



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

10-12 settembre 2026

LIX Congresso internazionale di studi della Società di Linguistica Italiana

I DATI NELLE SCIENZE DEL LINGUAGGIO

Riassunti delle comunicazioni – Workshop

Sommario

<i>Workshop 1</i>	4
ALESSANDRO IANNELLA	
Tra artefatti plausibili e azioni incerte: quale postura per dialogare con l'IA?	4
ELISA PELLEGRINO E GA NAN	
La pronuncia in inglese LS tra parlato umano e sintesi vocale: uno studio sperimentale sull'accomodamento fonetico	6
DIEGO CORTÉS VELÁSQUEZ E DANIELA ZINI	
IA come <i>scaffolding</i> metacognitivo nella lettura plurilingue: un'esperienza di intercomprensione guidata	7
LUCA CIGNETTI	
La revisione testuale nella didattica della scrittura con l'IA generativa	8
SARAH DE VICO, DOMENICO DI RUSSO, AQUILINO LONGO ED ELISA MILILLI	
<i>Prompting</i> e riflessione metalinguistica. Percorsi di educazione linguistica nell'era dell'IA generativa per una valutazione formativa umana e autentica	9
ANDREA MAIELLO E YAHIS MARTARI	
La canzone d'autore (artificiale). Esperienze di insegnamento linguistico con canzoni scritte e prodotte dall'AI generativa	10
<i>Workshop 2</i>	12
FAHAD KHAN	
A Semantic View on Linguistic Linked Data	12
ROBERTA G. LEOTTA, ELEONORA LITTA, DOMENICO GIUSEPPE MUSCIANISI E RACHELE SPRUGNOLI	
Dall'annotazione sintattica al web semantico: il caso del siciliano	13
GIOVANNA TONAZZO	

Interrogare LiITA in linguaggio naturale: tre approcci a confronto	18
ALESSANDRO ALFREDO NANNINI, MARCO CARLO PASSAROTTI, GIOVANNI MORETTI, FRANCESCO MAMBRINI ED ELTON PRIFTI	
Il Lessico Etimologico Italiano (LEI) come Linked Open Data e il suo collegamento alle Knowledge Base LiITA e LiLa	22
LUCIA FRANICALANCI E FRANCESCO CORADESCHI	
La lessicografia storica nel Web Semantico: una sperimentazione sul TLIO	25
DANIELA VELLUTINO E CATERINA SARACCO	
Dati linguistici FAIR per l'addestramento supervisionato di un'IA. Problemi di annotazione e possibili soluzioni per dati aperti da tre a cinque stelle	32
ELIANA DI PALMA E VALERIO BASILE	
Emozioni in <i>Linked Open Data</i> : il caso di ELIta e Sentix	35
LORENZO BACCI E TOMMASO AGNOLONI	
Arricchimento strutturale e semantico delle pronunce della Corte Costituzionale e ripubblicazione come risorsa interoperabile	41
EMILIANO GIOVANNETTI, ANDREA BELLANDI, SIMONE MARCHI E MAFALDA PAPINI	
Towards an Interoperable Lexical Infrastructure for Italian: Linking and Enriching CompL-it	45
<i>Workshop 3</i>	50
NATALIYA LEVKOVYCH	
Dalla romanizzazione all'italianizzazione: prestito grammaticale nelle lingue non romane d'Italia	50
MARCO FORLANO	
<i>Finola o finarela?</i> L'integrazione variabile dei prestiti verbali italo-romani in sinto lombardo	53
ANDREA SCALA	
Il vocalismo atono in alcune varietà di romani d'Italia: fenomeni di convergenza verso l'italo-romanzo	56
MICHELE COSENTINO	
Il dialetto romanzo dei rom calabresi di Cosenza: strutture grammaticali e inquadramento teorico	58
NICOLA DUBERTI, ALESSANDRO FLECCHIA E RICCARDO REGIS	
<i>Na varietà bastarda</i> : la koinè piemontese tra focalizzazione e livellamento	61
CARLO ZIANO	
Il dialetto veneto dei sinti emiliani	64
NATALIYA LEVKOVYCH	
Il prestito delle congiunzioni italiane nelle lingue alloglotte d'Italia e nelle varietà romane	66
<i>Workshop 4</i>	68
MICHELA DOTA E MASSIMO PRADA	
Forme della coerenza e della coesione nelle homepage di YouTube, Instagram e TikTok	68
Sulla coerenza multimodale in testi neomediali: il caso del BookTok	70
LEONARDO ROMEI E LUCIANO PERONDI	
Sinsemia ovvero la dimensione compositiva della multimodalità: un'integrazione teorica e metodologica	72
ALESSANDRO BUTTINELLI	
Strategie di coesione in testi umani e sintetici: uno studio basato sull'analisi delle reti	75
ANTONIO MURA E JACQUELINE VISCONTI	
Coerenza e coesione in testi giudiziari italiani generati e riscritti dall'intelligenza artificiale	79

GIULIA MANTOVANI	
Coesione e coerenza nel testo mimetico delle favole ‘naturali’ vs ‘sintetiche’: un caso studio sui segnali discorsivi	82
LORENZO ZANASI E JENNIFER CARMEN FREY	
La coerenza alla prova dell’IA (e viceversa): dai testi umani alle riscritture sintetiche	85
<i>Workshop 5</i>	87
SILVIO CRUSCHINA	
On the relationship between deixis and mirativity: Evidence from Sicilian	87
ENRICO CASTRO	
<i>A ze morto me nona!</i> Una lettura mirativa per una struttura veneta	90
ELISABETTA MAGNI	
Gli usi di <i>bada e badali</i> : dal richiamo dell’attenzione alla <i>mock mirativity</i>	92
IVICA BAKMAZ, SARAH DESSÌ SCHMID, YANNICK PIECHOTTA ED EVELYN WIESINGER	
Expressing unexpectedness: The mirative use of <i>andare a</i> + infinitive in Italian	95
EMANUELA LI DESTRI	
La miratività nella perifrasi <i>venire a</i> + infinito nell’italiano contemporaneo	98
GIOVANNI DI PAOLA E ANDREA SANSÒ	
“E non ti arriva uno che deve passare?”: costruzioni dialogiche dell’inatteso in italiano	101
ILARIA FIORENTINI E LUISA TRONCONE	
<i>Non io che</i> : uno studio della miratività tra esperienti e negazione	104
CATERINA MAURI E ANTONIA RUSSO	
Risposte polari e miratività: uno studio sull’italiano	108
DANIELA VERONESI	
Esprimere la sorpresa per un pubblico di ascoltatori: l’esempio dell’incredulità ritualizzata nelle interazioni radiofoniche in italiano	112
FRANCESCO MORLEO	
L’espressione dell’inatteso nel portoghese europeo contemporaneo: strategie mirative nel parlato radiofonico	116

Workshop 1

Intelligenza artificiale ed educazione linguistica

Proponente: GISCEL

ALESSANDRO IANNELLA

Tra artefatti plausibili e azioni incerte: quale postura per dialogare con l'IA?

I sistemi di IA Generativa e Agentiva operano secondo una logica probabilistica che non distingue tra plausibile e vero, tra combinazione statistica e conoscenza fondata. Le tecnologie generative producono artefatti convincenti ma non necessariamente accurati, plausibili secondo i pattern appresi durante il loro addestramento ma non verificati rispetto alla realtà. Inoltre, il parametro della temperatura, che regola il grado di variabilità degli output, fa sì che, a parità di input, questi ultimi siano ogni volta diversi e, di conseguenza, imprevedibili. Sebbene tale natura, non deterministica e in particolare stocastica, conferisca al processo generativo un carattere chiaramente creativo e non referenziale, essa è poco nota agli utenti, i quali tendono ad assimilare queste tecnologie a motori di ricerca o fonti di consultazione, nel primo caso per l'interazione in linguaggio naturale e nel secondo per la qualità apparente dei risultati. Questa confusione induce frequentemente una sospensione del giudizio sull'accuratezza dei risultati, con una cessione della valutazione epistemica all'accettazione acritica (Quattrociochi, Capraro & Perc 2025; Roncaglia 2023). Le tecnologie agentive spostano la questione dal contenuto prodotto all'azione: svolgono procedure su dispositivi e nel Web in condizioni di incertezza, aprendo a un paradigma di delega operativa in cui sistemi coordinati eseguono sequenze di azioni multi-step in modo sempre più indipendente dalla supervisione umana (Aranda & Sugimoto 2026). Tale configurazione introduce un ulteriore livello di complessità rispetto alle tecnologie generative, in quanto il sistema non restituisce un output da valutare ma esegue direttamente azioni in ambienti reali, ampliando così il perimetro dei rischi potenziali.

Il termine con cui è definito il secondo gruppo di tecnologie di IA, agentivo, è ironicamente lo stesso che la psicologia cognitiva riserva alla capacità propriamente umana di agire intenzionalmente (Bandura 2006). Ed è questa tensione che la documentazione educativa europea e globale chiede di abitare per impiegare l'IA con senso critico e ricavarne un beneficio autentico nella vita personale e professionale (Miao & Cukurova, 2024; Miao, Shiohira & Lao 2024; OECD & Commissione Europea, 2025). Farlo significa adottare una postura metacognitiva che si fonda su autoefficacia,

intenzionalità e responsabilità: riconoscere da quale posizione si agisce, per quale scopo si utilizza l'IA, che cosa le si delega e di che cosa si rimane autori (Iannella, Boldrini & Merlini 2026).

Offrendo una lettura dell'innovazione tecnologica non come un rischio da contenere, ma come un'opportunità da attraversare con consapevolezza, l'intervento discute le pratiche concrete di insegnamento e apprendimento che l'IA Generativa e l'IA Agentiva possono sostenere a partire dalla postura delineata, postura che si declina in modo inevitabilmente diverso per chi insegna e per chi apprende, perché diversi sono gli obiettivi, diversa è la posta in gioco cognitiva e diversi sono i rischi e le opportunità.

Riferimenti bibliografici

- Aranda, Luis & Sugimoto, Kasumi 2026. The agentic AI landscape and its conceptual foundations. *OECD Artificial Intelligence Papers*, 56. Paris: OECD Publishing. <https://doi.org/10.1787/396cf758-en>
- Bandura, Albert 2006. Toward a psychology of human agency. *Perspectives on Psychological Science* 1(2), 164–180. <https://doi.org/10.1111/j.1745-6916.2006.00011.x>
- Iannella, Alessandro, Boldrini, Elena & Merlini, Fabio 2026. A dare senso è sempre qualcuno e mai qualcosa. *Skilled* 2. 16–17.
- Miao, Fengchun & Cukurova, Mutlu 2024. *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>
- Miao, Fengchun, Shiohira, Kelly & Lao, Natalie 2024. *AI competency framework for students*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- OECD & Commissione Europea 2025. *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education* (Review draft). Paris: OECD Publishing. https://ailiteracyframework.org/wp-content/uploads/2025/05/AILitFramework_ReviewDraft.pdf
- Quattrociocchi, Walter, Capraro, Valerio & Perc, Matjaž 2025. Epistemological fault lines. <https://doi.org/10.48550/arXiv.2512.19466>
- Roncaglia, Gino (2023). *L'architetto e l'oracolo*. Roma/Bari: Laterza.

La pronuncia in inglese LS tra parlato umano e sintesi vocale: uno studio sperimentale sull'accomodamento fonetico

Il miglioramento della pronuncia in una lingua straniera (LS) richiede l'esposizione a modelli di parlato vicini a quelli dei parlanti nativi (Krashen, 1985; LaScotte et al., 2021), spesso poco accessibili in contesti a bassa immersione. I recenti sviluppi dell'intelligenza artificiale (IA) stanno però contribuendo a ridefinire questo scenario: i sistemi text-to-speech (TTS) e gli agenti vocali consentono di generare parlato (quasi) nativo, offrendo occasioni di pratica personalizzabile (Rogerson-Revell, 2021), ad integrazione o sostituzione dell'input vocale umano. In tali contesti, tuttavia, rimane aperta una questione fondamentale: l'*input* vocale dell'IA sostiene l'apprendimento della pronuncia allo stesso modo di quello umano? Il presente studio affronta tale quesito confrontando l'accomodamento fonetico di 28 apprendenti cinesi di inglese LS verso una voce generata dalla piattaforma *TTS Microsoft Azure* e la voce di una madrelingua in un compito di ascolto e ripetizione. L'analisi dell'accomodamento, condotta calcolando le distanze acustiche tra le produzioni degli apprendenti e quelle dei modelli prima e dopo l'esercitazione orale, si è focalizzata su: *Voice Onset Time* (VOT), realizzazione temporale e spettrale delle vocali tese e rilassate e pattern ritmici. I risultati evidenziano esiti differenziati, validati statisticamente: la voce umana favorisce l'imitazione del VOT, mentre quella prodotta tramite l'IA induce maggiori aggiustamenti nella durata vocalica e nei pattern ritmici. I dati suggeriscono punti di forza complementari dei due modelli, con implicazioni didattiche per l'insegnamento della pronuncia.

Riferimenti bibliografici

- LaScotte, Darren & Meyers, Coleen & Tarone, Elaine. 2021. Voice and mirroring in SLA: Top-down pedagogy for L2 pronunciation instruction. *RELC Journal* 52(1). 144–154. <https://doi.org/10.1177/0033688220953910>
- Krashen, Stephen. 1985. *The input hypothesis: Issues and implications*. London/New York: Longman.
- Rogerson-Revell, Pamela Mary. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, 52(1). 189–205. <https://doi.org/10.1177/0033688220977406>

DIEGO CORTÉS VELÁSQUEZ E DANIELA ZINI

IA come *scaffolding* metacognitivo nella lettura plurilingue: un'esperienza di intercomprensione guidata

Le strategie di lettura in contesto plurilingue restano spesso implicite e difficilmente osservabili. Nell'intercomprensione tra lingue romanze, la costruzione del senso richiede inferenze e confronti interlinguistici, processi spesso sostituiti, nella didattica, dalla traduzione immediata. Il contributo ripensa il ruolo dell'IA generativa come dispositivo di *scaffolding* metacognitivo (Akram et al. 2025) non come sostituto dell'attività interpretativa. Lo studio ha coinvolto circa 50 studenti universitari in un percorso di lettura guidata di testi in diverse lingue romanze. È stata progettata una chat personalizzata, ispirata alla Grammatica della lettura di EuRom (Bonvino et al. 2011), che non fornisce traduzioni dirette ma sollecita uso e verbalizzazione delle strategie e confronto tra lingue del repertorio. L'obiettivo era rendere visibile e regolabile il processo inferenziale, senza facilitare la comprensione in senso riduttivo. I dati raccolti consentono di osservare la varietà e la sequenzialità delle strategie attivate e l'eventuale riduzione della dipendenza traduttiva. In questa prospettiva, l'IA è ridefinita come strumento di emancipazione cognitiva a sostegno dell'autonomia interpretativa e si iscrive nei principi dell'educazione linguistica democratica (GISCEL 1975).

Riferimenti bibliografici

- Akram, Alia, Kiren, Ayesha, Sabir, Sabeen & Aslam, Sumaira. 2025. The effect of AI based scaffolding on problem solving and metacognitive awareness in learners. *The Critical Review of Social Sciences Studies* 3. 1074–1089. <https://doi.org/10.59075/ja7dvy78>
- Bonvino, Elisabetta, Caddéo, Sandrine, Vilagínés Serra, Eulalia & Pippa Salvador. 2011. *EuRom5-Ler e comprendre 5 línguas românicas–Leer y entender 5 lenguas románicas–Llegir i entendre 5 llengües romàniques–Leggere e capire 5 lingue romanze–Lire et comprendre 5 langues romanes*. Milano, Hoepli.
- GISCEL – Gruppo di Intervento e Studio nel Campo dell'Educazione Linguistica della Società di linguistica italiana. 1975. *Dieci tesi per l'educazione linguistica democratica*, <https://giscel.it/>

LUCA CIGNETTI

La revisione testuale nella didattica della scrittura con l'IA generativa

L'intervento propone una riflessione teorico-didattica sull'integrazione dell'IA generativa nei percorsi di insegnamento della scrittura, assumendo la revisione testuale come pratica privilegiata di osservazione e intervento.

Gran parte della letteratura sulla scrittura assistita dall'IA si è finora concentrata sulla progettazione dei prompt, sulla produzione di testi coerenti e stilisticamente adeguati e sul feedback automatico degli errori linguistici (Mezzadri & Paita, 2025; Su et al., 2023); marginale resta invece l'attenzione al ruolo della revisione, intesa come pratica in grado di identificare e correggere *bias*, distorsioni informative e “allucinazioni” del modello, nonché di attivare processi di metariflessione cognitiva.

Assumendo come quadro teorico la teoria di Bereiter e Scardamalia (1995), nella quale la revisione è concepita come processo consapevole di analisi, valutazione e ristrutturazione testuale, il contributo mostrerà, attraverso un'ampia esemplificazione, come un percorso guidato e consapevole di revisione permetta agli studenti liceali di intercettare e correggere distorsioni informative e incoerenze logiche.

Si intende così contribuire a definire modalità di integrazione dell'IA nella didattica della scrittura coerenti con i principi del GISCEL (Loiero & Lugarini, 2019), in cui autonomia e pensiero critico rappresentano assi portanti del processo formativo.

Riferimenti bibliografici

Bereiter, Carol & Scardamalia, Marlene. 1995. *La psicologia della composizione scritta*. Firenze: La Nuova Italia.

Loiero, Silvana & Lugarini, Edoardo (a cura di). 2019. *Tullio De Mauro: Dieci tesi per una scuola democratica*. Firenze: Cesati.

Mezzadri, Marco & Paita, Mariapaola. 2025. Integrare il prompt engineering tra le abilità trasversali per l'educazione linguistica: Uno studio di caso. *Italiano LinguaDue* 17(2). 634–658. <https://doi.org/10.54103/2037-3597/30456>

Su, Yanfang, Lin, Yun & Lai, Chun. 2023. Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing* 57. 1–13.

SARAH DE VICO, DOMENICO DI RUSSO, AQUILINO LONGO ED ELISA MILILLI

***Prompting* e riflessione metalinguistica. Percorsi di educazione linguistica nell'era dell'IA generativa per una valutazione formativa umana e autentica**

Il contributo illustra i risultati di una ricerca-azione, condotta da un gruppo di insegnanti del GISCEL Abruzzo della Scuola Primaria e della Scuola Secondaria di primo e secondo grado, volta a esplorare il potenziale del *prompt-design* come dispositivo per l'educazione linguistica.

La richiesta di formulare *prompt* descrittivi per la generazione di immagini più simili possibili a quelle fornite in classe è stata occasione per gli studenti di trasformare l'errore in opportunità di riflessione metalinguistica e di sperimentare forme di valutazione tra pari, implementando i *feedback* sulle rubriche di valutazione precedentemente elaborate e socializzate.

I dati raccolti hanno evidenziato come il *prompting*, se inserito in una cornice di valutazione tra pari, potenzi le funzioni metalinguistiche e critiche della scrittura, essenziali per una cittadinanza digitale consapevole così come auspicata dal *framework DigComp 3.0* (European Commission, Joint Research Centre 2025), che individua nella gestione responsabile dell'interazione uomo-macchina una competenza chiave per la cittadinanza digitale

Riferimenti bibliografici

Colombo, Adriano. 2011. «A ME MI». *Dubbi, errori, correzioni nell'italiano scritto*. Milano: Franco Angeli.

European Commission, Joint Research Centre. 2025. *DigComp 3.0. The Digital Competence Framework for Citizens – Advanced Integration of Generative AI and Data Literacy*, Luxembourg: Publications Office of the European Union.

GISCEL. 2019. *Dieci tesi per un'educazione linguistica democratica*. In Loiero, Silvana & Lugarini Edoardo (a cura di), *Tullio De Mauro: Dieci tesi per una scuola democratica*, 37-50. Firenze Franco Cesati Editore. <https://giscel.it/dieci-tesi-perleducazione-linguistica-democratica/>

La canzone d'autore (artificiale). Esperienze di insegnamento linguistico con canzoni scritte e prodotte dall'AI generativa

La possibilità di integrare strumenti di AI generativa nella didattica delle lingue apre scenari particolarmente promettenti per la didattica delle lingue (lingua madre e lingue straniere) nella scuola primaria e secondaria. Tra gli altri, piattaforme come *MakeSong* (<https://www.makesong.com/it>), che consentono di generare canzoni a partire da prompt testuali, offrono ai docenti di lingua l'opportunità di trasformare contenuti linguistici e metalinguistici in esperienze multimodali ad alto coinvolgimento emotivo e cognitivo (Gourkar 2025; Botlík Nuc & Klusáčková 2025). La creazione di brani musicali attraverso l'AI può essere orientata sia al consolidamento di specifiche forme linguistiche e item lessicali (ad esempio tempi verbali, connettivi testuali, campi semantici, lessico specialistico), sia alla rielaborazione di nozioni metalinguistiche di diversa natura: in questo secondo spazio si colloca l'impiego di tali strumenti per esplorare le cosiddette figure retoriche — metafora, similitudine, anafora, ossimoro etc. — che, grazie alla dimensione ritmica e sonora del testo cantato, diventano oggetto di riflessione attiva e produzione creativa.

Si presentano in proposito i dati tratti da alcune esperienze di insegnamento linguistico — realizzate in diversi contesti nazionali e internazionali — con canzoni generate da AI e raccolte dal CRDI - Centro di ricerca per la Didattica dell'Italiano dell'Università di Bologna. Il focus sarà poi posto sul lavoro didattico in una seconda classe di scuola secondaria di primo grado. La classe (che accoglie 8 partecipanti con BES — tra cui 4 NAI) è stata divisa in due gruppi equilibrati da dieci studenti (A e B). A ciascuno dei gruppi è stata sottoposta, specularmente, una canzone generata con *MakeSong* su di una figura retorica (ossimoro o litote), oppure una lezione tradizionale con appunti sullo stesso argomento, cosicché ogni gruppo fungesse da gruppo di controllo dell'altro (Tab. 1). Nell'intervento discuteremo delle difficoltà di costruzione del prompt adeguato; della qualità dei prodotti multimediali generati; della ricezione e del gradimento da parte di insegnanti e apprendenti; dell'efficacia della tecnica glottodidattica. In generale, cercheremo di comprendere se davvero, in questa prospettiva, l'AI si può configurare come mediatore didattico capace di favorire motivazione e apprendimento significativo.

	GRUPPO A	GRUPPO B
OSSIMORO	LEZIONE CON AI	LEZIONE TRADIZIONALE
LITOTE	LEZIONE TRADIZIONALE	LEZIONE CON AI

Tabella 1: Somministrazione materiali didattici.

Riferimenti bibliografici

- Gourkar, Madhumati. 2025. AI-generated songs as a didactic tool in teaching German as a foreign language. *Metaversalize* 2(2). 107–121. <https://doi.org/10.22105/metaverse.v2i2.74>
- Botlík Nuc, Tomas & Klusáčková, Marie. 2025. Using AI-Generated Songs in Foreign Language Learning. *Pedagogika* 75(4). 401–422.

Workshop 2

Verso l'interoperabilità: Linked Open Data e risorse linguistiche per l'italiano

Proponenti: **CLAUDIO IACOBINI, ELEONORA LITTA, MARCO PASSAROTTI E RACHELE SPRUGNOLI**

FAHAD KHAN

A Semantic View on Linguistic Linked Data

In this talk we look at the Linguistic Linked Data from a semantic point of view, examining the various different ways that we can understand the term ‘semantic’ in this context. We will start by outlining the original vision behind the Semantic Web as a version of the web in which the meaning of data would be explicit enough for machines to be able to easily process, and ask how we can understand this vision given the current popularity of intelligent agents. Then we will focus on the role of so called semantic artifacts such as ontologies and taxonomies, in helping to implement this vision in the context of Linguistic Linked Data and especially with regard to their status as shared formal conceptualizations. Finally, we look at Ontolex-Lemon, perhaps one of the most successful domain specific semantic artifacts in the area of Linguistic Linked Data and which was originally based around the idea of using ontologies to model word meaning using the so-called semantics by reference approach.

Riferimenti bibliografici

- Khan, Anas Fahad, Chiarcos, Christian, Declerck, Thierry, Gifu, Daniela, González-Blanco García, Elena, Gracia, Jorge, Ionov, Maxim, Mambrini, Francesco, McCrae, John, Pagé-Perron, Émilie, Passarotti, Marco, Ros Muñoz, Salvador & Truică, Ciprian-Octavian. 2022. When Linguistics Meets Web Technologies. Recent Advances in Modelling Linguistic Linked Data. *Semantic Web* 13 (6): 987–1050. <https://doi.org/10.3233/SW-222859>.
- Khan, Anas Fahad. 2025. Linguistic Linked (Open) Data. In Bański, Piotr, Heid, Ulrich, & Herzberg, Laura (a cura di), *Harmonizing Language Data*, 281–302. De Gruyter. <https://doi.org/10.1515/9783112208212-012>.
- Cimiano, Philipp, Chiarcos, Christian, McCrae, John P. & Gracia, Jorge. 2020. *Linguistic Linked Data. Representation, Generation and Applications*. Cham: Springer.

Dall'annotazione sintattica al web semantico: il caso del siciliano

La creazione e la rappresentazione semantica di risorse linguistiche per varietà di lingua sottorappresentate, come i dialetti italiani, costituiscono un ambito di ricerca ancora poco esplorato ma di grande rilevanza per la loro valorizzazione e per il loro studio sistematico. La realizzazione di una base di conoscenza del siciliano (Sprugnoli et al. 2025), modellata secondo i principi dei *Linked Open Data* (Cimiano et al. 2020), analogamente a quanto sviluppato anche per il dialetto di Parma (Sprugnoli & Muscianisi 2025), ha rappresentato un primo passo verso la costruzione di un ecosistema di risorse dialettali interoperabili. In continuità con questo approccio, l'obiettivo principale del nostro lavoro è duplice: (1) modellare e collegare in RDF una *treebank* del siciliano alla corrispettiva base di conoscenza lessicale; (2) realizzare la connessione semantica dei lemmi siciliani con le traduzioni italiane presenti nella base di conoscenza della lingua italiana sviluppata nel progetto *LiITA: Linking Italian* (Litta et al. 2024).

Il siciliano è un ottimo caso d'uso in questo contesto. Esso, infatti, all'interno dei dialetti d'Italia occupa una posizione particolare grazie alla sua antica tradizione letteraria, attestata già nel XIII secolo presso la corte di Federico II e riconosciuta da Dante come fondamentale per la nascita della letteratura in volgare. È necessario distinguere il siciliano letterario, che a partire dalla Grammatica di Giuseppe Pitre (2008 [1875]) è dotato di ortografia "storica" latineggiante con recenti proposte di normalizzazione (Accademia Siciliana 2024), dai siciliani parlati, frutto di tradizioni orali, con scritture non standardizzate e fortemente influenzate dall'italiano, lingua tetto nazionale. Dal punto di vista linguistico, il siciliano appartiene al gruppo dei dialetti meridionali estremi, insieme al salentino e al calabrese meridionale. Una caratteristica distintiva è il sistema vocalico "siciliano" (De Angelis 2024), influenzato anche da elementi ellenici e caratterizzato da un pentavocalismo tonico (IPA: /ɪ e a ɔ u/) e un trivocalismo atono (IPA: /ɪ a u/). Le varietà parlate in Sicilia presentano differenze legate alla geografia e alla presenza di popolazioni alloglotte. Dal punto di vista socio- ed etnolinguistico, i principali processi sono l'italianizzazione e la regionalizzazione dell'italiano, accompagnati da persistenti pregiudizi verso il dialetto. In questo contesto, il lessico rappresenta il livello più dinamico, riflettendo influenze diverse, la scomparsa di termini e la formazione di neologismi, in linea con la naturale evoluzione linguistica (Ruffino 2001).

La *treebank Sicilian-STB* (Cappello et al. 2025) è un *dataset* di testi paralleli italiano-siciliano annotato secondo le linee guida dell'iniziativa *Universal Dependencies* (De Marneffe et al. 2021).

