

CLiC-it 2024

The Tenth Italian Conference on Computational Linguistics

Proceedings of the Conference

December 4-6, 2024

Copyright ©2024 for the individual papers by the papers' authors.
Copyright ©2024 for the volume as a collection by its editors.
This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

These proceedings have been published in the CEUR Workshop Proceedings series.
The original papers are available at: <https://ceur-ws.org/Vol-3878>.
The papers are mirrored in the ACL Anthology.

ISBN 979-12-210-7060-6

Table of Contents

Preface to the CLiC-it 2024 Proceedings

Felice Dell’Orletta, Alessandro Lenci, Simonetta Montemagni and Rachele Sprugnoli 1

Lifeless Winter without Break: Ovid’s Exile Works and the LiLa Knowledge Base

Aurora Alagni, Francesco Mambrini and Marco Passarotti 4

Exploring the Use of Cohesive Devices in Dementia within an Elderly Italian Semi-spontaneous Speech Corpus

Giorgia Albertin and Elena Martinelli 13

SimilEx: The First Italian Dataset for Sentence Similarity with Natural Language Explanations

Chiara Alzetta, Felice Dell’orletta, Chiara Fazzone and Giulia Venturi 20

Data Augmentation for Low-Resource Italian NLP: Enhancing Semantic Processing with DRS

Muhammad Saad Amin, Luca Anselma and Alessandro Mazzei 29

ItaEval and TweetyIta: A New Extensive Benchmark and Efficiency-First Language Model for Italian

Giuseppe Attanasio, Pieter Delobelle, Moreno La Quatra, Andrea Santilli and Beatrice Savoldi 39

LLaMAntino against Cyber Intimate Partner Violence

Pierpaolo Basile, Marco Degemmis, Marco Polignano, Giovanni Semeraro, Lucia Siciliani, Vincenzo Tamburrano, Fabiana Battista and Rosa Scardigno 52

Taking Decisions in a Hybrid Conversational AI Architecture Using Influence Diagrams

Roberto Basile Giannini, Antonio Origlia and Maria Di Maro 59

KEVLAR: The Complete Resource for EuroVoc Classification of Legal Documents

Lorenzo Bocchi, Camilla Casula and Alessio Palmero Aprosio 66

Title Is (Not) All You Need for EuroVoc Multi-Label Classification of European Laws

Lorenzo Bocchi and Alessio Palmero Aprosio 74

Exploring the Dissociated Nucleus Phenomenon in Semantic Role Labeling

Tommaso Bonomo, Simone Conia and Roberto Navigli 81

Data Augmentation through Back-Translation for Stereotypes and Irony Detection

Tom Bourgeade, Silvia Casola, Adel Mahmoud Wizan and Cristina Bosco 90

Community-based Stance Detection

Emanuele Brugnoli and Donald Ruggiero Lo Sardo 98

Towards a Hate Speech Index with Attention-based LSTMs and XLM-RoBERTa

Mauro Bruno, Elena Catanese and Francesco Ortame 106

Written Goodbyes: How Genre and Sociolinguistic Factors Influence the Content and Style of Suicide Notes

Lucia Busso and Claudia Roberta Combei 114

Argument Mining in BioMedicine: Zero-Shot, In-Context Learning and Fine-tuning with LLMs

J  r  mie Cabessa, Hugo Hernault and Umer Mushtaq 122

Multisource Approaches to Italian Sign Language (LIS) Recognition: Insights from the MultiMedaLIS Dataset

Gaia Caligiore, Raffaele Mineo, Concetto Spampinato, Egidio Ragonese, Simone Palazzo and Sabina Fontana 132

<i>Combining Universal Dependencies and FrameNet to Identify Constructions in a Poetic Corpus: Syntax and Semantics of Latin Felix and Infelix in Virgilian Poetics</i>	
Giulia Calvi, Riccardo Ginevra and Federica Iurescia	141
<i>Lost in Disambiguation: How Instruction-Tuned LLMs Master Lexical Ambiguity</i>	
Luca Capone, Serena Auriemma, Martina Miliani, Alessandro Bondielli and Alessandro Lenci	148
<i>BaBIEs: A Benchmark for the Linguistic Evaluation of Italian Baby Language Models</i>	
Luca Capone, Alice Suozzi, Gianluca Leboni and Alessandro Lenci	157
<i>Beyond Headlines: A Corpus of Femicides News Coverage in Italian Newspapers</i>	
Eleonora Cappuccio, Benedetta Muscato, Laura Pollacci, Marta Marchiori Manerba, Clara Punzi, Chandana Mala, Margherita Lalli, Gizem Gezici, Michela Natilli and Fosca Giannotti	171
<i>Women's Professions and Targeted Misogyny Online</i>	
Alessio Cascione, Aldo Cerulli, Marta Marchiori Manerba and Lucia Passaro	182
<i>DWUGs-IT: Extending and Standardizing Lexical Semantic Change Detection for Italian</i>	
Pierluigi Cassotti, Pierpaolo Basile and Nina Tahmasebi	190
<i>History Repeats: Historical Phase Recognition from Short Texts</i>	
Fabio Celli and Valerio Basile	198
<i>Emojilingo: Harnessing AI to Translate Words into Emojis</i>	
Francesca Chiusaroli, Federico Sangati, Johanna Monti, Maria Laura Pierucci and Tiberio Uricchio	205
<i>Towards an ASR System for Documenting Endangered Languages: A Preliminary Study on Sardinian</i>	
Ilaria Chizzoni and Alessandro Vietti	214
<i>Controllable Text Generation to Evaluate Linguistic Abilities of Italian LLMs</i>	
Cristiano Ciaccio, Felice Dell'orletta, Alessio Miaschi and Giulia Venturi	221
<i>A Modal Sense Classifier for the French Modal Verb Pouvoir</i>	
Anna Colli, Diego Rossini and Delphine Battistelli	233
<i>Topic Similarity of Heterogeneous Legal Sources Supporting the Legislative Process</i>	
Michele Corazza, Leonardo Zilli and Monica Palmirani	244
<i>Join Together? Combining Data to Parse Italian Texts</i>	
Claudia Corbetta, Giovanni Moretti and Marco Passarotti	251
<i>Using Large Speech Models for Feature Extraction in Cross-Lingual Speech Emotion Recognition</i>	
Federico D'asaro, Juan José Márquez Villacís, Giuseppe Rizzo and Andrea Bottino	258
<i>Building a Pragmatically Annotated Diachronic Corpus: The DIADIta Project</i>	
Irene De Felice and Francesca Strik Lievers	266
<i>Building CorefLat. a Linguistic Resource for Coreference and Anaphora Resolution in Latin</i>	
Eleonora Delfino, Roberta Leotta, Marco Passarotti and Giovanni Moretti	273
<i>Is Explanation All You Need? An Expert Survey on LLM-generated Explanations for Abusive Language Detection</i>	
Chiara Di Bonaventura, Lucia Siciliani, Pierpaolo Basile, Albert Merono Penuela and Barbara McGillivray	280

