

CLARIN Annual Conference Proceedings

2023

Edited by

Krister Lindén, Jyrki Niemi, and Thalassia Kontino

16 – 18 October 2023

Leuven, Belgium

Please cite as:

CLARIN Annual Conference Proceedings, 2023. ISSN 2773-2177 (online).

Eds. Krister Lindén, Jyrki Niemi, and Thalassia Kontino.

Leuven, Belgium, 2023.

Programme Committee

Chair:

- Krister Lindén, University of Helsinki (FI)

Members:

- Starkaður Barkarson, Árni Magnússon Institute for Icelandic Studies (IS)
- Lars Borin, University of Gothenburg (SE)
- António Branco, Universidade de Lisboa (PT)
- Tomaž Erjavec, Jožef Stefan Institute (SI)
- Cristina Grisot, University of Zurich (CH)
- Eva Hajičová, Charles University Prague (CZ)
- Monica Monachini, Institute of Computational Linguistics “A. Zampolli” (IT)
- Karlheinz Mörrh, Austrian Academy of Sciences (AT)
- Costanza Navarretta, University of Copenhagen (DK)
- Maciej Piasecki, Wrocław University of Science and Technology (PL)
- Stelios Piperidis, Institute for Language and Speech Processing (ILSP), Athena Research Center (GR)
- Gijsbert Rutten, Leiden University (NL)
- Kiril Simov, IICT, Bulgarian Academy of Sciences (BG)
- Inguna Skadiņa, University of Latvia (LV)
- Koenraad De Smedt, University of Bergen (NO)
- Marko Tadič, University of Zagreb (HR)
- Jurgita Vaičenonienė, Vytautas Magnus University (LT)
- Vincent Vandeghinste, Instituut voor de Nederlandse Taal (Dutch Language Institute), the Netherlands & KU Leuven (BE)
- Tamás Váradi, Research Institute for Linguistics, Hungarian Academy of Sciences (HU)
- Joshua Wilbur, University of Tartu (EE)
- Andreas Witt, University of Mannheim (DE)
- Friedel Wolff, South African Centre for Digital Language Resources, North-West University (ZA)
- Martin Wynne, University of Oxford (UK)

Reviewers:

- Starkaður Barkarson, IS
- Lars Borin, SE
- António Branco, PT
- Tomaž Erjavec, SI
- Cristina Grisot, CH
- Eva Hajičová, CZ
- Krister Lindén, FI
- Monica Monachini, IT
- Karlheinz Mörth, AT
- Costanza Navarretta, DK
- Maciej Piasecki, PL
- Stelios Piperidis, GR
- Gijsbert Rutten, NL
- Kiril Simov, BG
- Inguna Skadiņa, LV
- Koenraad De Smedt, NO
- Marko Tadić, HR
- Jurgita Vaičenonienė, LT
- Vincent Vandeghinste, BE
- Tamás Váradi, HU
- Joshua Wilbur, EE
- Andreas Witt, DE
- Friedel Wolff, ZA
- Martin Wynne, UK

Subreviewers:

- Ilze Auzina, LV
- Federico Boschetti, IT
- Riccardo Del Gratta, IT
- Amelie Dorn, AT
- Maria Gavriilidou, GR
- Luís Gomes, PT
- Marissa Griesel, ZA
- Kinga Jelencsik-Mátyus, HU
- Mateja Jemec Tomazin, SI
- Fahad Khan, IT
- Penny Labropoulou, GR
- László János Laki, HU
- Kristine Levane-Petrova, LV
- Noémi Ligeti-Nagy, HU
- Amália Mendes, PT
- Hannes Pirker, AT
- Valeria Quochi, IT
- João Rodrigues, PT
- Rodrigo Santos, PT
- João Silva, PT
- Juan Steyn, ZA
- Benito Trollip, ZA
- Menno van Zaanen, ZA

CLARIN 2023 submissions, review process and acceptance

- Call for abstracts: 23 January 2023 first call published on CLARIN website, disseminated, and submission system open
- Submission deadline: 28 April 2023
- In total 52 submissions were received and reviewed (three reviews per submission)
- Virtual PC meeting: 9 June 2023
- Notifications to authors: 30 June 2023
- 37 accepted submissions

More details on the paper selection procedure and the conference can be found at <https://www.clarin.eu/event/2023/clarin-annual-conference-2023>.