L'annotazione include informazioni su lemmi, categorie grammaticali, tratti morfologici e relazioni di dipendenza sintattica. I lemmi costituiscono l'elemento centrale per il nostro lavoro, poiché rappresentano il nodo principale sia nella base di conoscenza del siciliano sia in quella dell'italiano. Quest'ultima è stata sviluppata nell'ambito del progetto PRIN *LiITA: Linking Italian* (<http://www.liita.it>), che mira a garantire l'interoperabilità tra risorse lessicali e testuali eterogenee tramite l'impiego degli standard de facto creati dalla comunità scientifica dei *Linked Open Data* (Litta et al. 2025).

Entrambe le basi di conoscenza sono modellate utilizzando ontologie standard come OntoLex-Lemon (McCrae et al. 2017) e costituiscono il fondamento per il collegamento semantico tra risorse lessicali e risorse testuali.

Il primo passo del lavoro consiste nell'estrazione di informazioni dalla *treebank UD Sicilian-STB*, con particolare attenzione a: lemma, parti del discorso, caratteristiche morfologiche e allineamenti tra siciliano e italiano forniti nella porzione parallela del corpus.

Tali informazioni vengono successivamente rappresentate in RDF, modellando i lemmi, le relazioni tra lemmi e *token* e le traduzioni italiane, che possono essere collegate alle corrispondenti entità lessicali presenti in LiITA. La modellazione di *treebank* è stata già adottata con successo sia nel caso dell'*Index Thomisticus Treebank* (Mambrini et al. 2022), contenente testi di Tommaso d'Aquino, sia nel caso di *UDante* (Passarotti et al. 2021), la risorsa sintatticamente annotata delle opere latine di Dante Alighieri.

La fase successiva riguarda il linking tra la rappresentazione RDF della *treebank* siciliana e la base di conoscenza lessicale. Da un lato, ogni lemma siciliano nella *treebank* viene associato all'entrata corrispondente nella base di conoscenza del siciliano; dall'altro, le traduzioni italiane consentono di creare un bridge semantico verso i lemmi presenti nella base di conoscenza dell'italiano.

I dati ottenuti attraverso le attività di modellazione e linking saranno resi pubblicamente disponibili, al fine di favorirne la consultazione, il riuso e l'integrazione in ulteriori studi. In particolare, le informazioni saranno pubblicate sotto forma di triple in formato RDF, secondo i principi dei *Linked Open Data*, garantendo così interoperabilità e accessibilità. Tali dati potranno essere interrogati tramite lo SPARQL endpoint del progetto LiITA, che consentirà agli utenti di esplorare e combinare dinamicamente le informazioni provenienti dalle diverse risorse collegate. Inoltre, sarà messo a disposizione un insieme di query SPARQL di esempio, con l'obiettivo di supportare gli utenti nelle loro ricerche e di illustrare possibili casi d'uso, mostrando come le

informazioni lessicali e testuali interoperabili possano essere integrate per analisi linguistiche più articolate.

Il linking tra la *treebank UD* per il siciliano, la base di conoscenza del siciliano e la *LiITA Knowledge Base* rappresenta un passo significativo verso l'integrazione di varietà linguistiche minoritarie all'interno di un'infrastruttura di dati linguistici interoperabili. Questo approccio consente di facilitare ricerche avanzate e approfondimenti linguistici, nonché di favorire nuovi scenari di utilizzo delle risorse in ambito lessicografico, tipologico e computazionale. È però importante sottolineare come l'interoperabilità aumenti con il crescere del numero e della varietà delle risorse collegate, consentendo analisi più ricche e affidabili attraverso diversi dataset. Per questo motivo, si auspica l'integrazione di un numero sempre maggiore di risorse, sia lessicali sia testuali, che coprano la più ampia gamma possibile di varietà dialettali. L'ampliamento della rete di dati dialettali interconnessi favorirebbe non solo la reperibilità e il riuso delle risorse, ma anche lo sviluppo di studi più sistematici e comparativi sulla variazione linguistica.

Riferimenti bibliografici

- Cademia Siciliana. 2024. Documento per l'ortografia del siciliano / Documentu pi l'ortografia dū sicilianu. 2^aed. <https://cademiasiciliana.org/wp-content/uploads/Cademia-Siciliana-DOS24-v.2.1.pdf>
- Cappello, Caterina Maria, D'Alì, Sabrina, Guglielmetti, Mario, Di Nuovo, Elisa & Bosco, Cristina 2025. Arbuli sunnu: A Sicilian-Italian parallel treebank. In Bosco, Cristina, Jezek, Elisabetta, Polignano, Marco & Sanguinetti, Manuela (a cura di), *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, Cagliari, 24-26 maggio 2025, 155–168. Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2025.clicit-1.17/>
- Cimiano, Philipp, Chiarcos, Christian, McCrae, John P. & Gracia, Jorge. 2020. *Linguistic Linked Data. Representation, Generation and Applications*. Cham: Springer.
- De Angelis, Alessandro. 2024. Sulla genesi del vocalismo “siciliano”. In Cennamo, Michela, Dovetto, Francesca Maria, Perri, Antonio, Schirru, Giancarlo & Sornicola, Rosanna (a cura di), *Linguistica e filologia tra Oriente e Occidente*, 245–275. Napoli: Il Calamo.
- De Marneffe, Marie-Catherine, Manning, Christopher D., Nivre, Joakim & Zeman, Zeman. 2021. Universal Dependencies. *Computational Linguistics* 47(2). 255–308. https://doi.org/10.1162/coli_a_00402

- Litta, Eleonora, Passarotti, Marco, Basile, Valerio, Bosco, Cristiana, Di Fabio, Andrea & Brasolin, Paolo. 2025. LiITA: A knowledge base of interoperable resources for Italian. In Alam, Mehwish, Tchechmedjiev, Andon, Gracia, Jorge, Gromann, Dagmar, di Buono, Maria Pia, Monti, Johanna & Ionov, Maxim (a cura di), *Proceedings of the 5th Conference on Language, Data and Knowledge, Napoli, 9-12 settembre 2025*, 130–135. Napoli: UniorPress. <https://doi.org/10.6093/978-88-6719-333-2>
- Litta, Eleonora, Passarotti, Marco, Brasolin, Paolo, Moretti, Giovanni, Basile, Valerio, Di Fabio, Andrea & Bosco, Cristiana. 2024. The lemma bank of the LiITA knowledge base. In Dell’Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*, Pisa, 4-6 dicembre 2024, 517–522, Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2024.clicit-1.0/>
- Mambrini, Francesco, Passarotti, Marco, Moretti, Giovanni & Pellegrini, Matteo. 2022. The Index Thomisticus Treebank as linked data. In Calzolari, Nicoletta, Frédéric, Béchet, Blache, Philippe, Choukri, Khalid, Cieri, Christopher, Declerck, Thierry, Goggi, Sara, Isahara, Hitoshi, Maegaard, Bente, Mariani, Joseph, Mazo, Hélène, Odijk, Jan & Piperidis, Stelios (a cura di), *Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, 20-25 giugno 2022*, 4022–4029. Paris: European Language Resources Association (ELRA). <https://aclanthology.org/2022.lrec-1/>
- McCrae, John P., Bosque-Gil, Julia, Gracia, Jorge, Buitelaar, Paul & Cimiano, Philipp. 2017. The OntoLex-Lemon model: Development and applications. In Kosem, Iztok, Tiberius, Carole, Jakubíček, Miloš, Kallas, Jelena, Krek, Simon & Baisa, Vít (a cura di), *Proceedings of the eLex 2017 Conference, Leiden, 19–21 settembre 2017*, 587–597. Brno: Lexical Computing CZ s.r.o.
- Passarotti, Marco, Cecchini, Flavio Massimiliano, Sprugnoli, Rachele & Moretti, Giovanni 2021. UDante. L’annotazione sintattica dei testi latini di Dante. *Studi Danteschi* 86. 309–338. <https://zenodo.org/records/7244878>
- Pitrè, Giuseppe. 2008 [1875]. *Grammatica siciliana*. Palermo: Sellerio (2008).
- Ruffino, Giovanni. 2001. *Sicilia*. Roma/Bari: Laterza.
- Sprugnoli, Rachele & Muscianisi, Domenico Giuseppe. 2025. Linked data and Italian dialectology. *L’Analisi Linguistica e Letteraria* 33(3). 105–126. [10.69117/ALL.2025.3.07](https://doi.org/10.69117/ALL.2025.3.07)

Sprugnoli, Rachele, Moretti, Giovanni, Muscianisi, Domenico Giuseppe & Litta, Eleonora. 2025. Ciallabacialla! Modeling and linking a regional lexical resource. In Bosco, Cristina, Jezek, Elisabetta, Polignano, Marco, Sanguinetti, Manuela (a cura di), Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025), 24-26 settembre 2025, 1093–1101. Aachen: CEUR Workshop Proceedings.

Interrogare LiITA in linguaggio naturale: tre approcci a confronto

L'accesso e l'interrogazione delle risorse linguistiche pubblicate come *Linked Open Data* (LOD) rappresentano ancora una barriera significativa per una parte rilevante della comunità di ricerca, a causa della necessità di padroneggiare formalismi come SPARQL e di conoscere in dettaglio la struttura dei grafi di conoscenza sottostanti. Tali difficoltà emergono in modo particolarmente evidente nel caso di risorse linguistiche articolate, che integrano e rendono interoperabili dati eterogenei secondo modelli ontologici condivisi.

In questo contesto si colloca LiITA (<https://www.liita.it/>) (Litta et al. 2024), una *knowledge base* aperta per la lingua italiana che mira a rendere interoperabili corpora, lessici, dizionari, glossari ed edizioni digitali attraverso l'applicazione dei principi del paradigma LOD. Le risorse integrate in LiITA sono collegate mediante una raccolta centrale di lemmi e modellate attraverso vocabolari ampiamente adottati nel dominio dei *Linked Open Data* linguistici, tra cui OntoLex-Lemon (Cimiano et al. 2020). Se da un lato questa architettura consente interrogazioni ricche e trasversali su risorse linguistiche eterogenee, dall'altro essa accentua la complessità delle modalità di accesso e interrogazione, rendendo necessaria una conoscenza approfondita delle strutture ontologiche e delle relazioni tra le diverse componenti della *knowledge base*.

Questo contributo presenta una ricerca volta a facilitare l'interrogazione della *knowledge base* LiITA attraverso interfacce che consentono la traduzione automatica di interrogazioni espresse in linguaggio naturale in *query* SPARQL corrette e semanticamente valide.

La ricerca si è concentrata sulla progettazione, implementazione e confronto di tre modelli distinti per la traduzione *Natural Language to SPARQL*, ciascuno caratterizzato da un diverso grado di integrazione tra approcci simbolici e modelli linguistici di grandi dimensioni (*Large Language Models*, LLM).¹ Tutti i modelli sono stati sviluppati tenendo conto delle specificità architetturali di LiITA e delle risorse linguistiche interoperabili in essa integrate.

NL2SPARQL implementa un approccio basato sul paradigma del *Retrieval-Augmented Generation* (Avila et al. 2025; Lewis et al. 2021): *retrieval* semantico (*sentence transformers* + database vettoriale) e algoritmi di *ranking* selezionano esempi rilevanti da fornire come *few-shot*

¹ I tre sistemi sviluppati supportano i seguenti LLM tramite API: *OpenAI GPT* (<https://platform.openai.com/>), *Anthropic Claude* (<https://platform.claude.com/docs>), *Mistral AI* (<https://docs.mistral.ai>), *Google Gemini* (<https://ai.google.dev>). È inoltre supportata l'esecuzione locale di modelli open-source tramite *Ollama* (<https://ollama.com>), senza dipendenze da servizi esterni.

learning a un LLM, che genera la *query* SPARQL completa. Il sistema include validazione sintattica, esecuzione contro l'*endpoint* LiITA e auto-correzione degli errori. È disponibile anche una variante *agent-based* che implementa un *workflow* di analisi, generazione, verifica e auto-correzione iterativa secondo il paradigma ReAct (Yao et al. 2023). Il sistema è inoltre accessibile tramite MCP (*Model Context Protocol*) Server, permettendo l'integrazione con *client* compatibili.

PRISMA (*Pattern-based Rules for Intent-driven SPARQL with Multiple-resource Assembly*) utilizza un LLM esclusivamente per la classificazione dell'intento dell'utente. Questo approccio è motivato dall'evidenza che i modelli linguistici ottengono risultati migliori quando ricevono informazioni ontologiche mirate piuttosto che quando generano *query* strutturate direttamente (Allemang & Sequeda 2024; Civili et al. 2025), mantenendo la generazione SPARQL completamente deterministica e basata su template, nella tradizione di Unger et al. (2012) e Shekarpour et al. (2013). Questa separazione garantisce trasparenza e correttezza strutturale, delegando all'LLM solo il compito di identificare l'intento ed estrarre i parametri dalla *query*.

MoSAIC (*Modular SPARQL Assembler for Interlinked Corpora*) adotta un'architettura *block-based* in cui pattern SPARQL riusabili vengono assemblati componendo *query* complete, nella tradizione dei sistemi *template-based* (Shekarpour et al., 2013; Unger et al. 2012). Il sistema offre due modalità: (a) deterministica, basata su regole e *pattern matching* senza utilizzo di LLM, garantendo prevedibilità e trasparenza; (b) agentica, in cui un LLM decompone la *query* in operazioni strutturate che vengono poi tradotte in blocchi SPARQL tramite regole deterministiche, secondo il paradigma ReAct (Yao et al. 2023).

Il contributo confronta questi tre approcci lungo diverse dimensioni: (1) trasparenza e interpretabilità del processo di traduzione; (2) accuratezza nella generazione di *query* SPARQL valide; (3) capacità di gestire *query* complesse che coinvolgono risorse multiple. L'analisi evidenzia i *trade-off* tra flessibilità linguistica (favorita da approcci *LLM-based end-to-end*) e garanzie di correttezza strutturale (favorita da approcci *pattern-based*), fornendo indicazioni per la progettazione di interfacce in linguaggio naturale per risorse linguistiche in LOD.

Questo lavoro si inserisce nella tematica (a) del workshop, trattando specificamente di strumenti a supporto dell'utilizzo di risorse linguistiche in LOD e dell'accesso e interrogazione di tali risorse, con particolare riferimento all'integrazione tra *Large Language Models* e conoscenza linguistica esplicita rappresentata in grafi di conoscenza LOD.

Riferimenti bibliografici

- Allemang, Dean & Sequeda, Juan. 2024. Increasing the accuracy of LLM question-answering systems with ontologies. In Demartini, Gianluca, Hose, Katja, Acosta, Maribel, Palmonari, Matteo, Cheng, Gong, Skaf-Molli, Hala, Ferranti, Nicolas, Hernández, Daniel, Hogan, Aidan (a cura di), *The Semantic Web – ISWC 2024. 23rd International Semantic Web Conference, Proceedings, Part III, Hanover, MD, USA 11 novembre 2024*, 324–339. Cham: Springer.
- Avila, Caio Viktor S., Vidal, Vânia M. P., Franco, Wellington & Casanova, Marco A. 2025. Few-shot learning or RAG in LLM-based text-to-SPARQL? Why not both? In *Proceedings of the 2025 IEEE 19th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 3-5 febbraio 2025*, 76–79. Los Alamitos: IEEE Computer Society.
- Cimiano, Philipp, Chiacos, Christian, McCrae, John P. & Gracia, Jorge. 2020. *Linguistic Linked Open Data: Representation, Generation and Applications*. Cham: Springer.
- Civili, Cristina, Sherkhonov, Evgeny, Stirewalt, Kurt. 2025. Using off-the-shelf LLMs to query enterprise data by progressively revealing ontologies. In Curry, Edward, Presutti, Valentina, McCrae, John P., Alam, Mehwish, Colpaert, Pieter, Parreira, Josiane Xavier, Collarana, Diego, Sabou, Marta, Harth, Andreas, Lisena, Pasquale (a cura di), *The Semantic Web: ESWC 2025 Satellite Events*, 305–309. Cham: Springer.
- Lewis, Patrick, Perez, Ethan, Piktus, Aleksandra, Petroni, Fabio, Karpukhin, Vladimir, Goyal, Naman, Küttler, Heinrich, Lewis, Mike, Yih, Wen-tau, Rocktäschel, Tim, Riedel, Sebastian & Kiela, Douwe. 2021. Retrieval-augmented generation for knowledge-intensive NLP tasks. In Ranzato, Marc’Aurelio, Beygelzimer, Alina, Dauphin, Yann, Liang, Percy S. & Vaughan, Jennifer W. (a cura di), *Advances in Neural Information Processing Systems*, vol. 34, 9459–9474. Red Hook (NY): Curran Associates.
- Litta, Eleonora, Passarotti, Marco, Brasolin, Paolo, Moretti, Giovanni, Basile, Valerio, Di Fabio, Andrea & Bosco, Cristina. 2024. The lemma bank of the LiITA knowledge base of interoperable resources for Italian. In Dell’Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, 4-6 dicembre 2024*, 517–522. Aachen: CEUR Workshop Proceedings.
- Shekarpour, Saeedeh, Auer, Sören, Ngonga Ngomo, Axel-Cyrille, Gerber, Daniel, Hellmann, Sebastian & Stadler, Claus. 2013. Generating SPARQL queries using templates. *Web Intelligence and Agent Systems* 11(3). 283–295. <https://doi.org/10.3233/WIA-13027>

Unger, Christina, Bühmann, Lorenz, Lehmann, Jens, Ngonga Ngomo, Axel-Cyrille, Gerber, Daniel & Cimiano, Philipp. 2012. Template-based question answering over RDF data. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12), Lyon France, 16 – 20 aprile 2012*, 639–648. New York: Association for Computing Machinery.

Yao, Shunyu, Zhao, Jeffrey, Yu, Dian, Du, Nan, Shafran, Izhak, Narasimhan, Karthik & Cao, Yuan. 2023. ReAct: Synergizing reasoning and acting in language models. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023), Kigali Rwanda, 1-5 maggio 2023*, 1-33 Kigali: OpenReview.net.

Il Lessico Etimologico Italiano (LEI) come Linked Open Data e il suo collegamento alle Knowledge Base LiITA e LiLa

Il contributo mira a presentare la pubblicazione del *Lessico Etimologico Italiano* (LEI) in *Linked Open Data* (LOD) e il suo *linking* con la *Lemma Bank* di due *Knowledge Base* di risorse LOD, ovvero LiITA (*Linking Italian*)² e LiLa (*Linking Latin*)³.

Il LEI, dizionario storico-etimologico delle varietà italo-romanze fondato da Max Pfister e diretto da Elton Prifti e Wolfgang Schweickard, pubblicato a partire dal 1979, contiene al suo interno sia materiale latino (limitatamente alla sezione dei Latinismi)⁴ che lessico italiano. Infatti, dal punto di vista strutturale, il LEI mette a lemma le basi etimologiche, mentre nel corpo degli articoli si distribuiscono, in base a diversi criteri d'ordine, le forme italo-romanze.

Negli ultimi anni il LEI ha intrapreso un processo di digitalizzazione (progetto *LEI Digitale*) che interessa sia il materiale edito in formato analogico che l'infrastruttura redazionale per la produzione di nuovi articoli lessicografici *born-digital* (Prifti 2019; 2022).

La fase di digitalizzazione del cartaceo ha finora prodotto una grande mole di dati lessicografici in formato XML proprietario, consultabili online alla pagina <https://online.lei-digitale.it>. La disponibilità dei dati in formato digitale ha permesso di effettuare sperimentazioni in ottica di un'eventuale conversione in RDF del dizionario (Nannini 2023; 2025). La presente comunicazione rappresenta un ulteriore passo in questa direzione.

L'estrazione e la modellazione dei dati del *LEI Digitale* secondo il modello OntoLex-Lemon (McCrae et al. 2017) sono i primi passi necessari al fine di creare le condizioni per l'interoperabilità con altre risorse linguistiche nel Web Semantico.

La successiva fase di *linking* ha, infatti, come obiettivo il collegamento con le *Lemma Bank* di LiLa e di LiITA, rispettivamente per le basi etimologiche e per le entrate italiane.

Una prima analisi condotta sui primi 8 volumi del LEI (A-B) ha evidenziato 2543 basi etimologiche e un totale di 57874 entrate contrassegnate come *it.* (italiano) e *it.a.* (italiano antico). Questi numeri saranno poi filtrati nella fase di matching con i lemmi di LiLa e LiITA.

² <https://www.liita.it/> (ultimo accesso: 14/02/2026).

³ <https://lila-erc.eu/> (ultimo accesso: 14/02/2026).

⁴ Dal punto di vista macrostrutturale, il LEI si articola in diverse sezioni relative all'origine degli etimi (cf. Aprile 2004).

Le *Knowledge Base* LiLa (Passarotti et al. 2020) e LiITA (Litta et al. 2025) sono fondate su una architettura che ha come propria componente fondamentale una *Lemma Bank*, ovvero una raccolta di forme canoniche di citazione delle parole appartenenti rispettivamente al lessico latino (LiLa) e a quello italiano (LiITA). Dal punto di vista ontologico, le *Lemma Bank* sono insiemi di istanze appartenenti alla classe `lila:Lemma`, che è una sottoclasse di `ontolex:Form`. Il lavoro di *linking* tra i LEI e le due *Knowledge Base* consiste, dunque, nel produrre triple che, registrate nel grafo RDF del LEI, rappresentano la relazione tra la voce lessicale (latina o italiana) nel LEI e il corrispondente lemma nelle *Lemma Bank* di LiLa e LiITA. Questa relazione è rappresentata tramite il ricorso alla Proprietà `ontolex:canonicalForm`.

Una volta realizzare le triple di collegamento tra il LEI e LiLa/LiITA e pubblicato il LEI come RDF, i suoi contenuti risulteranno essere interoperabili con quelli delle altre risorse linguistiche pubblicate come LOD nelle due *Knowledge Base*, consentendo così ricerche federate su più grafi incrociando dati e metadati di diverse risorse sia lessicali che testuali

Riferimenti bibliografici

Aprile, Marcello. 2004. *Le strutture del lessico etimologico italiano*. Galatina: Congedo.

LEI = Prifti, Elton & Schweickard, Wolfgang (a cura di). 1979–. *LEI. Lessico Etimologico Italiano. Fondato da Max Pfister*. Wiesbaden: Reichert. *LEI digitale*: <https://lei-digitale.it> (ultimo accesso: 14/02/2026).

Litta, Eleonora, Passarotti, Marco Carlo, Basile, Valerio, Bosco, Cristina, Di Fabio, Andrea & Brasolin, Paolo. 2025. LiITA: A Knowledge Base of Interoperable Resources for Italian. In Alam, Mehwish, Tchechmedjiev, Andon, Gracia, Jorge, Gromann, Dagmar, di Buono, Maria Pia, Monti, Johanna & Ionov, Maxim (a cura di), *Proceedings of the 5th Conference on Language, Data and Knowledge, Napoli, 9-12 settembre 2025*, 130–135. Napoli: UniorPress. <https://doi.org/10.6093/978-88-6719-333-2>

McCrae, John P., Bosque-Gil, Julia, Gracia, Jorge, Buitelaar, Paul & Cimiano, Philipp. 2017. The OntoLex-Lemon Model: Development and Applications. In Kosem, Iztok, Tiberius, Carole, Jakubíček, Miloš, Kallas, Jelena, Krek, Simon & Baisa, Vít (a cura di), *Proceedings of the eLex 2017 Conference, Leiden, 19–21 settembre 2017*, 587–597. Brno: Lexical Computing CZ s.r.o.

Nannini, Alessandro Alfredo. 2023. La mappatura semantica del «Lessico Etimologico Italiano (LEI)». Possibilità, metodi e prospettive. In Becker, Lidia, Kuhn, Julia, Ossenkop, Christina,

- Polzin-Haumann, Claudia & Prifti, Elton (a cura di), *Digitale romanistische Sprachwissenschaft. Akten des Romanistischen Kolloquiums XXXIV*, 119–145. Tübingen: Narr.
- Nannini, Alessandro Alfredo. 2025. Sulla rappresentazione delle informazioni semantiche nei dizionari storici digitali. *Umanistica Digitale* 9(20). 221–260. <https://doi.org/10.6092/issn.2532-8816/21250>.
- Passarotti, Marco Carlo, Mambrini, Francesco, Franzini, Greta, Cecchini, Flavio Massimiliano, Litta, Eleonora, Moretti, Giovanni, Ruffolo, Paolo & Sprugnoli, Rachele. 2020. Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin. *Studi e Saggi Linguistici* 58(1). 177–212. <https://doi.org/10.4454/ssl.v58i1.277>.
- Prifti, Elton. 2019. Lo stato della digitalizzazione del «LEI». Un resoconto. In Leonardi, Lino & Squillaciotti, Paolo (a cura di), *Italiano antico, italiano plurale*, 233–242. Alessandria: Edizioni dell’Orso.
- Prifti, Elton. 2022. Il «LEI digitale». Un resoconto, con particolare attenzione alla dialettologia. In Cortelazzo, Michele A., Morgana, Silvia & Prada, Massimo (a cura di), *Lessicografia storica dialettale e regionale*, 293–314. Firenze: Franco Cesati.

La lessicografia storica nel Web Semantico: una sperimentazione sul TLIO

Il contributo presenta una sperimentazione di modellizzazione in *Linked Open Data* (LOD) del *Tesoro della Lingua Italiana delle Origini* (TLIO), vocabolario storico dell'italiano antico elaborato presso l'Opera del Vocabolario Italiano (OVI-CNR). Il TLIO descrive il lessico dell'italiano delle origini sulla base dell'intera documentazione testuale disponibile, dalle prime attestazioni fino al limite convenzionale del 1375 (anno della morte di Boccaccio). Una caratteristica rilevante del TLIO è il trattamento unitario delle diverse varietà linguistiche dell'italiano antico, comprese quelle che oggi vengono definite dialettali, in coerenza con il quadro storico precedente alla formazione di una lingua nazionale. Il vocabolario presenta inoltre una struttura redazionale articolata e una ricca stratificazione delle informazioni lessicografiche, ed è pubblicato progressivamente in formato digitale, configurandosi come una risorsa complessa, di ampia estensione e in continua evoluzione.

Negli ultimi anni, le tecnologie del Web Semantico e il paradigma dei *Linked Open Data* (LOD) sono stati progressivamente applicati alla pubblicazione e all'interoperabilità di risorse linguistiche e lessicografiche (Krek *et al.* 2018). Esperienze che hanno sperimentato la rappresentazione dei dizionari come *Linked Data* includono, tra gli altri, il DEAF – *Dictionnaire étymologique de l'ancien français* (Tittel & Chiarcos 2018, Tittel *et al.* 2018), iniziative di modellizzazione del LEI – *Lessico Etimologico Italiano* (Prifti 2019) e infrastrutture come LiLa – *Linking Latin* (Mambrini *et al.* 2020, Passarotti *et al.* 2020). Tali lavori hanno assunto il modello OntoLex-Lemon (Cimiano *et al.* 2016; McCrae *et al.* 2017) come standard di riferimento per trasformare i dizionari in rappresentazioni formali strutturate, permettendo di separare la descrizione delle unità lessicali dalla loro organizzazione concettuale. È da notare che, sebbene tale modello sia stato applicato con successo a varie risorse lessicografiche, il suo impiego per l'italiano antico risulta ancora relativamente poco esplorato. Un caso significativo e termine di confronto rilevante è costituito dal progetto VocaBo – *Vocabolario di Boccaccio Online*, la cui formalizzazione è discussa in Bellandi (2025). Tra le iniziative recenti per l'italiano si segnala anche il progetto LiIta – *Linking Italian* (Litta *et al.* 2024, 2025a, 2025b), che, in continuità con LiLa, sviluppa una *Knowledge Base* di risorse linguistiche interoperabili, il cui componente centrale è costituito da un repertorio di lemmi (*Lemma Bank*), rappresentato secondo il modello OntoLex-Lemon. Si tratta di un sistema orientato all'integrazione tra risorse lessicali e risorse testuali; la *Lemma Bank* non costituisce una raccolta di voci lessicali, ma un insieme di forme canoniche di citazione. Il presente contributo si colloca invece sul piano della modellizzazione della struttura interna delle voci lessicografiche.

A partire da questo quadro, il contributo propone un caso di studio esplorativo volto a valutare in che misura il paradigma LOD consenta di rappresentare in modo adeguato la complessità strutturale e semantica del TLIO, preservandone le scelte redazionali e rendendolo al tempo stesso interoperabile con altre risorse linguistiche.

