



Prima edizione del ciclo di conferenze

Digitale historische Wortforschung (DhW) "Testi e parole in rete: nuovi accessi digitali"

27 e 28 novembre 2025

Lehrstuhl für Romanische Philologie Universität des Saarlandes, Campus Saarbrücken, Geb. E1 7, R. 0.01



Link per la partecipazione online

Il ciclo di conferenze Digitale historische Wortforschung (DhW) organizzato dalla cattedra di Romanische Philologie dell'Università del Saarland, coinvolge la linguistica storica, la lessicografia, la filologia testuale e le digital humanities. Il progetto si inserisce in una tradizione di ricerca consolidata, rappresentata dai progetti di ricerca a lungo termine Lessico Etimologico Italiano (LEI), Wissensnetze in der mittelalterlichen Romania (ALMA) e Archivio Carlo Tagliavini (ACT), e guarda a future collaborazioni interdisciplinari e internazionali.

L'edizione inaugurale del ciclo, dal titolo *Testi e parole in rete: nuovi accessi digitali*, dedicata alla filologia testuale e alla lessicografia storica digitali, si articolerà in due giornate: la prima, aperta al pubblico, prevede interventi di studiose e studiosi internazionali su nuovi strumenti e metodi digitali per l'ecdotica e l'analisi linguistica; la seconda sarà dedicata a workshop di carattere pratico, organizzati nell'ambito dei diversi progetti di ricerca.

Il ciclo di conferenze DhW è un'occasione di aggiornamento sulle tecnologie emergenti, di riflessione sulle pratiche di edizione digitale e analisi linguistica nonché di consolidamento delle reti di collaborazione internazionali e interdisciplinari.





Programma

27 novembre 2025

Introduzione

09:00-10:00 | Elton Prifti, Giulia Barison, Marcus Husar, Alessandro Alfredo Nannini, Yasmine Posillipo (Universität des Saarlandes / Mannheim), Apertura della conferenza e presentazione dei progetti Lessico Etimologico Italiano (LEI), Wissensnetze in der mittelalterlichen Romania (ALMA) e Archivio Carlo Tagliavini.

Sessione di Filologia Testuale (10:00 – 12:45)

10:00-10:45 | Elena Pierazzo (Université de Tours), Filologia strutturale e quantitativa per testualità fluide.

10:45-11:15 | Pausa caffè

11:15-12:00 | Chantal Pivetta (Lunds universitet), L'edizione sinottica di Uberto e Filomena, poema in ottave del XV secolo, realizzata con Digital Philology for Dummies e il database relazionale integrato.

12:00-12:45 | Roberto Rosselli Del Turco (Università degli Studi di Torino), Evoluzione degli approcci all'edizione digitale in EVT 3: nuove funzionalità e tipologie di edizione.

12:45-14:00 | Pausa Pranzo

Sessione di Linguistica Storica (14:00 – 17:30)

14:00-14:45 | Marco Passarotti (Università Cattolica del Sacro Cuore), A Comprehensive Journey of Language Resources for Latin (and Beyond).

14:45-15:30 | Sara Aakrout e Sabine Tittel (Heidelberger Akademie der Wissenschaften), The Mortality of the grande mortalité: the Medieval Plaque from a Linguistic Point of View.

15:30-16:00 | Pausa caffè

16:00-16:45 | Stefania Degaetano-Ortlieb (Universität des Saarlandes), Modeling Change in Language Use with Information Theory: The Case of the Chemical Revolution.

16:45-17:30 | Nicoletta Puddu (Università degli Studi di Cagliari), La digitalizzazione di testi sardi di età moderna: il Corpus EModSar.

17:30-18:15 | Renato Caenaro (SilentWave SRLS), Analisi relazionale su testo medico medievale come caso studio per il progetto ALMA.