Table of Contents

Corpora

| | |
|--|----|
| <i>A Spoken Academic Belgian Dutch Corpus</i> Vincent Vandeghinste, Jolien Mathysen, Elke Peters and Patrick Wambacq | 1 |
| <i>NGT-HoReCo and GoSt-ParC-Sign: Two new Sign Language - Spoken Language parallel corpora</i> Mirella De Sisto, Dimitar Shterionov, Lien Soetemans, Vincent Vandeghinste and Caro Brosens .. | 6 |
| <i>Teaching Syntax with Clarin Corpora and Resources</i> Antonio Balvet | 10 |
| <i>Workflows for Semantic Change Research with CLARIN Resource Families</i> Paola Marongiu, Fahad Khan and Barbara McGillivray | 15 |

Infra I

| | |
|--|----|
| <i>Standards Information System for CLARIN Centres and Beyond</i> Piotr Banski and Eliza Margaretha Illig | 20 |
| <i>The CLARIN:EL Infrastructure</i> Maria Gavriilidou, Stelios Piperidis, Dimitris Galanis, Juli Bakagianni, Penny Labropoulou, Athanasia Kolovou, Dimitris Gkoumas, Miltos Deligiannis, Kanella Pouli, Iro Tsiouli, Leon Voukoutis and Katerina Gkirtzou | 25 |
| <i>NB DH-LAB: a Corpus Infrastructure for Social Sciences and Humanities Computing</i> Magnus Breder Birkenes, Lars Johnsen and Andre Kåsen | 30 |

Infra II

| | |
|--|----|
| <i>CORLI CLARIN K Centre: Development and Perspectives</i> Christophe Parisse and Céline Poudat | 35 |
| <i>The SSH Open Marketplace and CLARIN</i> Alexander König, Laure Barbot, Cristina Grisot, Michael Kurzmeier and Edward J. Gray | 39 |
| <i>CLARIN-IT: Texts, Documents and New Contexts</i> Federico Boschetti, Angelo Mario Del Grosso, Riccardo Del Gratta, Francesca Frontini and Monica Monachini | 44 |

Meta Data and Annotation

| | |
|---|----|
| <i>Documenting Corpus Annotation in CMDI: State of Affairs</i> Jakob Lenardič | 48 |
| <i>Do Chatbots Dream of Copyright? Copyright in AI-generated Language Data</i> Pawel Kamocki, Toby Bond, Krister Lindén and Thomas Margoni | 53 |
| <i>Between Lexicon and Grammar: Towards Integrated Valencies for Bulgarian</i> Petya Osenova and Kiril Simov | 57 |

ParlaMint

| | |
|---|----|
| <i>The ParlaMint Project: Ever-growing Family of Comparable and Interoperable Parliamentary Corpora</i> Maciej Ogrodniczuk, Petya Osenova, Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Çağrı Çöltekin, Matyáš Kopp, Katja Meden and Taja Kuzman | 62 |
| <i>Workflow and Metadata Challenges in the ParlaMint Project: Insights from Building the ParlaMint-UA Corpus</i> Anna Kryvenko and Matyáš Kopp | 67 |
| <i>Adding Political Orientation Metadata to ParlaMint Corpora</i> Tomaž Erjavec, Katja Meden and Jure Skubic | 71 |

Tools

| | |
|--|----|
| <i>MATEO: Machine Translation Evaluation for Users and Developers</i> Bram Vanroy | 76 |
| <i>Domain-Specific Languages for Epigraphy: The Case of ItAnt</i> Federico Boschetti, Luca Rigobianco and Valeria Quochi | 80 |
| <i>Finding Dutch Multiword Expressions</i> Jan Odijk, Martin Kroon, Tijmen Baarda, Ben Bonfil and Sheean Spoel | 85 |
| <i>Automatic Anonymization of Human Faces in Images of Authentic Social Interaction: A web application</i> André Frank Krause, Anne Fergner and Karola Pitsch | 90 |