In linea con la letteratura scientifica più recente sulla modellizzazione di risorse lessicografiche in LOD, la metodologia proposta adotta come riferimento il modello OntoLex-Lemon per la descrizione delle unità lessicali, integrandolo con il modulo Lexicog (Bosque-Gil & Gracia 2019) per la rappresentazione formale della struttura delle voci e dell'ordine dei significati, e con SKOS - *Simple Knowledge Organisation System* (Miles & Bechhofer 2009) per la formalizzazione dei concetti e delle relazioni semantiche. In tale prospettiva, ciascuna accezione, sotto-accezione e sotto-definizione viene associata a uno o più concetti distinti, successivamente mappati su una struttura tassonomica esistente. Il modello ontologico sviluppato è denominato TLIOntology.

Per mantenere la sperimentazione contenuta e metodologicamente controllabile, lo studio si concentra su un sottoinsieme limitato di voci. Tale scelta consente di osservare in modo mirato alcune criticità di modellizzazione che risultano generalizzabili anche ad altri casi del TLIO. In particolare, sono state selezionate circa un centinaio di voci riconducibili al campo semantico della *paura*, ampiamente indagato nella riflessione dantesca e negli studi sul lessico delle emozioni nella lirica medievale (cfr. Rea 2012, 2014, 2015). Tale dominio, caratterizzato da una forte presenza di significati astratti, da una fitta rete di slittamenti semantici e da una polireferenzialità che oscilla tra dimensione psicologica e fisica, assume particolare rilevanza nella fase di mappatura tassonomica dei concetti. Le voci del TLIO presentano una struttura interna fortemente articolata: ciascuna voce include l'insieme delle forme grafiche riconducibili al lemma, informazioni etimologiche, la distribuzione geolinguistica delle attestazioni, nonché un sistema strutturato di definizioni, sotto-definizioni e sotto-accezioni, ciascuna accompagnata da esempi tratti dal corpus ed eventualmente da marche di varia natura (grammaticale, d'uso e semantiche). Tale complessità, dovuta alla presenza di informazioni di varia natura all'interno delle voci, solleva interrogativi sulla copertura effettiva dei diversi campi offerta dai modelli disponibili. L'uso di estensioni come il modulo Lexicog consente di formalizzare l'architettura editoriale delle voci, mantenendo la distinzione tra definizioni, sotto-definizioni e sotto-accezioni. La sperimentazione ha tuttavia messo in evidenza alcune criticità nell'applicazione del modello, in particolare per quanto riguarda la rappresentazione delle marche d'uso e delle marche semantiche, solo parzialmente coperte dagli standard disponibili, nonché la gestione di unità definitorie prive di una glossa esplicita. Nell'ambito della formalizzazione sono stati pertanto definiti diversi vocabolari controllati associati all'ontologia, relativi ai riferimenti bibliografici dei testi indicati negli esempi, alla datazione, alle aree geolinguistiche, oltre a tre distinti

set di marche. Tali vocabolari contribuiscono a rendere esplicite e normalizzate le informazioni presenti nelle voci, favorendone l'interoperabilità con altre risorse linguistiche.

L'elenco delle voci effettivamente modellate è dinamico e in evoluzione. Al momento comprende poco più di 130 voci, una percentuale ridotta rispetto all'estensione complessiva del TLIO, per il quale sono previste circa 59.000 voci, di cui oltre 53.000 già redatte e circa 46.000 (79%) attualmente disponibili online. Nonostante la limitatezza del campione, il dataset selezionato consente una copertura totale delle configurazioni strutturali presenti nel vocabolario, tra cui accezioni organizzate in modo gerarchico, rinvii etimologici, ambiguità legate alla categoria grammaticale dei lemmi, presenza di marche specifiche, annotazioni metalinguistiche.

Da un punto di vista tecnico, l'integrità strutturale delle voci del TLIO è garantita da norme redazionali validate in modo semi-automatico mediante la conversione del testo redatto in un formato XML specifico per il TLIO. L'esistenza di questo livello di validazione formale consente una mappatura ontologica sistematica e automatica dell'intera struttura delle voci. Il modello costruito sul dataset di prova opera infatti sulla struttura formale, e non sulle informazioni specifiche delle singole entrate, realizzando una conversione a livello di schema da XML a RDF. La conversione è automatizzata mediante script in Python, con Turtle come formato di destinazione; il grafo finale è validato usando regole definite in formato SHACL. Ciò implica che, una volta definite e verificate le regole di mappatura e di validazione, la mappatura strutturale possa essere applicata a qualunque voce del TLIO. Il lavoro sul dataset di sperimentazione ha quindi come obiettivo la modellizzazione della struttura complessiva del vocabolario nel suo complesso.

Parallelamente alla modellizzazione della struttura delle voci, i concetti sono mappati su una tassonomia onomasiologica esterna, l'*Historical Thesaurus of English* – HTE (Kay et al. 2021), adottata in modo esplorativo e non normativo, come strumento per mettere in luce scarti concettuali, asimmetrie e limiti dell'allineamento tra significati storicamente situati e sistemi di organizzazione semantica sviluppati in contesti linguistici e culturali differenti.

Un precedente significativo in questa direzione è rappresentato dal progetto DHistOntology (Giuliani 2024; Giuliani & Molina Sangüesa 2020), che ha sperimentato l'allineamento a posteriori tra le voci del TLIO e la tassonomia dell'HTE nel dominio medico, evidenziando problematiche analoghe legate alla granularità concettuale e al rischio di anacronismo.

Anche la mappatura semantica e concettuale presenta criticità rilevanti: il confronto con sistemi concettuali esterni comporta infatti il rischio di anacronismo e mette in luce la non perfetta sovrapponibilità tra i concetti ricostruiti dal dizionario e le categorie proposte da tassonomie o

ontologie sviluppate in contesti diversi, come evidenziato da studi su lessici emotivi storici rappresentati come LOD (cfr. Khan *et al.* 2016). Nel caso di studio tali problemi sono particolarmente evidenti, poiché i valori semantici del concetto di *paura* sono fortemente dipendenti dal contesto storico, culturale e discorsivo, e difficilmente riconducibili a definizioni astratte o universalizzanti. Nel Medioevo, la paura era infatti spesso concepita in stretta relazione con la fisiologia del corpo. La mappatura richiede quindi di distinguere accuratamente tra la paura come emozione e le sue manifestazioni fisiche o mediche, evitando di imporre categorie psicologiche moderne a testi del XIV secolo.

A partire dal caso del TLIO, il contributo intende offrire una riflessione sull'uso del paradigma LOD per l'interoperabilità delle risorse lessicografiche storiche dell'italiano, discutendone potenzialità e limiti. La sperimentazione ha consentito di rappresentare in OntoLex-Lemon la struttura interna delle voci del TLIO, rendendo espliciti i diversi livelli informativi che le compongono. Su questa base diventa possibile effettuare interrogazioni strutturate dei dati, che si affiancano alla ricerca tradizionale full-text già presente nel sistema. È così possibile estendere la ricerca a tutti i singoli campi del vocabolario e alla loro combinazione, ad esempio selezionando tutte le voci riferibili a un dato arco temporale o a una specifica varietà o area geolinguistica. Il lavoro svolto è parallelo e integrato con lo sviluppo della piattaforma Pluto (Piattaforma Lessicografica Unica del Tesoro delle Origini), sistema per la redazione, la gestione e la pubblicazione del TLIO (Arcidiacono 2022), che prevede una versione XML-TEI delle voci e nuove modalità di interrogazione (ad esempio la ricerca per etimi). Le funzionalità di ricerca rese possibili dalla mappatura LOD del TLIO saranno progressivamente integrate con le altre risorse informatiche sviluppate all'OVI e messe a disposizione online.

Riferimenti bibliografici

- Arcidiacono, Salvatore. 2022. *Lessicografia elettronica e italiano delle origini*. Palermo: Centro di studi filologici e linguistici siciliani.
- Bellandi, Andrea. 2025. Lexicographic Resources in the Semantic Web: Models, Tools, and Case Studies. *Umanistica Digitale* 9 (21), 107–140. <https://doi.org/10.60923/issn.2532-8816/22199>
- Bosque-Gil, Julia & Gracia, Jorge (a cura di). 2019. *The OntoLex Lemon Lexicography Module*. W3C Community Group Final Report. (Report) <https://www.w3.org/2019/09/lexicog/>

- Cimiano, Philipp, McCrae, John P. & Buitelaar, Paul. 2016. *Lexicon Model for Ontologies: Community Report*. W3C Community Group Final Report. (Report) <https://www.w3.org/2016/05/ontolex/>
- Giuliani, Mariafrancesca. 2024. Dai repertori storici alle “cose” del passato passando per le parole: risorse, domande e strategie della ricerca. *Bollettino dell’Atlante lessicale degli antichi volgari italiani* 17. 13–34. [10.19272/202409601002](https://doi.org/10.19272/202409601002)
- Giuliani, Mariafrancesca & Molina Sangüesa, Itziar. 2020. Hacia una taxonomía integrada en la redacción y revisión de diccionarios históricos. *Bollettino dell’Opera del Vocabolario Italiano* 25. 325–374.
- Kay, Christian, Roberts, Jane, Samuels, Michael, Wotherspoon, Irené & Alexander, Marc (a cura di). 2020. *The Historical Thesaurus of English*, versione 5.0. Glasgow: University of Glasgow.
- Khan, Fahad, Díaz Vera, Javier E. & Monachini, Monica. 2016. The Representation of an Old English Emotion Lexicon as Linked Open Data. In McCrae, John P., Chiarcos, Christian, Cimiano, Philipp, Declerck, Thierry & Gracia, Jorge (a cura di), *Proceedings of the LREC 2016 Workshop “LDL 2016 - 5th Workshop on Linked Data in Linguistics”*, Portorož, 24 maggio 2016, 73–76. Paris: European Language Resources Association.
- Krek, Simon, Kosem, Iztok, McCrae, John P., Navigli, Roberto, Pedersen, Bolette S., Tiberius, Carole & Wissik, Tanja. 2018. European Lexicographic Infrastructure (ELEXIS). In Čibej, Jaka, Gorjanc, Vojko, Kosem, Iztok & Krek, Simon (a cura di), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. 881–892. Ljubljana: Ljubljana University Press.
- Litta, Eleonora, Passarotti, Marco, Brasolin, Paolo, Moretti, Giovanni, Basile, Valerio, Di Fabio, Andrea & Bosco, Cristina. 2024. The Lemma Bank of the LiITA Knowledge Base of Interoperable Resources for Italian. In Dell’Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024)*, Pisa, 4-6 dicembre 2024, 517–522. Aachen: CEUR Workshop Proceedings.
- Litta, Eleonora, Passarotti, Marco, Basile, Valerio, Bosco, Cristina, Di Fabio, Andrea & Brasolin, Paolo. 2025a. LiITA: a Knowledge Base of Interoperable Resources for Italian. In Alam, Mehwish, Tchechmedjiev, Andon, Gracia, Jorge, Gromann, Dagmar di Buono, Maria Pia, Monti, Johanna & Ionov, Maxim (a cura di), *Proceedings of the 5th Conference on Language*,

Data and Knowledge, Napoli, 9-12 settembre 2025, 130–135. Napoli: UniorPress.
<https://doi.org/10.6093/978-88-6719-333-2>

- Litta, Eleonora, Passarotti, Marco, Brasolin, Paolo, Basile, Valerio, Bosco, Cristina & Di Fabio, Andrea. 2025b. LiITA, una Knowledge Base di risorse interconnesse per l'italiano. In Rebora, Simone, Rospocher, Marco & Bazzaco, Stefano (a cura di), *Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica nell'Era dell'Intelligenza Artificiale. Atti del XIV Convegno Annuale dell'AIUCD, Verona, 11-13 giugno 2025*, 460–465. Verona: Quaderni di Umanistica Digitale.
- Mambrini, Francesco, Cecchini, Flavio Massimiliano, Franzini, Greta, Litta, Eleonora, Passarotti, Marco Carlo & Ruffolo, Paolo. 2020. LiLa: Linking Latin. Risorse linguistiche per il latino nel Semantic Web. *Umanistica Digitale* 8. 63–78. <https://doi.org/10.6092/issn.2532-8816/9975>
- McCrae, Jhon P., Bosque-Gil, Julia, Gracia, Jorge, Buitelaar, Paul & Cimiano, Philipp. 2017. The OntoLex-Lemon Model: Development and Applications. In Kosem, Iztok, Tiberius, Carole, Jakubíček, Miloš, Kallas, Jelena, Krek, Simon & Baisa, Vít (a cura di), *Electronic lexicography in the 21st century: Lexicography from scratch. Proceedings of eLex 2017*, 587–597. Brno: Lexical Computing CZ s.r.o.
- Miles, Alistair & Bechhofer, Sean. 2009. *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation. <https://www.w3.org/TR/skos-reference/>
- Passarotti, Marco, Mambrini, Francesco, Franzini, Greta, Cecchini, Flavio Massimiliano, Litta, Eleonora, Moretti, Giovanni, Ruffolo, Paolo & Sprugnoli, Rachele. 2020. Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin. *Studi e Saggi Linguistici* 58 (1), 177–212. <https://doi.org/10.4454/ssl.v58i1.277>
- Prifti, Elton. 2019. Lo stato della digitalizzazione del LEI. Un resoconto. In Lino, Leonardi & Paolo, Squillaci (a cura di), *Italiano antico, italiano plurale. Testi e lessico del Medioevo nel mondo digitale: Atti del convegno internazionale in occasione delle 40.000 voci del TLIO, Firenze, 13-14 settembre 2018*, 233–242. Alessandria: Edizioni dell'Orso.
- Rea, Roberto. 2012. Psicologia ed etica della paura nel primo canto dell'Inferno: la compunctio timoris. *Dante Studies* 130. 183–206.
- Rea, Roberto. 2014. «Di paura il cor compunto»: teologia della Paura nel prologo dell'Inferno. In Canettieri, Paolo & Punzi, Arianna (a cura di), *Dai pochi ai molti. Studi in onore di Roberto Antonelli*, vol. II, 1433–1446. Roma: Viella.

- Rea, Roberto. 2015. Appunti sul lessico e la psicologia della Paura nei *Rerum vulgarium fragmenta*. In Decaria, Alessio & Leonardi, Lino (a cura di), *Ragionar d'amore. Il lessico delle emozioni nella lirica medievale, Atti del convegno di Siena, Siena, 17-19 aprile 2013*, 259–280. Firenze: Edizioni del Galluzzo.
- Tittel, Sabine, Chiarcos, Christian. 2018. Historical Lexicography of Old French and Linked Open Data: Transforming the resources of the Dictionnaire étymologique de l'ancien français with OntoLex-Lemon. In McCrae, John P., Chiarcos, Christian, Declerck, Thierry, Gracia, Jorge & Klimek, Bettina (a cura di), *Proceedings of the 6th Workshop on Linked Data in Linguistics: Towards Linguistic Data Science, co-located with LREC2018, Miyakaki, 12 maggio 2018*, xx–xx. Parigi: European Language Resources Association (ELRA).
- Tittel, Sabine, Bermúdez-Sabel, Helena & Chiarcos, Christian. 2018. Using RDFa to Link Text and Dictionary Data for Medieval French. In McCrae, John P., Chiarcos, Christian, Declerck, Thierry, Gracia, Jorge & Klimek, Bettina (a cura di), *Proceedings of the 6th Workshop on Linked Data in Linguistics (LDL-2018), co-located with LREC 2018, Miyakaki, 12 maggio 2018*, 30–38. Parigi: European Language Resources Association.

Dati linguistici FAIR per l'addestramento supervisionato di un'IA. Problemi di annotazione e possibili soluzioni per dati aperti da tre a cinque stelle

Il contributo presenta i problemi di annotazione e le soluzioni trovate per il progetto PRIN 2022 *Human-Centered Artificial Intelligence for Gender Sensitive Linguistic Simplification in German, English and Italian Institutional Communication* (HCAI4Gend) per la creazione di dataset linguistici FAIR da tre a cinque stelle del modello “*Open Stars*” di Tim Berenes Lee (Bizer, Heath & Berners-Lee 2009).

Il progetto PRIN 2022 HCAI4Gend ha affrontato lo studio di un modello linguistico multilingua per lo sviluppo di un editor di testo *open source* basato su un'intelligenza artificiale (IA) addestrata per rilevare e generare soluzioni linguistiche relative all'uso dei generi grammaticali e sociali nei testi istituzionali in lingua italiana, inglese, tedesca.

Il modello linguistico è stato progettato per l'addestramento supervisionato di un'IA specializzata per lo specifico dominio d'uso della comunicazione istituzionale amministrativa d'ateneo. Per questo modello linguistico sono state progettate e create risorse linguistiche annotate manualmente, quali corpora di testi istituzionali comparabili, lessici e grammatiche, corredate da metadati documentali con informazioni socio-pragmalinguistiche (Vellutino 2024) per migliorare la precisione, la reperibilità, interoperabilità e la riutilizzabilità secondo i principi FAIR (Wilkinson et al. 2016).

Per la costruzione di questi dataset linguistici abbiamo dovuto affrontare problemi relativi alle annotazioni multilingui e trovare possibili soluzioni. I dati linguistici prodotti nell'ambito del progetto sono oggettivi e sono stati raccolti in prospettiva sincronica in contesti comunicativi reali.

Per garantire coerenza nell'analisi linguistica tra le lingue, per la creazione di corpora di testi scritti istituzionali di 283 atenei (133 di lingua inglese, 34 di lingua italiana, 116 di lingua tedesca) sono state individuate diverse tipologie testuali scelte proprio perché comparabili: statuti, regolamenti, moduli, contratti e bandi di concorso. Pertanto, le risorse linguistiche sono costituite da testi autentici in formato di documento *Portable Document Format*, (PDF), vale a dire dati a una stella, che sono stati poi convertiti in formato di file di testo semplice (TXT) per avere dati grezzi a tre stelle. utili per essere processati attraverso la piattaforma di annotazione linguistica *open source* Doccano che consente la creazione di dataset linguistici formato *JavaScript Object Notation Lines*, (JSONL), uno standard aperto, leggero e basato su testo, ampiamente utilizzato per la rappresentazione e lo scambio di dati aperti (*open data*) strutturati.

Una prima difficoltà è stata proprio la scelta della piattaforma di annotazione manuale che avrebbe dovuto avere le funzionalità di gestione delle etichette di annotazione e di assicurare il lavoro collaborativo sui testi. Altra difficoltà, a seguito dell'impresa dell'accordo di annotazione, è stata proprio la gestione delle etichette di annotazione. Infatti, la segmentazione del testo richiesta dal software è risultata talvolta inadeguata rispetto alla complessità dei fenomeni linguistici osservati.

Nel caso dei testi istituzionali, ad esempio, alcune strategie discorsive o scelte lessicali legate alla rappresentazione del genere emergono solo a livello di frase complessa o di sequenze testuali più ampie (come la diatesi passiva), mentre Doccano consente di operare su unità più ristrette. Un'ulteriore criticità riguarda la definizione delle etichette. Nella pratica dell'annotazione è spesso necessario trovare un equilibrio tra precisione teorica e semplicità operativa. Categorie troppo dettagliate possono generare ambiguità e ridurre l'accordo tra annotatori, mentre categorie eccessivamente generiche rischiano di impoverire l'analisi linguistica.

Nel nostro caso l'uso di Doccano, ad esempio, oltre all'accordo di annotazione per le diverse lingue, ha previsto di considerare che la necessità di selezionare etichette predefinite (come ASTERISCO per tutti i simboli non facenti parte dell'ortografia standard della lingua tedesca) può portare a semplificazioni che non sempre riflettono adeguatamente la complessità dei fenomeni analizzati (*Student/in* rispetto a *Student*in* riflette una binarietà di genere e non una pluralità). A queste difficoltà si aggiunge il problema della coerenza inter-annotatore. Quando il lavoro di annotazione coinvolge più ricercatori/trici, diventa fondamentale definire un accordo tra annotatori/trici, vale a dire condividere linee guida e procedure di validazione che permettano di misurare e migliorare il livello di accordo. Anche in presenza di strumenti digitali avanzati, infatti, l'annotazione rimane un processo interpretativo, fortemente influenzato dalle competenze linguistiche e dalle scelte metodologiche dei/delle ricercatrici.

Il nostro studio, nato dalla stretta collaborazione tra linguisti e informatici, ci ha comunque consentito di sperimentare un modello linguistico basato su risorse linguistiche che possono essere considerate “scatole aperte” per affermare il paradigma *Human-Centered Artificial Intelligence* (Shneiderman 2020).

Riferimenti bibliografici

Bizer, Christian, Heath, Tom & Berners-Lee, Tim. 2009. Linked Data – The Story So Far. In Seneviratne, Oshani & Hendler, James (a cura di) *Linking the World's Information: Essays on*

Tim Berners-Lee's Invention of the World Wide Web, 115–143. New York: Association for Computing Machinery. <https://doi.org/10.1145/3591366.3591378>

Shneiderman, Ben. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction* 36. 1–10. <https://doi.org/10.1080/10447318.2020.1741118>

Wilkinson, Mark D, Dumontier, Michel, Aalbersberg, IJsbrand Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, da Silva Santos, Luiz Bonino, Bourne, Philip E, Bouwman, Jildau, Brookes, Anthony J, Clark, Tim, Crosas, Mercè, Dillo, Ingrid, Dumon, Olivier, Edmunds, Scott, Evelo, Chris T, Finkers, Richard, Gonzalez-Beltran, Alejandra, Gray, Alasdair JG, Groth, Paul, Goble, Carole, Grethe, Jeffrey S, Heringa, Jaap, 't Hoen, Peter AC, Hooft, Rob, Kuhn, Tobias, Kok, Ruben, Kok, Joost, Lusher, Scott J, Martone, Maryann E, Mons, Albert, Packer, Abel L, Persson, Bengt, Rocca-Serra, Philippe, Roos, Marco, van Schaik, Rene, Sansone, Susanna-Assunta, Schultes, Erik, Sengstag, Thierry, Slater, Ted, Strawn, George, Swertz, Morris A, Thompson, Mark, van der Lei, Johan, van Mulligen, Erik, Velterop, Jan, Waagmeester, Andra, Wittenburg, Peter, Wolstencroft, Katherine, Zhao, Jun, Mons, Barend. 2016. The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3. 160018. <https://doi.org/10.1038/sdata.2016.18>

Vellutino, Daniela. 2024. Quali varietà d'italiano insegnare all'IA? Sviluppo di un modello linguistico. *Studi Italiani di Linguistica Teorica e Applicata*, LIII (3). 558–573.

Emozioni in *Linked Open Data*: il caso di ELIta e Sentix

L'intreccio tra parole ed emozioni, il modo in cui siamo in grado di veicolare i nostri stati emotivi attraverso l'espressione linguistica, è un aspetto affascinante della lingua che ha lungamente riscosso l'interesse dei linguisti. Bally (1913) si è interrogato, ad esempio, sull'origine dell'emozione, valutando se essa sia intrinseca alla parola oppure se sia invece il contesto, la realtà che ci circonda, a esserne fattore determinante.

È scevro da dubbi che le lingue abbiano un proprio repertorio lessicale specifico per la descrizione degli stati emotivi⁵. Tuttavia, il rapporto tra parole ed emozioni non si esaurisce nella dimensione puramente descrittiva: le emozioni, infatti, possono trasparire dalla descrizione di reazioni somatiche, fisiologiche o dall'uso di linguaggio figurato; anche quando non vi è un riferimento diretto, o indiretto, è il contesto d'uso a giocare un ruolo determinante nel "colorare" le parole di connotazioni emotive (cfr. Partington 2004). La polarità e la carica emotiva associate a un'unità lessicale risultano soggette all'uso concreto, in una accezione affettiva dell'ipotesi distribuzionale (cfr. Lenci, Lebani & Passaro 2018).

Sulla base teorica dell'associazione tra parole e testo, il passo successivo ha visto l'utilizzo di questa associazione per inferire dalle parole l'emotività del testo. Questa prospettiva ha aperto la strada all'analisi computazionale emotiva di testi, campo dalle molte potenziali applicazioni pratiche. La ricerca linguistica computazionale si è concentrata, dunque, sull'identificazione della polarità e delle emozioni nei testi, portando alla proliferazione di numerose risorse linguistiche volte a registrare l'associazione tra emozione e parola, o la connotazione positiva o negativa della parola (ovvero la sua polarità). Tuttavia, nonostante la ricchezza di questo panorama, la frammentazione dei formati e dei criteri di annotazione ostacola spesso una reale sinergia tra gli strumenti disponibili per l'analisi emotiva dei testi.

Il presente studio nasce proprio dalla necessità di superare tale isolamento, proponendo l'adattamento di risorse esistenti in un'ottica di *Linked Open Data* (LOD) per renderle non solo interconnesse, ma anche semanticamente interoperabili in un ecosistema interconnesso (Berners-Lee, Hendler & Lassila 2001). Uno dei pilastri del paradigma LOD è, infatti, il Semantic Web, che promuove l'impiego di schemi di dati comuni, quali ontologie e vocabolari, e l'adozione di standard come RDF (*Resource Description Framework*) per supportare una rappresentazione coerente dei dati e l'integrazione di basi di conoscenza eterogenee all'interno di un'unica rete di

⁵ In letteratura *emotion label*, cfr. Johnson-laird & Oatley (1989).

significati condivisi.

Nel panorama italiano delle risorse affettive, abbiamo preso in considerazione ELIta (Di Palma 2024) e Sentix (Basile et al. 2025), con l'obiettivo di integrarle nella struttura della Lemma Bank di LiITA (Litta et al. 2024).

ELIta (*Emotion Lexicon for Italian*) è un lessico annotato tramite crowdsourcing che incorpora sia le emozioni di base di Plutchik, 1980 (Tristezza, Disgusto, Rabbia, Paura, Sorpresa, Fiducia e Aspettativa), sia le dimensioni VAD⁶. Il bacino principale da cui attinge è il “Nuovo Vocabolario di Base” (NVdB) (De Mauro 2016)⁷.

Sentix, invece, è un lessico affettivo per la lingua italiana nato nel 2013 generato automaticamente a partire dall'allineamento di diverse risorse lessicali (Basile & Nissim 2013). L'evoluzione della risorsa ha portato a Sentix 2.0, in cui i punteggi di polarità sono stati ricalcolati e aggregati sulla base delle frequenze del corpus SemCor (Langone, Haskell & Miller 2004). Il rilascio della risorsa è stato accompagnato da MAL (*Morphologically-inflected Affective Lexicon*), una versione del lessico che estende il punteggio di polarità alle forme flesse dei lemmi tramite Morph-it! (Vassallo et al. 2019; Zanchetta & Baroni 2005). Successivamente, gli autori hanno pubblicato WMAL (*Weighted-Morphologically-inflected Affective Lexicon*) i cui punteggi integrano le frequenze d'uso del corpus TWITA per mitigare lo sbilanciamento della polarità (Vassallo et al. 2020). L'aggiornamento a Sentix 3.0 (Di Palma et al. 2025) armonizza queste tre risorse (Sentix, MAL, WMAL) per garantirne la coerenza complessiva, offrendo un set espanso di 68.190 voci nella versione 3.1.1 con classificazioni categoriali in Positivo, Neutro e Negativo (Basile et al. 2025).

Ispirato da progetti come LiLa (*Linking Latin*) (Passarotti et al. 2020), il nostro lavoro mira a integrare e formalizzare due risorse lessicali affettive per l'italiano, rendendole interoperabili attraverso il collegamento alla Lemma Bank del progetto LiITA (*Linking ITALian*) (Litta et al. 2024).

Seguendo la metodologia stabilita per il lessico affettivo del latino all'interno del progetto LiLa, *LatinAffectus* (Sprugnoli et al. 2020), sono state impiegate per la resa LOD di ELIta e Sentix le ontologie Ontolex-Lemon (McCrae et al. 2017) per descrivere i dati lessicali, e MARL (Westerski, Iglesias & Rico 2011) per rappresentare le proprietà relative alla polarità. Ogni entrata lessicale è associata a un'etichetta, a una `ontolex:canonicalForm`. La proprietà

⁶ Ovvero, Valenza (l'attrattiva intrinseca o l'avversione per un evento, oggetto o situazione), Attivazione (il livello di attivazione fisiologica, dal sonno all'eccitazione) e Dominanza (il grado di controllo percepito su una situazione) (Russell & Mehrabian 1977).

