ANNUAL CONFERENCE 2018
Pisa, Italy

CLARIN

CLARIN ANNUAL CONFERENCE 2018 IS ORGANISED BY

CLARIN ERIC
Common Language Resources and
Technology Infrastructure
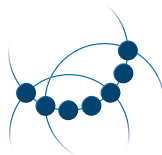
IN COLLABORATION WITH

The Institute for Computational Linguistics (ILC) that is part of the Department of Social Science and Humanities, Cultural Heritage (DSU) of the National Research Council of Italy (CNR)

Connect with CLARIN

Website: **www.clarin.eu**

Email: **clarin@clarin.eu**

Twitter: **@CLARINERIC** | CLARIN Annual Conference 2018 hashtag: **#CLARIN2018**

Facebook: **www.facebook.com/CLARINERIC**

YouTube: CLARIN ERIC

VideoLectures.NET: **www.videolectures.net/clarin**

GitHub: **www.github.com/CLARIN-ERIC**

Key to the letter codes in the programme:

- 🟡 CLARIN in relation to other infrastructures

- 🟣 Language resources and tools

- 🟠 Knowledge infrastructure and dissemination

- 🌸 Multimedia & multimodality

- 🟢 Web services and applications for SSH users

- 🔵 Access, cataloguing and browsing

- 🟢 Interoperability and technical issues

- 🔴 Design and construction of the CLARIN infrastructure

- 🟤 Restricted access and IPR issues

09:00-10:30	**Committee meetings I** (members and invitees only)	
	CLARIN National Coordinators' Forum (NCF)	Vip Hall
	Centre Assessment Committee (CAC)	Mig Hall
10:30-11:00	Coffee break	Gallilei Foyer
11:00-13:00	**Committee meetings II** (members and invitees only)	
	CLARIN National Coordinators' Forum (NCF)	Vip Hall
	User Involvement Group	Tornado Hall
	Standing Committee on CLARIN Technical Centres (SCCTC)	Mig Hall
	Taskforce CMDI + Metadata Curation	Lear Hall
13:00-14:00	Lunch	Restaurant
14:00-16:00	**Committee meetings III** (members and invitees only)	
	Knowledge Sharing Infrastructure Committee (KSI)	Lear Hall
	CLARIN Standards Committee (CSC)	Mig Hall
	CLARIN Legal Issues Committee (CLIC)	Vip Hall
	Taskforce PID	Concorde Hall
	Taskforce FCS + AAI	Concorde Hall

14:00-16:00	**Seminar** (registered participants only)	
	CLARIN Social Media Session	Tornado Hall
	Nataša Briški (Meta's list, Slovenia)	
16:00-16:30	Coffee break	Gallilei Foyer
16:30-16:50	**CLARIN 2018 opening ceremony**	Concorde Hall
	Franciska de Jong (Executive Director)	
	Inguna Skadiņa (Programme Chair)	
	Monica Monachini (Local Organizer)	
16:50-17:40	**Invited talk I**	Concorde Hall
	Chair: Inguna Skadiņa	
	CLAMS: Computational Linguistic Applications for Multimedia Services Image Archival Data	
	James Pustejovsky (Brandeis University, USA)	

17:45-18:45   **PARALLEL SESSIONS**   Concorde Hall /Tornado Hall

1: CLARIN in Relation to Other Infrastructures and Projects
     Chair: Lars Borin

Using Linked Data Techniques for Creating an IsiXhosa Lexical Resource - a Collaborative Approach.
Thomas Eckart, Bettina Klimek, Sonja Bosch and Dirk Goldhahn.

A Platform for Language Teaching and Research (PLT&R).
Maria Stambolieva, Valentina Ivanova and Mariyana Raykova.

Curating and Analyzing Oral History Collections.
Cord Pagenstecher.

2: CLARIN Knowledge Infrastructure, Legal Issues and Dissemination
     Chair: Kadri Vider

New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure.
Pawel Kamocki, Erik Ketzan, Julia Wildgans and Andreas Witt.

Processing personal data without the consent of the data subject for the development and use of language resources.
Aleksei Kelli, Krister Lindén, Kadri Vider, Pawel Kamocki, Ramūnas Birštonas, Silvia Calamai, Chiara Kolletzek, Penny Labropoulou and Maria Gavrilidou.

Toward a CLARIN Data Protection Code of Conduct.
Pawel Kamocki, Erik Ketzan, Julia Wildgans and Andreas Witt.