<i>Scalable Query Understanding for E-commerce: An Ensemble Architecture with Graph-based Optimization</i>	
Giuseppe Di Fabrizio, Evgeny Stepanov, Ludovico Frizziero and Filippo Tessaro	289
<i>ELIta: A New Italian Language Resource for Emotion Analysis</i>	
Eliana Di Palma	297
<i>Comparing Large Language Models Verbal Creativity to Human Verbal Creativity</i>	
Anca Dinu and Andra Florescu	308
<i>ItGraSyll: A Computational Analysis of Graphical Syllabification and Stress Assignment in Italian</i>	
Liviu Dinu, Ioan-Bogdan Iordache, Simona Georgescu, Alina Maria Cristea and Bianca Guita	316
<i>Generation and Evaluation of English Grammar Multiple-Choice Cloze Exercises</i>	
Nicolò Donati, Matteo Periani, Paolo Di Natale, Giuseppe Savino and Paolo Torroni	325
<i>ReCLAIM Project: Exploring Italian Slurs Reappropriation with Large Language Models</i>	
Lia Draetta, Chiara Ferrando, Marco Cuccarini, Liam James and Viviana Patti	335
<i>You Write like a GPT</i>	
Andrea Esuli, Fabrizio Falchi, Marco Malvaldi and Giovanni Puccetti	343
<i>Constructing a Multimodal, Multilingual Translation and Interpreting Corpus: A Modular Pipeline and an Evaluation of ASR for Verbatim Transcription</i>	
Alice Fedotova, Adriano Ferraresi, Maja Miličević Petrović and Alberto Barrón-Cedeño	349
<i>Exploring YouTube Comments Reacting to Femicide News in Italian</i>	
Chiara Ferrando, Marco Madeddu, Viviana Patti, Mirko Lai, Sveva Pasini, Giulia Telari and Beatrice Antola	356
<i>Automatic Error Detection: Comparing AI vs. Human Performance on L2 Italian Texts</i>	
Irene Fioravanti, Luciana Forti and Stefania Spina	366
<i>Explainability for Speech Models: On the Challenges of Acoustic Feature Selection</i>	
Dennis Fucci, Beatrice Savoldi, Marco Gaido, Matteo Negri, Mauro Cettolo and Luisa Bentivogli	373
<i>Recurrent Networks Are (Linguistically) Better? An (Ongoing) Experiment on Small-LM Training on Child-Directed Speech in Italian</i>	
Achille Fusco, Matilde Barbini, Maria Letizia Piccini Bianchessi, Veronica Bressan, Sofia Neri, Sarah Rossi, Tommaso Sgrizzi and Cristiano Chesi	382
<i>On Cross-Language Entity Label Projection and Recognition</i>	
Paolo Gajo and Alberto Barrón-Cedeño	390
<i>NYTAC-CC: A Climate Change Subcorpus of New York Times Articles</i>	
Francesca Grasso, Ronny Patz and Manfred Stede	403
<i>Task-Incremental Learning on Long Text Sequences</i>	
Natalia Graziuso, Andrea Zugarini and Stefano Melacci	410
<i>The Vulnerable Identities Recognition Corpus (VIRC) for Hate Speech Analysis</i>	
Ibai Guillén-Pacho, Arianna Longo, Marco Antonio Stranisci, Viviana Patti and Carlos Badenes-Olmedo	417
<i>The Self-Contained Italian Negation Test (SCIN)</i>	
Viola Gullace, David Kletz, Thierry Poibeau, Alessandro Lenci and Pascal Amsili	425

<i>La Non Canonica L'hai Studiata? Exploring LLMs and Sentence Canonicity in Italian</i>	
Claudiu Hromei, Danilo Croce, Rodolfo Delmonte and Roberto Basili	431
<i>Enhancing Job Posting Classification with Multilingual Embeddings and Large Language Models</i>	
Hamit Kavas, Marc Serra-Vidal and Leo Wanner	440
<i>Divergent Discourses: A Comparative Examination of Blackout Tuesday and #BlackLivesMatter on Instagram</i>	
Aenne Knierim, Michael Achmann-Denkler, Ulrich Heid and Christian Wolff	451
<i>THAVQA: A German Task-oriented VQA Dataset Annotated with Human Visual Attention</i>	
Moritz Kronberger and Viviana Ventura	459
<i>Are You a Good Assistant? Assessing LLM Trustability in Task-oriented Dialogues</i>	
Tiziano Labruna, Sofia Brenna, Giovanni Bonetta and Bernardo Magnini	470
<i>Comparative Evaluation of Computational Models Predicting Eye Fixation Patterns During Reading: Insights from Transformers and Simpler Architectures</i>	
Alessandro Lento, Andrea Nadalini, Nadia Khlif, Vito Pirrelli, Claudia Marzi and Marcello Ferro	478
<i>Hits or Misses? A Linguistically Explainable Formula for Fanfiction Success</i>	
Giulio Leonardi, Dominique Brunato and Felice Dell'orletta	488
<i>A Novel Multi-Step Prompt Approach for LLM-based Q&As on Banking Supervisory Regulation</i>	
Daniele Licari, Canio Benedetto, Praveen Bushipaka, Alessandro De Gregorio, Marco De Leonardis and Tommaso Cucinotta	496
<i>Lupus Alberto: A Transformer-Based Approach for SLE Information Extraction from Italian Clinical Reports</i>	
Livia Lilli, Laura Antenucci, Augusta Ortolan, Silvia Laura Bosello, Maria Antonietta D'agostino, Stefano Patarnello, Carlotta Masciocchi and Jacopo Lenkiewicz	510
<i>The Lemma Bank of the LiITA Knowledge Base of Interoperable Resources for Italian</i>	
Eleonora Litta, Marco Passarotti, Paolo Brasolin, Giovanni Moretti, Valerio Basile, Andrea Di Fabio and Cristina Bosco	517
<i>Multimodal Chain-of-Thought Prompting for Metaphor Generation</i>	
Sofia Lugli and Carlo Strapparava	523
<i>Leveraging Advanced Prompting Strategies in LLaMA3-8B for Enhanced Hyperpartisan News Detection</i>	
Michele Maggini and Pablo Gamallo Otero	531
<i>Understanding High-complexity Technical Documents with State-of-Art Models</i>	
Bernardo Magnini and Roberto Zanolini	540
<i>Temporal Word Embeddings in the Study of Metaphor Change over Time and across Genres: A Proof-of-concept Study on English</i>	
Veronica Mangiaterra, Chiara Barattieri Di San Pietro and Valentina Bambini	548
<i>Fine-grained Sexism Detection in Italian Newspapers</i>	
Federica Manzi, Leon Weber-Genzel and Barbara Plank	556
<i>Towards a More Comprehensive Evaluation for Italian LLMs</i>	
Luca Moroni, Simone Conia, Federico Martelli and Roberto Navigli	584