⁷ Per ulteriori informazioni sulla costruzione del lessico cfr. Di Palma (2025).

`marl:hasPolarity` collega ogni entrata lessicale ad una categoria interna a `marl:Polarity` (ovvero positivo, negativo o neutro), mentre `marl:polarityValue` è stata utilizzata per descrivere i dati di Sentix 3.0, permettendo di assegnare valori continui tra -1 e 1.

Per l'annotazione categoriale delle emozioni di base è stata invece utilizzata una nuova ontologia, snellendo la struttura della precedente Onyx (Carlos et al. 2017) progettata per assegnare l'emotività ai testi, e adattandola ai lemmi: *elita ontology*.

Il nucleo dell'ontologia è costituito dalla classe `elita:Emotion`, che funge da categoria generale per tutte le emozioni di base di Plutchik, includendo anche la diade Amore. Le emozioni specifiche, come Gioia, sono modellate come individui (istanze) di questa classe. Per associare un elemento lessicale a un'emozione, l'ontologia definisce la proprietà dell'oggetto `elita:HasEmotion`. Questo `owl:ObjectProperty` collega un soggetto (la parola) a un'istanza di `elita:Emotion`, esprimendo così il contenuto emotivo attribuito a quell'elemento (il diagramma della classe è visibile in Figura 1).

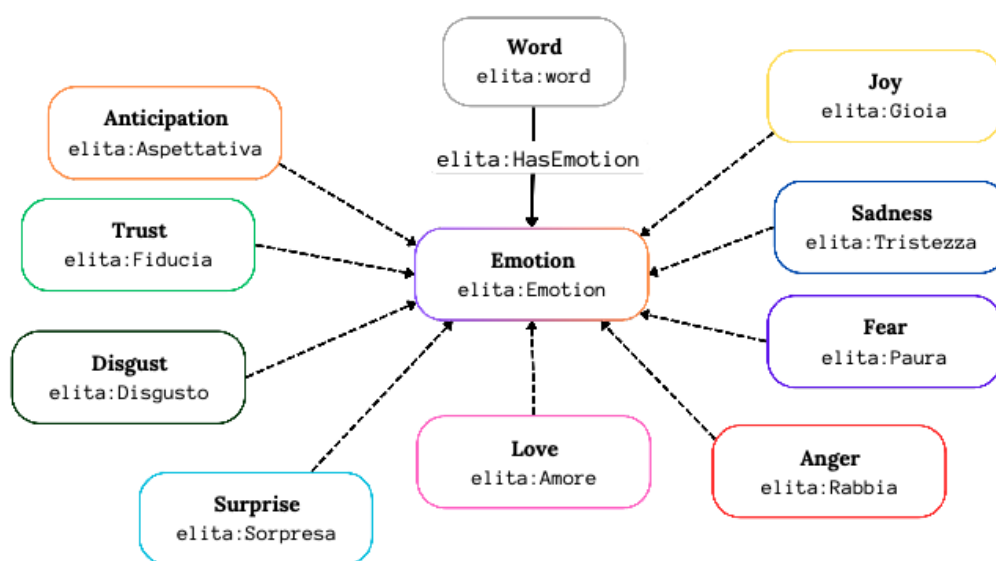


Figura 1. Diagramma di elita.

```
word:abisso rdfs:label "abisso";
elita:HasEmotion elita:Paura,
elita:Tristezza ;
marl:Polarity "Negative" ;
marl:hasPolarityValue "-0.5"^^xsd:float.
```

Una volta stabilite le ontologie di riferimento e riformulati i lessici in un formato RDF, è stato effettuato il collegamento tra l'URI univoco di ogni entrata lessicale e i lemmi della *Lemma Bank* di LiITA, arrivando a una copertura del 70% per ELIta e del 36% per Sentix (dato dovuto alla numerosa presenza di espressioni multiparola all'interno del lessico).

L'integrazione di risorse affettive come ELIta e Sentix 3.0 all'interno dell'ecosistema LiITA permetterà di eseguire interrogazioni trasversali tra i dati nativi di LiITA e le risorse lessicali e testuali interconnesse, aprendo la strada a nuove frontiere della ricerca linguistica sull'intreccio tra parola ed emozione, soprattutto in un ecosistema in cui la parola è alla base della conoscenza.

Riferimenti bibliografici

Bally, Charles. 1913. *Le langage et la vie*. Genève: Droz.

Basile, Valerio & Nissim, Malvina. 2013. Sentiment analysis on Italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 100–107. Atlanta, Georgia: Association for Computational Linguistics.

Basile, Valerio, Nissim, Malvina, Bosco, Cristina, Vassallo, Marco & Gabrieli, Giuliano. 2025. *Sentix (Versione 3.1.0) [Dataset]*.

Berners-Lee, Tim, Hendler, James & Lassila, Ora. 2001. The semantic web. *Scientific american* 284(5). 34–43.

De Mauro, Tullio. 2016. Il Nuovo vocabolario di base della lingua italiana. *Internazionale*, 23 dicembre 2016. <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>

Di Palma, Eliana. 2024. ELIta: a new Italian language resource for emotion analysis. In Dell'Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the 10th italian conference on computational linguistics (clic-it 2024)*, 297–307. Pisa, Italy. CEUR Workshop Proceedings.

Di Palma, Eliana. 2025. *Un lessico emotivo per l'italiano: dalla teoria linguistica alla pratica computazionale*. Roma: Sapienza, Università di Roma. (Tesi di dottorato.)

Di Palma, Eliana, Basile, Valerio, Vardanega, Agnese, Gabrieli, Giuliano & Vassallo, Marco. 2025. Linking emotions: affective and lexical resources for Italian in linked open data. In Bosco,

- Cristina, Jezek, Elisabetta, Polignano, Marco & Sanguinetti, Manuela (eds.), *Proceedings of the eleventh italian conference on computational linguistics (clit-it 2025)*, 384–396. Cagliari, Italy. CEUR Workshop Proceedings.
- Iglesias, Carlos, A., Sánchez-Rada, José Fernando, Vulcu, Gabriela & Buitelaar, Paul. 2017. Linked data models for sentiment and emotion analysis in social networks. In Pozzi, Federico Alberto, Fersini, Elisabetta, Messina, Enza & Liu, Bing (eds.), *Sentiment analysis in social networks*, 49–69. Boston: Morgan Kaufmann.
- Johnson-laird, P. N. & Oatley, Keith. 1989. The language of emotions: an analysis of a semantic field. *Cognition and Emotion* 3(2). 81–123. <https://doi.org/10.1080/02699938908408075>
- Langone, Helen, Haskell, Benjamin R. & Miller, George A. 2004. Annotating wordnet. In *Proceedings of the workshop on frontiers in corpus annotation*, 63–69. Boston: ACL.
- Lenci, Alessandro, Lebani, Gianluca E. & Passaro, Lucia C. 2018. The emotions of abstract words: a distributional semantic analysis. *Topics in Cognitive Science* 10(3). 550–572. <https://doi.org/10.1111/tops.12335>
- Litta, Eleonora, Passarotti, Marco, Brasolin, Paolo, Moretti, Giovanni, Basile, Valerio, Di Fabio, Andrea & Bosco, Cristina. 2024. The lemma bank of the LiITA knowledge base of interoperable resources for Italian. In Dell’Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the 10th italian conference on computational linguistics (CliC-it 2024)*, 517–522. Pisa, Italy. CEUR Workshop Proceedings.
- McCrae, John P, Bosque-Gil, Julia, Gracia, Jorge, Buitelaar, Paul & Cimiano, Philipp. 2017. The ontolex-lemon model: development and applications. In *Proceedings of elex 2017 conference*, 19–21.
- Partington, Alan. 2004. “Utterly content in each other’s company”: semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9(1). 131–156. <https://doi.org/10.1075/ijcl.9.1.07par>
- Passarotti, Marco, Mambrini, Francesco, Franzini, Greta, Cecchini, Flavio Massimiliano, Litta, Eleonora, Moretti, Giovanni, Ruffolo, Paolo & Sprugnoli, Rachele. 2020. Interlinking through lemmas. the lexical collection of the lila knowledge base of linguistic resources for latin. *Studi e Saggi Linguistici* 58(1). 177–212. <https://doi.org/10.4454/ssl.v58i1.277>
- Plutchik, Robert. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, 3–33. Elsevier.

- Russell, James A & Mehrabian, Albert. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11(3). 273–294.
- Sprugnoli, Rachele, Mambrini, Francesco, Moretti, Giovanni & Passarotti, Marco. 2020. Towards the modeling of polarity in a Latin Knowledge Base. In *Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020) Co-located with 15th Extended Semantic Web Conference (ESWC 2020)*, 59–70. Heraklion, Greece. CEUR Workshop Proceedings. <https://doi.org/10.5281/zenodo.4067813>
- Vassallo, Marco, Gabrieli, Giuliano, Basile, Valerio & Bosco, Cristina. 2019. The tenuousness of lemmatization in lexicon-based sentiment analysis. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, 520–525 Bari, Italy. CEUR Workshop Proceedings.
- Vassallo, Marco, Gabrieli, Giuliano, Basile, Valerio & Bosco, Cristina. 2020. Polarity Imbalance in Lexicon-based Sentiment Analysis. In Dell’Orletta, Felice, Monti, Johanna & Tamburini, Fabio (a cura di), *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020 : Bologna, Italy, March 1-3, 2021* (Collana Dell’Associazione Italiana Di Linguistica Computazionale), 457–463. Torino: Accademia University Press.
- Westerski, Adam, Iglesias, Carlos Angel & Rico, Fernando Tapia. 2011. Linked opinions: describing sentiments on the structured web of data. In *Proceedings of the 4th International Workshop Social Data on the Web*, 10-21. Bonn, Germany. CEUR Workshop Proceedings.
- Zanchetta, Eros & Baroni, Marco. 2005. Morph-it! a free corpus-based morphological resource for the italian language. In *Proceedings of corpus linguistics conference series*, vol. 1, 1–12. Birmingham: University of Birmingham.

Arricchimento strutturale e semantico delle pronunce della Corte Costituzionale e ripubblicazione come risorsa interoperabile

Introduzione

La disponibilità di testi giuridici in formato digitale apre opportunità rilevanti per la linguistica computazionale a condizione che tali testi siano pubblicati in forme che ne consentano l'interrogazione e il riuso in contesti eterogenei. Le pronunce della Corte Costituzionale italiana rappresentano in questo senso un patrimonio testuale di eccezionale valore: un corpus di oltre 21.000 documenti distribuiti in un arco temporale di quasi settant'anni.

La Corte Costituzionale ha avviato negli ultimi anni un'iniziativa *Open Data* che mette a disposizione l'intero archivio periodicamente aggiornato delle pronunce e delle massime in formati processabili (XML, JSON, CSV), con licenza *Creative Commons* (cfr. Corte Costituzionale 2024). Nonostante questa apprezzabile apertura, l'analisi del formato XML originale rivela limiti significativi sul piano dell'annotazione: la gerarchia concettuale del documento non è rispecchiata dalla struttura dei tag, le sezioni principali (fatto, diritto, dispositivo) sono marcate in modo incompleto o del tutto assente in molti casi, i paragrafi numerati non sono identificati, le liste non sono strutturate, e i riferimenti normativi e giurisprudenziali non sono annotati né normalizzati.

Il progetto *Marker-Cost*, sviluppato presso il CNR-IGSG, si propone di colmare queste lacune attraverso un arricchimento automatico delle annotazioni strutturali e semantiche e la ripubblicazione del corpus in un formato XML più ricco, auto-esplicativo e orientato all'interoperabilità.

Il corpus e la sua struttura testuale

Le pronunce (sentenze e ordinanze) della Corte Costituzionale presentano una macrostruttura testuale stabile e ben definita, articolata in tre parti principali: una *testa*, contenente la presentazione del collegio giudicante; un *corpo*, articolato nelle sezioni *epigrafe*, *fatto*, *diritto* e *dispositivo*; e una conclusione con date e firme. All'interno delle sezioni fatto e diritto i paragrafi possono presentare una numerazione mentre nell'epigrafe e nel dispositivo possono comparire liste letterali e numeriche.

Il corpus copre fenomeni di grande interesse per la linguistica italiana: la sintassi del periodo lungo e ricco di subordinate, l'uso tecnico-specialistico del lessico giuridico, le formule cristallizzate che introducono le sezioni ("Ritenuto in fatto", "Considerato in diritto"), la citazione di fonti

normative e giurisprudenziali come pratica testuale ricorrente, il testo tra virgolette come marcatore di discorso riportato. Proprio su questi fenomeni si concentra parte del lavoro di annotazione semantica sviluppato nel progetto.

Annotazione strutturale e semantica

Il formato di annotazione XML prodotto da *Marker-Cost* (cfr. Bacci 2025) rispecchia fedelmente la struttura gerarchica delle pronunce e comprende l'identificazione esplicita delle sezioni epigrafe, fatto, diritto e dispositivo come elementi contenitore, l'annotazione dei paragrafi numerati all'interno delle sezioni fatto e diritto, con assegnazione di identificatori univoci che codificano sia la sezione di appartenenza sia la posizione nel testo nonché la demarcazione di liste numeriche e letterali e infine la strutturazione di eventuali ordinanze allegate alle pronunce.

In aggiunta all'analisi strutturale sono inoltre introdotte annotazioni relative a:

- *riferimenti normativi e giurisprudenziali*: ogni citazione a leggi, decreti, sentenze e altri atti giuridici viene identificata, annotata con un tag dedicato e normalizzata in forma strutturata (cfr. Bacci et al. 2019), con attributi che codificano il tipo di documento, l'autorità emanante, il numero, la data e la partizione citata (articolo, comma, etc.);
- *esiti delle sentenze*: le formule testuali del dispositivo che esprimono il giudizio finale della Corte (dichiarazioni di illegittimità, inammissibilità, infondatezza, restituzione degli atti, etc.) vengono riconosciute e annotate. L'annotazione degli esiti ha già consentito di produrre statistiche su larga scala sull'intero corpus storico: si contano ad esempio oltre 6.700 esiti di inammissibilità, oltre 4.900 di illegittimità e oltre 4.400 di non fondatezza nell'intero dataset dal 1956 al 2023;
- *testo tra virgolette*: le porzioni di testo riportate tra virgolette nelle sentenze costituiscono un fenomeno linguisticamente rilevante, in quanto segnalano quasi sempre discorso riportato da atti normativi, sentenze precedenti o dichiarazioni delle parti. L'annotazione di questi elementi, robusta rispetto alle varianti tipografiche delle virgolette presenti nel corpus, consente analisi sistematiche delle pratiche di citazione testuale nel genere giuridico.

Il corpus arricchito comprende infine le massime giurisprudenziali associate a ciascuna pronuncia, anch'esse strutturate e annotate con riferimenti normativi e giurisprudenziali normalizzati, e collegate alle pronunce di riferimento.

Verso l'interoperabilità

Il formato arricchito del corpus delle pronunce della Corte Costituzionale, oltre a facilitare la conversione da XML a HTML con incorporazione di metadati semantici nel markup HTML attraverso RDFa, costituisce una solida base per una piena integrazione nell'ecosistema dei *Linked Open Data*.

I metadati bibliografici e semantici delle pronunce (tipo di pronuncia, data, presidente, relatore, esito, parametri costituzionali invocati) verranno arricchiti, linkati ai testi annotati ed esposti come triple RDF, consentendo il collegamento con altre risorse del web semantico giuridico, in particolare con i dataset di Normattiva.it e con le ontologie del dominio giuridico come ELI (*European Legislation Identifier*), ECLI (*European Case Law Identifier*) e Akoma Ntoso e sarà infine disponibile uno SPARQL *Endpoint* per l'interrogazione diretta del corpus (cfr. CIRSFD 2024; Council of the European Union 2011, 2017; Palmirani et al. 2011).

Un ulteriore sviluppo riguarda l'utilizzo del corpus arricchito in applicazioni di *Information Retrieval* e architetture RAG (*Retrieval-Augmented Generation*): la granularità dell'annotazione a livello di sezione e paragrafo consente di passare dal *document retrieval* a un *passage retrieval* semanticamente motivato, con possibilità di pesare differientemente le informazioni provenienti da sezioni distinte (es. dare maggiore peso alle citazioni nella sezione diritto rispetto alla sezione fatto).

Conclusioni

L'arricchimento strutturale e semantico delle pronunce della Corte Costituzionale fornisce un contributo sia sul fronte dell'interoperabilità che su quello della ricerca linguistica sul corpus giuridico italiano. Da un lato, trasforma un dataset già disponibile in *Open Data* ma strutturalmente povero in una risorsa testuale annotata, interoperabile e integrata nell'ecosistema LOD. Dall'altro, mette a disposizione della comunità dei linguisti un corpus di italiano giuridico di dimensioni ampie, con annotazione strutturale e semantica che consente interrogazioni a grana fine e studi sistematici sulla struttura retorica e argomentativa delle sentenze e su altri fenomeni linguistici di rilievo (cfr. Mortara Garavelli 2001).

Riferimenti bibliografici

Bacci, Lorenzo. 2025. Marker-Cost: *Arricchimento strutturale e semantico delle pronunce della Corte Costituzionale e ripubblicazione in formato XML*. Rapporto tecnico CNR-IGSG n. 2/2025.

- Bacci, Lorenzo, Agnoloni, Tommaso, Marchetti, Carlo & Battistoni, Roberto. 2019. Improving public access to legislation through legal citations detection: The Lincoln project at the Italian Senate. In xxx (a cura di), *Knowledge of the Law in the Big Data Age, Frontiers in Artificial Intelligence and Applications*, vol. 317, 149–158. Amsterdam: IOS Press.
- CIRSFID - Università di Bologna. 2024. *Pronunce della Corte Costituzionale*. <http://bach.cirsfid.unibo.it/ldms-cortecostituzionale>
- Corte Costituzionale. 2024. *Open Data delle pronunce e delle massime*. <https://dati.cortecostituzionale.it>
- Council of the European Union. 2011. Council conclusions inviting the introduction of the European Case Law Identifier (ECLI) and a minimum set of uniform metadata for case law . *Official Journal of the European Union*, C 127 , 29/04/2011, 1–7, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:127:0001:0007:EN:PDF>
- Council of the European Union. 2017. Council conclusions on the European Legislation Identifier (2017/C 441/05). *Official Journal of the European Union*. C 441/8, 22.12.2017, 1–5.
- [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01))
- Mortara Garavelli, Bice. 2001. *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*. Torino: Einaudi.
- Palmirani, Monica & Vitali, Fabio. 2011. Akoma-Ntoso for legal documents. In Sartor, Giovanni, Palmirani, Monica, Francesconi, Elnrico & Biasiotti, Maria Angela (a cura di), *Legislative XML for the Semantic Web* 75–100. Dordrecht: Springer.

Towards an Interoperable Lexical Infrastructure for Italian: Linking and Enriching CompL-it

CompL-it is an open computational lexicon for contemporary Italian, formally modelled according to the OntoLex-Lemon standard and released as Linguistic Linked Open Data (Sciolette et al., 2024). This contribution discusses the evolution of CompL-it from a standalone lexical resource into a hub within an interoperable ecosystem of Italian linguistic data.

The guiding idea behind this evolution is that a computational lexicon, once formalised in RDF and aligned with community standards, can function as a pivot for cross-resource linking, domain specialisation, diachronic integration and controlled automatic enrichment. We argue that interoperability is not merely a technical property but a methodological strategy for building sustainable lexical infrastructures.

CompL-it as a LLOD Hub

CompL-it currently includes 101,795 lexical entries and 791,541 word forms, associated with 55,713 lexical senses and 86,577 instances of 137 types of semantic relations. These figures position the resource among the largest freely available Italian lexical datasets structured according to Linked Open Data principles.

Unlike many lexica that encode primarily semantic relations (e.g. WordNet-style resources), CompL-it integrates morphology and semantics within a unified OntoLex-Lemon framework. This dual-layer richness makes it particularly suitable as a “pivot resource”: it can connect corpus-based annotations, knowledge bases and terminological datasets at the lemma, form and sense levels. The interoperability strategy developed in the current phase rests on two main axes: i) linking to general lexical knowledge bases; ii) linking to specialised and domain-specific terminologies.

Linking CompL-it to LiITA

A first major step in this strategy is the linking of CompL-it to the LiITA resource, as described in Litta et al. (2025). LiITA constitutes a large-scale interoperable knowledge base for Italian, centred on a Lemma Bank aligned with other Linguistic Linked Open Data resources (Litta et al., 2024).

The alignment between CompL-it and LiITA required reconciling different part-of-speech tagsets (LexInfo in CompL-it and UPOS in LiITA) and performing systematic matching on lemma–PoS pairs. The process resulted in three main scenarios (1:1, 1:N, and 1:0 matches), requiring specific handling strategies. In cases of multiple potential matches (1:N), a pragmatic disambiguation criterion based on lemma identifiers in the Lemma Bank was adopted (ibid.). The linking is implemented at the LexicalEntry level through the `ontolex:canonicalForm` relation, which connects CompL-it lexical entries to corresponding lemmas in the LiITA Lemma Bank.

The result is that CompL-it entries can now be embedded within a broader network of interlinked lexical resources. This integration increases query expressivity and enables federated SPARQL queries across datasets, strengthening the Italian LLOD ecosystem.

3. Linking General Lexicon and Specialised Terminology: the ITALMUD Case

A particularly significant demonstration of CompL-it’s pivot role is its integration with ITALMUD, a multilingual computational terminology, still under construction, derived from the Italian translation of the Babylonian Talmud.⁸ ITALMUD is being built from glossaries curated over more than ten years within the Babylonian Talmud Translation Project⁹ and subsequently formalised according to OntoLex-Lemon. Currently, the resource comprises 1,613 Talmudic terms, including 823 Italian, 773 Hebrew and 17 Aramaic entries.

Entries are classified as “Word” or “Multiword Expression” and enriched with definitions, PoS information and translation links. However, the glossary-derived terminology initially lacked an internal semantic network dense enough to support advanced querying. To address this, Italian entries in ITALMUD were anchored to CompL-it, which provides an already structured morpho-semantic superstructure.

The linking is implemented via a dedicated linkset, acting as a bridge resource, and involves three main types of relations:

1. *Sense-to-sense linking (same lemma)*

ITALMUD senses are connected to corresponding CompL-it senses using SKOS properties `skos:exactMatch` and `skos:closeMatch`). The choice between these two properties is made on a case-by-case basis, according to the degree of conceptual overlap between the ITALMUD

⁸ The Babylonian Talmud is one of the foundational texts of the Jewish tradition. Written mainly in Jewish Babylonian Aramaic, with Hebrew elements, it consists of a complex interweaving of legal, narrative, exegetical and sapiential discussions, characterized by strong terminological density, polysemy and contextual variability.

⁹ <https://www.talmud.it>.

and CompL-it senses and to the intended use of the linkset. `skos:exactMatch` is adopted when the two senses can be treated as referring to the same concept within the scope of the resource, whereas `skos:closeMatch` is used when the Talmudic sense introduces domain-specific nuances or contextual restrictions.

2. *Sense-to-sense linking (different lemmas)*

In cases involving transliterations or semantically related but formally distinct terms, linking uses SKOS or CompL-it semantic relations (e.g., hyponymy, `madeOf`, `objectOfActivity`).

3. *Sense-to-entry linking (new specialised sense)*

When ITALMUD introduces a specialised meaning absent in CompL-it, the Talmudic sense is linked to the CompL-it lexical entry via `ontolex:isSenseOf`, effectively extending the general lexicon with a domain-specific sense.

Currently, 227 ITALMUD terms are linked to CompL-it: 158 via SKOS sense-level relations and 69 via `ontolex:isSenseOf`.

This architecture transforms a specialised terminology into a semantically navigable network by inheriting relational pathways from the general lexicon. Through federated SPARQL queries, it becomes possible to retrieve Talmudic terms via semantic chains (e.g., hyponymy hierarchies or usage relations), illustrating how interoperability enhances both linguistic and domain-specific research.

Controlled Enrichment through Large Language Models

Recent work has focused on extending CompL-it’s semantic layer by integrating structured information inherited from PSC (Parole-Simple-Clips), namely semantic traits and semantic types (Giovannetti et al., 2026). A total of 2,702 trait–sense relations were extracted from the original PSC’s relational database and converted into RDF triples through the introduction of a dedicated property (`complit:trait`) linking `ontolex:LexicalSense` instances to formally defined trait classes. In parallel, 254 semantic types were modelled as an external OWL class hierarchy and connected to lexical senses via the property `hasSemanticType`, resulting in 54,487 sense–type associations.

In addition, an experiment was conducted to automatically generate missing definitions using small to medium-sized Large Language Models. The test focused on 235 senses, corresponding to 229 lexical entries, selected among the 9,158 senses lacking definitions but still provided with examples, semantic relations and semantic types. Three multilingual models from the Gemma 3, Phi-4 and Mistral-Small families generated one Italian definition for each sense, for a total of 705

definitions. These outputs were evaluated through an LLM-as-a-Judge protocol, using three independent judge models and a 0–10 scoring scale. The evaluation showed generally usable results, with Phi selected most often as the best-performing generator and examples emerging as the most useful contextual feature for definition quality.

This hybrid enrichment strategy combines curated semantic structure with scalable machine-supported completion, strengthening both interpretability and interoperability within the LLOD framework.

From Resource to Infrastructure

CompL-it is progressively evolving from a lexical dataset into an interoperable infrastructure for Italian, within a broader strategy of cross-resource linking and AI-supported enrichment. Integrations such as those with LiITA and ITALMUD exemplify this approach, positioning the resource as a morpho-semantic backbone for specialised terminologies and as a node within the Italian Linguistic Linked Open Data cloud. Interoperability is therefore not an accessory feature but the guiding principle of its development: by adhering to OntoLex-Lemon and Linked Open Data standards, CompL-it supports the structured alignment of heterogeneous resources while enabling their semantic integration at scale. The long-term objective is to consolidate a coherent lexical infrastructure in which general language, specialised terminology and computational modelling are formally connected within a shared, extensible framework.

References

- Giovannetti, Emiliano, Bellandi, Andrea, Marchi, Simone & Papini, Mafalda. 2026. Extending the Semantic Layer of the CompL-it Italian Lexicon: Traits, Semantic Types, and Definitions. In Piperidis, Stelios; Bel, Núria; van den Heuvel, Henk; Ide, Nancy; Krek, Simon; Toral, Antonio (a cura di), *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), Palma, Mallorca, 11-16 May 2026*, 7857–7866. Paris: European Language Resources Association (ELRA). <https://dx.doi.org/10.63317/3rvf2vbt4ier>
- Litta Modignani Picozzi, Eleonora Maria Gabriella, Passarotti, Marco Carlo, Brasolin, Paolo, Moretti, Giovanni, Mambrini, Francesco, Basile, Valerio, Di Fabio, Andrea & Bosco, Cristina. 2024. The lemma bank of the LiITA knowledge base of interoperable resources for Italian. In Dell’Orletta, Felice, Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (eds.),

Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024), Pisa, 4-6 December 2024, 517–522. Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2024.clicit-1.61/>

Litta Modignani Picozzi, Eleonora Maria Gabriella, Passarotti, Marco Carlo, Moretti, Giovanni, Brasolin, Paolo, Mambrini, Francesco, Basile, Valerio, Di Fabio, Andrea, Di Palma, Eliana, Giovannetti, Emiliano, Marchi, Simone, Bellandi, Andrea & Sciolette, Flavia. 2025. Linking CompL-it to the LiITA Knowledge Base. In Bosco, Cristina, Jezek, Elisabetta, Polignano, Marco & Sanguinetti, Manuela (eds.), *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025), Cagliari, 24-26 September 2025*, 595–602. Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2025.clicit-1.57/>

Sciolette, Flavia, Giovannetti, Emiliano, Marchi, Simone & Bellandi, Andrea. 2024. CompL-it: A computational lexicon of Italian language. *AIDA Informazioni* 42(3–4). 119–148. <https://doi.org/10.57574/596545646>

Workshop 3

Romaní e Romancisation: esiti del contatto tra i dialetti dei rom e dei sinti italiani e le varietà italo-romanze

Proponenti: **RICCARDO REGIS E ANDREA SCALA**

NATALIYA LEVKOVYCH

Dalla romanizzazione all'italianizzazione: prestito grammaticale nelle lingue non romanze d'Italia

Le lingue romanze costituiscono, direttamente o indirettamente, una fonte rilevante di diffusione di tratti indotti dal contatto linguistico in centinaia di lingue del mondo, come evidenziato da Stolz (2008) nel suo contributo programmatico sulla romanizzazione linguistica (*Romancisation*). Tale processo riguarda non soltanto il lessico, inclusi prestiti lessicali, neologismi e internazionalismi, ma anche la diffusione di strutture grammaticali, marcatori discorsivi, schemi sintattici e categorie funzionali in lingue geneticamente non imparentate e tipologicamente eterogenee. In questo quadro, Gardani (2022) sintetizza l'influenza delle lingue romanze ai diversi livelli della grammatica – dalla fonologia e prosodia fino alla morfologia e alla sintassi – sottolineando al contempo le implicazioni teoriche di tali fenomeni per i modelli della linguistica del contatto.

