Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022)

Held at LREC 2022

Marseille, France
25 June 2022

Editors:

Rachele Sprugnoli
Marco Passarotti

ISBN: 978-1-7138-6127-0

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© (2022) by European Language Resources Association (ELRA)
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-NC. (https://creativecommons.org/licenses/by-nc/4.0/)

Printed with permission by Curran Associates, Inc. (2023)

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

Identifying Cleartext in Historical Ciphers
Maria-Elena Gambardella, Beata Megyesi and Eva Pettersson 1

Detecting Diachronic Syntactic Developments in Presence of Bias Terms
Oliver Hellwig and Sven Sellmer 10

Accurate Dependency Parsing and Tagging of Latin
Sebastian Nehrdich and Oliver Hellwig 20

Annotating "Absolute" Preverbs in the Homeric and Vedic Treebanks
Luca Brigada Villa, Erica Biagetti and Chiara Zanchi 26

CHJ-WLSP: Annotation of 'Word List by Semantic Principles' Labels for the Corpus of Historical Japanese
Masayuki Asahara, Nao Ikegami, Tai Suzuki, Taro Ichimura, Asuko Kondo, Sachi Kato and Makoto Yamazaki 31

The IKUVINA Treebank
Mathieu Dehouck 38

Machine Translation of 16Th Century Letters from Latin to German
Lukas Fischer, Patricia Scheurer, Raphael Schwitter and Martin Volk 43

A Treebank-based Approach to the Supprema Constructio in Dante's Latin Works
Flavio Massimiliano Cecchini and Giulia Pedonese 51

From Inscriptions to Lexica and Back: A Platform for Editing and Linking the Languages of Ancient Italy
Valeria Quochi, Andrea Bellandi, Fahad Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi and Cesare Zavattari 59

BERToldo, the Historical BERT for Italian
Alessio Palmero Aprosio, Stefano Menini and Sara Tonelli 68

In Search of the Flocks: How to Perform Onomasiological Queries in an Ancient Greek Corpus?
Alek Keersmaekers and Toon Van Hal 73

Contextual Unsupervised Clustering of Signs for Ancient Writing Systems
Michele Corazza, Fabio Tamburini, Miguel Valério and Silvia Ferrara 84

Towards the Creation of a Diachronic Corpus for Italian: A Case Study on the GDLI Quotations
Manuel Favaro, Elisa Guadagnini, Eva Sassolini, Marco Biffi and Simonetta Montemagni 94

Automatic Translation Alignment for Ancient Greek and Latin
Tariq Yousef, Chiara Palladino, David J. Wright and Monica Berti 101

Handling Stress in Finite-State Morphological Analyzers for Ancient Greek and Ancient Hebrew
Daniel Swanson and Francis Tyers 108

From Inscription to Semi-automatic Annotation of Maya Hieroglyphic Texts
Cristina Vertan and Christian Prager 114

Multilingual Named Entity Recognition for Medieval Charters Using Stacked Embeddings and Bert-based Models.
Sergio Torres Aguilar 119

Linguistic Annotation of Neo-Latin Mathematical Texts: A Pilot-Study to Improve the Automatic Parsing of the Archimedes Latinus
Margherita Fantoli and Miryam de Lhoneux 129

The First International Ancient Chinese Word Segmentation and POS Tagging Bakeoff: Overview of the EvaHan 2022 Evaluation Campaign
Bin Li, Yiguo Yuan, Jingya Lu, Minxuan Feng, Chao Xu, Weiguang QU and Dongbo Wang . . 135

Automatic Word Segmentation and Part-of-Speech Tagging of Ancient Chinese Based on BERT Model
Yu Chang, Peng Zhu, Chaoping Wang and Chaofan Wang 141

Ancient Chinese Word Segmentation and Part-of-Speech Tagging Using Data Augmentation
Yanzhi Tian and Yuhang Guo 146

BERT 4EVER@EvaHan 2022: Ancient Chinese Word Segmentation and Part-of-Speech Tagging Based on Adversarial Learning and Continual Pre-training
Hailin Zhang, Ziyu Yang, Yingwen Fu and Ruoyao Ding 150

Construction of Segmentation and Part of Speech Annotation Model in Ancient Chinese
Longjie Jiang, Qinyu C. Chang, Huyin H. Xie and Zhuying Z. Xia 155

Simple Tagging System with RoBERTa for Ancient Chinese
Binghao Tang, Boda Lin and Si Li 159

The Uncertainty-based Retrieval Framework for Ancient Chinese CWS and POS
Pengyu Wang and Zhichen Ren 164

Data Augmentation for Low-resource Word Segmentation and POS Tagging of Ancient Chinese Texts
Yutong Shen, Jiahuan Li, Shujian Huang, Yi Zhou, Xiaopeng Xie and Qinxin Zhao 169

A Joint Framework for Ancient Chinese WS and POS Tagging Based on Adversarial Ensemble Learning
Shuxun Yang 174

Glyph Features Matter: A Multimodal Solution for EvaHan in LT4HALA2022
Wei Xinyuan, liu Weihao, Qing Zong, zhang shao qing and Baotian Hu 178

Overview of the EvaLatin 2022 Evaluation Campaign
Rachele Sprugnoli, Marco Passarotti, Flavio Massimiliano Cecchini, Margherita Fantoli and Giovanni Moretti 183

An ELECTRA Model for Latin Token Tagging Tasks
Wouter Mercelis and Alek Keersmaekers 189

Transformer-based Part-of-Speech Tagging and Lemmatization for Latin
Krzysztof Wróbel and Krzysztof Nowak 193