<i>A Study on the Soundness of Closed-ended Evaluation of Large Language Models Adapted to the Italian Language</i>	
Elio Musacchio, Lucia Siciliani, Pierpaolo Basile, Edoardo Michielon, Marco Pasqualini, Asia Beatrice Ubaldi and Giovanni Semeraro	600
<i>Understanding the Future Green Workforce through a Corpus of Curricula Vitae from Recent Graduates</i>	
Francesca Nannetti and Matteo Di Cristofaro	612
<i>Exploring Italian Sentence Embeddings Properties through Multi-tasking</i>	
Vivi Nastase, Giuseppe Samo, Chunyang Jiang and Paola Merlo	620
<i>Exploring Syntactic Information in Sentence Embeddings through Multilingual Subject-verb Agreement</i>	
Vivi Nastase, Giuseppe Samo, Chunyang Jiang and Paola Merlo	631
<i>Dynamic Prompting: Large Language Models for Task Oriented Dialog</i>	
Jan Nehring, Akhil Juneja, Adnan Ahmad, Roland Roller and Dietrich Klakow	644
<i>Exploring Text-Embedding Retrieval Models for the Italian Language</i>	
Yuri Noviello and Fabio Tamburini	654
<i>Introducing MultiLS-IT: A Dataset for Lexical Simplification in Italian</i>	
Laura Occhipinti	662
<i>Enhancing Lexical Complexity Prediction in Italian through Automatic Morphological Segmentation</i>	
Laura Occhipinti	670
<i>Measuring Bias in Instruction-Following Models with ItaP-AT for the Italian Language</i>	
Dario Onorati, Davide Venditti, Elena Sofia Ruzzetti, Federico Ranaldi, Leonardo Ranaldi and Fabio Massimo Zanzotto	679
<i>Minerva LLMs: The First Family of Large Language Models Trained from Scratch on Italian Data</i>	
Riccardo Orlando, Luca Moroni, Pere-Lluís Huguet Cabot, Simone Conia, Edoardo Barba, Sergio Orlandini, Giuseppe Fiameni and Roberto Navigli	707
<i>Benchmarking the Semantics of Taste: Towards the Automatic Extraction of Gustatory Language</i>	
Teresa Paccosi and Sara Tonelli	720
<i>Nominal Class Assignment in Swahili: A Computational Account</i>	
Giada Palmieri and Konstantinos Kogkalidis	728
<i>Did Somebody Say 'Gest-IT'? A Pilot Exploration of Multimodal Data Management</i>	
Ludovica Pannitto, Lorenzo Albanesi, Laura Marion, Federica Martines, Carmelo Caruso, Claudia Bianchini, Francesca Masini and Caterina Mauri	735
<i>Confronto tra Diversi Tipi di Valutazione del Miglioramento della Chiarezza di Testi Amministrativi in Lingua Italiana</i>	
Mariachiara Pascucci and Mirko Tavosanis	746
<i>Towards an Automatic Evaluation of (In)coherence in Student Essays</i>	
Filippo Pellegrino, Jennifer Frey and Lorenzo Zanasì	757
<i>MONICA: Monitoring Coverage and Attitudes of Italian Measures in Response to COVID-19</i>	
Fabio Pernisi, Giuseppe Attanasio and Debora Nozza	766
<i>Unraveling the Enigma of SPLIT in Large-Language Models: The Unforeseen Impact of System Prompts on LLMs with Dissociative Identity Disorder</i>	
Marco Polignano, Marco De Gemmis and Giovanni Semeraro	774

<i>The limits of Italian in Reasoning Tasks</i>	
Leonardo Ranaldi, Giulia Pucci, Federico Ranaldi, Elena Sofia Ruzzetti and Fabio Massimo Zanzotto	781
<i>How Far Does the Sequence of Compositions Impact Multilingual Pre-Training?</i>	
Leonardo Ranaldi, Giulia Pucci and Fabio Massimo Zanzotto	796
<i>From 'It's All Greek to Me' to 'Nur Bahnhof Verstehen': An Investigation of mBERT's Cross-Linguistic Capabilities</i>	
Aria Rastegar and Pegah Ramezani	805
<i>Is Sentence Splitting a Solved Task? Experiments to the Intersection between NLP and Italian Linguistics</i>	
Arianna Redaelli and Rachele Sprugnoli	813
<i>From Explanation to Detection: Multimodal Insights into Disagreement in Misogynous Memes</i>	
Giulia Rizzi, Paolo Rosso and Elisabetta Fersini	821
<i>To Click It or Not to Click It: An Italian Dataset for Neutralising Clickbait Headlines</i>	
Daniel Russo, Oscar Araque and Marco Guerini	829
<i>AI vs. Human: Effectiveness of LLMs in Simplifying Italian Administrative Documents</i>	
Marco Russodivito, Vittorio Ganfi, Giuliana Fiorentino and Rocco Oliveto	842
<i>Assessing the Asymmetric Behaviour of Italian Large Language Models across Different Syntactic Structures</i>	
Elena Sofia Ruzzetti, Federico Ranaldi, Dario Onorati, Davide Venditti, Leonardo Ranaldi, Tommaso Caselli and Fabio Massimo Zanzotto	854
<i>Morphological vs. Lexical Antonyms in Italian: A Computational Study on Lexical Competition</i>	
Martina Saccomando, Andrea Zaninello and Francesca Masini	864
<i>Multimodal Attention Is All You Need</i>	
Marco Saioni and Cristina Giannone	873
<i>Assessing Italian Large Language Models on Energy Feedback Generation: A Human Evaluation Study</i>	
Manuela Sanguinetti, Alessandro Pani, Alessandra Perniciano, Luca Zedda, Andrea Loddo and Maurizio Atzori	880
<i>Non Verbis, Sed Rebus: Large Language Models Are Weak Solvers of Italian Rebuses</i>	
Gabriele Sarti, Tommaso Caselli, Malvina Nissim and Arianna Bisazza	888
<i>Leveraging Large Language Models for Fact Verification in Italian</i>	
Antonio Scaiella, Stefano Costanzo, Elisa Passone, Danilo Croce and Giorgio Gambosi	898
<i>A Gentle Push Funziona Benissimo: Making Instructed Models in Italian via Contrastive Activation Steering</i>	
Daniel Scalena, Elisabetta Fersini and Malvina Nissim	909
<i>Subcategorization of Italian Verbs with LLMs and T-PAS</i>	
Luca Simonetti, Elisabetta Jezek and Guido Vetere	921
<i>Unipa-GPT: A Framework to Assess Open-source Alternatives to Chat-GPT for Italian Chat-bots</i>	
Irene Siragusa and Roberto Pirrone	929
<i>Annotation and Detection of Emotion Polarity in 'I Promessi Sposi': Dataset and Experiments</i>	
Rachele Sprugnoli and Arianna Redaelli	940