All'interno di questo più ampio quadro di diffusione di tratti romanzi, un caso particolarmente rilevante è rappresentato dall'italianizzazione, ossia l'influsso dell'italiano e delle varietà italo-romanze sulle lingue in contatto con esse. Questo fenomeno si manifesta in particolare nelle lingue storicamente parlate sul territorio italiano. Il contatto prolungato e intenso con l'italiano ha favorito processi di convergenza che coinvolgono anche elementi grammaticali e discorsivi, quali i connettivi. Studi precedenti hanno mostrato che il prestito di parole funzionali, e in particolare delle congiunzioni dall'italiano come lingua donatrice, è ampiamente attestato nelle lingue alloglotte d'Italia e nel maltese (Stolz 2005, 2007; Levkovich 2025). Tuttavia, tali studi non prendono in esame questo fenomeno nelle varietà romaní parlate sul territorio italiano.

D'altra parte, è noto che le lingue romaní sono particolarmente suscettibili al prestito di congiunzioni, come mostra Matras (1998, 2002). Si osservano, ad esempio, prestiti di congiunzioni tedesche nelle varietà lovari, manush e sinti; di congiunzioni provenienti da diverse lingue slave nelle varietà ajia varvara, gurbet, bugurdži, lovari e polska roma; nonché di congiunzioni greche nelle varietà ajia varvara e bugurdži. Il prestito di congiunzioni da una lingua romanza, il francese, è

attestato inoltre nel lovari e nel manush, come nel caso dell'avversativo *mais* 'ma'. Forlano (2025) documenta infine il prestito di congiunzioni italiane nel sinti lombardo. Le congiunzioni prese in prestito nelle lingue romaní non sembrano colmare lacune funzionali, ma tendono piuttosto a sostituire le congiunzioni di origine indoaria.

In questa prospettiva, il prestito delle congiunzioni può essere interpretato come uno degli elementi del più ampio processo di romanizzazione. Il presente contributo esamina tale fenomeno nelle varietà romaní in contatto con l'italiano, ponendo a confronto con quanto osservato nelle lingue alloglotte non romanze d'Italia. Verranno inoltre discussi casi paralleli di romanizzazione, quali il prestito di congiunzioni dal francese o dal rumeno. Infine, il contributo intende verificare la validità dei modelli teorici relativi alle gerarchie di prestito delle congiunzioni proposti da Matras (2007).

Riferimenti bibliografici

- Forlano, Marco. 2025. *I sinti lombardi di Pavia. Un'indagine sociolinguistica*. Pavia: Università degli Studi di Pavia (Tesi di dottorato.) <https://hdl.handle.net/10446/299749>.
- Gardani, Francesco. 2022. Contact and borrowing. In Ledgeway, Adam & Maiden, Martin (a cura di), *The Cambridge handbook of Romance linguistics*, 845–869. Cambridge: Cambridge University Press.
- Levkovich, Nataliya. 2025. Italian loan conjunctions in the alloglottic non-Romance languages of Italy. *STUF - Language Typology and Universals* 78(2). 265–319. <https://doi.org/10.1515/stuf-2025-2009>
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2). 281–331.
- Matras, Yaron. 2002. *Romani. A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2007. The borrowability of structural categories. In Matras, Yaron & Sakel, Jeanette (a cura di), *Grammatical borrowing in cross-linguistic perspective*, 31–74. Berlin, New York: Mouton de Gruyter.
- Stolz, Thomas. 2005. Italianisierung in den alloglotten Sprachen Italiens. In Bidese, Ermenegildo Dow, James R. & Stolz, Thomas (a cura di), *Das Zimbrische zwischen Germanisch und Romanisch*, 43–68. Bochum: Universitätsverlag Brockmeyer.

- Stolz, Thomas. 2007. *Allora*: On the recurrence of function-word borrowing in contact situations with Italian as donor language. In Rehbein, Jochen, Hohenstein, Christiane & Pietsch, Lukas (a cura di), *Connectivity in grammar and discourse*, 75–100. Amsterdam: John Benjamins.
- Stolz, Thomas. 2008. Romancisation worldwide. In Stolz, Thomas, Bakker, Dik & Salas Palomo, Rosa (a cura di), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes*, 1–42. Berlin, Boston: De Gruyter.

Finola o finarela? L'integrazione variabile dei prestiti verbali italoromanzi in sinto lombardo

Le varietà di romaní si caratterizzano per un'elevata incidenza di prestiti lessicali da lingue di contatto passate o correnti, anche in ottica interlinguistica (Adamou 2016; Elšik 2009). I prestiti lessicali in romaní coinvolgono tanto parole funzione quanto parole contenuto, come nomi e verbi, i quali tendono ad andare incontro a svariati processi di integrazione strutturale (cfr. Elšik 2020, 168–170).

Il presente contributo analizza i processi di integrazione morfologica dei prestiti verbali italoromanzi in sinto lombardo, una varietà di romaní parlata in Lombardia. Recentemente, Meli (2023) ha mostrato che i parlanti di sinto lombardo ricorrono a due strategie di integrazione, entrambe riconducibili a meccanismi di *indirect insertion* (nella tipologia di Wohlgemuth 2009): la prima prevede l'integrazione della radice lessicale italoromanza nei paradigmi flessivi romaní attraverso il verbalizzatore di origine indoaria *-ar-* (p.es. *perdun-ar-ava* 'perdono'); la seconda prevede l'impiego del verbalizzatore di origine indoaria *-o-*, derivato da un originario **-ov-* (p.es. *cuminci-o-la* 'comincia' < **cuminci-ov-ela*). Sebbene l'alternanza tra i due morfemi in romaní sia tradizionalmente ricondotta alla valenza del verbo, con *-ar-* associato ai verbi transitivi e *-o-* agli intransitivi (Matras 2002, 133), Meli (2023) nota che in sinto lombardo *-ar-* si sta espandendo anche ai verbi intransitivi a scapito di *-o-* (p.es. *tusarava* 'tossisco').

Muovendo da queste osservazioni, il presente contributo mira a indagare in maniera sistematica i fattori che regolano l'integrazione variabile dei prestiti verbali italoromanzi in sinto lombardo. L'analisi si basa su un corpus di parlato (semi)spontaneo di circa 50.000 parole raccolto nella città di Pavia da un campione sociolinguisticamente stratificato di locutori. Ai fini dell'analisi, vengono estratti dal corpus tutti i verbi italoromanzi integrati nella morfologia romaní, per un totale (allo stato attuale) di 528 occorrenze.

L'analisi procede in due fasi. In una prima fase, viene analizzata la distribuzione di *-ar-* e *-o-* in relazione alla valenza del verbo, distinguendo tra transitivi, intransitivi e riflessivi/reciproci (cfr. Ćirković & Mirić 2022). I risultati mostrano che i verbi transitivi e riflessivi/reciproci vengono integrati stabilmente attraverso il suffisso *-ar-*, mentre gli intransitivi esibiscono variazione tra *-ar-* e *-o-*, selezionati in proporzioni simili.

In una seconda fase, l'analisi si concentra esclusivamente sui verbi intransitivi. L'obiettivo è valutare la relazione tra selezione del verbalizzatore e diversi predittori linguistici e sociolinguistici, considerati sia individualmente (tramite test del chi-quadrato), sia congiuntamente (mediante un

modello di regressione logistica a effetti misti e un *Conditional Inference Tree*). Si considerano in particolare: (a) il grado di agentività del soggetto; (b) la persona del verbo; (c) il tempo del verbo; (d) la frequenza del verbo; (e) la fascia d'età del parlante. Nella regressione logistica, il singolo parlante e il lemma sono inclusi come effetti casuali. L'agentività del soggetto è operazionalizzata come variabile binaria, distinguendo tra soggetti con proprietà agentive e soggetti con proprietà di tema-paziente (Dowty 1991). Caratterizzazioni semantiche dei verbi intransitivi secondo tassonomie a grana più fine (Sorace 2000) sono discusse separatamente in un'analisi esplorativa.

Tra i risultati preliminari, l'analisi univariata evidenzia un'associazione significativa tra selezione del verbalizzatore e almeno due predittori: l'agentività del soggetto ($\chi^2(1) = 56.61$, $p < .001$) e la persona del verbo ($\chi^2(1) = 57.65$, $p < .001$; 1^a e 2^a persona raggruppate), sebbene non sia da escludere un'interrelazione tra i due fattori, potenzialmente mediata dall'agentività del soggetto. In particolare, i verbi con soggetto agentivo e quelli alla 1^a/2^a persona tendono a selezionare *-ar-*, mentre i verbi con soggetto non agentivo e quelli alla 3^a persona favoriscono *-o-*. La variazione nelle strategie di integrazione in base all'agentività del soggetto è anche evidente a livello di uno stesso lemma, come negli esempi di seguito:

- (1) na mancaram mai da rispeto ke mengar genitori
 'non abbiamo mai mancato di rispetto ai nostri genitori'
- (2) u tut na mancolas miga parché is u garamja
- (3) 'il latte non mancava mai perché c'erano le mucche'

Complessivamente, questo studio mostra che le strategie di integrazione dei prestiti verbali italo-romanzi in sinto lombardo sono condizionate da una pluralità di fattori che prescindono, almeno in parte, dalla tradizionale distinzione tra verbi transitivi e intransitivi, come il grado di agentività del soggetto (ma cfr. Ćirković & Mirić 2022 per osservazioni simili su una varietà gurbet in Serbia). Allo stesso tempo, il fatto che, in sinto lombardo, *-ar-* sia anche attestato con verbi intransitivi non agentivi sembra indicarne una progressiva affermazione come verbalizzatore di *default* (cfr. anche Meli 2023).

Riferimenti bibliografici

Adamou, Evangelia. 2016. *A Corpus-Driven Approach to Language Contact: Endangered Languages in a Comparative Perspective*. Berlino: Mouton de Gruyter.

- Ćirković, Svetlana & Mirić, Mirjana. 2022. Serbian Loanverb Adaptation in the Gurbet Romani Variety in Eastern Serbia. *Mediterranean Language Review* 29. 89–127.
- Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67. 547–619. <https://doi.org/10.2307/415037>
- Elšik, Viktor. 2009. *Loanwords in Selice Romani, an Indo-Aryan language of Slovakia*. In Haspelmath, Martin & Tadmor, Uri (a cura di), *Loanwords in the World's Languages: A Comparative Handbook*, 260–303. Berlino: Mouton de Gruyter.
- Elšik, Viktor. 2020. *Romani Morphology*. In Matras, Yaron & Tenser, Anton (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*, 155–186. Londra: Palgrave Macmillan.
- Matras, Yaron. 2002. *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Meli, Giulia, 2023. *L'integrazione morfologica dei prestiti romanzi in sinto piemontese di Francia e in sinto lombardo*. In Faraoni, Vincenzo, Filipponio, Lorenzo, Paciaroni, Tania & Schmid, Stephan (a cura di), *Prospettive di ricerca in linguistica italiana e romanza: studi offerti a Michele Loporcaro dagli allievi e dai collaboratori zurighesi*, 407–424. Pisa: ETS.
- Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language*, 76(4). 859–890. <https://doi.org/10.2307/417202>
- Wohlgemuth, Jan. 2009. *A Typology of Verbal Borrowings*. Berlino: Mouton de Gruyter.

Il vocalismo atono in alcune varietà di romaní d'Italia: fenomeni di convergenza verso l'italo-romanzo

Una delle manifestazioni più significative del contatto tra romaní e dialetti italo-romanzi nello spazio linguistico italiano è senz'altro la convergenza fonologica della prima verso i secondi. Un esame della fonologia delle diverse varietà di romaní parlate nella Penisola restituisce infatti un quadro di forte innovatività, soprattutto a livello di fonologia sintagmatica, più che paradigmatica; quindi più a livello di regole fonologiche che di unità fonologiche. In altre parole se sembrano scarsi e complessivamente marginali i casi di induzione di fonema nei dialetti romaní d'Italia, assai meglio testimoniati sono i casi di calco di regole fonologiche. Per fare un esempio tra i meglio studiati, la romaní d'Abruzzo (Soravia 1977, 84–91, 2019) presenta la propagginazione, una regola molto rara, che finora è stata trovata solo nei dialetti dell'Italia meridionale e in lingue nigerio-kordofaniane come il birom (Tuttle 1985; Schirru 2008, 2013). Tale regola si ritrova a carico del lessico indoario della romaní d'Abruzzo, cfr. [u (k)'kwaf̥tə] 'il legno' ma [ni (k)'kaftə] 'un ramo, un legno', [u (g)'gwa:tə] 'la camicia' ma [ni (g)'ga:tə] 'una camicia', [u xwa'ro] 'il pugnale' ma [ni 'xa'ro] 'un pugnale'; vista la rarità della regola l'unica spiegazione persuasiva è che essa sia stata importata dai dialetti coterrioriali in cui è ampiamente testimoniata. Le regole importate dalla romaní d'Abruzzo che hanno come modello l'abruzzese sono almeno una decina e forse alcune sono ancora da scoprire (cfr. e.g. Scala 2015, 2018). Per quanto è a mia conoscenza, ad esempio, finora non è stata mai segnalata la completa convergenza del vocalismo pretonico con quello dei dialetti abruzzo-molisani. Il vocalismo tonico della romaní è a 5 timbri (/a/, /e/, /i/, /o/, /u/) e nella Early-Romani anche quello pretonico era identico. Oggi nella romaní d'Abruzzo il vocalismo tonico è immutato, mentre quello pretonico è ridotto a 4 timbri con /a/ e /u/ conservati, /o/ → [u], /i, e/ → [ə], ma → /i/ prima di consonante palatale (cfr. Early-Romani [ma'ro] 'pane' → RA [ma'ro], ER [go'no] 'sacco' → RA [gu'no], ER [du'ral] 'lontano' → RA [du'ra:lə], ER [di'ves] 'giorno' → RA [də'ves], ER [ke'rel] 'egli fa' → RA [kə're:lə], ma ER [ʃe'ro] 'testa' → RA [ʃi'ro], ER [kin'd(j)as] 'vendette' → RA [kip'na]). La situazione è identica in abruzzese e molisano, soprattutto nell'area di Chieti e Pescara, dove si trova la stessa distribuzione dei timbri e lo stesso influsso delle palatali (Giammarco 1979, 40–41).

Spostando lo sguardo su un altro dialetto romaní si osserva qualcosa di analogo: il sinto lombardo parlato a Piacenza e a Pavia, benché alquanto innovativo dal punto di vista prosodico, presenta un significativo mutamento nel vocalismo pretonico. Qui il vocalismo atono prevede le seguenti innovazioni: /a/, /i/ e /u/ sono stabili, invece /o/ → [u] ed /e/ → [i], ma → [a] se davanti a /r/ (ER [ka'mav] 'io amo' → SL [ka'mava], ER [di'k^hav] 'io vedo' → SL [di'kava], ER [ʃu'nav] 'io dormo'

→ SL [su'nava], ER [ro'dav] 'io cerco' → SL [ru'dava], ER [p^he'nav] 'io dico' → SL [pi'nava], ma ER [ke'rav] 'io faccio' → SL [ka'rava] 'io faccio'). Anche in questo caso la distribuzione dei timbri vocalici in sillaba pretonica è identica a quella dei dialetti coterrioriali, inclusi gli allofoni combinatori innescati da suoni vicini.

Il vocalismo atono della romaní non prevedeva restrizione di timbri (e ciò è conservato ancora in molti dialetti sia dei sinti che dei rom, che hanno 5 vocali toniche e 5 atone coincidenti). Nel caso della romaní d'Abruzzo e del sinto lombardo la situazione è radicalmente mutata. Il numero di timbri disponibili in sillaba pretonica è diminuito, la distribuzione a base prosodica è identica a quella dei dialetti coterrioriali e si allinea ad essi anche negli esiti allofonici condizionati a livello segmentale. Oggigiorno il vocalismo pretonico di romaní d'Abruzzo e sinto lombardo sembra dunque gestito dalla stessa regola che agisce nei dialetti coterrioriali.

Riferimenti bibliografici

- Giammarco, Ernesto. 1979. *Abruzzo*. Pisa: Pacini.
- Scala, Andrea. 2015. Propagginazione e romaní d'Abruzzo: un caso di importazione di regola fonologica. *L'Italia dialettale* 76. 181–209.
- Scala, Andrea. 2018. Italo-Romance Phonological Rules and Indo-Aryan Lexicon: the case of Abruzzian Romani. In D'Alessandro, Roberta & Pescarini, Diego (a cura di), *Advances in Italian Dialectology*, 165–187. Leida: Brill.
- Schirru, Giancarlo. 2008. Propagginazione e categorie nominali in un dialetto del Molise. In De Angelis, Alessandro (a cura di), *I dialetti meridionali tra arcaismo e interferenza, Atti del Convegno Internazionale di Dialettologia, Messina, 4-6 giugno 2008*, 291–309. Palermo: Centro Studi Filologici e Linguistici Siciliani.
- Schirru, Giancarlo. 2013. La propagginazione in un dialetto della Val di Comino. In Avolio, Francesco (a cura di), *Lingua e dialetto tra l'Italia centrale e l'Italia meridionale. I dialetti della Media Valle del Liri e delle zone limitrofe*, 78–91. Roccasecca: Arte Stampa Editore.
- Soravia, Giulio. 1977. *Dialetti degli Zingari Italiani*. Pisa: Pacini.
- Soravia, Giulio. 2019. *La lingua romanès d'Abruzzo*. Bologna: Bonomo.
- Tuttle, Edward F. 1985. Assimilazione «permansiva» negli esiti centro-meridionali di *a* tonica, *L'Italia dialettale* 98. 1–34.

Il dialetto romanzo dei rom calabresi di Cosenza: strutture grammaticali e inquadramento teorico

Le varietà romanze parlate dai rom e dai sinti di antico insediamento in Italia costituiscono un interessante terreno di ricerca per il linguista. Da poco più di un decennio, infatti, è stato rilevato che alcune di queste varietà, originariamente riservate alla comunicazione esocomunitaria, stanno via via rimpiazzando o hanno ormai soppiantato la romanì come L1 endocomunitaria (Scala 2012). È emerso, inoltre, come i dialetti neolatini di rom e sinti non siano sovrapponibili ai dialetti coterrioriali dei *gagè* (= non rom), ma, anzi, rappresentino delle *koinè* nate dal contatto serrato tra parlanti di varietà dialettali diverse e in cui sono chiaramente riconoscibili caratteristiche strutturali di varia provenienza geografica. In alcuni casi (ci riferiamo, ad esempio, al piemontese dei sinti piemontesi o al veneto parlato dai sinti emiliani), il contatto tra codice indo-ario e codici romanzi ha dato luogo a delle vere e proprie lingue miste con grammatica romanza e lessico romanì (Duberti 2010; Scala 2015; Ziano 2025). Gran parte di queste nuove varietà romanze o, per lo meno, a base romanza, resta, tuttavia, ancora da descrivere.

Ciò considerato, scopo principale della presente comunicazione sarà quello di discutere alcune caratteristiche salienti della varietà neolatina parlata dai rom di Cosenza, lo *zingariškə*, su cui solo in anni recenti si è cominciato a far luce – le uniche informazioni finora pubblicate sono in Grottola (2006, 85-89) – e che, pertanto, può ben essere inclusa tra le varietà italo-romanze dei rom tuttora semineglette.

Nello specifico, si tenterà anzitutto di dimostrare che questo dialetto, in linea con quanto si riscontra in altre comunità rom e sinti (v. sopra), non può essere assimilato ad alcun dialetto calabrese o lucano parlato dai non rom. A tale scopo, si offrirà una breve descrizione di alcuni fenomeni dello *zingariškə* (p. es.: la presenza di voci a vocalismo sardo, la mancata retroflessione dei nessi latini -tr- e -str-, la conservazione della desinenza verbale -[sə] < -s di 2sg, la tendenza a generalizzare l'ausiliare 'avere' in tutte le classi verbali), evidenziandone la palese distanza dal cosentino dei *gagè* e da gran parte degli altri dialetti calabresi della Valle del Crati. In secondo luogo, dopo averne proposto una classificazione geolinguistica, si cercherà di appurare se lo *zingariškə* sia una *koinè* dai caratteri stabilmente definiti o, come sembrerebbero indicare l'alto grado di variazione interfamiliare e l'episodica coesistenza di alternative fonetiche, morfologiche e sintattiche a livello idiolettale, una lingua comune in corso di formazione, vale a dire una *koinè* in cui i fenomeni di livellamento e lo

sviluppo di forme ibride interdialettali sono ancora parziali (Kerswill & Trudgill 2005; Siegel 2001; Trudgill 2004).

Si tenterà, inoltre, di motivare teoricamente due fenomeni (segnatamente, il rifacimento analogico dei possessivi tonici di 1pl e 2pl sulle corrispettive forme del singolare e l'uso pleonastico del possessivo enclitico di 3sg come elemento “inanalizzat[o] e agglutinat[o] con il costituente testa” cfr. Andorno et al. 2017, 124) che lo *zingariškə* non sembra aver mutuato da nessuna varietà italo-romanza e che potrebbero rappresentare delle strutture meglio inquadrabili all'interno di una lingua di contatto come un pidgin o un creolo (Turchetta 2009). In effetti, sebbene il calabrese dei rom non sia in alcun modo accostabile alle due tipologie di lingue appena citate, è vero che nella formazione delle *koinà* e dei creoli “*the same processes of mixing and levelling seem to be in operation*” (Siegel 2001, 192). Ma, mentre nelle *koinà* le varianti sottoposte a livellamento afferiscono a varietà diverse di una medesima lingua o di lingue tipologicamente affini già esistenti, nei pidgin e nei creoli queste varianti costituiscono “a large proportion of newly modified versions of a language resulting from transfer and restructuring in second language learning and from model simplification” (Siegel 2001, 193). Proprio quest'ultimo processo, almeno in una certa fase, potrebbe aver avuto luogo in *zingariškə*, se si ammette che i due ibridismi presi in esame si siano sviluppati quando i rom calabresi avevano ancora come L1 la romanì e come L2 una varietà romanza che, pur rappresentando il mezzo di comunicazione con i *gagè*, era stata appresa in condizioni di input assai ridotto e, pertanto, doveva esibire fenomeni di semplificazione / ristrutturazione grammaticale progressivamente “nativizzati”, compatibili con le caratteristiche di un pidgin o di un creolo (Siegel 2001, 186). Tale circostanza, se provata, oltre a fornire nuovi dati empirici sulle lingue di contatto, consentirebbe di chiarire meglio il processo di formazione dello *zingariškə* e, più in generale, le condizioni in cui dovette formarsi il repertorio bilingue degli attuali rom cosentini.

Riferimenti bibliografici

- Andorno, Cecilia, Bernini, Giuliano, Giacalone Ramat, Anna & Valentini, Ada. 2017. Sintassi. In Giacalone Ramat, Anna (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 116–178. Roma: Carocci.
- Duberti, Nicola. 2010. Trin kamlé tikné: studenti sinti a Rocca de' Baldi (Cuneo). *Bollettino dell'Atlante Linguistico Italiano* 34. 37–78.

- Grottola, Pierluigi. 2006. *La lingua della comunità rom di Cosenza*. In Berlingieri, Rubina, Cicala, Antonio, Dionesalvi, Claudio & Grottola, Pierluigi (a cura di), *Rubbina: un racconto sugli zingari di Cosenza*, 73–125. Soveria Mannelli: Città Calabria Edizioni.
- Kerswill, Paul & Peter Trudgill. 2005. *The birth of new dialects*. In Auer, Peter, Hinskens, Frans & Kerswill, Paul (a cura di), *Dialect change: Convergence and divergence in European languages*, 196–220. Cambridge: Cambridge University Press.
- Scala, Andrea. 2012. *Purché la lingua non sia una sola... Trasformazione dei repertori e conservazione del plurilinguismo presso i Sinti italiani dall'Unità ad oggi*. In Telmon, Tullio, Raimondi, Gianmario & Revelli, Luisa (a cura di), *Coesistenze linguistiche nell'Italia pre- e postunitaria*. Atti del XLV Congresso Internazionale della Società di linguistica italiana (SLI), Aosta, Bard, Torino, 26-28 settembre 2011, 437–446. Roma: Bulzoni.
- Scala, Andrea. 2015. *Dal sinto piemontese al piemontese sinto: sulle tracce di una lingua mista*. In Bruno, Carla, Casini, Simone, Gallina, Francesca & Siebetchu, Raymond (a cura di), *Plurilinguismo e sintassi. Atti del XLVI Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI)*, Siena, 27-29 settembre 2012, 255–267. Roma: Bulzoni.
- Siegel, Jeff. 2001. Koine formation and creole genesis. In Smith, Norval & Veenstra, Tonjes (a cura di), *Creolization and Contact*, 175–197. Amsterdam/Philadelphia: Benjamins.
<https://doi.org/10.1075/cll.23.08sie>
- Trudgill, Peter. 2004. *New-dialect formation: the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Turchetta, Barbara. 2009. Genesi, diffusione e crescita di una lingua di contatto. In Turchetta, Barbara (a cura di), *Pidgin e creoli. Introduzione alle lingue di contatto*, 17–68. Roma: Carocci.
- Ziano, Carlo. 2025. Sul dialetto veneto dei sinti emiliani (e sull'origine di *cantàvino*, *cantábanos* 'cantavamo'). Relazione orale presentata all'*VIII Convegno Internazionale di Dialettologia Progetto A.L.Ba.: Dialetti: per parlare e parlarne*, Potenza, 8-10 maggio 2025.

Na varietà bastardaja: la koinè piemontese sinta tra focalizzazione e livellamento

Nel panorama italo-romanzo il contatto delle popolazioni sinte e rom con i parlanti delle varietà locali ha instaurato dinamiche molto articolate e complesse. In questo contesto, il Piemonte costituisce di certo un caso interessante. Come già evidenziato da Scala (2012), mentre i sinti lombardi hanno mantenuto la *romaní* in funzione di *we-code*, quelli piemontesi si sono orientati sul dialetto locale in funzione endocomunitaria. Nello specifico, il piemontese dei sinti (PdS) si presenta come un'unica varietà, diffusa nei campi della regione e con caratteristiche perlopiù condivise. Tale idea di unità linguistica, che ben si sposa con il già accennato carattere di *we-code* associato a questa varietà, si realizza per mezzo di una varietà mescidata, livellata e parzialmente semplificata (cfr. Trudgill 2004): una koinè che raggruppa e armonizza elementi di provenienza diatopica (e forse anche diastratica) differente.