19:30   Shuttle bus departure to the Conference welcome reception location

20:00   Welcome reception (drinks + food): Villa Comel

09:00-09:40	**State of the Infrastructure** CLARIN ERIC Board of Directors	Concorde Hall
09:40-10:00	**Presentation by CLARIN Italy**	Concorde Hall
10:00-10:30	Coffee break	Gallilei Foyer

Thematic session: Multimedia, Multimodality, Speech

10:30-11:20	**Invited talk II** **Chair: Maciej Piasecki**	Concorde Hall

Speech and gestures: computational linguistic studies.
Costanza Navarretta (University of Copenhagen, Denmark)

11:20-13:00	**THEMATIC PAPERS** **Chair: Jan Odijk**	Concorde Hall

EXMARaLDA meets WebAnno.
Steffen Remus, Hanna Hedeland, Anne Ferger, Kristin Bührig and Chris Biemann.

Human-human, human-machine communication: on the HuComTech multimodal corpus.
Laszlo Hunyadi, Tamás Váradi, István Szekrényes, György Kovács, Hermina Kiss and Karolina Takács.

8

Oral History and Linguistic Analysis. A Study in Digital and Contemporary European History.
Florentina Armaselu, Elena Danescu and François Klein.

The Acorformed Coprus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions.
Magalie Ochs, Philippe Blache, Grégoire Montcheuil, Jean-Marie Pergandi, Roxane Bertrand, Jorane Saubesty, Daniel Francon and Daniel Mestre.

Media Suite: Unlocking Archives for Mixed Media Scholarly Research.
Roeland Ordelman, Liliana Melgar, Carlos Martinez-Ortiz and Julia Noordegraaf.

13:00-14:30	Lunch	Restaurant
14:30-15:00	**PhD students: poster slam** **Chair: Darja Fišer**	Concorde Hall

Automatic genre identification with machine learning methods.
Andrea Dömötör.

Improving OCR of historical newspapers and journals published in Finland by adding Swedish training data.
Senka Drobac.

Utilising Large Quantities of Found Audio Data.
Per Fallgren.

Dramatic Languages: Foreign Languages in the Writings of Peter Handke.
Vanessa Hannesschläger.

9

Continued from previous page

Instrument of parliamentary discourse analysis – Saeima debate corpus.
Aija Kažoka.

Automatization of Detection of Information-Dense Texts.
Danguolė Kalinauskaitė.

Revealing relationships between Old Norse texts using computer-assisted methods.
Katarzyna Anna Kapitan.

Speech Analysis in the Clarin-PL Project.
Mariusz Kleć.

Corpus-driven conversational agents: tools and resources for multimodal dialogue systems development.
Maria Di Maro.

Archival Dynamics : the Langues de France project and the building of the Judeo-Spanish Oral Archive (JSOA).
Panteleimon Mavrogiannis.

Modeling Lexical Knowledge for Natural Language Processing.
Alexander Popov.

Completing the BLARK for Portuguese with finely-tuned Distributional Semantic Models.
João Rodrigues.

10

Studying the ins and outs of external possession with the CLARIN infrastructure.
Jolien Scholten.

Word Embeddings for Cross-Language Learning in Low-Resource Languages.
Tadej Škvorc.

Automatic Collocation Identification Using Word Embeddings.
Yana Strakatova.

Using the Dutch Parallel Corpus to Calculate English Dutch Word Translation Entropy.
Bram Vanroy.

Regional variation in spoken Russian.
Benedikte Vardøy.

15:00-16:30	**PhD students: poster session**	Gallilei Foyer
16:30-17:00	Coffee break	Gallilei Foyer
17:00-17:40	**Short reports on Committee meetings (by the respective chairs)**	Concorde Hall
17:40-18:40	**CLARIN Bazaar**	Gallilei Foyer
19:30	Shuttle bus departure to the Conference dinner location	
20:00	Conference dinner: Stazione Leopolda	

11

09:00-10.40   **PARALLEL SESSIONS**   Concorde Hall /Tornado Hall

3: Use of the CLARIN infrastructure
Chair: Antonio Branco

From Language Learning Platform to Infrastructure for Research on Language Learning.
David Alfter, Lars Borin, Ildikó Pilán, Therese Lindström Tiedemann and Elena Volodina.

Bulgarian Language Technology for Digital Humanities: a focus on the Culture of Giving for Education.
Kiril Simov and Petya Osenova.

Multilayer Corpus and Toolchain for Full-Stack NLU in Latvian.
Normunds Grūzītis and Artūrs Znotiņš.

(Re-)Constructing "public debates" with CLARIAH MediaSuite tools in print and audiovisual media.
Berrie van der Molen, Jasmijn van Gorp and Toine Pieters.

Improving Access to Time-Based Media through Crowdsourcing and CL Tools: WGBH Educational Foundation and the American Archive of Public Broadcasting.
Karen Cariani and Casey Davis-Kaufman.

4: Design and construction of the CLARIN infrastructure
Chair: Tomaž Erjavec

Discovering software resources in CLARIN.
Jan Odijk.

Towards a protocol for the curation and dissemination of vulnerable people archives.
Silvia Calamai, Chiara Kolletzek and Aleksei Kelli.