<i>Complexifying BERT Using LoRA Adapters</i>	
Fabio Tamburini	948
<i>How Do We Counter Hate Speech in Italy?</i>	
Vittoria Tonini, Simona Frenda, Marco Antonio Stranisci and Viviana Patti	955
<i>Nesciun Lengaz Lascià Endò: Machine Translation for Fassa Ladin</i>	
Giovanni Valer, Nicolò Penzo and Jacopo Staiano	967
<i>Neutral Score Detection in Lexicon-based Sentiment Analysis: The Quartile-based Approach</i>	
Marco Vassallo, Giuliano Gabrieli, Valerio Basile and Cristina Bosco	976
<i>Sensitivity of Syllable-Based ASR Predictions to Token Frequency and Lexical Stress</i>	
Alessandro Vietti, Domenico De Cristofaro and Picciau Sara	983
<i>Modelling Filled Particles and Prolongation Using End-to-end Automatic Speech Recognition Systems: A Quantitative and Qualitative Analysis.</i>	
Vincenzo Norman Vitale, Loredana Schettino and Francesco Cutugno	990
<i>Implicit Stereotypes: A Corpus-Based Study for Italian</i>	
Wolfgang Wolfgang Schmeisser-Nieto, Giacomo Ricci, Simona Frenda, Mariona Taule and Cristina Bosco	997
<i>SLIMER-IT: Zero-Shot NER on Italian Language</i>	
Andrew Zamai, Leonardo Rigutini, Marco Maggini and Andrea Zugarini	1005
<i>Harnessing LLMs for Educational Content-Driven Italian Crossword Generation</i>	
Kamyar Zeinalipour, Achille Fusco, Asya Zanollo, Marco Maggini and Marco Gori	1013
<i>Voice Activity Detection on Italian Language</i>	
Shibingfeng Zhang, Gloria Gagliardi and Fabio Tamburini	1024
<i>Topic Modeling for Auditing Purposes in the Banking Sector</i>	
Alessandro Giaconia, Valeria Chiariello and Marco Passarotti	1030
<i>IDRE: AI Generated Dataset for Enhancing Empathetic Chatbot Interactions in Italian Language.</i>	
Simone Manai, Laura Gemme, Roberto Zanolli and Alberto Lavelli	1036
<i>Multimodal Online Manipulation: Empirical Analysis of Fact-Checking Reports</i>	
Olga Uryupina	1043
<i>Life and Death of Fakes: On Data Persistence for Manipulative Social Media Content</i>	
Olga Uryupina	1049
<i>CALAMITA: Challenge the Abilities of Language Models in ITALian</i>	
Giuseppe Attanasio, Pierpaolo Basile, Federico Borazio, Danilo Croce, Maria Francis, Jacopo Gili, Elio Musacchio, Malvina Nissim, Viviana Patti, Matteo Rinaldi and Daniel Scalena	1054
<i>ItaEval: A CALAMITA Challenge</i>	
Giuseppe Attanasio, Moreno La Quatra, Andrea Santilli and Beatrice Savoldi	1064
<i>PERSEID - Perspectivist Irony Detection: A CALAMITA Challenge</i>	
Valerio Basile, Silvia Casola, Simona Frenda and Soda Marem Lo	1074
<i>TRACE-it: Testing Relative clAuses Comprehension through Entailment in ITalian: A CALAMITA Challenge</i>	
Dominique Brunato	1082

<i>MAGNET - MACHines GeNERating Translations: A CALAMITA Challenge</i>	
Mauro Cettolo, Andrea Piergentili, Sara Papi, Marco Gaido, Matteo Negri and Luisa Bentivogli	
	1089
<i>GATTINA - GenerAtion of TiTles for Italian News Articles: A CALAMITA Challenge</i>	
Maria Francis, Matteo Rinaldi, Jacopo Gili, Leonardo De Cosmo, Sandro Iannaccone, Malvina Nissim and Viviana Patti	
	1094
<i>GFG - Gender-Fair Generation: A CALAMITA Challenge</i>	
Simona Frenda, Andrea Piergentili, Beatrice Savoldi, Marco Madeddu, Martina Rosola, Silvia Casola, Chiara Ferrando, Viviana Patti, Matteo Negri and Luisa Bentivogli	
	1106
<i>VeryFIT - Benchmark of Fact-Checked Claims for Italian: A CALAMITA Challenge</i>	
Jacopo Gili, Viviana Patti, Lucia Passaro and Tommaso Caselli	
	1116
<i>AMELIA - Argument Mining Evaluation on Legal documents in ItAlian: A CALAMITA Challenge</i>	
Giulia Grundler, Andrea Galassi, Piera Santin, Alessia Fidelangeli, Federico Galli, Elena Palmieri, Francesca Lagioia, Giovanni Sartor and Paolo Torroni	
	1125
<i>BLM-It - Blackbird Language Matrices for Italian: A CALAMITA Challenge</i>	
Chunyang Jiang, Giuseppe Samo, Vivi Nastase and Paola Merlo	
	1135
<i>DIMMI - Drug InforMation Mining in Italian: A CALAMITA Challenge</i>	
Raffaele Manna, Maria Pia Di Buono and Luca Giordano	
	1144
<i>GITA4CALAMITA - Evaluating the Physical Commonsense Understanding of Italian LLMs in a Multi-layered Approach: A CALAMITA Challenge</i>	
Giulia Pensa, Ekhi Azurmendi, Julen Etxaniz, Begoña Altuna and Itziar Gonzalez-Dios	
	1153
<i>ABRICOT - ABstRactness and Inclusiveness in CONtext: A CALAMITA Challenge</i>	
Giovanni Puccetti, Claudia Collacciani, Andrea Amelio Ravelli, Andrea Esuli and Marianna Bolognesi	
	1161
<i>INVALSI - Mathematical and Language Understanding in Italian: A CALAMITA Challenge</i>	
Giovanni Puccetti, Maria Cassese and Andrea Esuli	
	1168
<i>Termite Italian Text-to-SQL: A CALAMITA Challenge</i>	
Federico Ranaldi, Elena Sofia Ruzzetti, Dario Onorati, Fabio Massimo Zanzotto and Leonardo Ranaldi	
	1176
<i>Multi-IT Multiple Choice Questions on Multiple Topics in Italian: A CALAMITA Challenge</i>	
Matteo Rinaldi, Jacopo Gili, Maria Francis, Mattia Goffetti, Viviana Patti and Malvina Nissim	
	1184
<i>EurekaRebus - Verbalized Rebus Solving with LLMs: A CALAMITA Challenge</i>	
Gabriele Sarti, Tommaso Caselli, Arianna Bisazza and Malvina Nissim	
	1202
<i>GEESE - Generating and Evaluating Explanations for Semantic Entailment: A CALAMITA Challenge</i>	
Andrea Zaninello and Bernardo Magnini	
	1209
<i>ITA-SENSE - Evaluate LLMs' ability for ITALian word SENSE disambiguation: A CALAMITA Challenge</i>	
Pierpaolo Basile, Elio Musacchio and Lucia Siciliani	
	1217
<i>BEEP - BEst DrivEr's License Performer: A CALAMITA Challenge</i>	
Fabio Mercorio, Daniele Potertì, Antonio Serino and Andrea Seveso	
	1222

<i>PejorativITy - In-Context Pejorative Language Disambiguation: A CALAMITA Challenge</i>	
Arianna Muti	1228
<i>MACID - Multimodal Action Identification: A CALAMITA Challenge</i>	
Andrea Amelio Ravelli, Rossella Varvara and Lorenzo Gregori	1234
<i>ECWCA - Educational CrossWord Clues Answering: A CALAMITA Challenge</i>	
Andrea Zugarini, Kamyar Zeinalipour, Achille Fusco and Asya Zanollo	1239