Posters

| | |
|--|-----|
| <i>Building and Consolidating a FAIR-compliant Ecosystem of Infrastructures</i> Noah Bubenhofer, Andrea Malits, Stefanie Strebel, Johannes Graën, Stefan Buerli and Cristina Grisot | 95 |
| <i>The ACoDe Project: Creating a Dementia Corpus for Icelandic</i> Elena Callegari, Agnes Sólmundsdóttir and Anton Karl Ingason | 100 |

| | |
|--|-----|
| <i>A Continous Integration (CI) Workflow for Quality Assurance Checks for Corpora of Multimodal Interaction</i> | |
| Anne Ferger, André Frank Krause and Karola Pitsch | 106 |
| <i>Swissdox@LiRI – a Large Database of Media Articles Made Accessible to Researchers</i> | |
| Johannes Graěn, Igor Mustač, Nikolina Rajović, Jonathan Schaber, Gerold Schneider and Noah Bubenhofer | 111 |
| <i>Dynamically Chaining APIs: from Dracor to TEITOK</i> | |
| Maarten Janssen | 116 |
| <i>Analyses of Information Security Standards on Data Crawled from Company Web Sites using SweClarin Resources</i> | |
| Arne Jönsson, Subhomoy Bandyopadhyay, Svyetlana Pantic Dragisic and Andrea Fried | 120 |
| <i>Linguistic Resources and Tools for Ukrainian: Grounds for Creating a K-center</i> | |
| Olha Kanishcheva and Maria Shvedova | 125 |
| <i>Korpusnik: a Corpus Summarizing Tool for Slovene</i> | |
| Iztok Kosem, Jaka Čibej, Kaja Dobrovoljc and Simon Krek | 129 |
| <i>Sharing the Finnish Dark Web Marketplace Corpus (FINDarC)</i> | |
| Krister Lindén, Teemu Ruokolainen, Lasse Hämäläinen and Tuomas Harviainen | 134 |
| <i>The making of the CLARIN Resource Family for Oral History: Lessons Learned from ‘Voices from Ravensbrück’</i> | |
| Stefania Scagliola, Silvia Calamai, Henk Van Den Heuvel and Christoph Draxler | 140 |
| <i>The LiRI Corpus Platform</i> | |
| Jonathan Schaber, Johannes Graěn, Daniel McDonald, Igor Mustač, Nikolina Rajović, Gerold Schneider and Noah Bubenhofer | 145 |
| <i>Topics in Swedish News on Climate Change: A timeline 2016 - 2023</i> | |
| Maria Skeppstedt | 150 |
| <i>DBBErt: Part-of-Speech Tagging of Pre-Modern Greek Text</i> | |
| Colin Swaelens, Els Lefever and Ilse De Vos | 155 |
| <i>Emotion and Abstractness in Austrian Parliamentary Discourse</i> | |
| Tanja Wissik and Klaus Hofmann | 159 |
| <i>Libraries as Data Infrastructures</i> | |
| Martin Wynne, Andreas Witt, Peter Leinen and Sally Chambers | 164 |
| <i>A Multilingual Database for Icelandic L2 Flashcards</i> | |
| Xindan Xu, Þórunn Arnardóttir and Anton Karl Ingason | 168 |

Developing Manually Annotated Corpora for Teaching and Learning Purposes of Brazilian Portuguese, Dutch, Estonian, and Slovene (the CrowLL Project)

Tanara Zingano Kuhn, Carole Tiberius, Špela Arhar-Holdt, Kristina Koppel, Iztok Kosem, Rina Zviel Girshin and Ana R. Luís 173

Domain-Specific Languages for Epigraphy: the Case of ItAnt

Federico Boschetti

CNR-ILC

URT Venezia, Italy

federico.boschetti@ilc.cnr.it

Luca Rigobianco

Dipartimento di Studi Umanistici

Università Ca' Foscari Venezia

Venezia, Italy

luca.rigobianco@unive.it

Valeria Quochi

CNR-ILC

Pisa, Italy

valeria.quochi@ilc.cnr.it

Abstract

ItAnt is a research project devoted to the languages and cultures of ancient Italy witnessed by a digital collection of inscriptions. This contribution illustrates how the definition of a Domain-Specific Language can support the activity of the epigraphists involved in the project by increasing the human readability of the encoded data without sacrificing the compliance to standard models and formats. Finally, an example of concrete use of the encoded texts within the CLARIN-IT DigItAnt platform will be briefly described.