Versioning with Persistent Identifiers.
Martin Matthiesen and Ute Dieckmann.

Interoperability of Second Language Resources and Tools.
Elena Volodina, Maarten Janssen, Therese Lindström Tiedemann, Nives Mikelic Preradovic, Silje Karin Ragnhildstveit, Kari Tenfjord and Koenraad de Smedt.

Tweak Your CMDI Forms to the Max.
Rob Zeeman and Menzo Windhouwer.

10:40-11:00   Coffee break   Gallilei Foyer

11:00-11:30   **Poster slam**   Concorde Hall
                     Chair: Kiril Simov

CLARIN Data Management Activities in the PARTHENOS Context.
Marnix van Berchum and Thorsten Trippel.

Integrating language resources in two OCR engines to improve processing of historical Swedish text.
Dana Dannélls and Leif-Jöran Olsson.

Looking for hidden speech archives in Italian institutions.
Vincenzo Galatà and Silvia Calamai.

Setting up the PORTULAN / CLARIN centre.
Luís Gomes, Frederico Apolónia, Ruben Branco, João Silva and António Branco.

LaMachine: A meta-distribution for NLP software.
Maarten van Gompel and Iris Hendrickx.

XML-TEI-URS: using a TEI format for annotated linguistic ressources.
Loïc Grobol, Frédéric Landragin and Serge Heiden.

Visible Vowels: a Tool for the Visualization of Vowel Variation.
Wilbert Heeringa and Hans Van de Velde.

ELEXIS - European lexicographic infrastructure.
Milos Jakubicek, Iztok Kosem, Simon Krek, Sussi Olsen
and Bolette Sandford Pedersen.

Sustaining the Southern Dutch Dialects: the Dictionary of the Southern Dutch Dialects (DSDD) as a case study for CLARIN and DARIAH.
Van Keymeulen Jacques, Sally Chambers, Veronique De Tier, Jesse de Does, Katrien Depuydt, Tanneke Schoonheim, Roxane Vandenberghe
and Lien Hellebaut.

SweCLARIN – Infrastructure for Processing Transcribed Speech.
Dimitrios Kokkinakis, Kristina Lundholm Fors and Charalambos Themistokleous.

TalkBankDB: A Comprehensive Data Analysis Interface to TalkBank.
John Kowalski and Brian MacWhinney.

L2 learner corpus survey – Towards improved verifiability, reproducibility and inspiration in learner corpus research.
Therese Lindström Tiedemann, Jakob Lenardič and Darja Fišer.

DGT-UD: a Parallel 23-language Treebank.
Nikola Ljubešić and Tomaž Erjavec.

Continued from previous page

11:00-11:30	**Poster slam**	Concorde Hall

DI-ÖSS - Building a digital infrastructure in South Tyrol.
Verena Lyding, Alexander König and Elisa Gorgaini.

Linked Open Data and the Enrichment of Digital Editions: the Contribution of CLARIN to the Digital Classics.
Monica Monachini, Francesca Frontini, Anika Nicolosi and Fahad Khan.

How to use DameSRL: A framework for deep multilingual semantic role labeling.
Quynh Ngoc Thi Do, Artuur Leeuwenberg, Geert Heyman and Marie-Francine Moens.

Speech Recognition and Scholarly Research: Usability and Sustainability.
Roeland Ordelman.

Towards TICCLAT, the next level in Text-Induced Corpus Correction.
Martin Reynaert, Maarten van Gompel, Ko van der Sloot and Antal van den Bosch.

SenSALDO: a Swedish Sentiment Lexicon for the SWE-CLARIN Toolbox.
Jacobo Rouces, Lars Borin, Nina Tahmasebi and Stian Rødven Eide.

Error Coding of Second-Language Learner Texts Based on Mostly Automatic Alignment of Parallel Corpora.
Dan Rosén, Mats Wirén and Elena Volodina.

Using Apache Spark on Hadoop Clusters as Backend for WebLicht Processing Pipelines.
Soheila Sahami, Thomas Eckart and Gerhard Heyer.

UWebASR – Web-based ASR engine for Czech and Slovak.
Jan Švec, Martin Bulín, Aleš Pražák and Pavel Ircing.

