Nijs, Valeria Irene Boano, Alek Keersmaekers, Mark Depauw, Tim Van de Cruys and Margherita Fantoli	152
<i>Capturing Ancient Chinese Sense Induction with Automatic Pipelines</i>	
Guan-Yu Tseng, Chunki Lim, Chih-Han Lin, Tung-Le Pan, Yu-Chieh Wang, Lang-Ching Yeh and Shu-Kai Hsieh	163
<i>A Computational Evaluation of Syllabic Hypotheses for Rongorongo: Evidence from N-gram Analysis</i>	
Evgeniya Korovina	177
<i>Building a Corpus and Database for Rare and Undeciphered Scripts</i>	
Beata Megyesi, Rune Rattenborg, Benedek Láng, Michelle Waldispühl and Mihály Héder	184
<i>A Layered Annotation Workflow for Semitic Epigraphy</i>	
Tal Bernstein, Shai Gordin and Letizia Cerqueglini	197
<i>Overview of the Dependency Parsing Task at EvaLatin 2026</i>	
Federica Iurescia, Marco Passarotti and Rachele Sprugnoli	211
<i>THIVLVC: Retrieval Augmented Dependency Parsing for Latin</i>	
Luc Pommeret, Thibault Wagret and Jules Deret	219
<i>Overview of the Named Entity Recognition Task at EvaLatin 2026</i>	
Valeria Irene Boano, Eleonora Litta and Matteo Romanello	226
<i>Transfer Learning for Named Entity Recognition of Classical Latin through LLM Prompting</i>	
Callum Chan	234
<i>Overview of EvaHan2026: The First International Evaluation of Ancient Chinese OCR and Layout Analysis</i>	
Dongbo Wang, Dongmei Zhu, Jieqiong Li, Chang Liu, Ruifeng Wu, Fan Yang, Zhu Yue, Xin Gao, Zhao Zhixiao, Xue Zhao, Zhongzheng Wu, Liu Liu and Li Bin	244
<i>A Multi-Stage System for Ancient Chinese OCR and Layout Understanding in the EvaHan2026 Shared Task</i>	
KeYan Liang and Meiling Liu	263
<i>A Multi-Modal Recognition Framework for Ancient Books Integrating DoRA-DPO Text Recognition and YOLO Layout Analysis</i>	
Chaokun Zhang, Xin Wen and Tongtong Zhou	268
<i>Beijing Normal University at EvaHan 2026: Enhancing Ancient Chinese Character Recognition and Layout Analysis via VLM Fine-Tuning and Linguistic Post-Processing</i>	
Yihuan Yin and Qian Zhao	273
<i>A Dual-Modality Framework for Ancient Document Layout Analysis and Text Recognition</i>	
Qi Fan, Jieming Hu and Chen Ye	277
<i>EvaHan 2026 Ancient Books Multimodal OCR and Layout Analysis System Technical Report</i>	
Chenrui Zheng	288

<i>A Parameter-Efficient and Data-Centric Framework for Ancient Chinese Text Recognition and Layout Analysis</i>	
Yuchun Meng	294
<i>LVLM Optimization for Ancient Chinese Book Image Analysis with Task-specific Augmentation and Instruction Tuning</i>	
Tian Xia, Yulong Liu, Yilin Wang, Yumeng Yang, Dongheng Cai, Yuyang Tan and Menghui Yang	299
<i>Data-Centric Strategies for Ancient Chinese Text Recognition: Augmentation, Annotation Refinement, and Style Transfer in EvaHan 2026</i>	
Li Chengfei, Zhang Yunjie, Li XiaoYi, Quan Changshun, Cao Taihe and Liu Bin	305
<i>AnandaSky: A Vision–Language Model for Line-Level Transcription of Historical Sinographic Documents</i>	
Colin Brisson, Ayoub Kahfy, Frédéric Constant and Marc Bui	311
<i>Multimodal Ancient Document Parsing: Technical Report for EvaHan2026 Competition</i>	
Liqi He, Qiwei Li, Ziyang Yang and Zuchao Li	322
<i>Multi-Task Learning Trade-offs in Vision–Language Models for Ancient Chinese OCR: An Empirical Analysis of Parameter-Efficient Adaptation</i>	
Yuhan Shu and Huizi Zhou	330
<i>Building Character(s): Synthetic Data and In-Context Learning Strategies for Few-Shot Ancient Chinese Recognition</i>	
Denise Atzori, Marie Bizais-Lillig, Mathias Garnier, Maxime Létoffé, Charles Planque, Tianjie Yin and Chahan Vidal-Gorène	339
<i>The UD_Latin-PROIEL as Linked Open Data: Integrating a Latin Treebank into the LiLa Knowledge Base</i>	
Lucas Consolin Dezotti, Marco Passarotti, Federica Iurescia and Giovanni Moretti	353
<i>Language Models for the Restoration of Latin Legal Manuscripts</i>	
Shibingfeng Zhang, Edoardo Caraffa, Annafelicia Zuffrano, Maddalena Modesti and Giovanni Colavizza	361
<i>Evaluating Hierarchical Aggregation and LLM-Based Matching for Synset Selection in Ancient Greek</i>	
Luca Brigada Villa, Marco Passarotti, Chiara Zanchi, Riccardo Ginevra, Erica Fratellini and Eleonora Litta	368
<i>Miktub: A Manuscript Dataset of Historical Maltese for Handwritten Text Recognition</i>	
Thomas Koppens and Claudia Borg	380
<i>Smelling the Past: Investigating Historical Models for Olfactory Event Extraction</i>	
Teresa Paccosi and Marijn Koolen	389
<i>Contemporizing 20-th Century Estonian</i>	
Heiki-Jaan Kaalep	400
<i>Cost-Aware Pre-Annotation Strategies for Nested NER in Historical Latin Notarial Deeds</i>	
Charlene Ellul, Vanessa Buhagiar, Claudia Borg and Charlie Abela	407