1 Introduction

The project *Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models* (ItAnt hereafter) is an initiative funded by the Italian Ministry of University and Research and involves a consortium comprising the Ca' Foscari University of Venice, the University of Florence, and the Institute for Computational Linguistics "A. Zampolli" of the National Research Council of Italy¹. This project aims at investigating the languages of Ancient Italy combining the methods of historical linguistics with digital technologies specifically designed to create a set of interrelated resources, particularly critical digital editions of inscriptions, lexica and bibliographies.

With the sole exception of Roman Latin, the languages of Ancient Italy (8th century BC-1st century AD) are fragmentary languages, that is to say dead languages attested through a highly restricted corpus of texts. Specifically, their evidence consists almost exclusively of epigraphic texts, which often present problems relating to the reading, segmentation into words, linguistic analysis, and interpretation. Therefore, one of the key challenges of the ItAnt project is to adapt the digital tools, practices, and methodologies of digital epigraphy and computational lexicography to the highly fragmentary nature of such a documentation.

The main objectives of the project are to create and interlink one another a digital archive of (critical editions of) inscriptions and a multilingual computational lexicon, as well as to encode project information using CIDOC CRM and its extensions, namely CRMtex², CRMInf³, and FRBRoo/LRMoo⁴. With regard to the digital archive, the inscriptions are being encoded in XML according to the XML-TEI/EpiDoc schema⁵. Furthermore, the edition of the inscriptions is enriched with standard metadata, thus allowing for an accurate description of each of them as both a linguistic and a material object.

2 The Digital Edition of the Inscriptions

As mentioned above, the project envisages the inscriptions being encoded according to the XML-TEI/EpiDoc schema. Such a schema is the result of an international effort aimed at customising the Text Encoding Initiative's standard for the representation of ancient documents according to the Leiden

¹This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

²<https://www.prin-italia-antica.unifi.it/>

³<https://www.cidoc-crm.org/crmtext/home-8>

⁴<https://www.cidoc-crm.org/crmInf/>

⁵<https://cidoc-crm.org/frbroo/>

⁶<http://www.stoa.org/epidoc/g1/latest/>

Conventions. In particular, XML-TEI/EpiDoc provides mark-ups for the text (edition, apparatus, translation, commentary, bibliography) as well as the materiality and history of the object on which the text appears (repository, support, layout, hand, place and date of origin, provenance). Furthermore, thanks to the extensibility of XML and the versatility of XML-TEI/EpiDoc, ItAnt has proposed solutions for managing specific issues arising from the nature of the languages of Ancient Italy as fragmentary languages and their specific epigraphic features, with particular reference to word division, shape and reuse of the support, opisthography, line dimensions, description of linguistic elements, and description of languages and scripts (Murano et al., 2023).

With the goal of data integration, ItAnt makes use of widely used vocabularies and gazetteers, in particular The Art and Architecture Thesaurus provided by the The Getty Research Institute for the object type ⁶, material, and writing technique ⁷, the EAGLE vocabulary for the type of inscriptions (dedicatory, funerary, etc.)⁸, and Pleiades and GeoNames for ancient and modern names respectively⁹. In addition, Trismegistos IDS are used, when available, to identify the text¹⁰ and bibliographical records are also linked through a specific library built up by using Zotero¹¹.

3 How ItAntDSL Facilitates the Encoding

Encoding epigraphic contextual metadata and textual data in XML-TEI/EpiDoc is a complex, error-prone task. Indeed, XML-TEI is quite verbose (because element names, attributes and values must be written in full) and redundant (because opening and closing tags repeat the element names). The percentage of informative and structural contents is unbalanced. XML-TEI ensures data interchange among software applications and promotes machine actionability and interpretability, but human readability of an encoded document decreases rapidly as complexity increases.

In ItAnt linguistic, philological and prosopographical data are highly entangled. Each word is associated to its part of speech, conjectural integrations to textual gaps (lacunae) are recorded, and named entities are identified. These chunks of information often overlap: for instance a lacuna in a line of text may extend between the end of the third token and the beginning of the fourth one, whereas a named entity defined by *praenomen* (partially conjectured), *gentilicium* and *patronymicus* may extend from the fourth to the sixth token.