Pictograph Translation Technologies.
Vincent Vandeghinste, Leen Sevens and Ineke Schuurman.

| 11:30-13:00	**Poster and demo session**	Gallilei Foyer
13:00-14:00	Lunch	Restaurant
14:00-14:30	**State of the Technical Infrastructure** Dieter Van Uytvanck	Concorde Hall
14:30-14:45	**Award ceremony**	Concorde Hall
14:45-15:30	**Looking forward:** Board of Directors **Acknowledgements and goodbye:** Inguna Skadiņa	Concorde Hall
15:30	End of conference	

DAY 4

09:00-10:30	Scientific Advisory Board (closed meeting)	Lear Hall
10:30-11:00	Coffee break	Lear Hall
11:00-13:00	Scientific Advisory Board (closed meeting)	Lear Hall
13:00	Lunch	Restaurant

Hotel Galilei Congress

Address:
Via Darsena
1 - 56121 Pisa (PI)
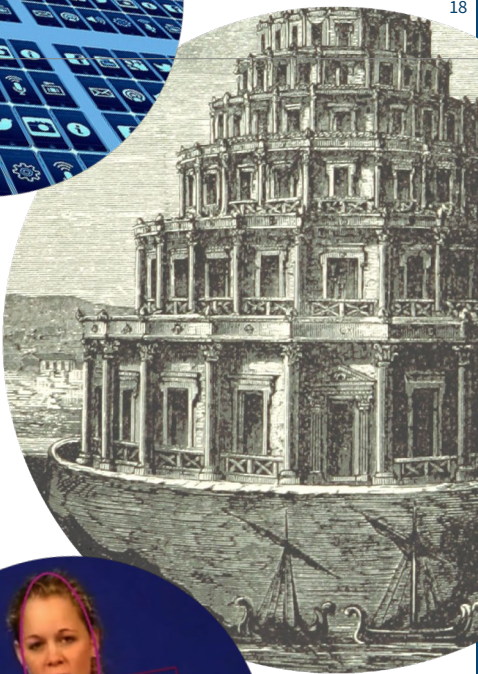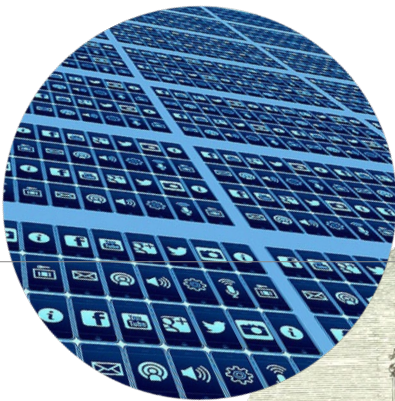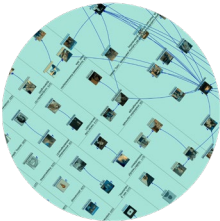
http://www.hotelgalileipisa.it/
Phone: +39 050 507111

20

"A modern 'crystal giant' watching over Pisa's rooftops, Hotel Galilei is the place to go for your business trips or for a holiday, whatever you want to see in Tuscany. Located at the south gate to the city, near its international airport, it is the best 'launchpad' to the whole world."

Francesco de Rosa
Hotel Manager

21

RECEPTION

Villa Comel / Domus Comeliana

Address:
Via Cardinale Maffi 48
Pisa

http://www.domuscomeliana.com/
Phone: +39 02 0062051

Domus Comeliana's site, located between Piazza del Duomo and the 1155 medieval city walls was already identified in the first City plan (late 1500), and was occupied by gardens until 1870. An engraving dated 1872 shows a street perpendicular to the walls, and a long building classified as a 'house' surrounded by a 'vegetable garden with water reserves.'

An open gallery, a barn, and a shed were later added and in 1947 the complex became the home and office of the University of Pisa Dermatology Clinic director Marcello Comel. From 1992, the scientific foundation carried out extensive refurbishment of the facilities and made the original villa with its terrace and gallery and the two gardens available to the city for events.

DINNER

Stazione Leopolda

Address:
Piazza Guerrazzi
56125 Pisa

http://www.leopolda.it/
Phone: +36 050 21531

The Leopolda Station is a cultural centre, housed in an ancient railway station that was built at the end of the 19th century. It promotes events, exhibitions, shows, conferences, workshops and leisure activities.

The structure management has been entrusted by the Municipality of Pisa to Casa della Citta Leopolda, a no-profit association of second level founded in 1999 by the associations that promoted the project of transformation of the old marketplace and worked together on the European project for its recovery works.

The cultural centre opened in 2002 and it includes: the historical hall, a conference room, a multimedia centre, a theatre and dance rehearsal room, a shop of short food supply chain products and an inner square.

ACCOMMODATION

Hotel Galilei Congress

Address:
Via Darsena, 1, 56121 Pisa

http://www.hotelgalileipisa.it/
Phone: +39 050 507111

Hotel Repubblica Marinara

Address:
Via Matteucci, 81, Pisa

http://www.hotelrepubblicamarinara.it/
Phone: +39 050 3870100

AC Hotel Pisa

Address:
Via delle Torri, 20, Pisa

http://ac-pisa.pisa-hotels.com/
Phone: +39 050 575395

G	**Hotel Galilei**
M	**Hotel Repubblica Marinara**
P	**AC Hotel Pisa**
D	**Dinner	Stazione Leopolda**
R	**Reception: Domus Comeliana**