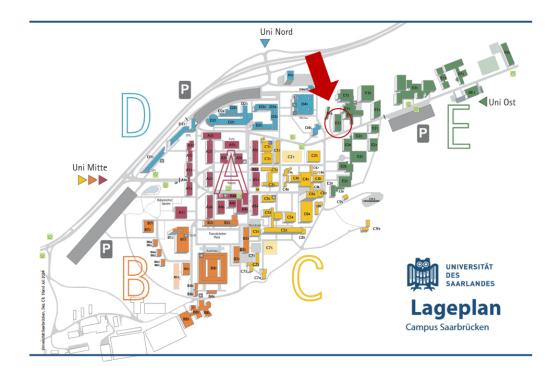


28 novembre 2025 Workshop

09:30-12:00 | Renato Caenaro (SilentWave SRLS), Chantal Pivetta (Lunds universitet), Roberto Rosselli Del Turco (Università degli Studi di Torino), Digital Philology for Dummies ed Edition Visualization Technology.

12:00-13:30 | Pausa Pranzo

13:30-15:00 | Luigi Talamo (Universität des Saarlandes), How to do linguistic analysis with the Universal Dependencies (UD).



La partecipazione ai workshop avviene solo previa registrazione. È possibile registrarsi fino al 24 novembre 2025 scrivendo una mail a:

Giulia Barison giulia.barison@uni-saarland.de

Alessandro Alfredo Nannini alessandro.nannini@uni-saarland.de

Yasmine Posillipo yasmin.posillipo@uni-saarland.de.





DhW-1

Book of abstracts

Renato Caenaro (SilentWave SRLS)

Analisi relazionale su testo medico medievale come caso studio per il progetto ALMA

Questo contributo presenta una prima prova di analisi computazionale su un testo medico medievale privo di codifica strutturale, condotta nell'ambito del progetto ALMA. Il documento analizzato, un inventario chirurgico attribuito a Guy de Chauliac (Montpellier, 1363), è stato sottoposto a una procedura di text mining e visualizzazione relazionale basata su Python e Google Colab. L'obiettivo è testare la costruzione automatica di grafi semantici a partire da liste di parole chiave (cura/anatomia), esplorando la co-occorrenza in unità testuali non sintatticamente definite.

La rete generata evidenzia relazioni disassortative tra tecniche terapeutiche e parti anatomiche, con cluster coerenti e comunità individuate tramite il metodo di Lovanio. L'analisi della degree centrality suggerisce nodi di rilevanza clinica e semantica, mentre la struttura comunitaria offre indizi su modelli terapeutici ricorrenti. Il confronto tra reti "inter-lista" e "intra-lista" mostra potenziale per la ricostruzione di procedure mediche e macroaree anatomiche.

Questa sperimentazione si colloca nel quadro metodologico di ALMA, integrando approcci onomasiologici, analisi semantica e modellazione RDF. I risultati offrono spunti per la selezione di testi da codificare in TEI/XML, per l'estrazione di glossari tematici e per la mappatura di concetti medici medievali in ontologie storicizzate. In prospettiva, il metodo può contribuire alla costruzione di reti lessicali panromanze e al riconoscimento di strutture semantiche stabili attraverso epoche e lingue.

Stefania Degaetano-Ortlieb (Universität des Saarlandes)

Modeling Change in Language Use with Information Theory: The Case of the Chemical Revolution

We trace the development of English scientific writing during the period of the chemical revolution showing linguistic changes favoring communicative efficiency. Our empirical basis is a curated subset of texts from the Royal Society of London (RSC Corpus, Fischer et al. 2020), which provides a long, internally coherent record of expert discourse. We employ two information-theoretic notions, entropy and surprisal, to model the temporal dynamics of change in language use. First, we use a data-driven method by applying relative entropy, which allows us to compare entire linguistic levels over time without the need for pre-selecting specific features and to detect periods of change (Degaetano-Ortlieb & Teich 2019). Second, building on the concept of surprisal, our approach models linguistic variation based on the amount of information a linguistic unit conveys in its preceding context. This method allows us to understand how certain words or phrases







become more or less predictable over time within scientific discourse, reflecting the evolving norms and expectations of the scientific community. For example, we observe cycles of change with alternating peaks and throughs in predictability at both the lexical and grammatical level (e.g., post-modification such as 'iron that has reacted with dephlogisticated air' or 'oxide of iron' vs. pre-modification 'iron oxide'), indicating a move towards greater communicative efficiency by using compression strategies (Biber & Gray 2016). Through analytical techniques informed by information theory, our research aims to shed light on the complex mechanisms that have shaped the language of scientific communication over the centuries.