<i>From Lemmatization to Legal Terminology: Assessing an Hybrid Pipeline on Justinian's Digest</i> Paola Marongiu and Eva Sassolini	418
<i>UppsalaNLP at EvaLatin 2026: Multilingual parsing for Latin</i> Sara Stymne	429
<i>Contextual Probing for Low-Resource Named Entity Recognition in Latin</i> Maria Mihaela Trusca, Mark Depauw, Violet Soen, Ine de Daele, Kevin Verbruggen and Tim Van de Cruys	437
<i>Classificatio Sine lactu – That Is, Zero-Shot NERC in Latin</i> Luisa Ripoll-Alberola, Fernando Nicolás-Flores and Francisco Javier Muñoz Acebes .	443
<i>Extending omnes flores for the EvaLatin 2026 Dependency Parsing Tasks</i> Hiroshi Matsuda and Masayuki Asahara	448
<i>Pre-Editorial Normalization for Automatically Transcribed Medieval Manuscripts in Old French and Latin</i> Thibault Clérice, Rachel Bawden, Anthony Glaise, Ariane Pinche and David Smith ...	453
<i>OldBERTur: Named Entity Recognition for Medieval Icelandic</i> Pontus Henningsson, Eva Pettersson and Erik Lenas	469
<i>Neural Machine Translation for Coptic-French: Strategies for Low-Resource Ancient Languages</i> Nasma Chaoui and Richard Khoury	482

Workshop Program

Monday, May 11, 2026

- 9:00–9:15 *Welcome and introduction*
Marco Passarotti and Rachele Sprugnoli
- 09:15–10:30 Oral Session 1**
- 09:15–09:30 *Morphological Annotation of Old Serbian in Universal Dependencies*
Vladimir Polomac and Silvie Cinkova
- 09:30–09:45 *Tracing Morph Origins in Czech: A Computational Approach to Morph-Level Etymology*
Aleš Manuel Manuel Papáček and Zdeněk Žabokrtský
- 09:45–10:00 *Uncovering Work from Words: LLM-Based Information Extraction from Historical Petitions*
Ellinor Lindqvist, Eva Pettersson and Joakim Nivre
- 10:00–10:15 *Extracting Volcanological Knowledge from Historical Texts: A Language-Technology Pipeline for Diachronic Geovisualization*
Costanza Marini, Gianluca Casagrande, Alessio Palmero Aprosio and Claudia Principe
- 10:15–10:30 *When Lexicographic Quotations Become a Corpus: To Deduplicate or Not to Deduplicate?*
Manuel Favaro, Elisa Guadagnini, Eva Sassolini, Marco Biffi and Simonetta Montemagni
- 10:30–11:30 Poster Session + Coffee Break**
- 10:30–11:30 *A New State-of-the-Art BERT Model for Judeo-Arabic*
Elisha Rosensweig, Yitzchak Lindenbaum, Hillel Gershuni, Vered Raziel-Kretzmer, Daniel Caine and Avi Shmidman
- 10:30–11:30 *BEReshiT: an Ancient Hebrew Model based on DictaBERT*
Iglika Nikolova-Stoupak, Maxime Amblard and Frédérique Rey
- 10:30–11:30 *Automatic Detection of Metaphorical Expressions in Classical Japanese Using WLSP-Enhanced BERT*
Hang Zhu, Mitoki Ohara, Rei Kikuchi, Kanako Komiya, Masayuki Asahara and Sachi Kato
- 10:30–11:30 *Domain-Aware Error Correction for Citation NER in Medieval Hebrew Responsa*
Shmuel Liebeskind, Maayan Zhitomirsky-Geffet, Binyamin Katzoff, Nati Ben-Gigi and Jonathan Schler

