2nd Workshop on Ancient Language Processing (ALP 2025)

Albuquerque, New Mexico, USA
4 May 2025

ISBN: 979-8-3313-1910-6

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Automatic Text Segmentation of Ancient and Historic Hebrew
Elisha Rosensweig, Benjamin Resnick, Hillel Gershuni, Joshua Guedalia, Nachum Dershowitz and Avi Shmidman 1

Integrating Semantic and Statistical Features for Authorial Clustering of Qumran Scrolls
Yonatan Lourie, Jonathan Ben-Dov and Roded Sharan 12

Assignment of account type to proto-cuneiform economic texts with Multi-Class Support Vector Machines
Piotr Zadworny and Shai Gordin 22

Accessible Sanskrit: A Cascading System for Text Analysis and Dictionary Access
Giacomo De Luca 31

A Dataset of Ancient Chinese Math Word Problems and an Application for Research in Historic Mathematics
Florian Keßler 40

Using Cross-Linguistic Data Formats to Enhance the Annotation of Ancient Chinese Documents Written on Bamboo Slips
Michele Pulini and Johann-Mattis List 52

Towards an Integrated Methodology of Dating Biblical Texts: The Case of the Book of Jeremiah
Martijn Naaijer and Aren Wilson-Wright 59

The Development of Hebrew in Antiquity – A Computational Linguistic Study
Hallel Baitner, dimid duchovny, Lee-Ad Gottlieb, Amir Yorav, Nachum Dershowitz and Eshbal Ratzon 65

Evaluating Evaluation Metrics for Ancient Chinese to English Machine Translation
Eric R. Bennett, HyoJung Han, Xinchen Yang, Andrew Schonebaum and Marine Carpuat 71

Neural Models for Lemmatization and POS-Tagging of Earlier and Late Egyptian (Supporting Hieroglyphic Input) and Demotic
Aleksi Sahala and Eliese-Sophia Lincke 77

Bringing Suzhou Numerals into the Digital Age: A Dataset and Recognition Study on Ancient Chinese Trade Records
Ting-Lin Wu, Zih-Ching Chen, Chen-Yuan Chen, Pi-Jhong Chen and Li-Chiao Wang 83

The Historian's Fingerprint: Computational Stylometric Analysis of the Zuo Commentary and Discourses of the States
Wenjie Hua 90

Overview of EvaHan2025: The First International Evaluation on Ancient Chinese Named Entity Recognition
Bin Li, Bolin Chang, Ruilin Liu, Xue Zhao, Si Shen, Lihong Liu, Yan Zhu, Zhixing Xu, Weiguang QU and Dongbo Wang 97

Exploring the Application of 7B LLMs for Named Entity Recognition in Chinese Ancient Texts
yicheng zhu and han bi 106

Construction of NER Model in Ancient Chinese: Solution of EvaHan 2025 Challenge
Yi Lu and Minyi Lei 112

LLM's Weakness in NER Doesn't Stop It from Enhancing a Stronger SLM
Weilu Xu, Renfei Dang and Shujian Huang 117

Named Entity Recognition in Context: Edit_Dunhuang team Technical Report for Evahan2025 NER Competition
Colin Brisson, Ayoub Kahfy, Marc Bui and Frédéric Constant 123

Simple Named Entity Recognition (NER) System with RoBERTa for Ancient Chinese
Yunmeng Zhang, Meiling Liu, Hanqi Tang, Shige Lu and Lang Xue 129

Make Good Use of GujiRoBERTa to Identify Entities in Ancient Chinese
Lihan Lin, Yiming Wang, Jiachen Li, Huan Ouyang and Si Li 136

GRoWE: A GujiRoBERTa-Enhanced Approach to Ancient Chinese NER via Word-Word Relation Classification and Model Ensembling
tian Xia, yilin wang, xinkai Wang, Qun Zhao and Menghui Yang 141

When Less Is More: Logits-Constrained Framework with RoBERTa for Ancient Chinese NER
Wenjie Hua and Shenghan Xu 146

Multi-Strategy Named Entity Recognition System for Ancient Chinese
Wenxuan Dong and Meiling Liu 151

Multi-Domain Ancient Chinese Named Entity Recognition Based on Attention-Enhanced Pre-trained Language Model
Qi Zhang, Zhiya Duan, Shijie Ma, ShengYu Liu, Zibo Yuan and RuiMin Ma 159

EvaCun 2025 Shared Task: Lemmatization and Token Prediction in Akkadian and Sumerian using LLMs
Shai Gordin, Aleksi Sahala, Shahar Spencer and Stav Klein 164

Lemmatization of Cuneiform Languages Using the ByT5 Model
Pengxiu Lu, Yonglong Huang, Jing Xu, Minxuan Feng and Chao Xu 173

Finetuning LLMs for EvaCun 2025 token prediction shared task
Josef Jon and Ondřej Bojar 182

Beyond Base Predictors: Using LLMs to Resolve Ambiguities in Akkadian Lemmatization
Frederick Riemenschneider 187

A Low-Shot Prompting Approach to Lemmatization in the EvaCun 2025 Shared Task
John Sbur, Brandi Wilkins, Elizabeth Paul and Yudong Liu 193

From Clay to Code: Transforming Hittite Texts for Machine Learning
Emma Yavasan and Shai Gordin 198

Towards Ancient Meroitic Decipherment: A Computational Approach
Joshua N. Otten and Antonios Anastasopoulos 208

Detecting Honkadori based on Waka Embeddings
Hayato Ogawa, Kaito Horio and Daisuke Kawahara 220

Incorporating Lexicon-Aligned Prompting in Large Language Model for Tangut–Chinese Translation
Yuxi Zheng and Jingsong Yu 228

ParsiPy: NLP Toolkit for Historical Persian Texts in Python
Farhan Farsi, Parnian Fazel, Sepand Haghighi, Sadra Sabouri, Farzaneh Goshtasb, Nadia Hajipour, Ehsaneddin Asgari and Hossein Sameti 238