Combining Universal Dependencies and FrameNet to identify constructions in a poetic corpus: syntax and semantics of Latin *felix* and *infelix* in Virgilian poetics

Giulia Calvi¹, Riccardo Ginevra¹ and Federica Iurescia¹

¹ Università Cattolica del Sacro Cuore, 20123 Milano, Italy

Abstract

The paper is a pilot study which argues for a constructionist and computer-based approach to the syntactic and semantic analysis of a poetic corpus in Latin. We focus on the terms *felix* and on its opposite *infelix* and perform manual annotation of their occurrences in Virgil's poems using Universal Dependencies for the syntactic analysis and FrameNet for the semantic one. Integrating the approaches of Dependency Syntax and Construction Grammar, we analyze the linguistic contexts in which the two terms occur and identify the different "constructions" (pairings of form and function) that they instantiate. Our methodology is language-independent and has the potential to aid scholars in the comparative analysis of poetic texts, allowing for the detection of hidden parallels in the style and poetics of different texts and authors.

Keywords

Universal Dependencies, FrameNet, Construction Grammar, Frame Semantics, Latin, Virgil.

1. Introduction

The aim of the present study is to demonstrate the potential of a constructionist and computer-based approach to the analysis of syntax and semantics in a Latin poetic corpus. Our corpus comprises Virgil's (70–19 BCE) literary works, namely (in chronological order of composition) the *Eclogues* (*Ecl.*) or *Bucolics*, the *Georgics* (*Georg.*), and the *Aeneid* (*Aen.*). We focus on two lemmas that have been studied as key terms in Virgil's poetics (e.g. [1]; [2]): *felix* 'productive, auspicious, fortunate, lucky, happy' and its opposite *infelix* 'unproductive, unlucky, ill-fated, miserable'.¹

Bellincioni [1] analyzed the meanings of the two terms in Virgil's works and detected differences in their poetic uses. On the one hand, *felix* is attested in a variety of contexts, ranging from its (likely original) concrete senses 'productive', 'fruitful' to more figurative senses linked with prosperity and well-being (granted by divine will). When it qualifies humans, *felix* takes the religious nuance of 'favored' by gods and fate. Gagliardi [2] also stressed the polysemy of *felix* in the Virgilian corpus: the lemma may refer to fecundity, propitious benevolence,

or happiness, acquiring new connotations thanks to innovative uses in Virgil's poetics. On the other hand, according to Bellincioni [1] *infelix* is rarely used in the technical sense of 'infertile' or in the senses 'helpless' and 'inauspicious', and in the majority of cases it rather seems to be used to qualify human beings as 'ill-fated'.

In order to identify patterns of the use of these terms in context, we combine a syntactic analysis with a semantic one. Following Osborne and Groß [4] and Osborne, Putnam and Groß [5], we integrate the approaches of Dependency Syntax and Construction Grammar. In doing so, we rely on the Universal Dependencies (UD) framework for the syntactic analysis and on the FrameNet approach for the semantic one, drawing inspiration from previous studies along these lines (e.g. [6]; [7]).

This integrated approach allows us, on the one hand, to identify the linguistic contexts in which *felix* and *infelix* occur in Virgil's corpus and, on the other hand, to analyze correspondences between the syntactic and the semantic levels of the Virgilian passages where these two terms are employed.

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 – 06, 2024, Pisa, Italy

✉ giulia.calvi02@unicatt.it (G. Calvi);
riccardo.ginevra@unicatt.it (R. Ginevra);
federica.iurescia@unicatt.it (F. Iurescia)

0000-0002-6731-6494 (R. Ginevra); 0000-0001-5100-5539 (F. Iurescia);



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹ We rely on translations provided by [3].

By combining syntactic and semantic analyses, we explore the potential of an approach that integrates Universal Dependencies with FrameNet. In doing so, we aim at demonstrating that ours is a viable methodology to retrieve the contexts in which the two terms occur in Virgil’s corpus and to study the correspondences between their syntactic and semantic uses.

2. Theoretical framework

2.1. Construction Grammar and Frame Semantics

The term “Construction Grammar” encompasses a series of approaches to grammar, which share the premise that all levels of grammatical analysis involve so-called “constructions”, i.e. “learned pairings of form and function”, including “morphemes or words, idioms, partially lexically filled and fully general phrasal patterns” ([8], p. 5). Within this framework, no rigid division between lexicon and syntax is assumed: constructions are rather arranged along the lexicon-syntax continuum, varying in their degree of internal complexity and schematicity.² The different instances of constructions (i.e. their tokens in a type-token distinction) are called “constructs”.

Construction Grammar is in turn the formal counterpart of Frame Semantics, originally developed by Fillmore [10], which posits that word meanings are understood through the “semantic frames” they evoke. A semantic frame may be defined as “any system of concepts related in such a way that to understand any of them you have to understand the whole structure in which it fits” ([10], p. 111). The presence in a text of words evoking specific frames reveals different ways in which the speaker conceptualizes the situation.

2.2. Dependency Syntax and Universal Dependencies

In order to identify the constructions instantiated by *felix* and *infelix* within Virgil’s corpus, the relevant occurrences were analyzed within the framework of Dependency Syntax. This choice aligns with Osborne and Groß’s [4] claim that Dependency Syntax is more compatible with Construction Grammar’s theoretical

assumptions and practical goals, compared to Phrase Structure (or Constituency) Syntax.³

Osborne, Putnam and Groß ([5], p. 354) introduced the concept of “catena” to refer to “a word or a combination of words that is continuous with respect to dominance”, and proposed to regard it as the fundamental unit of syntax. As argued by Osborne and Groß [4], most constructions discussed within the framework of Construction Grammar can be analyzed as catenae, i.e. as chains of words linked together by dependencies.

Given the high compatibility of Dependency Syntax with Construction Grammar, we adopt the UD framework [12] to perform the syntactic annotation of sentences in Virgil’s corpus which included occurrences of *felix* and *infelix*. The annotation served as a basis for the identification of catenae and of the corresponding constructions.

3. Data and methods

3.1. Corpus and annotation task

Our corpus of Virgil’s texts originates from the *Opera Latina* corpus [13] developed by the LASLA research centre in Liège.⁴ The *Opera Latina* corpus is enhanced with sentence-splitting, tokenization, lemmatization, PoS-tagging and the annotation of morphological features according to a format developed by the LASLA team. The texts in the corpus were converted from the LASLA format into the CoNLL-U format, and into the UD formalism [14].⁵ This textual resource is included among the linguistic resources for Latin that are made interoperable through their linking to the LiLa Knowledge Base.⁶ The interlinking of the *Opera Latina* corpus in the LiLa Knowledge Base allowed us to build upon the existent annotation in order to add a further layer. Thanks to the LiLa Interactive Search Platform (LISP), one of the online services designed to query the Knowledge Base [15],⁷ we were able to retrieve all occurrences of *felix* and *infelix* in Virgil’s works: 90 tokens distributed across 89 sentences (see Table 1 in the Appendix).