Monday, May 11, 2026 (continued)

- 10:30–11:30 *From Lemmas to Links: A Lemma Bank for Ancient Greek*
Colin Swaelens, Francesco Mambrini and Marco Passarotti
- 10:30–11:30 *Across Generations: A Comparative Analysis of NER for Latin Inscriptions from Classical Machine Learning to LLMs*
Wenhui Cui and Phillip Benjamin Ströbel
- 10:30–11:30 *POS Tagging with Generative LLMs for Historical Germanic Low-Resource Languages: An Evaluation Against Fine-Tuned BERT*
Irene Miani, Gregory Darwin and Sara Stymne
- 10:30–11:30 *From Manuscript to Model: Developing HTR for Medieval Greek*
Nicklas Sindlev Andersen, Byron MacDougall, Tariq Yousef and Aglae Pizzone
- 10:30–11:30 *I, RE:Claudius 256: Towards Linking Classical Latin Person Mentions to a Domain-specific Knowledge Base*
Marijke Beersmans, Evelien de Graaf, Julie Nijs, Valeria Irene Boano, Alek Keersmaekers, Mark Depauw, Tim Van de Cruys and Margherita Fantoli
- 10:30–11:30 *Capturing Ancient Chinese Sense Induction with Automatic Pipelines*
Guan-Yu Tseng, Chunki Lim, Chih-Han Lin, Tung-Le Pan, Yu-Chieh Wang, Lang-Ching Yeh and Shu-Kai Hsieh
- 11:30–12:15 Oral Session 2**
- 11:30–11:45 *A Computational Evaluation of Syllabic Hypotheses for Rongorongo: Evidence from N-gram Analysis*
Evgeniya Korovina
- 11:45–12:00 *Building a Corpus and Database for Rare and Undeciphered Scripts*
Beata Megyesi, Rune Rattenborg, Benedek Láng, Michelle Waldispühl and Mihály Héder
- 12:00–12:15 *A Layered Annotation Workflow for Semitic Epigraphy*
Tal Bernstein, Shai Gordin and Letizia Cerqueglini

Monday, May 11, 2026 (continued)

12:15–13:00 **EvaLatin**

- 12:15–12:25 *Overview of the Dependency Parsing Task at EvaLatin 2026*
Federica Iurescia, Marco Passarotti and Rachele Sprugnoli
- 12:25–12:35 *THIVLVC: Retrieval Augmented Dependency Parsing for Latin*
Luc Pommeret, Thibault Wagret and Jules Deret
- 12:35–12:45 *Overview of the Named Entity Recognition Task at EvaLatin 2026*
Valeria Irene Boano, Eleonora Litta and Matteo Romanello
- 12:45–12:55 *Transfer Learning for Named Entity Recognition of Classical Latin through LLM Prompting*
Callum Chan

14:00–15:15 **EvaHan**

- 14:00–14:02 *Opening Remarks*
Li Bin
- 14:02–14:07 *Invited talk*
Feng Zhiwei
- 14:07–14:22 *Overview of EvaHan2026: The First International Evaluation of Ancient Chinese OCR and Layout Analysis*
Dongbo Wang, Dongmei Zhu, Jieqiong Li, Chang Liu, Ruifeng Wu, Fan Yang, Zhu Yue, Xin Gao, Zhao Zhixiao, Xue Zhao, Zhongzheng Wu, Liu Liu and Li Bin
- 14:23–14:26 *A Multi-Stage System for Ancient Chinese OCR and Layout Understanding in the EvaHan2026 Shared Task*
KeYan Liang and Meiling Liu
- 14:27–14:31 *A Multi-Modal Recognition Framework for Ancient Books Integrating DoRA-DPO Text Recognition and YOLO Layout Analysis*
Chaokun Zhang, Xin Wen and Tongtong Zhou
- 14:32–14:35 *Beijing Normal University at EvaHan 2026: Enhancing Ancient Chinese Character Recognition and Layout Analysis via VLM Fine-Tuning and Linguistic Post-Processing*
Yihuan Yin and Qian Zhao

Monday, May 11, 2026 (continued)