Il presente contributo si propone di descrivere tale varietà, evidenziando il diverso apporto in campo fonetico, morfologico e lessicale delle diverse varietà di piemontese e – pur in misura più limitata – della *romaní*. Fra i tratti ascrivibili al piemontese orientale che sembra aver fornito l'apporto più consistente alla formazione del PdS, menzioniamo ad esempio:

- la terminazione in *-i* del plurale dei femminili (*bòci* 'bocce') e della desinenza di seconda persona singolare (*vèni* 'vieni'). L'esito in *-i*, che è pressoché generalizzato nei dialetti piemontesi orientali (cfr. Ferrarotti 2022), si oppone a quello in *-e* proprio invece delle varietà del Piemonte occidentale (che tuttavia, a eccezione del torinese, manifestano generalmente *-i* nell'infinito dei verbi della seconda e terza coniugazione, come nelle varietà orientali di piemontese e nel PS: *seurti* 'uscire', *deurmi* 'dormire'). La pervasività della terminazione *-i* giunge addirittura ad intaccare l'articolo femminile plurale, che è *li* (es. *li bòci* 'le bocce'), riscontrabile in poche varietà periferiche di piemontese;
- il sistema di selezione degli ausiliari, che prevede *avere* con tutti gli intransitivi: es. *l'avìa andait* (lett. 'aveva andato'), *l'an mòrt tuti* (lett. 'hanno morti tutti'), *l'an asparii* (lett. 'hanno sparite'). Tuttavia, contrariamente a quanto accade ad esempio in biellese o in astigiano, l'ausiliare è sempre *essere* con i riflessivi: *l'è perdissi* 'si è perso'.

A questo elenco di tratti fondamentalmente orientali, se ne affiancano altri riconducibili a varietà del Piemonte occidentale, quali per esempio le forme verbali *avìa*, *fasìa*, *vorìa* ('aveva', 'faceva', 'voleva') o la desinenza di prima singolare del presente indicativo in *-o* ([u]), estesa anche all'imperfetto (come in cuneese): *i parlo* 'parlo', *i ero* 'ero'. Tali forme, comunque, alternano al

presente con le varianti più tipicamente orientali e meridionali con uscita in consonante: es. *i gir* ‘giro’.

Vi sono infine alcune caratteristiche che sembrano essere proprie del solo PdS, come l’accordo di genere nei participi femminili dei verbi inaccusativi (es. *rivaja* ‘arrivata’ vs *rivà* ‘arrivato’), caratteristica propria anche di altre varietà (alessandrine *in primis*) ma che nel PdS arriva a interessare anche i transitivi: *l’a mangiaja* ‘(lei) ha mangiato’. O, ancora, l’uso della preposizione *da* per *əd/ad* ‘di’ (es. *bicer da vin* ‘bicchiere di vino (e non da vino)’). Anche questo tratto può però occorrere in alternanza con participi femminili che non manifestano l’accordo di genere (es. *rivà* ‘arrivata, arrivato’).

Alla luce dei dati presentati, si dimostrerà come il processo di koinizzazione avvenuto in questa varietà sia del tutto peculiare. Da un lato, il PdS è una varietà certamente focalizzata o stabilizzata (Trudgill 2004), poiché esso è passato, nel corso del Novecento, da esoletto e endoletto (Scala 2015), arrivando a soppiantare prima la varietà romaní di sinto piemontese, poi il piemontese sinto (*pararomaní*, cfr. Bakker 2020). Dall’altro lato, il fatto che persista nel PdS un’alternanza così complessa fra tratti occidentali e orientali indirizza verso un livellamento che ha agito su piani e in momenti differenti. La peculiarità del processo di livellamento nel PdS (che, canonicamente, precede la fase di focalizzazione, o stabilizzazione; cfr. Siegel 2001 e Trudgill 2004) risiede nello specifico contesto in cui è avvenuta. Se, infatti, la focalizzazione – con annessa sostituzione del PdS alla precedente varietà endoletale – è un fenomeno relativamente recente, il livellamento, al contrario, deve essersi verificato in una fase storica in cui le popolazioni sinte del Piemonte si trovavano ancora in una fase nomade, o perlomeno semi-sedentarizzata. Ciò si riflette nella presenza delle varianti in competizione sopra citate, a cui va aggiunta la desinenza *-[a]* alla IIIsg nei verbi della seconda classe (es. *a perda* ‘perde’), che alterna con quella in consonante (*a perd*). Quando due varianti sopravvivono entrambe al livellamento, può verificarsi un fenomeno chiamato *riallocazione* (Britain 2018, 151), mediante il quale esse vengono rifunzionalizzate, assumendo ruoli nuovi e diversi all’interno del sistema. Per esempio, l’accordo di genere nel participio dei transitivi contraddistingue, in certe aree del Piemonte, la varietà di alcune famiglie sinte (assumendo pertanto una funzione diastratica). Infine, all’interno della riallocazione, interviene non di rado un processo di accomodazione, ovvero di adattamento alle varietà di piemontese parlate dai non sinti nelle aree in cui si trovano i campi, che svolgono così un’azione di rinforzo rispetto all’una o all’altra delle varianti (*contact-induced stability*, Kühl & Braunmüller 2014).

Riferimenti bibliografici

- Bakker, Peter. 2020. Para-Romani varieties. In Matras, Yaron & Tenser, Anton (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*, 353–386. Cham: Palgrave Macmillan.
- Britain, David. 2018. Dialect contact and new dialect formation. In: Boberg, Charles, Nerbonne, John & Watt, Dominic (a cura di), *The Handbook of Dialectology*, 143–158. Oxford: Wiley-Blackwell.
- Kühl, Karoline & Braunmüller, Kurt. 2014. Linguistic stability and divergence: an extended perspective on language contact. In Braunmüller, Kurt, Höder, Steffen & Kühl, Karoline (a cura di), *Stability and divergence in language contact. Factors and mechanisms*, 13–38. Amsterdam/Philadelphia: Benjamins.
- Ferrarotti, Lorenzo. 2022. *I dialetti del Piemonte orientale*. Berlino/Boston: Mouton de Gruyter.
- Scala, Andrea. 2012. Purché la lingua non sia una sola... Trasformazione dei repertori e conservazione del plurilinguismo presso i Sinti italiani dall'Unità ad oggi. In Telmon, Tullio, Raimondi, Gianmario & Revelli, Luisa (a cura di), *Coesistenze linguistiche nell'Italia pre- e postunitaria. Atti del XLV Congresso Internazionale della Società di linguistica italiana (SLI), Aosta/Bard/Torino, 26-28 settembre 2011*, 437–446. Roma: Bulzoni.
- Scala, Andrea. 2015. *Dal sinto piemontese al piemontese sinto: sulle tracce di una lingua mista*. In Bruno, Carla, Casini, Simone, Gallina, Francesca & Siebetchu, Raymond (a cura di), *Plurilinguismo e sintassi. Atti del XLVI Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI), Siena 27-29 settembre 2012*, 255–267. Roma: Bulzoni.
- Siegel, Jeff. 2001. Koine formation and creole genesis. In Smith, Norval & Veenstra, Tonjes (a cura di), *Creolization and Contact*, 175–197. Amsterdam/Philadelphia: Benjamins.
<https://doi.org/10.1075/cil.23.08sie>
- Trudgill, Peter. 2004. *New-dialect formation: the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Il dialetto veneto dei sinti emiliani

Nell'intervento si vorrebbe analizzare la varietà italo-romanza parlata dai Sinti insediati nelle province di Reggio Emilia, Modena e Bologna (Trevisan 2009), ricostruendone il profilo linguistico e le implicazioni storico-migratorie. Il gruppo, caratterizzato da forti legami parentali e da una lunga tradizione di attività itineranti (circhi e spettacoli viaggianti), presenta oggi un repertorio linguistico complesso: accanto a una varietà romaní di tipo lombardo (detta *sinto stretto*), ormai parlata solo da pochi anziani, è in uso generale una varietà italo-romanza che i parlanti stessi chiamano *sinto* (Scala 2020).

L'analisi fonologica e morfologica mostra che questa varietà non è emiliana, bensì riconducibile all'area veneta centrale. Tra i fenomeni diagnostici si segnalano: conservazione delle vocali finali; generalizzazione del morfo di 1pl *-emo*; sincretismi verbali (tra 2sg e 2pl nell'imperfetto, tra 3sg e 3pl di tutti i modi e tempi); presenza dell'infixo velare nella 1sg del presente indicativo (*vago, fago*); il morfo della 1pl del non-presente in *-no* (*dormivino* 'dormivamo'); l'uso di *iera* 'era'; i pronomi clitici obliqui (*ma, ta, ga, sa*); 3sg.ind.prs. in *-a* (*vegna, dorma*); la congiunzione *quanto* 'quando'; le forme accorciate *favo* 'facevo', *femo* 'facciamo'. Questi tratti non appaiono riconducibili a un singolo centro, ma a una zona compresa tra Padova, Verona e Mantova, che fu probabilmente usata dalle famiglie sinte come base per spostamenti più ampi attraverso l'intera Italia settentrionale e centrale.

Il caso rappresenta un nuovo esempio di "eteroglossia interna" (Telmon 1992): una minoranza insediata stabilmente in Emilia conserva una varietà alloglotta, distinta dai dialetti locali, grazie a forte coesione sociale e una netta separazione dal gruppo maggioritario (*gagi*). I dati linguistici consentono inoltre di ricostruire le traiettorie migratorie del gruppo: l'arrivo in Emilia sarebbe avvenuto da est (Veneto), non da ovest (Piacenza), dove si trovano altri sinti lombardi. Infine, si può ipotizzare che la perdita progressiva della romaní tra anni '50 e '60 — parallela all'ingresso nel mondo dello spettacolo viaggiante e a dinamiche di assimilazione — sia correlata all'adozione della varietà veneta, che — in Emilia — poteva funzionare come codice identitario (*we-code*).

Riferimenti bibliografici

Scala, Andrea. 2020. La romaní in Italia tra rappresentazione e legittimazione. *Language Problems and Language Planning* 44. 346–370. <https://doi.org/10.1075/lplp.00070.sca>

Telmon, Tullio. 1992. *Le minoranze linguistiche in Italia*. Alessandria: Edizioni dell'Orso.

Trevisan, Paola. 2009. *Etnografia di un libro. Scritture, politiche e parentela in una comunità di Sinti*.
Roma: CISU.

Il prestito delle congiunzioni italiane nelle lingue alloglotte d'Italia e nelle varietà romaní

Il prestito di connettori costituisce un fenomeno diffuso nelle lingue del mondo (Matras 2007, 2020). Le congiunzioni possono essere oggetto di prestito già nelle fasi iniziali del contatto linguistico (Thomason 2001). Gli studi precedenti mostrano che il prestito di congiunzioni non rappresenta un fenomeno casuale, bensì riflette regolarità strutturali: il prestito di congiunzioni appartenenti a una determinata funzione precede quello di altre tipologie funzionali. Matras (2007) e Grant (2012) propongono a questo proposito implicazioni funzionali, ovvero le gerarchie di prestito. Secondo tali gerarchie, se in una lingua sono presenti congiunzioni disgiuntive in prestito, sono attestate anche congiunzioni avversative in prestito e a loro volta nelle lingue dotate di congiunzioni copulative in prestito si riscontrano anche congiunzioni disgiuntive e avversative ($ma > o > e$). Nell'ambito dei subordinatori avverbiali, quelli logici (come causali, concessivi, ecc.) precedono il prestito di altri tipi di congiunzioni, quali quelle temporali. Inoltre, il prestito delle congiunzioni coordinanti implica quello dei subordinatori o complementatori.

Tali gerarchie, tuttavia, sono state elaborate sulla base di campioni interlinguistici relativamente limitati, comprendenti soltanto alcune decine di lingue. La verifica di queste gerarchie mediante campioni più ampi e maggiormente rappresentativi delle diverse macroaree linguistiche mostra che esse costituiscono piuttosto tendenze generali che veri e propri universali linguistici (Stolz & Levkovich 2022, 2023).

Per quanto riguarda le lingue parlate sul territorio italiano, Levkovich (2025) conferma le gerarchie di prestito per le lingue alloglotte non romanze, seppure con alcune limitazioni. Il campione di Matras (2007), sul quale si basano le gerarchie proposte, comprende due varietà romaní: il domari e il rumungro. Anche il comportamento di ulteriori varietà romaní rispetto al prestito di congiunzioni (Matras 1998) conferma almeno alcune gerarchie. Tuttavia, le varietà romaní in contatto con l'italiano non sono state finora oggetto di ricerca sotto questo aspetto.

La principale domanda di ricerca del presente contributo è la seguente: in che modo si comportano le varietà romaní in contatto con l'italiano rispetto alle lingue alloglotte non-romanze parlate sul territorio italiano e rispetto ad altre varietà romaní in contatto con lingue non italiane per quanto riguarda il prestito di congiunzioni? Inoltre, il contributo intende verificare la validità delle gerarchie di prestito relative alle congiunzioni.

Riferimenti bibliografici

- Grant, Anthony P. 2012. Contact, convergence, and conjunctions: A cross-linguistic study of borrowing correlations among certain kinds of discourse, phasal adverbial, and dependent clause markers. In Claudine Chamoreau & Isabelle Léglise (a cura di), *Dynamics of Contact-Induced Language Change*, 311–358. Berlino/Boston: Mouton de Gruyter.
- Levkovych, Nataliya. 2025. Italian loan conjunctions in the alloglottic non-Romance languages of Italy. *STUF - Language Typology and Universals* 78(2). 265–319. <https://doi.org/10.1515/stuf-2025-2009>
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2). 281–331.
- Matras, Yaron. 2007. The borrowability of structural categories. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 31–74. Berlino/New York: Mouton de Gruyter.
- Matras, Yaron. 2020. *Language Contact*. Cambridge: Cambridge University Press.
- Stolz, Thomas & Nataliya Levkovich. 2022. On loan conjunctions: A comparative study with special focus on the languages of the former Soviet Union. In Nataliya Levkovich (a cura di), *Susceptibility vs. Resistance: Case studies on different structural categories in language-contact situations*, 259–392. Berlin: Mouton de Gruyter.
- Stolz, Thomas & Nataliya Levkovich. 2023. Travellers in time and space: Tracing loan conjunctions in the replica languages of the Arabo-Persian sphere of linguistic influence (with special focus on the shores of the Indian Ocean). In Nataliya Levkovich (a cura di), *Diversity in Contact*, 327–366. Berlin: Mouton de Gruyter.
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Workshop 4

La coerenza e la coesione nella comunicazione digitale, tra testi sintetici e multimodalità

Proponenti: ANNA-MARIA DE CESARE, GIULIANA FIORENTINO E FILIPPO PECORARI

MICHELA DOTA E MASSIMO PRADA

Forme della coerenza e della coesione nelle homepage di YouTube, Instagram e TikTok

Il contributo analizza i modi e i principi semiotici generali che operano nelle *homepage* di YouTube, Instagram e TikTok, al fine di garantirne il funzionamento e la percezione da parte degli utenti come testi coesi e coerenti. L'analisi si colloca nell'ambito dell'approccio sistemico-funzionale all'analisi multimodale del discorso e integra strumenti della semiotica sociale (Kress & Van Leeuwen 2001; Djonov & Knox 2014; Jewitt 2014; Jewitt, Bezemer & O'Halloran 2025).

L'intervento intende mostrare come le *homepage* dei tre servizi non siano semplici interfacce di navigazione, ma autentici artefatti comunicativi complessi, a pieno titolo assimilabili a testi (Prada 2022a; 2022b). In tali ambienti digitali, le tre metafunzioni hallidayane della comunicazione – ideativa, interazionale e testuale – trovano realizzazione attraverso la cooperazione di più modi semiotici.

Questa sinergia coinvolge il modo verbale scritto, impiegato in etichette, titoli e didascalie; il modo iconico, attraverso immagini e miniature; nonché il modo grafico e tipografico, inclusi gli aspetti cromatici e disposizionali. In particolare, il *layout* svolge un ruolo centrale nel guidare la lettura dell'*homepage*, che può essere analizzata come una struttura articolata in cluster e sottounità, organizzate secondo una testualità prototipica fatta di unità riconoscibili e gerarchicamente ordinate.

Le gerarchie informative sono ulteriormente segnalate dalle scelte cromatiche e tipografiche, mentre miniature e didascalie tendono a disporsi in insiemi coerenti. Anche pulsanti e icone contribuiscono alla coesione testuale, grazie a funzioni prevedibili e convenzionali condivise dagli utenti. La cooperazione di questi diversi modi semiotici concorre così a garantire la coerenza e la coesione complessive dell'*homepage*.

Infine, le *homepage* analizzate si configurano come una forma di rimediazione di generi testuali convenzionali e fortemente formalizzati, quali la bacheca, il catalogo, l'elenco e il sommario, risultando perciò immediatamente riconoscibili. Al contempo, esse realizzano una rete di relazioni tipica della testualità ipermediale contemporanea: includono puntatori ipertestuali verso altri testi

(video, canali, *playlist*), generano metatesti dinamici costruiti dagli algoritmi in base alla navigazione degli utenti e presentano connessioni interne, come commenti che rimandano a specifici momenti dei video, citazioni e riferimenti incrociati.

Riferimenti bibliografici

- Djonov, Emilia & Knox, John. 2014. How to analyze webpages. In Norris, Sigrid & Maier, Carmen Daniela (a cura di), *Interactions, images and texts: a reader in Multimodality*, 171–193. Berlin: De Gruyter.
- Kress, Gunther & Van Leeuwen, Theo. 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Hodder.
- Jewitt, Carey. 2014. *The Routledge Handbook of Multimodal Analysis*. London/New York: Routledge.
- Jewitt, Carey, Bezemer, Jeff & O'Halloran, Kay. 2025. *Introducing Multimodality, II ed.* London/New York: Routledge.
- Prada, Massimo. 2022a. Letture multimodali per l'educazione linguistica. *Italiano LinguaDue* 14(2). 85–130. <https://doi.org/10.54103/2037-3597/19651>
- Prada, Massimo. 2022b. Non solum per verba. Minime considerazioni su coesione (rinvio, coriferimento, sostituzione), struttura tematica e dinamismo informativo in testi multimodali telematici. *Lid'O. Lingua italiana d'oggi* XIX. 23–43.

Sulla coerenza multimodale in testi neomediali: il caso del BookTok

Il contributo analizza la coerenza nei prodotti multimodali (cfr., *inter alia*, Bateman 2014; Kong 2006; Royce 2007; Stöckl & Pflaeging 2022) pubblicati su TikTok, piattaforma finora marginalmente considerata nel panorama degli studi italiani sulla multimodalità; in particolare, mostra come risorse semiotiche eterogenee – verbali, cinesiche, iconiche, musicali – interagiscano per assicurare l'unità e la continuità discorsive nei video prodotti dalla comunità del BookTok (Dera 2024; Martens et al. 2022), una delle più cospicue sulla piattaforma, che si caratterizza per la brevità e la densità informativa dei suoi contenuti. La comunità del BookTok è costituita prevalentemente da adolescenti e giovani adulti, i quali realizzano contenuti incentrati sui libri. Il corpus d'indagine è costituito da 50 video della durata complessiva di circa tre ore, in cui i booktoker recensiscono i libri letti o documentano l'esperienza di lettura nel corso del suo svolgimento, descrivendo le reazioni emotive da essa suscitate.

Lo studio, che adotta un approccio sociosemiotico all'analisi multimodale (Kress & Van Leeuwen 2001) basato sulla grammatica sistemico-funzionale di Halliday (1978), distingue due principali classi di relazioni logico-semantiche tra le risorse semiotiche impiegate nei video (Martinec & Salway 2005; Van Leeuwen 2005): relazioni di elaborazione e quelle di espansione. L'analisi del corpus ha evidenziato una netta prevalenza di relazioni di elaborazione: risorse verbali (in forma scritta), cinesiche e registiche riformulano o precisano quanto espresso nel parlato. Ad esempio, la variazione dell'inquadratura, il gesto deittico, il titolo del libro in sovraimpressione si configurano come correlati transmodali del referente introdotto in forma orale: si tratta di una ripetizione multimodale che rafforza la coerenza del messaggio audiovisivo, necessaria in un ambiente comunicativo caratterizzato da una rapida fruizione dei contenuti. Si riscontrano relazioni di espansione quando uno o più modi veicolano informazioni ulteriori rispetto a quelle fornite dalle altre risorse semiotiche: nei video in cui il creator commenta oralmente i libri letti compaiono frequentemente valutazioni sintetiche espresse per via iconico-grafica (mediante una scala a più livelli, da una a cinque stelle), oppure indicazioni (in forma scritta) relative al genere del libro, alla sua collocazione all'interno di una saga o alla tipologia di narratore.

L'indagine mira, infine, a individuare le configurazioni di integrazione tra i diversi modi più diffuse all'interno del BookTok in modo da far luce sui canoni comunicativi e sulle convenzioni espressive proprie di questa comunità.

Riferimenti bibliografici

- Bateman, John. 2014. Multimodal coherence research and its application. In Gruber, Helmut & Redeker, Gisela (a cura di), *The Pragmatics of Discourse Coherence*, 145–177. Amsterdam/Philadelphia: John Benjamins.
- Dera, Jeroen. 2024. BookTok: A Narrative Review of Current Literature and Directions for Future Research. *Literature Compass* 21 (10-12). e70012. <https://doi.org/10.1111/lic3.70012>
- Halliday, Michael A. K. 1978. *Language as Social Semiotics: The Social Interpretation of Language and Meaning*. London: Arnold.
- Kong, Kenneth. 2006. A taxonomy of the discourse relations between words and visual. *Information Design Journal* 14. 207–230. <https://doi.org/10.1075/idj.14.3.04kon>
- Kress, Gunther & van Leeuwen, Theo. 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.
- Martens, Marianne, Balling, Gitte & Higgason, Kristen A. 2022. #BookTokMadeMeReadIt: young adult reading communities across an international, sociotechnical landscape. *Journal of Information and Learning Sciences* 123 (11-12). 705–722. <https://doi.org/10.1108/ILS-07-2022-0086>
- Martinec, Radan & Salway, Andrew. 2005. A system for image-text relations in new (and old media). *Visual Communication* 4. 337–371. <https://doi.org/10.1177/14703572050559>
- Royce, Terry. 2007. Intersemiotic complementarity: a framework for multimodal discourse analysis. In Royce, Terry & Bowcher, Wendy L. (a cura di), *New Directions in the Analysis of Multimodal Discourse*, 63–109. Mahwah (NJ): Lawrence Erlbaum Associates.
- Stöckl, Hartmut & Pflaeging, Jana. 2022. Multimodal Coherence Revisited: Notes on Move from Theory to Data in Annotating Print Advertisements. *Frontiers in Communication* 7. 900994 <https://doi.org/10.3389/fcomm.2022.900994>
- van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. Londra: Routledge.

Sinsemia ovvero la dimensione compositiva della multimodalità: un'integrazione teorica e metodologica

Il contributo propone un confronto e una integrazione teorica e metodologica tra multimodalità e sinsemia a partire dalla distinzione tra coerenza e coesione nei testi complessi. Assumendo che ogni comunicazione sia intrinsecamente multimodale (Adami 2015; Kress & van Leeuwen 1996, 40–44, 175–214) e che i modi concorrano alla costruzione del significato attraverso relazioni strutturate, la multimodalità fornisce il quadro comunicativo generale per descrivere l'interazione tra risorse semiotiche. La sinsemia, intesa come disposizione deliberata e regolare degli elementi nello spazio per produrre senso non ambiguo, offre invece un modello progettuale specifico per l'organizzazione della scrittura nella sua dimensione spaziale (Lussu 2019).

Gli studi sulla coesione nei documenti multimodali mostrano come diverse forme di *signaling* (*directive, descriptive*, assente) incidano sull'integrazione tra testo e figure e sui processi cognitivi di lettura (Acartürk, Taboada & Habel 2013), confermando che la coesione non è esclusivamente linguistica ma anche grafica e topografica.

In questa prospettiva, la metafunzione compositiva (Kress & van Leeuwen 1996; Sindoni 2022: 25) costituisce il punto di contatto con la dimensione strutturale della scrittura spaziale.

La tesi è che la coerenza nei testi sinsemici efficaci sia primariamente una proprietà del *reference frame* – ossia della struttura spaziale che orienta l'interpretazione – mentre la coesione si distribuisce tra marcatori linguistici, variabili visive e dispositivi topografici, inclusi i meccanismi di *cross-referencing* descritti negli studi sulla coesione multimodale.

L'analisi si concentrerà su testi sinsemici di provata efficacia, non selezionati per esemplarità retorica ma per stabilità d'uso e capacità di produrre risultati cognitivi condivisi – ad esempio il piano cartesiano come dispositivo di *graphicacy* avanzata (Ciccione & Dehaene 2025). Il campione consente di osservare come gerarchie, aggregazioni e regolarità delle variabili visive configurino un campo di variazione strutturato che rende possibile l'inferenza. La lettura e l'uso dei diagrammi cartesiani (ma anche i diagrammi di Eulero) ad esempio possono essere interpretati come un caso di rielaborazione neurale di competenze culturali recenti, in cui circuiti evolutivamente deputati alla percezione e all'orientamento spaziale vengono riutilizzati e riconfigurati per funzioni simbolico- astratte (Ciccione *et al.* 2023).

Una volta elaborato il *framework* su artefatti/sistemi di provata efficacia, si può usare il quadrifoglio sinsemico come framework di analisi per altri artefatti come predittore di efficacia, in particolare con riferimento a testi in cui la dicotomia parola-figura è sfumata (ad esempio il *Liber Figurarum* di Gioacchino da Fiore, i manoscritti di Leonardo da Vinci, i diagrammi per le sequenze del film *Aleksandr Nevskij* di Sergej M. Ėjzenštejn).

Attraverso un modello sinsemico – che includa variabili visive, gerarchia, aggregazioni, *reference frame* – si mostrerà come la sinsemia possa essere intesa come la dimensione compositiva della multimodalità, fornendo strumenti analitici e progettuali per descrivere e costruire testi multimodali coerenti. Il contributo intende così chiarire differenze e complementarità tra i due ambiti, proponendo una cornice integrata per l'analisi e la progettazione della scrittura spaziale.

Riferimenti bibliografici

- Acartürk, Cengiz, Taboada, Maite & Habel, Christopher. 2013. Cohesion in multimodal documents: Effects of cross-referencing. *Information Design Journal* 20(2). 98–110. <https://doi.org/10.1075/idj.20.2.02aca>
- Adami, Elisabetta. 2016. Multimodality. In García, Ofelia, Flores, Nelson & Spotti, Massimiliano (a cura di), *The Oxford Handbook of Language and Society*. Oxford: Oxford University Press.
- Ciccione, Lorenzo, Sablé-Meyer, Mathilde, Boissin, Emmanuel, Josserand, Marine, Potier-Watkins, Clara, Caparos, Serge & Dehaene, Stanislas. 2023. Trend judgment as a perceptual building block of graphicacy and mathematics, across age, education, and culture. *Scientific Reports* 13(1). 10266. <https://doi.org/10.1038/s41598-023-37172-3>
- Ciccione, Lorenzo & Dehaene, Stanislas. 2026. The neural bases of graphical perception: A novel instance of cultural recycling? *Journal of Cognitive Neuroscience* 38(1). 71–88. <https://doi.org/10.1162/JOCN.a.81>
- Kress, Gunther & van Leeuwen, Theo. 1996. *Reading Images. The Grammar of Visual Design* (II ed.). Londra: Routledge..
- Lussu, Giovanni. 2019. Writing is image. In Cicalò, Enrico (a cura di), *Proceedings of the 2nd International and Interdisciplinary Conference on Image and Imagination - IMG 2019. (Advances in Intelligent Systems and Computing, vol 1140), Alghero, 4-5 luglio 2019*, 15–20. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-41018-6_3

Sindoni, Maria Grazia. 2022. Traiettorie della multimodalità: gli snodi teorici e i modelli applicativi. In Dota, Michela, Polimeni, Giuseppe & Prada, Massimo (a cura di), *Multimedialità e multimodalità. Teoria, prassi e didattica dei testi complessi. Italiano LinguaDue* 14(2). 19–46. <https://doi.org/10.54103/2037-3597/19647>

Strategie di coesione in testi umani e sintetici: uno studio basato sull'analisi delle reti

Scopo di questo studio pilota è fornire un confronto tra le diverse strategie di coesione adottate da umani nella scrittura di testi giuridico-amministrativi e riscritture semplificate da parte di Large Language Models (LLMs). I primi sono caratterizzati da una complessità stigmatizzata per i problemi che crea nell'accesso alla giustizia da parte del cittadino (Cassese 1994, 9; Fortis 2005, 89). Gli LLM rappresentano degli strumenti efficienti per la semplificazione (Fiorentino 2025; Ganfi & Russodivito 2025; Nozza & Attanasio 2023; Ondelli & Santoro 2025; *inter alia*), ma la loro efficacia merita ancora adeguata valutazione.