The problem of overlapping hierarchies in TEI is well-known and many solutions are available, both through manual encoding of stand-off annotations in XML (Spadini & Turska, 2019) and through alternative representations (e.g. in json), currently or planned to be convertible in XML-TEI (Neill & Schmidt, 2021). An experimental solution adopted in ItAnt for encoding part of the corpus, is based on a domain-driven approach, which involves the epigraphists to co-design a Domain-Specific Language (Parr, 2009), named ItAntDSL, to encode data and metadata. The aims of this approach are twofold: a) optimising the human readability of the encoded documents, both during manual encoding and for subsequent uses of the documents, and b) complying with the EpiDoc abstract model. As shown in Fig. 1, in fact, the encoding of contextual metadata (on the left) and textual data (on the right) is very compact. ItAntDSL is defined by a Context-Free Grammar (CFG) available on github¹². The documents encoded in ItAntDSL are parsed by ANTLR (Parr, 2013), which first converts the Domain-Specific Language into XML with a proprietary schema (XML-ItAnt), based on the production rules of the CFG.

Then, a chain of xquery scripts and XSLT stylesheets transforms XML-ItAnt documents into XML-TEI/EpiDoc documents. The transformations are not limited to the translation of element names and to structural modifications, but extend to the integration of a) automatically generated IDs; b) default values omitted in ItAntDSL documents; c) expansion of complex structured data encoded in ItAntDSL

⁶The iDAI.thesauri provided by the Deutsches Archäologisches Institut (<http://thesauri.dainst.org/de.html>) is used as a supplement with regard to natural supports such as cliffs.

⁷<https://www.getty.edu/research/tools/vocabularies/aat/>

⁸<https://www.eagle-network.eu/resources/vocabularies/typeins/>

⁹<https://pleiades.stoa.org/>; <https://www.geonames.org/>

¹⁰<https://www.trismegistos.org/tm/>

¹¹<https://www.zotero.org/groups/2552746/>

¹²<https://github.com/CoPhi/itantdsl/>

Figure 1: ItAntDSL

documents by reference (between quotation marks) and retrieved from the XML documents stored in an eXist-db. A sample of the final result is shown in Fig. 2.

```

145 <tei:div type="edition" subtype="interpretative" xml:space="preserve">
146 <tei:div type="textpart" n="face_a" style="text-direction:r-to-l" rend="ductus:sinistrorse">
147 <tei:ab>
148 <tei:lb n="1" xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1"/>
149 <tei:w xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1_w_1">
150 <tei:expan><tei:abbr><tei:supplied reason="lost" evidence="previouseditor">m</tei:supplied></tei:abbr><tei:ex>ediküd</tei:ex></tei:expan>
151 </tei:w>
152 <tei:w xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1_w_2">
153 <tei:expan><tei:abbr><tei:supplied reason="lost" evidence="previouseditor">t</tei:supplied></tei:abbr><tei:ex>üvtküüd</tei:ex></tei:expan>
154 </tei:w>
155 <tei:name type="praenomen" xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1_w_3" ref="#p1">
156 <tei:expan><tei:abbr><tei:supplied reason="lost" evidence="previouseditor">n</tei:supplied></tei:abbr><tei:ex>umsüüd</tei:ex></tei:expan>
157 </tei:name>
158 <tei:name type="gentilicium" xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1_w_4" ref="#p1">
159 <tei:unclear>de</tei:unclear>
160 <tei:supplied reason="lost" evidence="previouseditor">kiti</tei:supplied>
161 <tei:unclear>d</tei:unclear><tei:supplied reason="lost" evidence="previouseditor">d</tei:supplied>
162 </tei:name>
163 <tei:name type="patronymic" xml:lang="osc-Ital-x-oscetr" xml:id="Osc_3_1_1_w_5" ref="#p1">
164 <tei:expan><tei:abbr><tei:supplied reason="lost" evidence="previouseditor">m</tei:supplied></tei:abbr><tei:ex>inisiis</tei:ex></tei:expan>
165 </tei:name>
166 <tei:pc unit="word">.</tei:pc>
167 </!- ... -!>