The sentences were collected into a separate CoNLL-U file that was then enriched with syntactic annotation, manually performed according to UD guidelines.⁸

² A single expression may instantiate both less complex and phonologically specific constructions (e.g. morphemes, words) and more complex and schematic constructions (e.g. syntactic constructions, such as the transitive one), as long as they may all be analyzed as pairings of form and meaning ([9], p. 7).

³ Constituency Syntax “views the links between the units of sentence structure as indirect” and “mediated by additional groupings that are present as additional nodes in the syntactic structures” ([11], p. 33), in contrast with construction-based approaches, where “no underlying syntactic nor semantic forms are posited” ([8], p. 7).

⁴ Laboratoire d’Analyse Statistique des Langues Anciennes. (https://www.lasla.uliege.be/cms/c_8508894/fr/lasla).

⁵ This conversion process was managed by the CIRCSE research center of the Università Cattolica del Sacro Cuore in Milan (<https://centridiricerca.unicatt.it/circse/en.html>).

⁶ <https://lila-erc.eu/>.

⁷ <https://lila-erc.eu/LiLaLisp/>, https://github.com/CIRCSE/LiLa_LISP.

⁸ This annotation will be released as expansion of the UD_Latin-CIRCSE treebank (https://github.com/UniversalDependencies/UD_Latin-CIRCSE).

3.2. Syntactic analysis and extraction of catenae

In order to detect the main catenae involving *felix* and *infelix* (see Section 2.2), we exploited TüNDRA, a web application for querying treebanks that allows users to upload their own CoNLL-U files.⁹

Table 2 and Table 3 in the Appendix provide an overview of the tokens' distributions according to their dependency relation¹⁰ (deprel) to their heads. Tokens sharing the same deprel were then systematically analyzed to identify recurrent catenae with varying degrees of extension and abstraction. The analysis took into account the relations between each token of *felix* or *infelix* and both the upper and the lower nodes of the trees, starting from the deprel of the token to its head.

In what follows, the identified catenae are conventionally represented using square brackets (as per [11], pp. 60–61), which indicate the degree of dependency between words:

DEPREL₁ [DEPREL₂ [DEPREL₃]]

According to this notation system, dependents are enclosed in more brackets than their head, thus 0 brackets for the root, 1 for its dependents, 2 for their own dependents, and so forth, as in the following example:¹¹

- (1) *Arma virumque cano* ‘Arms and the man I sing’ (*Aen.* I, 1)

[OBJ_{arma} [CONJ_{virum} [CC_{que}]]] ROOT_{cano}

3.3. Semantic analysis and identification of constructions

The instances of the recurrent catenae were then analyzed with respect to their semantic structure. Due to the lack of a resource specifically developed for Latin, the semantic analysis was based on FrameNet,¹² a lexical database of English grounded on Frame Semantics. Within this resource, each frame (e.g. APPLY_HEAT) describes a type of event, relationship or entity, along with the participants involved in it, referred to as “frame elements” (e.g. COOK, HEATING_INSTRUMENT, and FOOD), while the words that evoke a given frame are called “lexical units” (e.g. *cook*, *grill*, and *roast*). For the semantic analysis, an expert manually assigned Latin

lemmas to the same frames as their corresponding English translations.

For each instance of a recurrent catena in the corpus, we identified the semantic frames evoked by the tokens that occur with the same deprel within the catena. In what follows, the correspondences between the syntactic and semantic levels of analysis are illustrated by enhancing the notation of the catenae (as per [6], p. 132 and passim) in order to represent them as constructions, i.e. as form-meaning pairings, where frames are represented by superscripts preceding the lexical units that evoke them:

FRAME.A¹ DEPREL₁ [FRAME.B² DEPREL₂ [FRAME.C³ DEPREL₃]]

For instance, the semantic analysis of (1) would be:

[WEAPON^{OBJ}arma [PEOPLE^{CONJ}virum [CC_{que}]]]
COMMUNICATION_MANNER^{ROOT}cano

4. Results

4.1. Different constructions, different uses

The most recurrent constructions in which *felix* and *infelix* occur allow for the identification of different usages of these two terms in Virgilian poetics. As shown in Table 4 in the Appendix, both *felix* and *infelix* often occur as adjectival modifiers (amod) of a noun, but significant differences exist in their respective uses.

Felix is attested only once as amod of a subject (nsubj).¹³ In 5 out of 17 attestations as amod,¹⁴ *felix* rather occurs as amod of an oblique nominal (obl), i.e. of a non-core argument or adjunct of the verb, in a construction that may denote various entities (winds, tree branches, marriage, death, auspices) and thus evoke various semantic frames:

[WHEATER | PLANTS | FORMING_RELATIONSHIPS | DEATH |
EXPECTATION^{OBL} [AMOD_{felix}]]

In contrast, *infelix* predominantly occurs as amod of a nsubj, i.e. in 22 out of 40 instances. In 12 occurrences the nsubj refers to human characters,¹⁵ but it may also denote other entities.¹⁶ This use can be represented by the construction:

⁹ <https://weblicht.sfs.uni-tuebingen.de/Tundra/about>.

¹⁰ UD provides a list of syntactic relations available at <https://universaldependencies.org/u/dep/index.html>.

¹¹ English translations of Virgil texts are taken from [16].

¹² <https://framenet.icsi.berkeley.edu>.

¹³ *Georg.* I, 345.

¹⁴ *Aen.* III, 118–120; *Aen.* VII, 598–599; *Aen.* XI, 29–33; *Aen.* XII, 819–825; *Georg.* II, 78–82.

¹⁵ *Aen.* XI, 85; *Aen.* X, 730; *Aen.* II, 456; *Aen.* XI, 563; *Aen.* IV, 68; *Aen.* I, 749; *Aen.* IV, 450; *Aen.* XII, 870; *Aen.* I, 712; *Aen.* III, 50; *Aen.* VI, 618; *Aen.* XII, 641.

¹⁶ *Ecl.* V, 37; *Georg.* I, 154; *Georg.* II, 314; *Aen.* XII, 941; *Aen.* II, 772; *Aen.* VI, 521; *Aen.* XII, 608; *Georg.* III, 37; *Georg.* III, 498; *Georg.* II, 198.

[PEOPLE | PLANTS | ARTIFACT | ENTITY | ANIMALS |
POLITICAL_LOCALES] NSUBJ [AMOD *infelix*]

All in all, *infelix* is significantly more frequent than *felix* in our corpus (see Table 1). The distribution of the lemmas in terms of their most frequent dependency relations shows that *felix* tends to modify adjuncts, while *infelix* tends to modify subjects (see Table 4).¹⁷ *Infelix* even occurs with the *nsubj* deprel in 5 occurrences,¹⁸ whereas *felix* never does so.

In what follows we provide two case studies of particularly interesting constructions in which *felix* and *infelix* occur.