References

Biber, D., & Gray, B. (2016). Grammatical Complexity in Academic English: Linguistic Change in Writing. Cambridge University Press.

Degaetano-Ortlieb, S., & Teich, E. (2019). "Toward an Optimal Code for Communication: The Case of Scientific English". Corpus Linguistics and Linguistic Theory, 18(1), 175-207.

Fischer, S., Knappen, J., Menzel, K. and Teich, E. (2020). "The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study". In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 794–802, Marseille, France. European Language Resources Association.

Marco Passarotti (Università Cattolica del Sacro Cuore)

A Comprehensive Journey of Language Resources for Latin (and Beyond)

In this talk, I will discuss the issue of interoperability among language resources, presenting the use case of the *LiLa Knowledge Base* for resources of the Latin language. After introducing the basic principles of the *Linked Data* paradigm, on which *LiLa* is grounded, I will describe the architecture of the *Knowledge Base*. Then, I will focus on the modeling of a number of lexical and textual resources of different kinds currently interlinked through *LiLa*, particularly focusing on how etymological information is represented and published. Finally, I will present a few online services to query and populate the *Knowledge Base*.





Elena Pierazzo (Université de Tours)

Filologia strutturale e quantitativa per testualità fluide

L'emergere di sistemi computazionali avanzati permette oggi di trattare grandi quantità di dati e di dare risposta a problemi ecdotici accantonati da tempo. Uno di questi è il problema della testualità fluida che caratterizza i romanzi medievali, le raccolte di sermoni, ma anche testi moderni trasmessi da manoscritti per i quali la varia lectio sorpassa anche le capacità umane di contare. All'interno del progetto europeo *PRIMA* abbiamo creato strumenti digitali capaci di gestire questa complessità, e stiamo creando dei flussi editoriali innovativi. Questi sistemi ci spingono a ripensare il metodo filologico nella sua interezza, spingendolo decisamente in direzione quantitativa. Siamo pronti per questo cambiamento?

Chantal Pivetta (Lunds universitet)

L'edizione sinottica di Uberto e Filomena, poema in ottave del XV secolo, realizzata con Digital Philology for Dummies e il database relazionale integrato

Il contributo presenta il caso di studio dell'edizione sinottica di *Uberto e Filomena*, poema in ottava rima del XV secolo appartenente a una tradizione testuale attiva e complessa. L'edizione è stata sviluppata all'interno della piattaforma *Digital Philology for Dummies*, che ha consentito una codifica XML/TEI ottimizzata, grazie all'impiego di strumenti automatizzati di marcatura ("tag automatici") e di un algoritmo di corrispondenza testuale (Match), basato su similarità statistica applicabili a testi monolingui e multilingue.

L'integrazione con un database relazionale ha ulteriormente potenziato l'analisi filologica e linguistica dei testimoni, permettendo la generazione di un cladiogramma statistico delle varianti sostanziali e la visualizzazione delle relazioni tra i testimoni, sia nel loro insieme sia in prospettiva diacronica, per ciascuna redazione del testo. L'approccio quantitativo è stato applicato anche allo studio linguistico del testimone di base e del suo rapporto con gli altri, consentendo di formulare ipotesi statisticamente fondate sulle diverse facies linguistiche che caratterizzano la tradizione conservata di *Uberto e Filomena*.