```

Figure 2: XML-TEI/EpiDoc

4 ItAnt and CLARIN-IT: a Concrete Use-Case

ItAnt is developing a user-friendly web platform, DigItAnt, for creating, exploring and querying LOD-compliant lexica natively interlinked with critical editions of inscriptions, citations and bibliographic references, plus other external available salient resources. The DigItAnt editing application (*EpiLexO*) is meant to be useful especially for encoding lexical information of ancient languages and in assisting scholars in linking it to other relevant (re-)sources according to the semantic web principles. Particularly important and central to the platform is linking to the texts encoded in digital scholarly editions of relevant inscriptions in TEI-EpiDoc. The editions encoded with ItAntDLS and then converted to EpiDoc-XML as described above, are subsequently ingested by the platform, for linking to the lexicon and searching into the exploration platform (see Quochi, Bellandi, Mallia, et al. (2022) for reference). In details, the (historical linguist) user uploads one or more EpiDoc XML documents into the platform so that (s)he can then link the exact text loci to either existing or newly created lexical items (see Fig. 3 and Quochi, Bellandi, Khan, et al. (2022)).

Thanks to the EpiDoc-XML encoding, visualisation of contextual information as well as of the text according to the Leiden conventions is also possible, both in the Epilexo editor and in the exploration platform (*EpiLexO-search*, see an example in Fig. 4)

The platform is supported by CLARIN-IT and will become one of its services once the project is completed. The editing components of the platform are ready and in use within the project¹³, while the

¹³The code is available at <https://github.com/DigItAnt/Epilexo>

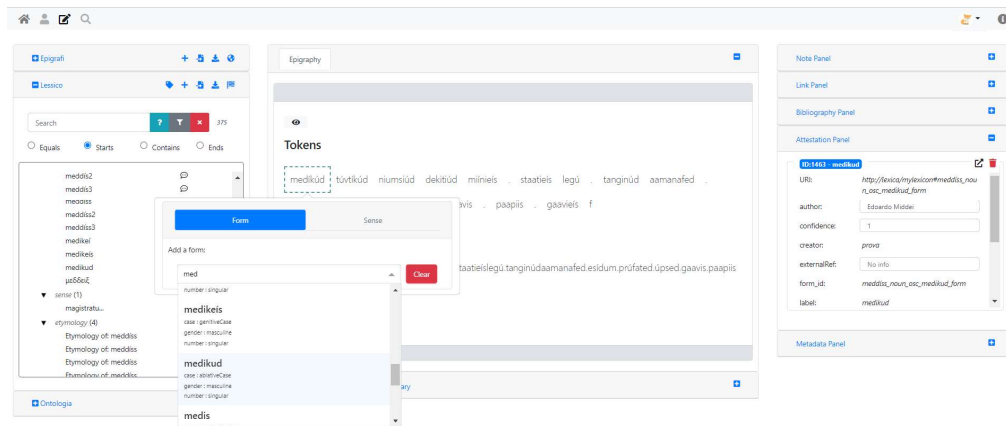


Figure 3: Linking texts loci to lexical forms in DigItAnt - EpiLexO

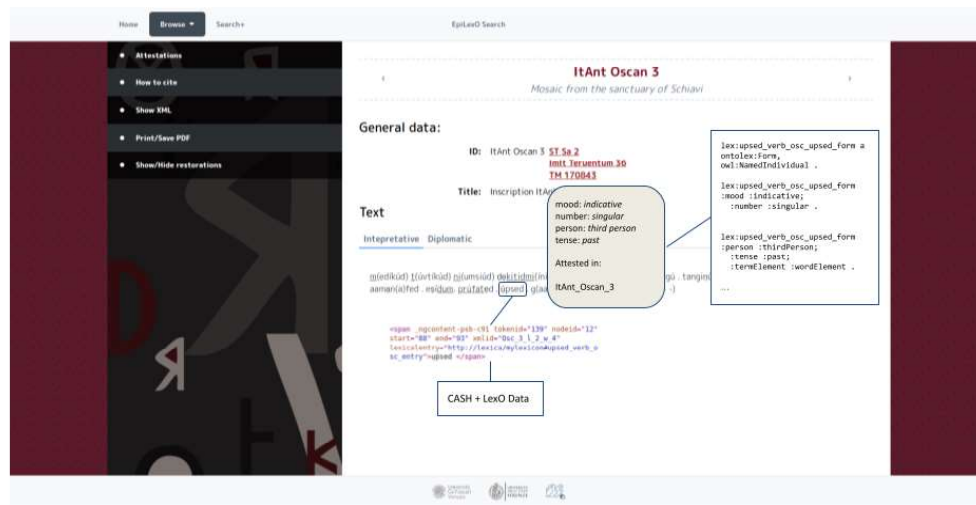


Figure 4: Linking texts loci to lexical forms in DigItAnt - EpiLexO

exploration platform is still in progress, continuously upgraded with new functionalities and improvements. An alpha version will be shown at the conference. Furthermore, as concerns the relations with CLARIN, all project resources, data and tools will be deposited into the ILC4CLARIN repository¹⁴, in order to ensure their FAIRness, long-term preservation and maximal exploitability by the community¹⁵.