4.2. Case study 1: vocative

When *infelix* and *felix* occur as *amod* of a vocative noun or as vocative themselves, they instantiate constructions with different functions, which point to different meanings for the two terms.

As for *infelix*, 4 occurrences attest the following catena:

[X_{verb}¹⁹ [VOCATIVE_{infelix} |²⁰ VOCATIVE
[AMOD_{infelix}]] [OBJ]] [NSUBJ | OBL [DET]]]

- (2) *a, virgo infelix, quae te dementia cepit!*. ‘Ah, unhappy girl, what a madness has gripped you!’ (Ecl. VI, 47)
- (3) *quid loquor? aut ubi sum? quae mentem insania mutat? / infelix Dido, nunc te facta impia tangunt?*. ‘What say I? Where am I? What madness turns my brain? Unhappy Dido, do only now your sinful deeds come home to you?’ (Aen. IV, 595-596)
- (4) “*infelix, quae tanta animum dementia cepit? / non vires alias conversaue numina sentis? / cede deo*”. ‘Unhappy man! How could such frenzy seize your mind? Do you not see the strength is another’s and the gods are changed? Yield to heaven!’ (Aen. V, 465-467)
- (5) *ut stetit et frustra absentem respexit amicum: / “Euryale infelix, qua te regione reliqui?”*. ‘when he halted and looked back in vain for his lost friend. “Unhappy Euryalus, where have I left you?”’ (Aen. IX, 389-390)

All these passages feature a rhetorical interrogative that conveys emotional turmoil (due either to despair or frenzy) experienced by the character addressed with the vocative. In (2), (3), and (4), the verb evokes the frames MANIPULATION or CAUSE_CHANGE, which describe the effect of madness on the state of mind of the vocative’s referent. The corresponding construction may be represented as follows:

[MANIPULATION | CAUSE_CHANGE] X_{verb} [VOCATIVE_{infelix} |
VOCATIVE [AMOD_{infelix}]] [PEOPLE | FEELING] [OBJ]
[MENTAL_PROPERTY] NSUBJ [DET]]]

As for *felix*, it occurs as *amod* of a vocative in two passages:

- (6) *dicite, felices animae, tuque, optime vates, / quae regio Anchisen, quis habet locus? illius ergo / venimus et magnos Erebi tranavimus amnis*. ‘Say, happy souls, and you, best of bards, what land, what place holds Anchises? For his sake are we come, and have sailed across the great rivers of Erebus.’ (Aen. VI, 669-671)
- (7) *ite meae, felix quondam pecus, ite capellae*. ‘Away, my goats! Away, once happy flock!’ (Ecl. I, 74)

Both passages attest a verb (*dicite* and *ite*, evoking the frames STATEMENT and MOTION, respectively) in the 2pl of the imperative present. The command is first addressed to a larger group (PEOPLE and AGGREGATE), evoked by a vocative (*animae* and *pecus*) and described as *felix*. Then, it is addressed to a specific entity within that group (PEOPLE_BY_VOCATION and ANIMALS), also evoked by a vocative (*vates* and *capellae*):

[STATEMENT | MOTION] X_{verb.2pl.imp.}²¹ [PEOPLE | AGGREGATE] VOCATIVE
[AMOD_{felix}] [PEOPLE_BY_VOCATION] CONJ_{vocative} |
[MOTION] CONJ_{verb.2pl.imp.} [ANIMALS] VOCATIVE]]]

This construction is in turn a subtype of a more general construction that also underlies the only instance of *felix* as vocative (8), whose head is a MOTION verb (*vade*) in the 2sg imperative:

¹⁷ With regard to the sentence depth, *infelix* tends to modify subjects with a sentence depth equal to one (ROOT [NSUBJ [AMOD *infelix*]]) in 15 out of 22 tokens), whereas *felix* tends to occur at lower levels of the syntactic tree.

¹⁸ Ecl. VI, 74-81; Aen. VII, 373-377; Aen. IX, 477-481; Aen. X, 424-425; Aen. X, 781-782.

¹⁹ In what follows, we use X to notate an element of the catena that may have any deprel, e.g. *cepit* has the *root* deprel in (2), *mutat* has *conj* in (3), whereas *cepit* and *reliqui* have *ccomp:reported* in (4) and (5), respectively.

²⁰ The pipe symbol within the notation is used to represent the two possible alternatives: *infelix* occurs either as an adjectival modifier of a vocative noun or as vocative itself.

²¹ The verb *dicite* has the *ccomp:reported* deprel in (6)(7) and *ite* has *root* in (7).

[STATEMENT | MOTION_X_{verb.2sg/pl.imp.} [VOCATIVE [AMOD_{felix}] | VOCATIVE_{felix}]]

- (8) ‘*vade*,’ *ait*, ‘*o felix nati pietate*’. ‘**Go forth**,’ he cries, ‘**blest** in your son’s love’ (*Aen.* III, 480)

As shown by these examples, different constructions are instantiated by *felix* and *infelix* when they occur as attributes of a vocative or as vocative themselves. Each construction has a specific function:

- the construction with *infelix* is employed to address the vocative’s referent in a rhetorical interrogative that emphasizes the pathos of the discourse;
- the construction with *felix* is employed to qualify the addressee of a command expressed in the imperative present.

4.3. Case study 2: *infelix Dido*

Infelix is used as epithet of Dido, queen of Carthage, in 8 occurrences within the *Aeneid*.²² In two of these, it instantiates the same complex catena:

ROOT_{verb.3sg.pres.} [NSUB]_{Phoenissa | Dido} [AMOD_{infelix}] [ACL [OBL | OBL:AGENT]] [CONJ]_{verb.3sg.pres.}

- (9) *praecipue infelix pesti deuota futurae/ expleri mentem nequit ardescitque tuendo Phoenissa* “Above all, the unhappy Phoenician, doomed to impending ruin, cannot satiate her soul, but takes fire as she gazes” (*Aen.* I, 712-714)
- (10) *Tum vero infelix fatis exterrita Dido/ mortem orat; taedet caeli convexa tueri* “Then, indeed, awed by her doom, luckless Dido prays for death; she is weary of gazing on the arch of heaven.” (*Aen.* IV, 450-451)

These two examples also attest common semantic features: they introduce the character of Dido, conveying the idea of her predestination to a fate of death and destruction. The passages correspond to critical points in the plot: in (9) Dido falls in love with Aeneas, whereas (10) describes her death. The corresponding construction may be represented as follows:

ROOT [NSUB]_{Phoenissa | Dido} [AMOD_{infelix}] [DESTINY | FEAR_{ACL} [DESTROYING_{OBL} | DESTINY_{OBL:AGENT}]] [EMOTION_HEAT | EXPERIENCER_FOCUSED_EMOTION_{CONJ}]

In both (9) and (10), Dido is the subject, modified not only by the attribute *infelix*, but also by a perfect participle (_{ACL}) that emphasizes her impending doom. More precisely, in (9), *devota* ‘doomed’ evokes the frame DESTINY, specified by the oblique nominal (_{OBL}) *pesti* ‘to ruin’; in (10) *exterrita* ‘awed’ evokes the frame FEAR, whereas DESTINY is evoked by the agent (_{OBL:agent}) *fatis* ‘by her doom’ causing the terror.