Nicoletta Puddu (Università degli Studi di Cagliari)

La digitalizzazione di testi sardi di età moderna: il Corpus EModSar

Il sardo di età moderna (XV-XVIII secolo), sinora poco studiato, è testimoniato, oltre che da un numero tutto sommato esiguo di testi letterari, anche da una notevole messe di documenti di tipo legale, sia di natura pubblica che privata (Cadeddu 2013, 2023; Murgia e Puddu 2024). Tali documenti si presentano spesso in forma plurilingue, includendo al loro interno le lingue alte del repertorio dell'epoca, ovvero il catalano, il castigliano e il latino (Pinto 2023). In questo contributo, presenteremo il corpus *EModSar*, che contiene un primo nucleo di testi della prima età moderna scritti in sardo, digitalizzati e annotati secondo lo standard TEIP5 (Puddu e Talamo 2020).





Come sottolinea Enrique Arias (2015), la digitalizzazione e la creazione di corpora di testi di questo tipo presenta numerosi vantaggi per l'analisi della lingua, a patto che, accanto all'annotazione linguistica, venga preservata l'informazione filologica. Tuttavia, come ricordano Jenset e McGillivray (2017), l'annotazione tramite l'uso di codifiche standard per la filologia è relativamente poco comune nei corpora storici.

Uno dei pochi strumenti che consente di combinare l'annotazione TEI con gli strumenti della linguistica dei corpora è la piattaforma TEITOK (Janssen 2016). L'utilizzo di tale piattaforma per la creazione del corpus *EModSar* rende possibile la navigazione tra diversi tipi di trascrizione (diplomatica, filologica e normalizzata), oltre alla visualizzazione dell'immagine digitale dei documenti. Inoltre, è stato possibile annotare e lemmatizzare i documenti, grazie a procedure semiautomatiche disponibili all'interno di TEITOK.

Bibliografia

- Cadeddu, Maria Eugenia (2013). "Scritture di una società plurilingue: note sugli atti parlamentari sardi di epoca moderna". In Krefeld, T., Oesterreicher, W., Schwägerl-Melchior, V. (eds.), Reperti di plurilinguismo nell'Italia spagnola, Berlin/Boston, De Gruyter: 13-26.
- Cadeddu, Maria Eugenia (2023). "Scrivere in castigliano, parlare in sardo. Esempi di contesti comunicativi in Ogliastra (XVIII secolo)". In Fresu R., Maninchedda P., Murgia G., Serra P., (eds.), Il "traffico delle lingue". Idiomi a contatto in Sardegna e nel Mediterraneo in età preunitaria, Cagliari, UNICApress: 149-174.
- Enrique Arias, Andrés (2015). "Edición digital y corpus diacrónicos: problemas en la edición de textos históricos producidos en un entorno bilingüe". In Sánchez Menendez, J.P. De la Torre, M., Codita, V. (eds.), Temas, problemas y métodos para la edición y el estudio de documentos hispánicos antiguos, Valencia, Tirant Humanidades: 395-406.
- Janssen, Maarten (2016). "Teitok: Text-faithful annotated corpora". In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H. Moreno, A., Odijk, J., Piperidis, S. (eds.), Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Paris, France.
- Jenset, Gard B., McGillivray, Barbara (2017). Quantitative historical linguistics, Oxford, Oxford University Press.
- Murgia, Giulia, Puddu, Nicoletta (2024). "Su alcuni binomi coordinati in un corpus di documenti sardi di età moderna". In Consani, C., Guazzelli, F., Perta., C. (eds.), *Gruppi professionali come fattore di innovazione linguistica. Evidenze documentarie in Europa tra Tarda Antichità e Medioevo*, Alessandria, Edizioni dell'Orso: 113-133.
- Pinto, Immacolata (2023). "The impact of Aragonese and Castilian Dominations on the Language and Literature of Sardinia". In Petrocchi A. Brown J. (eds.) Languages and cross-cultural Exchanges in Renaissance Italy, Turnhout, Brepols: 267-239.
- Puddu, Nicoletta, Talamo, Luigi (2020). "EModSar: A Corpus of Early Modern Sardinian Texts". In Marras, C., Passarotti, M., Franzini, G., Litta, E. (eds.) Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD). La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica, Quaderni di Umanistica Digitale: 210-215.