Proponiamo qui una metodologia basata sulle reti (Barabási 2016; Newman 2018). Il testo è rappresentato come una rete, i cui nodi sono gli enunciati e gli archi i diversi tipi di relazioni coesive tra enunciati (Halliday & Hasan 1976). In questa fase esplorativa, limitiamo l'analisi al regolamento TA.RI di Torino, confrontando l'originale con la riscrittura di Gemini 3.

Nella tabella 1 sono presentati i risultati in merito al numero di nodi e di archi nei due testi.

	Umano	Sintetico
Nodi	98	114
Archi	1496	765

Tabella 1. Numero di nodi e archi

L'originale è composto da meno enunciati, ma da più relazioni. Si confrontino i grafici 1 e 2.

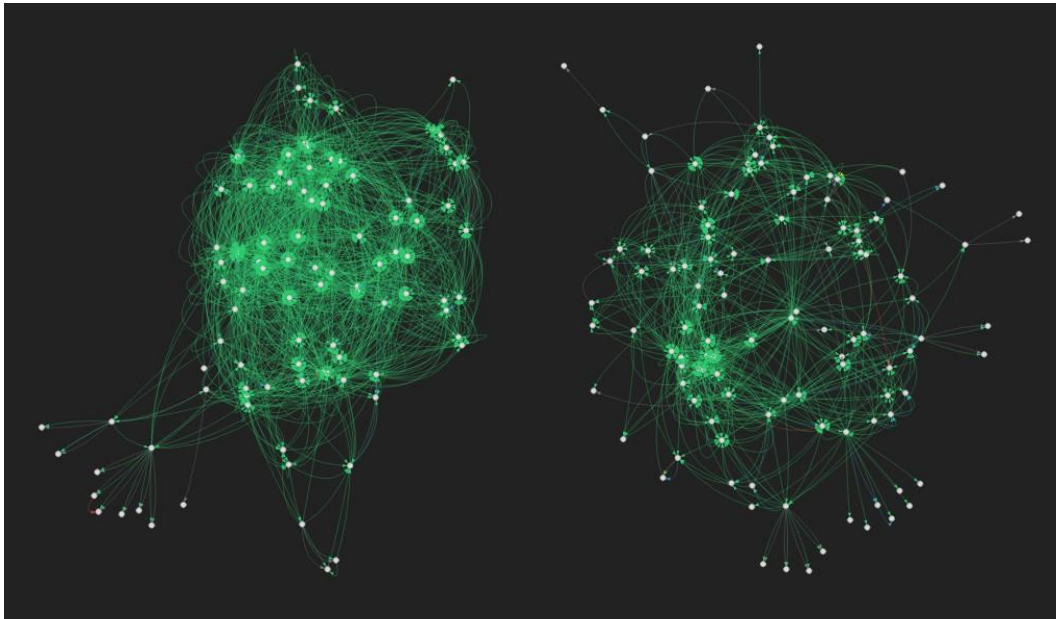


Grafico 1. Rete del testo umano

Grafico 2. Rete del testo sintetico

L'LLM riscrive un enunciato pesante ridistribuendo l'informazione:

- (1) Il presente regolamento disciplina l'applicazione della tassa comunale sui rifiuti (TARI) nel Comune di Torino di cui alla Legge 27 dicembre 2013 n. 147 e s.m.i. secondo le disposizioni previste dalla Autorità di Regolazione per l'Energia, Reti e Ambiente (ARERA) inerenti la regolazione della qualità del servizio di gestione dei rifiuti urbani.
- (2) Questo regolamento stabilisce le modalità di applicazione e gestione della Tassa sui Rifiuti (TARI) nel Comune di Torino.

Esso recepisce e attua le norme nazionali vigenti (in particolare la Legge n. 147/2013) e si conforma rigorosamente alle disposizioni dell'Autorità di Regolazione per l'Energia, Reti e Ambiente (ARERA), l'ente che vigila sulla correttezza delle tariffe e sulla qualità del servizio offerto ai cittadini.

Ciò ha un effetto anche sul numero di relazioni. Gli enunciati umani ricchi di informazione permettono di stabilire numerosi rimandi lungo tutto il testo. L'LLM costruisce invece il testo per mezzo di piccoli raggruppamenti di enunciati continui e molto coesi, ma poveri di rimandi a distanza.

Questo si quantifica osservando il numero di archi che si dipanano dal nodo, cioè il loro *grado*. Il grado medio dei nodi del testo umano è più del doppio di quello artificiale (15,27 vs 6,71). Il primo

mostra una struttura gerarchica dominata da *hub* fortemente interconnessi, il secondo appiattisce tale gerarchia distribuendo le relazioni uniformemente.

Per valutare il livello di organizzazione generale ricorriamo alla *global efficiency* e alla *small-worldness*. La prima misura quanto sia facile navigare dall'inizio alla fine della rete. La seconda valuta la modularità della rete, se questa è strutturata in isole coesive definite e collegate tra loro. Si veda la tabella 2.

	Umano	Sintetico
<i>Global efficiency</i>	0,61	0,45
<i>Small-Worldness</i>	1,15	1,78

Tabella 2. Metriche globali di rete

La maggiore *global efficiency* del testo umano dipende dalla costruzione globale della coesione, riflettendo una costruzione della coesione globalmente integrata. L'LLM favorisce invece modularità (alta *small-worldness*) e un procedere sequenziale in unità minori altamente coese al loro interno.

Sebbene la frammentazione possa essere ascritta all'operazione di semplificazione in generale, l'analisi di rete proposta ha permesso di caratterizzare l'esatta architettura coesiva prodotta dall'LLM. Sviluppi futuri includeranno testi semplificati da umani come termine di paragone per isolarne la specificità.

Riferimenti bibliografici

- Barabási, Albert-László. 2013. *Network Science*. Cambridge: Cambridge University Press.
- Cassese, Sabino (a cura di). 1994. *Codice di stile delle comunicazioni scritte ad uso delle amministrazioni pubbliche*. Roma: Istituto poligrafico e zecca dello stato.
- Fiorentino, Giuliana. 2025. Umano vs AI: alcune considerazioni sulla semplificazione manuale e automatica del corpus Italt. In Fiorentino, Giuliana, Cioffi, Alessandro & Simonelli, Maria Ausilia (a cura di), *Amministrazione attiva: semplicità e chiarezza per la comunicazione amministrativa*. 37–65. Firenze: Cesati.
- Fortis, Daniele. 2005. Il linguaggio amministrativo italiano. *Revista de Llengua i Dret* 43. 47–116.

- Halliday, Michael A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English*. Londra: Routledge.
- Ganfi, Vittorio & Russodivito, Marco. 2025. SEMPL-IT: un modello di intelligenza artificiale per la semplificazione dell'italiano. In Fiorentino, Giuliana, Cioffi, Alessandro & Simonelli, Maria Ausilia (a cura di), *Amministrazione attiva: semplicità e chiarezza per la comunicazione amministrativa*. 79–101. Firenze: Cesati.
- Newman, Mark. 2018. *Networks*. Oxford: Oxford University Press.
- Nozza, Debora & Attanasio, Giuseppe. 2023. Is it really that simple? Prompting large language models for automatic text simplification in Italian. In Boschetti, Federico, Lebani, Gianluca, Magnini, Bernardo & Novielli, Nicole (a cura di), *Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023), Venezia 30 novembre -2 dicembre 2023*, 322–333. Aachen: CEUR Workshop Proceedings.
- Ondelli, Stefano & Santoro, Alessia. 2025. Come usare Chatgpt per semplificare i testi amministrativi? Alcuni confronti tra intelligenza umana e intelligenza artificiale. *Italiano LinguaDue* 17(2), 1327–1377. <https://doi.org/10.54103/2037-3597/30514>

Coerenza e coesione in testi giudiziari italiani generati e riscritti dall'intelligenza artificiale

Mentre l'impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è al centro di un quadro normativo in rapida evoluzione (Regolamento UE 2024/1689 “*AI Act*”, legge italiana n. 132 del 2025), la compagine dei testi giudiziari italiani generati o riscritti dall'intelligenza artificiale è un ambito ancora inesplorato. Gli studi linguistici sulle proprietà testuali dei testi generati si sono infatti finora occupati di temi o tipi testuali diversi (cfr. De Cesare 2026; Fiorentino & Tivosanis 2024): biografie (De Cesare 2023, 2025), testi narrativi (Antonelli 2025; Cicero 2025), testi amministrativi (Pascucci & Tivosanis 2024; Ondelli & Santoro 2025).

In questo contributo si propone un primo confronto tra un campione di sentenze, di merito e legittimità, tratte da banche dati ufficiali e i corrispettivi testi generati da alcune piattaforme di intelligenza artificiale sulla base di *prompt* analoghi. La sentenza scritta dal giudice estensore viene raffrontata con il corrispettivo provvedimento riscritto da *chatbot*, in due versioni: una edita dal sistema rielaborando quella reale, l'altra generata dall'IA sulla base di dati (risultanze probatorie processuali, elementi di diritto e dispositivo) forniti dall'operatore.

L'obiettivo è quello di individuare, alla luce delle proprietà già note del genere sentenza (ad es. Dell'Anna 2013; Mortara Garavelli 2001; Visconti 2022), le peculiarità testuali delle sentenze riscritte o generate dalla IA. Oltre a fenomeni di coesione, quali l'impiego dei segnali interpuntivi, l'uso dei connettivi, l'impiego dei dispositivi di referenza nominale e della *variatio* (Palermo 2025), lo studio mette al centro la dimensione enunciativo-polifonica del testo (Ferrari 2014, 233), in una varietà, quale quella giudiziaria, ricca di movenze dialogiche e controargomentative, rese con il discorso diretto e indiretto, ma anche con segni paragrafematici, aggettivi e avverbi che segnalano ciò che viene detto come voce altrui, o con il condizionale «di riserva», o di «dissociazione» (Mortara Garavelli 1995): proprio un diverso uso, marcato, di questo tipo di condizionale nei testi giudiziari generati rivela come nel confronto tra testualità naturale e testualità sintetica nel testo giudiziario giochi un ruolo importante la sostanza semantico-pragmatica del testo, soprattutto nella sua dimensione polifonico-enunciativa.

Riferimenti bibliografici

Antonelli, Giuseppe. 2025. Storia brevissima (ma molto intensa) dell'IA-taliano. *Lingue e Culture dei media* 9(1). 4–50. <https://doi.org/10.54103/2532-1803/29341>

- Cicero, Francesco. 2025. Esercizi di stile. La narrativa per l'infanzia delle intelligenze artificiali. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 2(2). 1–21. <https://doi:10.62408/ai-ling.v2i2.19>.
- De Cesare, Anna-Maria. 2023. Assessing the quality of ChatGPT's generated output in light of human-written texts. A corpus study based on textual parameters. *CHIMERA. Romance Corpora and Linguistic studies* 10. 179–210. <https://revistas.uam.es/chimera/article/view/17979>
- De Cesare, Anna-Maria. 2025. LLM referential chain generation. A qualitative case study based on Italian biographies produced by GPT-4. *Linguistik Online* 136(4). 25–52. <https://doi.org/10.13092/8tppdv47>
- De Cesare, Anna-Maria. 2026. *L'italiano sintetico dell'intelligenza artificiale generativa*. Firenze: Cesati.
- Dell'Anna, Maria Vittoria. 2013. *In nome del popolo italiano. Linguaggio giuridico e lingua della sentenza in Italia*. Roma: Bonacci.
- Ferrari, Angela. 2014. *Linguistica del testo. Principi, fenomeni, strutture*. Roma: Carocci.
- Fiorentino, Giuliana & Tivosanis, Mirko. 2024. Chiaro, sintetico, e brillante: l'italiano dei testi redatti con l'IA funziona? *Lid'O. Lingua italiana d'oggi* XXI. 37–65.
- Mortara Garavelli, Bice. 2001. *Le parole e la giustizia*. Torino: Einaudi.
- Mortara Garavelli, Bice. 1995. Il discorso riportato. In Renzi, Lorenzo, Salvi, Giampaolo & Cardinaletti, Anna (a cura di), *Grande grammatica italiana di consultazione*, vol. III, 429–470. Bologna: Il Mulino.
- Ondelli, Stefano & Santoro, Alessia. 2025. Come possiamo usare ChatGPT per semplificare i testi amministrativi? *Italiano LinguaDue* 2. 1306–1356. <https://doi.org/10.54103/2037-3597/30514>
- Palermo, Massimo. 2025. *Tanto per cambiare. La coazione a variare nella storia dell'italiano*. Bologna: Il Mulino.
- Pascucci, Mariachiara & Tivosanis, Mirko. 2024. Confronto tra diversi tipi di valutazione del miglioramento della chiarezza di testi amministrativi in lingua italiana. In Dell'Orletta, Felice *et al.* (a cura di), In Dell'Orletta, Felice Lenci, Alessandro, Montemagni, Simonetta & Sprugnoli, Rachele (a cura di), *Proceedings of the Tenth Italian Conference*

on Computational Linguistics (CLiC-it 2024), Pisa, 4-6 dicembre 2024, 746–756. Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2024.clicit-1.81/>

Visconti, Jacqueline. 2022. *Studi su testi giuridici: norme, sentenze, traduzione*. Firenze: Accademia della Crusca.

Coesione e coerenza nel testo mimetico delle favole ‘naturali’ vs ‘sintetiche’: un caso studio sui segnali discorsivi

L'avvento dei modelli linguistici di grandi dimensioni (*Large Language Models* – LLM) ha inaugurato un nuovo paradigma nella produzione di testi, che sono sempre più spesso oggetto di analisi linguistiche (cfr. per esempio Cicero 2025; De Cesare 2024, 2025; Mantovani *in stampa*; Tavosanis 2024). Un fenomeno accennato da Cicero (2025, 10s.), ma che merita ancora di essere approfondito, è quello dei segnali discorsivi (SD), elementi essenziali per garantire la coesione e la coerenza del testo in quanto assumono funzioni metatestuali, interazionali e cognitive (cfr. Bazzanella 2001, 232–249, 2005, 140s.; Molinelli 2014, 195s.; Sansò 2020, 16–27).

Il contributo indaga le forme e le funzioni dei SD nei discorsi diretti all'interno di un corpus di favole naturali, ovvero scritte da esseri umani, a confronto con favole generate da ChatGPT (favole ‘sintetiche’, cfr. De Cesare 2026). L'obiettivo è verificare in che modo la lingua mimetica generata si distingue da quella realizzata da esseri umani per quanto riguarda l'uso dei SD. In particolare, le domande di ricerca sono le seguenti: (i) quali differenze ci sono fra i corpora in termini di frequenza d'uso (*token*) e di forme (*type*) dei SD? (ii) quali sono le funzioni svolte dai SD e in che modo il loro uso incide sulla coesione e sulla coerenza dei testi in questione? (iii) in quale misura i vasti dati in lingua inglese con cui sono stati addestrati i modelli GPT influiscono sui testi generati in lingua italiana?

Per rispondere a queste domande è stata condotta un'analisi semiautomatica in tre sottocorpora creati *ad hoc*:

- 200 favole scritte da esseri umani (di cui 70 estratte dalle *Favole al telefono* di Rodari e 130 redatte da utenti del web);
- 150 favole generate da ChatGPT-5.2 in italiano;
- 150 favole generate da ChatGPT-5.2 in inglese.

I risultati dell'analisi (cfr. Tab. 1) mostrano una frequenza d'uso e una varietà di SD più elevate nelle favole naturali (38,7 *token* e 10,9 *type* ogni 10.000 parole) rispetto ai dati riscontrati nei sottocorpora GPT-IT (22,5 *token* – 3,1 *type*) e GPT-EN (9,9 *token* – 3,5 *type*). I SD ‘naturali’ provengono tuttavia principalmente dal corpus di Rodari, che si rivela non solo più denso in termini di *token* rispetto alle favole scritte da utenti del web (67,2 vs 28,3), ma anche più vario in termini di *type* (23,4 vs 7,9). I

dati suggeriscono come la natura statistico-probabilistica di ChatGPT permetta di replicare solo in parte una scrittura caratterizzata da SD, senza tuttavia riuscire a riprodurre lo stile complesso e individuale proprio di una scrittura d'autore come quella rodariana.

Sottocorpus	N. favole	N. parole	f token (10.000 parole)	f type (10.000 parole)
HUM	200	97.653	38,7	10,91
Rodari	70	26.184	67,2	23,35
Utenti web	130	71.469	28,3	7,93
GPT-IT	150	78.945	22,5	3,19
GPT-EN	150	80.018	9,9	3,50

Tabella 1: Presenza dei SD nei tre sottocorpora

La scarsità e la limitata varietà nelle favole sintetiche di SD che, in quanto elementi pragmatico-testuali, “assicurano una certa fluidità all’interazione” (De Marco & Leone 2016, 118), impoverisce la mimesi dell’oralità, intaccando la coesione testuale e la costruzione di senso. Per quanto riguarda la terza di domanda di ricerca, l’influsso dei dati di addestramento inglesi sembrerebbe circoscritto all’uso sovraesteso del SD *oh* nei testi sintetici (GPT-IT: 7,7; GPT-EN: 5,1), che risulta nettamente superiore rispetto alla frequenza registrata nel corpus HUM (1,3).

Riferimenti bibliografici

- Bazzanella, Carla. 2001. I segnali discorsivi. In Renzi, Lorenzo, Salvi, Giampaolo & Cardinaletti, Anna (a cura di), *Grande grammatica italiana di consultazione*. Vol. III, 225–257. Bologna: Il Mulino.
- Bazzanella, Carla. 2005. Segnali discorsivi e sviluppi conversazionali. In Albano Leoni, Federico & Giordano, Rosa (a cura di), *Italiano parlato. Analisi di un dialogo*, 137–157. Napoli: Liguori.
- Cicero, Francesco. 2025. Esercizi di stile: La narrativa per l’infanzia delle intelligenze artificiali. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 2(2). 1–21.
<https://doi:10.62408/ai-ling.v2i2.19>

- De Cesare, Anna-Maria. 2024. Nuove dinamiche di contatto linguistico: Le “impronte digitali” dell’inglese nell’italiano generato da LLM. *Lid’o. Lingua Italiana d’Oggi* XXI. 67–91.
- De Cesare, Anna-Maria. 2025. LLM referential chain generation. A qualitative case study based on Italian biographies produced by GPT-4. *Linguistik Online* 136(4). 25–52.
<https://doi.org/10.13092/8tppdv47>
- De Cesare, Anna-Maria. 2026. *L’italiano sintetico dell’intelligenza artificiale generativa*. Firenze: Cesati.
- De Marco, Anna & Leone, Paola. 2016. L’uso dei segnali discorsivi in apprendenti di italiano L2. In Santoro, Elisabetta & Vedder, Ineke (a cura di), *Pragmatica e interculturalità in italiano lingua seconda*, 117–132. Firenze: Cesati.
- Mantovani, Giulia. *In stampa*. Forme sintetiche e forme analitiche per i diminutivi dei nomi: un confronto tra favole naturali e favole generate dall’intelligenza artificiale. In Lala, Letizia, Castro, Enrico & Garzonio, Jacopo (a cura di), *Il nome in italiano. Morfologia, sintassi, testualità*. Firenze: Cesati.
- Molinelli, Piera. 2014. Orientarsi nel discorso: segnali discorsivi e segnali pragmatici in italiano. In Pîrvu, Elena (a cura di), *Discorso e cultura nella lingua e nella letteratura italiana*, 195–208. Firenze: Cesati.
- Rodari, Gianni. 1995 [1962]. *Favole al telefono*. Trieste: Edizioni EL.
- Sansò, Andrea. 2020. *I segnali discorsivi*. Roma: Carocci.
- Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1(1). 1–24.
<https://doi.org/10.62408/ai-ling.v1i1.14>

La coerenza alla prova dell'IA (e viceversa): dai testi umani alle riscritture sintetiche

Nell'ambito della scrittura scolastica in italiano è stata recentemente proposta una classificazione di tratti relativi alla coerenza (Bienati & Zanasi 2025), fondata sul Modello basilese della testualità (MBT) (Ferrari 2014, 2019) e volta a sistematizzare circa vent'anni di ricerche empiriche. A essa sono collegati una scala di valutazione e un corpus di 635 testi argomentativi in italiano L1 annotato per tratti di coerenza e coesione. Tali strumenti hanno consentito di operationalizzare aspetti del MBT distinguendo segmentazione e fenomeni sui piani tematico-referenziale e logico-argomentativo e traducendo queste dimensioni in descrittori osservabili e annotati manualmente. In questo senso, scala e corpus non rappresentano soltanto strumenti valutativi, bensì un'infrastruttura descrittiva fondata su un modello teorico coerente, pensata per rendere empiricamente analizzabile un aspetto della competenza scritta, di fatto carente di strumenti dedicati.

Con l'emergere dell'intelligenza artificiale generativa, la produzione automatica di testi apparentemente ben strutturati solleva la questione se categorie e strumenti elaborati per la scrittura scolastica adolescenziale mantengano la loro pertinenza in un nuovo scenario comunicativo. L'IA costituisce non solo un nuovo oggetto di analisi, ma anche un banco di prova per verificare la tenuta di categorie descrittive nate per testi prodotti da scriventi umani. I modelli linguistici intervengono sugli stessi tratti di coerenza individuati dagli esperti? In che modo?

Il contributo parte da questa domanda di ricerca e si concentra su un caso di studio nel quale si confrontano testi originali del corpus con riscritture generate dall'IA a partire dalla consegna originaria, senza indicazioni relative alla scala. L'analisi si focalizza su tratti semantici non locali – ad esempio gerarchie logico-argomentative (come la contraddizione) o fenomeni di progressione dato-nuovo (come l'informazione nuova non sviluppata) già annotati manualmente. L'obiettivo è verificare se tali fenomeni vengano riconosciuti dai modelli linguistici nelle operazioni di riscrittura e in che modo vengano eventualmente risolti o trasformati. Per testare la stabilità dei risultati rispetto a diverse architetture, è previsto un confronto tra un modello commerciale di ultima generazione e uno o due modelli *open source* eseguibili localmente. I primi dati disponibili riguardano il *gold standard* umano e confermano come, nei 635 testi, la dimensione più critica sia la strutturazione logico-argomentativa ($M = 4,13$ su 10), a fronte di punteggi medi più elevati per segmentazione ($M = 6,08$) e comprensibilità ($M = 6,34$). Il confronto con l'IA mira ad accertare se tali aree di criticità si rispecchino anche nelle operazioni di riscrittura e in che misura i modelli intervengano sui fenomeni individuati dagli annotatori.

In questa prospettiva, l'analisi apre anche una riflessione di tipo valutativo: da un lato consente di osservare come l'IA tratti determinate categorie descrittive; dall'altro, tali categorie offrono una lente per interpretare il comportamento dei modelli, ad esempio rispetto alla loro capacità di riconoscere le stesse aree di incoerenza individuate dai lettori esperti e di proporre riscritture potenzialmente rilevanti in ambito didattico. Il contributo si inserisce così nel dibattito metodologico recente (Tavosanis 2024), affrontando il problema della trasferibilità di categorie e strumenti sviluppati per la scrittura L1 alla testualità sintetica.

Riferimenti bibliografici

- Bienati, Arianna & Zanasi, Lorenzo. 2025. Anomalie testuali nella scrittura scolastica in italiano: una proposta di classificazione. *SILTA. Studi italiani di linguistica teorica e applicata* LIV(2). 355–377.
- Ferrari, Angela. 2014. *Linguistica del testo. Principi, fenomeni, strutture*. Roma: Carocci.
- Ferrari, Angela. 2019. *Che cos'è un testo*. Roma: Carocci.
- Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1(1). 1–24.
<https://doi.org/10.62408/ai-ling.v1i1.14>

Workshop 5

Esprimere l'inatteso: forme, costruzioni e strategie mirative nelle lingue romanze

Proponente: SARAH DESSÌ SCHMID, ILARIA FIORENTINI, ANDREA SANSÒ

SILVIO CRUSCHINA

On the relationship between deixis and mirativity: Evidence from Sicilian

Across languages, verbs of visual perception such as *see* and *look* are especially likely to develop into discourse markers. Their primary function is typically to direct the addressee's attention, but they may also acquire mirative extensions:

- (1) Look, the keys were in my pocket all along!
- (2) Look at how much you've grown!

In Spanish and Italian, the mirative meaning may become prominent in certain specific constructions. In Spanish, mirative *mira* is used in structures where it is followed by the complementizer *que* 'that' (see Fuentes Rodríguez 2020; Sánchez López 2017). Italian *guarda te*, in the example in (4) includes the accusative second person singular pronoun *te* 'you', but it is analysed as a single mirative discourse marker (see Badan 2020):

- (3) *¡Mira que eres linda!* (Spanish, Sánchez López 2017, 490)
'How cute you are!'
- (4) *Guarda te è andato al ristorante senza dirmelo!* (Venetan Italian, Badan 2020, 306)
'Surprisingly, he's gone to the restaurant without saying a word to me!'

In this paper, I discuss two Sicilian discourse markers that have developed from perception verbs: *talè* (from *taliari* 'to look', see Cruschina 2025) and *bi* – also realized as *mbi* or *vih* (from *vidiri* 'to see'; based on Scivoletto 2023, 2025):

- (5) *Talè bella giuvina!*
'Look, what a beautiful young woman!'
- (6) *Mbiiiii chi priuuuu!* (Scivoletto 2025, 187)
'Oh, what a joy!'

I show that these discourse markers have developed a fully-fledged mirative function, which may complement their function as attention-getters or serve as their sole expressive and subjective

meaning. Unlike the perception-verb based makers in other Romance languages, the mirative function of *talè* and *bi* is therefore rather predominant and is not restricted to specific structures or contexts. The diachronic evolution of discourse markers derived from perception verbs has been widely examined in linguistic studies, with a general consensus that their grammaticalization into attention-getter is triggered by the inherent deictic properties of the (imperative) source verb (see, e.g., Fagard 2010, Ghezzi & Molinelli 2014, 2015; Waltereit 2002). Following this framework, we argue that *talè* and *bi* have undergone a similar grammaticalization path, to which the mirative function can be added as a further or more advanced stage, as shown in Table 1 (from Cruschina 2025, 94, adapted from Ghezzi & Molinelli 2015, 129):

a.	direct the sight toward something in the deictic context	DEICTIC	
b.	direct the attention toward the process of enunciation	ATTENTION GETTING	
c.	direct the attention toward the content of utterance		
d.	direct the attention toward the unexpected content of the utterance		MIRATIVE
e.	speaker's expression of surprise		

Table 1: The grammaticalization path from deixis to mirativity

From a cognitive standpoint, the link between deixis and mirativity poses a number of intriguing and complex issues. Building on Peterson (2016, 2020), I propose that mirativity may be understood as a form of deixis, in that the speaker's expression of surprise operates much like a deictic gesture of pointing or reference. While canonical deictic expressions anchor an utterance to a particular time, place, or participant, mirative markers instead “point to” an event or situation that the speaker construes as unexpected or surprising. Seen in this light, the intermediate attention-getting stage becomes conceptually clear: mirativity serves to focus the interlocutors' attention on information that is both surprising and contextually salient for the speaker.