Moreover, the coordinated verb (_{CONJ}) in both instances relates to Dido’s emotional state, which is different in the two examples: in (9) *ardescit* ‘takes fire’ marks the beginning of Dido’s love for Aeneas, whereas in (10) *taedet* ‘is weary’ evokes her attitude towards life.

The initial and the final moments of Dido’s story are thus expressed by means of the same catena, evoking her impending ruin. This construction seems to encapsulate the whole thematic arc of Dido’s role in the *Aeneid*, which is framed both at its inception and at its conclusion by a linguistic structure that highlights the inevitability of her fate.

5. Conclusion and future work

With the present study, we show the potential of a constructionist and computer-based approach in the analysis of a poetic corpus in Latin. By integrating syntactic information based on UD with semantic annotation grounded on FrameNet, we were able to identify recurrent constructions involving two key lemmas of Virgilian poetics, *felix* and *infelix*. This enabled us to uncover differences and parallels in the uses of these two terms within Virgil’s language.

The present work is a pilot study which may pave the way for future research. Our approach is language-independent, and may thus be applied to different corpora across various languages and historical periods, for instance to explore similarities in the poetics of various authors within different traditions. Our investigation relied on manual annotation for both the syntactic and semantic analyses due to the lack or poor performance of automatic annotation systems for Latin poetry at the time of writing. The feasibility and effectiveness of such systems can vary significantly across different languages, depending on the resources available. Future improvements in automatic annotation for Latin may allow us to scale up this approach to perform analyses of even larger corpora.

Virgil’s poems played a crucial role in shaping later poetic traditions for centuries: an interesting application of our integrated approach may thus be to investigate whether the same constructions attested in Virgil’s poems also occur in the works of later poets who are

²² *Aen.* I, 712; *Aen.* I, 749; *Aen.* IV, 68; *Aen.* IV, 450; *Aen.* IV, 529; *Aen.* IV, 596; *Aen.* V, 3; *Aen.* VI, 456.

known to have been influenced by him, both in Latin (e.g. Valerius Flaccus’s *Argonautica*, Silius Italicus’s *Punica*, Publius Papinius Statius’s *Thebaid*), as well as in other languages, such as Italian (e.g. Dante Alighieri’s *Commedia*).

References

- [1] Bellincioni, Maria. 1985. *felix/infelix*. In: *Enciclopedia Virgiliana* 2. 486-488. Roma: Istituto dell’Enciclopedia Italiana, Treccani.
- [2] Gagliardi, Paola. 2017. *Beatus, felix, fortunatus: il lessico virgiliano della felicità*. Roma: Carocci.
- [3] *Oxford Latin Dictionary*. 1968. Oxford, Oxford University Press.
- [4] Osborne, Timothy and Groß, Thomas. 2012. Constructions are catenae: Construction Grammar meets Dependency Grammar. *Cognitive Linguistics* 23(1). 165-216.
- [5] Osborne, Timothy, Putnam Michael, Groß Thomas. 2012. Catenae: Introducing a Novel Unit of Syntactic Analysis. *Syntax* 15 (4). 354-396.
- [6] Ginevra, Riccardo, and Francesco Mambrini. "The Old Norse FrameNet (ONoFN): Developing a New Digital Resource for the Study of Semantics and Syntax within a Medieval Germanic Tradition". *Filologia Germanica – Germanic Philology* 14 (2022), 119-140.
- [7] Brigada Villa, Luca, Erica Biagetti, Riccardo Ginevra, and Chiara Zanchi. 2023. Combining WordNets with Treebanks to study idiomatic language: A pilot study on Rigvedic formulas through the lenses of the Sanskrit WordNet and the Vedic Treebank. In *Proceedings of the 12th Global WordNet Conference*. Edited by German Rigau, Francis Bond and Alexandre Rademaker. Donostia-San Sebastian: Global Wordnet Association, pp. 133–39.
- [8] Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- [9] Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- [10] Fillmore, Charles J. 1982. Frame Semantics. In: *The Linguistic Society of Korea (edd.), Linguistics in the morning calm*, 111-137. Seoul: Hanshin Publishing Co.
- [11] Osborne, Timothy. 2019. *A Dependency Grammar of English. An Introduction and Beyond*. Amsterdam: John Benjamins.
- [12] De Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021.

Universal Dependencies. *Computational Linguistics* 47: 255–308.

- [13] Denooz, Joseph. 2004. "Opera Latina: Une Base de Données Sur Internet." *Euphrosyne* 32: 79–88.
- [14] Margherita Fantoli, Marco Passarotti, Francesco Mambrini, Giovanni Moretti, and Paolo Ruffolo. 2022. Linking the LASLA Corpus in the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin. In *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference*, pages 26–34, Marseille, France. European Language Resources Association.
- [15] Marco Passarotti, Francesco Mambrini, and Giovanni Moretti. 2024. The Services of the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin. In *Proceedings of the 9th Workshop on Linked Data in Linguistics @ LREC-COLING 2024*, pages 75–83, Torino, Italia. ELRA and ICCL.
- [16] H. Rushton Fairclough. 1916. *Virgil*. (2 vols.). Cambridge (MA), Harvard University Press.

6. Appendices

Table 1 provides an overview of the tokens’ distribution of *felix* and *infelix* across Virgil’s works:

Table 1
Occurrences of *felix* and *infelix* in Virgil’s works

	<i>Eclogues</i>	<i>Georgics</i>	<i>Aeneid</i>	<i>Total</i>	<i>Relative Frequency</i>
<i>Felix</i>	2	8	21	31	0,00035
<i>Infelix</i>	5	6	48	59	0,00067
<i>Total</i>	7	14	69	90	0,00104

Table 2 and Table 3 provide an overview of the tokens’ distributions according to their (deprel) to their heads ("query:edge" in the table) listed in decreasing order:

Table 2
The deprels of *felix*

query:pos	query:edge	query:lemma	occurrences
ADJ	amod	felix	17
ADJ	conj	felix	4
ADJ	root	felix	3
ADJ	advcl:pred	felix	2
ADJ	acl:relcl	felix	1
ADJ	xcomp	felix	1
ADJ	ccomp:reported	felix	1
ADJ	vocative	felix	1
ADJ	parataxis	felix	1

Table 3
The deprels of *infelix*

query:pos	query:edge	query:lemma	occurrences
ADJ	amod	infelix	40
ADJ	advcl:pred	infelix	7
ADJ	nsubj	infelix	4
ADJ	root	infelix	3
ADJ	vocative	infelix	3
ADJ	parataxis	infelix	1
ADJ	nsubj:pass	infelix	1

Table 4 provides an overview of the most frequent *catenae* for *felix* and *infelix*:

Table 4
Felix and *infelix*'s most frequent *catenae*

	Total	AMOD		NSUBJ	
		[NSUBJ [AMOD]]	[OBL [AMOD]]	OTHER	
<i>Felix</i>	31	1	5	11	/
<i>Infelix</i>	59	22	1	17	5