- 14:36–14:39 *A Dual-Modality Framework for Ancient Document Layout Analysis and Text Recognition*
Qi Fan, Jieming Hu and Chen Ye
- 14:40–14:43 *EvaHan 2026 Ancient Books Multimodal OCR and Layout Analysis System Technical Report*
Chenrui Zheng
- 14:44–14:47 *A Parameter-Efficient and Data-Centric Framework for Ancient Chinese Text Recognition and Layout Analysis*
Yuchun Meng
- 14:48–14:51 *LVLM Optimization for Ancient Chinese Book Image Analysis with Task-specific Augmentation and Instruction Tuning*
Tian Xia, Yulong Liu, Yilin Wang, Yumeng Yang, Dongheng Cai, Yuyang Tan and Menghui Yang
- 14:52–14:55 *Data-Centric Strategies for Ancient Chinese Text Recognition: Augmentation, Annotation Refinement, and Style Transfer in EvaHan 2026*
Li Chengfei, Zhang Yunjie, Li XiaoYi, Quan Changshun, Cao Taihe and Liu Bin
- 14:56–14:59 *AnandaSky: A Vision–Language Model for Line-Level Transcription of Historical Sinographic Documents*
Colin Brisson, Ayoub Kahfy, Frédéric Constant and Marc Bui
- 15:00–15:03 *Multimodal Ancient Document Parsing: Technical Report for EvaHan2026 Competition*
Liqi He, Qiwei Li, Ziye Yang and Zuchao Li
- 15:04–15:07 *Unified Vision–Language Framework for End-to-End Ancient Chinese OCR and Layout Analysis: Our Submission to EvaHan 2026*
Yuhan Shu and Huizi Zhou
- 15:08–15:11 *Building Character(s): Synthetic Data and In-Context Learning Strategies for Few-Shot Ancient Chinese Recognition*
Denise Atzori, Marie Bizais-Lillig, Mathias Garnier, Maxime Létoffé, Charles Planque, Tianjie Yin and Chahan Vidal-Gorène
- 15:11–15:15 *Closing remarks for EvaHan 2026*
Li Bin

Monday, May 11, 2026 (continued)

15:15–16:00 Oral Session 3

15:15–15:30 *The UD_Latin-PROIEL as Linked Open Data: Integrating a Latin Treebank into the LiLa Knowledge Base*
Lucas Consolin Dezotti, Marco Passarotti, Federica Iurescia and Giovanni Moretti

15:30–15:45 *Language Models for the Restoration of Latin Legal Manuscripts*
Shibingfeng Zhang, Edoardo Caraffa, Annafelicia Zuffrano, Maddalena Modesti and Giovanni Colavizza

15:45–16:00 *Evaluating Hierarchical Aggregation and LLM-Based Matching for Synset Selection in Ancient Greek*
Luca Brigada Villa, Marco Passarotti, Chiara Zanchi, Riccardo Ginevra, Erica Fratellini and Eleonora Litta

16:00–17:00 Poster Session + Coffee Break

16:00–17:00 *Miktub: A Manuscript Dataset of Historical Maltese for Handwritten Text Recognition*
Thomas Koppens and Claudia Borg

16:00–17:00 *Smelling the Past: Investigating Historical Models for Olfactory Event Extraction*
Teresa Paccosi and Marijn Koolen

16:00–17:00 *Contemporizing 20-th Century Estonian*
Heiki-Jaan Kaalep

16:00–17:00 *Cost-Aware Pre-Annotation Strategies for Nested NER in Historical Latin Notarial Deeds*
Charlene Ellul, Vanessa Buhagiar, Claudia Borg and Charlie Abela

16:00–17:00 *From Lemmatization to Legal Terminology: Assessing an Hybrid Pipeline on Justinian's Digest*
Paola Marongiu and Eva Sassolini

16:00–17:00 *UppsalaNLP at EvaLatin 2026: Multilingual parsing for Latin*
Sara Stymne

16:00–17:00 *Contextual Probing for Low-Resource Named Entity Recognition in Latin*
Maria Mihaela Trusca, Mark Depauw, Violet Soen, Ine de Daele, Kevin Verbruggen and Tim Van de Cruys

Monday, May 11, 2026 (continued)

- 16:00–17:00 *Classificatio Sine lactu – That Is, Zero-Shot NERC in Latin*
Luisa Ripoll-Alberola, Fernando Nicolás-Flores and Francisco Javier Muñoz Acebes
- 16:00–17:00 *Extending omnes flores for the EvaLatin 2026 Dependency Parsing Tasks*
Hiroshi Matsuda and Masayuki Asahara
- 17:00–17:45 Oral Session 4**
- 17:00–17:15 *Pre-Editorial Normalization for Automatically Transcribed Medieval Manuscripts in Old French and Latin*
Thibault Clérice, Rachel Bawden, Anthony Glaise, Ariane Pinche and David Smith
- 17:15–17:30 *OldBERTur: Named Entity Recognition for Medieval Icelandic*
Pontus Henningsson, Eva Pettersson and Erik Lenas
- 17:30–17:45 *Neural Machine Translation for Coptic-French: Strategies for Low-Resource Ancient Languages*
Nasma Chaoui and Richard Khoury
- 17:45–18:00 Closing**