Roberto Rosselli Del Turco (Università degli Studi di Torino)

Evoluzione degli approcci all'edizione digitale in EVT 3: nuove funzionalità e tipologie di edizione

L'intervento si concentrerà su alcune delle nuove funzionalità e sulle diverse tipologie di edizione rese possibili dalla più recente versione di EVT, ovvero EVT 3.

Per quanto riguarda le tradizioni mono-testimoniali, le edizioni di tipo diplomatico/interpretativo possono già beneficiare delle caratteristiche disponibili in EVT 2 (quali il facsimile digitale, il collegamento immagine-testo, e il supporto per le Named Entities con link a risorse di tipo Linked Open Data) cui si affiancheranno nuove funzionalità attualmente in fase di sviluppo. Tra queste: la possibilità di navigare il solo facsimile digitale, il confronto tra immagini differenti, il supporto per più descrizioni codicologiche del manoscritto e la visualizzazione geografica dei toponimi su mappa.

Le tradizioni multi-testimoniali, invece, hanno già visto un ampliamento delle potenzialità grazie all'integrazione del metodo double-end-point attachment previsto dallo standard TEI per il collegamento delle voci d'apparato al testo critico. Sviluppata nell'ambito del progetto $PRIN\ RETI$, questa funzionalità è ora disponibile per tutti i progetti editoriali che richiedano una marcatura stand-off dell'apparato critico. Tra i vantaggi, si segnala la possibilità di mantenere il testo critico privo di markup non strutturale o non essenziale, rendendolo così disponibile per ulteriori livelli di annotazione semantica. Al confine tra tradizioni mono- e multi-testimoniali si collocano infine le edizioni di filologia d'autore, introdotte per la prima volta in EVT. Grazie alla collaborazione con il progetto sull'edizione del $Canzoniere\ 1919$ di Umberto Saba, è stato definito un modello di codifica in grado di visualizzare in modo dinamico le modifiche autoriali, mettendole in relazione con le fasi correttorie identificate all'interno del manoscritto. L'introduzione del fattore "tempo" nella rappresentazione dell'apparato è resa possibile proprio dalle capacità di EVT, che offre un'alternativa dinamica rispetto alle tradizionali trascrizioni statiche dei testi d'autore.





Luigi Talamo (Universität des Saarlandes)

How to do linguistic analysis with the Universal Dependencies

Token-based analysis (Levshina 2019) is a research methodology in typological linguistics that involves the use of instances (tokens) of constructions, as observed in actual linguistic usage and approximated through the results of psycholinguistic experiments, data collected from questionnaires, and corpora, preferably parallel ones.

In this workshop, we will use a selection of data from Miniciep+, the public version of the Corpus of Indo-European Prose and more (CIEP+: Talamo & Verkerk 2022), a parallel corpus of over thirty languages developed since 2019 at Saarland University based on translations of fiction books. Thanks to its annotation according to the Universal Dependencies (UD: de Marneffe et al. 2021) framework, the corpus lends itself to quantitative morphological and morphosyntactic analysis. We will discuss and experiment with a series of scripts in Python and R for extracting data from the corpus relating to word order, comparative concepts (e.g., quantifiers, adjectives, etc.), verbal categories such as finiteness, and personal pronouns.

Sara Aakrout e Sabine Tittel (Heidelberger Akademie der Wissenschaften)
The Mortality of the grande mortalité: the Medieval Plague from a Linguistic Point of
View

The aim of this study is to examine the plague and its designations starting from one of the most influential medical texts from the French Middle Ages: the *Grande Chirurgie*, written in Latin by Gui de Chauliac in 1363 and subsequently translated into several languages. Gui de Chauliac is said to be the first physician to explicitly differentiate between the bubonic and pneumonic forms of plague. Starting from a French translation of the text recorded in a manuscript dated to the second third of the 15th century, we check to confirm this by investigating whether Gui clearly describes two concepts of the plague, what terms he uses to designate them, and in which other texts these terms can be found. We work empirically, based on a digital corpus of Old and Middle French medical texts. We support our results by drawing on the pertinent dictionaries of Old and Middle French. Our study concludes with a medical-historical interpretation of our linguistic findings.