Finally, the know-how related to the annotation of inscriptions through ItAntDSL will be shared through the Digital and Public Textual Scholarship Knowledge Centre¹⁶ (DiPText-KC) of CLARIN.

¹⁴<https://ilc4clarin.ilc.cnr.it/>

¹⁵Currently, we deposited the first versions of ItAntDSL (<http://hdl.handle.net/20.500.11752/ILC-1003>) and EpiLexO (<http://hdl.handle.net/20.500.11752/ILC-1004>). At the end of the project all software, inscription corpora and ancient lexicons will be preserved, discoverable and consumable via CLARIN channels.

¹⁶<https://diptext-kc.clarin-it.it/>

Videotutorials and other initiatives, such as webinars and workshops, are planned towards the end of the project and after.

5 Conclusion

VeDPH, CNR-ILC and ILC4CLARIN in the last years are collaborating on DH projects related to many kinds of resources, such as collections of literary texts (Boschetti et al., 2021) and collections of epigraphic sources (Vagionakis et al., 2022). ItAnt provides a good opportunity to develop methods and tools to facilitate the encoding activities of the epigraphists, which must deal with complex entangled data. CLARIN provides not only the infrastructure to deposit the research data, but also the instruments to share new practices adequate to the domain of the epigraphic studies.

Acknowledgments

This work is supported by the Italian Ministry of the University and Research with the Italian National Strategic Research Grant PRIN 2017XJLE8J for the project: *Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models*. The project is also supported by and contributing to CLARIN-IT.

References

- Boschetti, F., Del Grosso, A. M., & Spinazzè, L. (2021). La galassia musisque deoque: Storia e prospettive. In *Paulo maiora canamus - raccolta di studi per paolo mastandrea* (pp. 405–419). Edizioni CaFoscari. <https://edizionicafoscari.unive.it/media/pdf/books/978-88-6969-558-2/978-88-6969-558-2-ch-26.pdf>
- Murano, F., Quochi, V., Del Grosso, A. M., Rigobianco, L., & Zinzi, M. (2023). Describing Inscriptions of Ancient Italy. The ItAnt Project and Its Information Encoding Process. *Journal on Computing and Cultural Heritage*, 16, 1–14. <https://doi.org/10.1145/3593431>
- Neill, I., & Schmidt, D. (2021). SPEEDy. A Practical Editor for Texts Annotated with Standoff Properties. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, 15, 45.
- Parr, T. (2009). Language implementation patterns: create your own domain-specific and general programming languages. *Language Implementation Patterns*, 1–380.
- Parr, T. (2013). The definitive ANTLR 4 reference. *The Definitive ANTLR 4 Reference*, 1–326.
- Quochi, V., Bellandi, A., Khan, F., Mallia, M., Murano, F., Piccini, S., Rigobianco, L., Tommasi, A., & Zavattari, C. (2022). From Inscriptions to Lexicon and Back: A Platform for Editing and Linking the Languages of Ancient Italy. *Proceedings of Second Workshop on Language Technologies for Historical and Ancient Languages LT4HALA 2022*, 59–67.
- Quochi, V., Bellandi, A., Mallia, M., Tommasi, A., & Zavattari, C. (2022). Supporting Ancient Historical Linguistics and Cultural Studies with EpiLexO. *CLARIN Annual Conference Proceedings*, 39.
- Spadini, E., & Turska, M. (2019). XML-TEI Stand-off Markup: One Step Beyond. *Digital Philology: A Journal of Medieval Cultures*, 8(2), 225–239.
- Vagionakis, I., Del Gratta, R., Boschetti, F., Baroni, P., Del Grosso, A. M., Mancinelli, T., & Monachini, M. (2022). ‘Cretan Institutional Inscriptions’ Meets CLARIN-IT. *CLARIN Annual Conference*, 139–150.