Riferimenti bibliografici

- Badan, Linda. 2020. Italian discourse markers: The case of *guarda te*. *Studia Linguistica* 74(2): 303–336. <https://doi.org/10.1111/stul.12127>
- Cruschina, Silvio. 2025. The discourse marker *talè* ‘look’ in Sicilian and the development of mirativity. *Italian Journal of Linguistics* 37(2): 83–106. <https://doi.org/10.26346/1120-2726-254>

- Fagard, Benjamin. 2010. *É vida, olha...*: Imperatives as discourse markers and grammaticalization paths in Romance. *Languages in Contrast* 10(2): 245–267. <https://doi.org/10.1075/bct.44.07fag>
- Fuentes Rodríguez, Catalina. 2020. *Mira/mira que*: construcciones, contexto argumentativo y funciones relacionales. *Revista de Filología Hispanica RILCE* 36(3): 941–966.
- Ghezzi, Chiara & Piera Molinelli. 2014. Italian *guarda, prego, dai*. Pragmatic markers and the left and right periphery. In Beeching, Kate & Detges, Ulrich (eds.), *Discourse Functions at the Left and Right Periphery*, 117–150. Leiden: Brill. https://doi.org/10.1163/9789004274822_007
- Ghezzi, Chiara & Piera Molinelli. 2015. Segnali allocutivi di richiamo: percorsi pragmatici e sviluppi diacronici tra latino e italiano. *Cuadernos de Filología Italiana* 22: 21–47. http://dx.doi.org/10.5209/rev_CFIT.2015.v22.50950
- Peterson, Tyler. 2016. Mirativity as surprise: Evidentiality, information, and deixis. *Journal of Psycholinguistic Research* 45. 1327–1357. <https://doi.org/10.1007/s10936-015-9408-9>
- Peterson, Tyler. 2020. Mirativity in Morphology. *Oxford Research Encyclopedia of Linguistics*.
- Sánchez López, Cristina 2017. Mirativity in Spanish. The case of the particle *mira*. *Review of Cognitive Linguistics* 15(2): 489–514. <https://doi.org/10.1075/rc1.15.2.08san>
- Scivoletto, Giulio. 2023. Il siciliano *bi* e l'espressione della miratività. *Cuadernos de Filología Italiana* 30: 189–206. <https://doi.org/10.5209/cfit.83685>
- Scivoletto, Giulio. 2025. Mirativity as a semantic notion: The trajectory and status of Sicilian *bì*. In Susana Rodriguez Rosique (ed.), *Expressing Surprise at the Crossroads: Mirativity, Exclamativity and Cooptation in Romance Languages*, 179–202. Berlin: De Gruyter. <https://doi.org/10.1515/9783111386683-008>
- Waltereit, Richard 2002. Imperatives, interruption in conversation, and the rise of discourse markers: a study of Italian *guarda*. *Linguistics* 40: 987–1010. <https://doi.org/10.1515/ling.2002.041>

A ze morto me nona! Una lettura mirativa per una struttura veneta

Il contributo intende mostrare come una specifica configurazione morfosintattica del padovano – esemplificata in frasi come *A ze morto me nona!* {CL-TOPIC be.3 died.M.SG my grandmother} – costituisca un dispositivo grammaticale di codifica della miratività. La proposta integra due analisi precedenti, sviluppate indipendentemente: da un lato l'interpretazione del clitico preverbale *-a* in Benincà (1994), dall'altro l'analisi dell'accordo participiale mancato (APM) in strutture VS con inaccusativi e passivi (da ultimo: Schaeffer 2023). Nessuno dei due lavori tematizza esplicitamente la miratività; si mostrerà tuttavia che la loro combinazione consente di identificare una strategia strutturalmente sistematica di espressione dell'inatteso nel dominio romanzo.

Il primo elemento della configurazione è il clitico *-a*, che Benincà (1994) dimostra non essere un clitico soggetto canonico: esso può coesistere con clitici soggetto, precede la negazione, compare in contesti impersonali e non occupa la posizione sintattica del soggetto. Benincà propone che *-a* leghi una posizione alta di Topic quando questa è vuota. In altri termini, *-a* segnala l'assenza di un topic discorsivo espresso, marcando una struttura priva di ancoraggio tematico. Il secondo elemento è l'APM, osservabile in strutture VS con verbi inaccusativi e passivi, dove il participio compare nella forma di default maschile singolare indipendentemente dalle proprietà ϕ del DP postverbale, come in *(a) ze nato do zemei* {CLTOPIC be.3 born.M.SG two twins}. L'APM alterna con l'accordo pieno ma è soggetto a restrizioni: è escluso con soggetti nulli referenziali, con pronomi di prima e seconda persona e in presenza di clitico soggetto flesso; è invece obbligatorio con quantificatori plurali come *nisuni* e *qualcheduni*. Questi vincoli mostrano che l'APM non è semplice opzionalità morfologica, ma correlato allo statuto discorsivo del DP postverbale.

Si propone che la co-occorrenza di *-a* e APM realizzi una configurazione mirativa in senso tipologico (Aikhenvald 2012): la frase codifica l'informazione come inattesa rispetto alla rappresentazione epistemica del parlante. *-a* segnala l'assenza di topic (struttura non categoriale), mentre l'APM neutralizza l'accordo con il DP postverbale, impedendone la promozione a centro tematico. La struttura risultante è monolitica, eventiva e non bipartita, e veicola interpretazioni di sorpresa, scoperta improvvisa o aspettativa disattesa. Diversamente dalle strategie mirative romanze finora descritte (tempi verbali, perifrasi, particelle pragmatiche), la configurazione padovana mostra proprietà di sistematicità morfosintattica: non è mera estensione pragmatica di categorie esistenti, ma emerge dall'interazione strutturale tra un marcatore alto (*-a*) e un pattern di accordo difettivo (APM). In tal senso, essa contribuisce al dibattito sullo statuto grammaticale della miratività nelle lingue

romanze, mostrando che, pur in assenza di morfemi esclusivamente mirativi, l'*unexpectedness* può essere codificato tramite configurazioni strutturali solo apparentemente opzionali.

Il contributo si inserisce così nella riflessione tipologica sulla miratività nelle lingue romanze, proponendo che il padovano offra un caso di marcatura configurazionale dell'inatteso, dove morfosintassi e struttura informativa convergono nella grammaticalizzazione di un valore mirativo.

Riferimenti bibliografici

Aikhenvald, Alexandra 2012. The essence of mirativity. *Linguistic Typology* 16. 435–485.

Benincà, Paola 1994. Il clitico «a» nel dialetto padovano. In Benincà, Paola (xxx), *La variazione sintattica. Scritti di dialettologia romanza*. Bologna: Il Mulino.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33-52.

Schaefer, Silvia. 2023. *Nobody There to Agree with. A Non-Movement-Based Account of Past Participle Agreement Variation in North-Eastern Italian Dialects*. Frankfurt am Main: Johann Wolfgang Goethe-Universität (Tesi di dottorato.)

Gli usi di *bada* e *badali*: dal richiamo dell'attenzione alla *mock mirativity*

In molte lingue i verbi di percezione e cognizione si prestano a veicolare la miratività, soprattutto nell'uso come *attention getters* (Iliescu 2014). Come lo sp. *mira* (Maldonado & De la Mora 2020; Sánchez López 2017), anche *guarda*, *sentì*, *pensa* (Lo Baido 2021) intrecciano funzioni interazionali, valutative, evidenziali con valori enfatici e mirativi. L'imperativo di *badare* sviluppa caratteristiche analoghe a partire dal senso 'volgere lo sguardo e la mente': per ricostruirne gli usi attraverso il tempo e le varietà dell'italiano proporrò un'analisi qualitativa dei dati ricavati da vari corpora. Per l'italiano antico userò il database OVI; per le epoche successive selezionerò opere teatrali, raccolte epistolari e romanzi con parti dialogate dal catalogo BiBit; gli esempi ibridi tra scritto e parlato tratti da ItTenTen20 illustreranno l'impiego di *bada* e *badali* nell'area toscana. Lo studio diacronico e sincronico di queste forme consentirà di discutere l'origine delle strategie mirative e precisarne gli aspetti semantici, pragmatici e retorici alla luce dei concetti di rilevanza, (inter)soggettività e deissi (Calaresu 2023; Peterson 2016).

Dal Settecento *bada* ricorre spesso in atti richiestivi con una complementazione variegata, ma inizia anche ad emanciparsi dalla sintassi della frase, acquisendo funzioni pragmatico-discorsive:

(1) CAIO *Vo' saperlo.*

FULVIO *Bada, ti pentirai.* (Monti, *Caio Gracco* 2,8)

Come *guarda*, *bada* indicizza la struttura del discorso e mantiene la coesione testuale indirizzando l'attenzione dell'ascoltatore "verso l'atto di enunciazione che il parlante sta per compiere" (Ghezzi & Molinelli 2015, 28). La formula parentetica *bada bene* enfatizza la natura procedurale e interpersonale dell'espressione, che può riprendere e focalizzare segmenti di testo come 'marcatore di rilevanza' (Pons Bordería 1998, 219-220):

(2) *verrò da te alle nove dalla parte dell'anticamera: bada alle nove: sbrigati dal servitore con qualche ambasciata* (Foscolo, *Epistolario 1794-1804*, 42)

Quando non implica effetti perlocutivi come segnale di allerta o di interruzione 'competitiva' (Bazzanella 1991; Waltereit 2002), *bada* sollecita l'interesse dell'ascoltatore introducendo fatti inattesi o sorprendenti:

- (3) *Una volta mi parve finanche che non avesse voluto stringermi la mano... Bada, s'accorse che gliela porgevo: si finse distratto; era un po' strano veramente.* (Pirandello, *La paura*, 1014)

Elicitando una valutazione condivisa su informazioni nuove, *bada* modula l'interazione anche quando lo stupore è simulato retoricamente (*ma bada un po'*), o rovesciato nell'antifrasi:

- (4) *Ministro, te? o dove l'hai il talento?*

Bada lì: che credi che ci voglia un gran talento a finire i quattrini degli altri? (Collodi, *Cotti come tegoli*, 216)

In (4) l'avverbio deittico metaforizza lo spazio testuale, rimandando al co-testo che precede, e la distanza dall'interlocutore, implicando una divergenza di opinioni. Nel parlato e nello scritto informale dei toscani *badalì* diventa un tratto identitario (Castelli 2018, 9): un 'marcatore di irrilevanza' che attiva implicature ironiche, tra *mock disagreement* e *mock mirativity*:

- (5) *Quel computer come minimo è protetto da una password. Come tutti i portatili.*

Badalì, ci metto un attimo. (Malvaldi, *Argento vivo*, 28)

- (6) *Badalì se devi anda' dietro a una che ciavrà sedicianni, con tutte velle che c'è a giro. Se lo sapesse la tu' nonna...* (Malvaldi, *La briscola in cinque*, 77)

Riferimenti bibliografici

- Bazzanella, Carla 1991. Le interruzioni 'competitive' e 'supportive', verso una configurazione complessiva. In Stati, Sorin, Weigand, Edda & Hundsnurscher, Franz (a cura di), *Dialoganalyse III: Referate der 3. Arbeitstagung*, vol. 1, 283–292. Tübingen: Niemeyer.
- Calaresu, Emilia 2023. Indicare, rappresentare e dislocare. La *Deixis am Phantasma* di Karl Bühler e i modi del Discorso Riportato. *Globe. A Journal of Language, Culture and Communication* 17. 119–131. <https://doi.org/10.54337/ojs.globe.v17i.8204>
- Castelli, Renzo 2018. *Bada lì. Parole, modi di dire, soprannomi e proverbi pisani*. Pisa: ETS.
- Ghezzi, Chiara & Molinelli, Piera 2015. Segnali allocutivi di richiamo: percorsi pragmatici e sviluppi diacronici tra latino e italiano. *Cuadernos de Filología Italiana* 22: 21–47.
- Iliescu, Maria 2014. Call markers in French, Italian, and Romanian. In Ghezzi, Chiara & Molinelli, Piera (a cura di), *Discourse and pragmatic markers from Latin to the Romance languages*, 29–40. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199681600.003.0003>

- Lo Baido, Maria Cristina. 2021. L'allocuzione come veicolo di (inter)soggettività: tra enfasi e miratività. *Cuadernos de Filología Italiana* 28: 89–117.
- Maldonado, Ricardo & De la Mora, Juliana. 2020. *Mira*. Evidencialidad, miratividad y validación subjetiva. In Maldonado, Ricardo & De la Mora, Juliana (eds.), *Evidencialidad. Determinaciones Léxicas y Construccionales*, 289–316. México: UNAM/UAQ.
- Peterson, Tyler. 2016. Mirativity as surprise: Evidentiality, information, and deixis. *Journal of Psycholinguistic Research* 45. 1327–1357. <https://doi.org/10.1007/s10936-015-9408-9>
- Pons Bordería, Salvador 1998. *Oye y mira* a los límites de la conexión. In Martín Zorraquino, María Antonia & Montolío Durán, Estrella (a cura di), *Los marcadores del discurso*, 213–228. Madrid: Arco.
- Sánchez López, Cristina 2017. Mirativity in Spanish. The case of the particle *mira*. *Review of Cognitive Linguistics* 15(2): 489–514. <https://doi.org/10.1075/rcl.15.2.08san>
- Waltereit, Richard 2002. Imperatives, interruption in conversation, and the rise of discourse markers: a study of Italian *guarda*. *Linguistics* 40: 987–1010. <https://doi.org/10.1515/ling.2002.041>

Expressing unexpectedness: The mirative use of *andare a* + infinitive in Italian

Over the past decade, several studies that document and discuss a mirative use of the verbal periphrasis with the auxiliary verb GO/COME (+ *a* ‘to’) + infinitive in various Romance languages and varieties have been conducted (see e.g., Bres & Labeau 2012, 2013, Bres & Levie 2019, and Celle & Lansari 2015 for French; Garachana Camarero 2018 and 2019, Kornfeld 2019, and Rodríguez Rosique 2023 for Spanish). For contemporary Italian, the existence of a mirative function of the construction *andare a* + infinitive, as shown in examples (1) and (2) below, has been mentioned (see e. g. Sornicola 1976, 73 or Cruschina 2022, 5). However, detailed investigations on its characteristics and frequency of use are still lacking (but see Masini, Mattiola & Vecchi 2019, Cruschina 2022 and Cruschina & Bianchi 2021 for mirativity in related constructions, such as pseudocoordination, in Italian and its varieties).

- (1) *La mia [pianta] aveva circa 5 anni ed era sopravvissuta anche agli ultimi inverni rigidi e **mi va a morire** d'estate.*

‘Mine [my plant] was about 5 years old and had survived even the last few harsh winters and surprisingly it died (lit. to me it goes to die) in summer.’

(<https://forum.giardinaggio.it/threads/la-mimosa-si-e-seccata.174045/>) (Dessi Schmid,

Momma & Wiesinger 2025, 235)

- (2) *Durante l'azione, **non mi va a capitare** che una bomba da 100, probabilmente perché avevo schiacciato il tasto troppo leggermente, non mi si sgancia? ...* (itTenTen20, alieuomini.it)

‘During the action, a 100 bomb –would you believe it!?!– doesn’t drop (lit. does not go to me to happen that the bomb does not drop), probably because I had pressed the button too lightly.’

Our contribution, which relies on data extracted from the extensive itTenTen20 corpus (approximately 12.45 billion words), therefore aims to pursue two main objectives:

1. to verify the hypothesis that the Italian periphrasis, often supported by other linguistic elements (e.g. adverbs or pronouns as *non ... mica, proprio, mi, ti, vi*, as ethical dative), is also used to express unexpectedness – although, as is likewise the case in French and Spanish, this function does not appear to be very frequent;
2. to analyse the obtained data by relating them to the typology of unexpectedness developed in Dessi Schmid, Momma & Wiesinger (2025), which is based on the criterion of orientation,

i.e. how, under what conditions and, above all, *for whom* unexpectedness is expressed – for the speaker, for the addressee, or for both.

This will allow us to reflect on the connections between the form through which unexpectedness is expressed (specifically *andare a* + infinitive) and its potential polyfunctionality, i.e. its capacity to express one or more types of mirative orientation. Hereby we aim to contribute to the modelling of a more fine-grained analysis of non-default common ground updates that take place between speaker and addressee in communication.

We focus our analysis on Italian but also discuss equivalent periphrases in Romance using the corresponding large-scale corpora *frTenTen23* and *esTenTen23*.

Riferimenti bibliografici

- Bres, Jacques & Labeau, Emmanuelle 2012. De la grammaticalisation des formes itive (*aller*) et ventive (*venir*) : valeur en langue, emplois en discours. In de Saussure, Louis & Rihs, Alain (eds.), *Etudes de sémantique et pragmatique françaises*, 143–165. Bern: Peter Lang.
<https://doi.org/10.3726/978-3-0352-0125-3>
- Bres, Jacques & Labeau, Emmanuelle 2013. *Aller* et *venir*: des verbes de déplacement aux auxiliaires aspectuelstemporels-modaux. *Langue française* 179(3). 13–28.
- Bres, Jacques & Levie, Yordanka 2019. Miratif (en bulgare) et allure extraordinaire (en français) : *allez croire* que c'est du pareil au même ! *Faits de Langues* 49(2). 183–198.
<https://doi.org/10.1163/19589514-04902013>
- Celle, A. & Lansari, L. 2015. On the mirative meaning of *aller* + infinitive compared with its equivalents in English. *Cahiers Chronos* 27. 289–305.
https://doi.org/10.1163/9789004292772_015
- Cruschina, Silvio (2022). Gone unexpectedly: Pseudo-coordination and the expression of surprise. In Giusti, Giuliana, Di Caro, Vincenzo Nicolò & Ross, Daniel (eds.), *Pseudo-Coordination and Multiple Agreement Constructions*, 129–148. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/la.274.05cru>
- Cruschina, Silvio & Bianchi, Valentina 2021. Mirative implicatures at the syntax-semantics interface: A surprising association and an unexpected move. In Trotzke, Andreas & Villalba, Xavier (eds.), *Expressive Meaning across Linguistic Levels and Frameworks*, 86–107. Oxford: OUP.
<https://doi.org/10.1093/oso/9780198871217.003.0005>
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (ed.),

- Expressing Surprise at the Crossroads: Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111386683-010>.
- Garachana Camarero, Mar 2018. From movement to grammar: Spanish verbal periphrases derived from verbs of movement. *Syntaxe & Sémantique* 19. 115–146. <https://doi.org/10.3917/e.ss.019.0115>
- Garachana Camarero, Mar 2019. De cuando “*ir a + INFINITIVO*” no se dirige al futuro. Construcciones gramaticales de pasado, sentido completivo y focalizador. Una aproximación desde la gramática de construcciones. *Revista Española de Lingüística* 49. 119–146. <http://dx.doi.org/10.31810/RSEL.49.6>
- Kornfeld, Laura Malena 2019. Expresión de la sorpresa, miratividad y gramaticalización de verbos inacusativos en español. *Borealis: An International Journal of Hispanic Linguistics* 8(2). 165–197. <https://doi.org/10.7557/1.8.2.4913>
- Masini, Francesca, Mattioli, Simone & Vecchi, Greta (2019). La costruzione “*prendere e V*” nell’italiano contemporaneo. In Moretti, Bruno, Kunz, Aline, Natale, Silvia & Krakenberger, Etna Rosa (eds.), *Le tendenze dell’italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana, Berna, 6-8 settembre 2018*, 115–137. Milano: Officinaventuno. <https://doi.org/10.17469/O2102SLI000007>
- Rodríguez Rosique, Susana. 2023. Time after time: The mirative construction <*no ir_{Fut.} a + infinitive*> in Spanish. In Rodríguez Rosique, Susana & Antolí Martínez, Jordi M. (eds.), *Verb and Context: The impact of shared knowledge on TAME categories*, 263–290. Amsterdam/Philadelphia: John Benjamins.
- Sornicola, Rosanna. 1976. *Vado a dire, vaiu a ddicu*: problema sintattico o problema semantico. *Lingua Nostra* XXXVII. 65–74.

La miratività nella perifrasi *venire a* + infinito nell'italiano contemporaneo

È ormai riconosciuto che i verbi itivi possano assumere valori mirativi, e diversi studi mostrano che nelle lingue romanze, fra cui l'italiano, la costruzione GO (TO) + VERB può esprimere sorpresa rispetto a un evento inatteso (Cruschina 2022; Dessì Schmid, Momm & Wiesinger 2025; Di Caro, Massaro & Molinari 2025; Ross 2016, 2024). Se il valore mirativo di *andare a* è largamente indagato, alcuni studi propongono un valore non dissimile per la perifrasi *venire a* + infinito (1), definito “allure extraordinaire” (Levie 2013; Silletti 2018).

(1) *E non mi vengano a dire che Paolo aveva fumato uno spinello* (CORIS)

In frasi di questo tipo, secondo Levie e Silletti, il verbo ventivo conferisce all'evento espresso dall'infinito un carattere “disturbante” o “anomalo” rispetto all'ordine previsto delle cose, in modo analogo a quanto descritto per *aller/venir* + infinito in francese (Damourette & Pichon 1911-1936; Bres & Labeau 2013), valore che è già stato accostato alla miratività (Celle & Lansari 2015).

Per i verbi itivi, Ross (2024) spiega la lettura mirativa come il risultato di una deviazione dal percorso atteso del soggetto verso un punto di arrivo, che indica una reazione del parlante a un evento inatteso. L'uso mirativo dei verbi ventivi, invece, risulta tipologicamente raro e i meccanismi attraverso cui tale valore emerge sono ancora poco studiati in letteratura. L'obiettivo di questo lavoro è indagare la possibilità di una lettura mirativa per *venire a* + infinito e se l'ipotesi di Ross (2024) possa essere estesa ai verbi ventivi.

Un'analisi preliminare delle occorrenze della perifrasi in due corpora, CORIS, corpus di italiano scritto, e KIParla, corpus di parlato, mostra un suo uso limitato: in CORIS, su 592 occorrenze di *venire a* + infinito solo 7 appaiono compatibili con l'espressione mirativa; in KIParla, su 172 occorrenze di *venire*, 8 casi sembrano mostrare lo stesso valore. La distribuzione suggerisce una scarsa diffusione nello scritto e una presenza leggermente maggiore nel parlato.

Nelle occorrenze emergono comunque alcuni tratti ricorrenti, sebbene non obbligatori, la cui distribuzione merita un approfondimento sistematico: si osserva innanzitutto una preferenza per frasi non assertive, come le interrogative e l'imperativo negativo; si riscontra inoltre l'uso prevalente di verbi di dire, spesso accompagnati da un dativo argomentale pronominale. Quest'ultimo aspetto risulta particolarmente rilevante: la presenza del dativo può rendere trasparente la direzione deittica del verbo *venire* verso il parlante (o verso un referente coinvolto). Questo ancoraggio verso il centro deittico permette di interpretare l'evento non solo come deviazione da un comportamento atteso, ma come una minaccia o un atto invasivo nei confronti dello spazio del parlante, il che potrebbe suggerire

una possibile estensione dell'ipotesi formulata da Ross (2024) circa l'emergere del valore mirativo nella perifrasi con *venire* + infinito.

Infine, in vari dei casi esaminati, si rileva che ciò che risulta contro le aspettative non è un evento già realizzato, bensì un atto potenziale attribuito ad altri. Il parlante presuppone la possibilità di un comportamento altrui percepito come sgradito o assurdo e mostra la sua contrarietà al fatto che ciò accada. Questo rafforzerebbe l'ipotesi che possa trattarsi di una forma di *counterexpectation* (Aikhenvald 2012) o di espressione di *disappointment* (Ross 2024), interpretazioni mirative documentate tipologicamente. In questo senso, *venire a* + infinito sembra veicolare valori coerenti con la miratività, seppur distinti da quelli tipici di *andare a*.

Riferimenti bibliografici

- Aikhenvald, Alexandra 2012. The essence of mirativity. *Linguistic Typology* 16. 435–485.
- Bres, Jacques & Labeau, Emmanuelle. 2013. *Allez donc sortir des sentiers battus!* La production de l'effet de sense *extraordinaire* par *aller* et *venir*. *Journal of French Language Studies* 23(2). 151–177. <https://doi.org/10.1017/S0959269512000154>
- Celle, Agnes & Lansari, Laure. 2015. On the mirative meaning of *aller* + infinitive compared with its equivalents in English. *Cahiers Chronos* 27. 289–305. https://doi.org/10.1163/9789004292772_015
- Cruschina, Silvio. 2022. Gone unexpectedly: Pseudo-coordination and the expression of surprise. In Giusti, Giuliana, Di Caro, Vincenzo Nicolò & Ross, Daniel (eds.), *Pseudo-Coordination and Multiple Agreement Constructions*, 129–148. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/la.274.05cru>
- Damourette, Jacques & Pichon, Edouard. 1911–1936. *Des Mots a la Pensée: Essai de Grammaire de la Langue Française*, tome V. Paris: D'Artrey.
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111386683-010>.
- Di Caro, Vincenzo Nicolò, Massaro, Angelapia & Molinari, Luca. 2025. Ethical dative, pseudo-expletive negation, and mirative GO in Italian. A preliminary analysis of stackable markers of mirativity. *Revue roumaine de linguistique* 70(1-2). 89–123.

- Levie, Yorkanda. 2013. *Valla a capire la gente!* La production de l'effet de sens extraordinaire par *andare* et *venire* en italien. *Studii de Știință și Cultură* 2. 21–36.
- Ross, Daniel. 2016. Going to Surprise: the grammaticalization of itive as mirative. In Woźny, Jacek (a cura di), *Online Proceedings of Cognitive Linguistics in Wrocław Web Conference*. Wrocław: Polish Cognitive Linguistics Association & University of Wrocław.
- Ross, Daniel. 2024. Going to surprise: The grammaticalization of itive as mirative. In Pfadenhauer, Katrin & Wiesinger, Evelyn (a cura di), *Romance motion verbs in language change: Grammar, lexicon, discourse*, 311–346. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783111248141-013>
- Silletti, Alida. 2018. Les périphrases en aller et venir de l'italien contemporain: grammaticalisation et effets de sens. *Syntaxe & Sémantique* 19. 87–114

“E non ti arriva uno che deve passare?”: costruzioni dialogiche dell’inatteso in italiano

Il contributo intende analizzare una famiglia di costruzioni dell’italiano, esemplificate in (1)-(4):

- (1) *Il tabacco, base ed alimento del regno di Pietro Cova, rischia di sparire del tutto [...] Orbene, cosa ti combina, nel momento del pericolo, quell’ingrato di De Martino? Abbandona la nave che affonda, lascia nei guai il povero Cova, si mette in testa di salvarsi da solo* (Lino Jannuzzi, *Il duello Cova-Trabucchi*, *L’Espresso*, 1965; corpus Diacoris)
- (2) *Ero sul rivale del fiume con la mia amicaappena kavate le mutande (eravamo in un posto sperduto) non ti arriva uno ke deve passare?!? Ma porc.....* (Corpus ItTenTen 2020)
- (3) *Al via, il mio favorito prende subito la testa e vola. Al primo giro era ancora in testa, fremevo e già sognavo la vincita, ma quel figlio di un cane non mi va a rompere l’andatura a cinquanta metri dal traguardo e mi viene squalificato? Mondo boia.* (Corpus Coris 19802000; Sez. Narrativa)
- (4) *Mentre stasera leggevo su internet le cronache dal Senato, il digitalissimo fidanzatodi non ti va a sfruculiare il satellite, scoprendo Superpippa Channel? Un nuovo, meraviglioso canale erotico in chiaro* (Corpus ItWac)

Queste costruzioni ricorrono in narrazioni monologiche e mettono in scena un dialogo simulato. Il tipo in (1) realizza un’autentica coppia adiacente: domanda a un interlocutore fittizio (*cosa ti combina*) + risposta immediata (*abbandona la nave*), entrambe prodotte dallo stesso parlante. In (2), (3) e (4) l’evento inatteso (*l’arrivo di uno che deve passare, il cavallo che rompe l’andatura, ecc.*) è invece inserito direttamente in una domanda polare retorica negativa, un tipo di domanda noncanonica (nel senso di Trotzke 2024, 36) che, oltre a esprimere *bias* del parlante, veicola sorpresa (cfr. anche Celle et al. 2017, 223–225).

Seguendo la recente tipologia in Dessì Schmid, Momma & Wiesinger (2025), queste costruzioni codificano il terzo tipo di scenario mirativo, i.e. una disattesa delle aspettative *hearer-oriented*, in cui l’informazione è presentata come esagerata e potenzialmente inattesa per l’interlocutore. Ne è indizio anche l’assenza di *recency restriction*, dato che gli eventi narrati riportano una sorpresa passata. Centrale è la domanda retorica; possono tuttavia contribuire alla lettura mirativa, pur non obbligatori, altri elementi: a) il dativo “conversazionale” di seconda persona (*entro nel bar e sai chi ti vedo?*; Masini 2012), ma anche di prima persona, come in (3); e b) la perifrasi *andare* + infinito (*era sopravvissuta agli inverni rigidi e mi va a morire d’estate*; Dessì Schmid, Momma &

Wiesinger 2025; Di Caro et al. 2025), che profila l'evento come eccedente uno spazio di aspettative predefinite.

Obiettivo del lavoro è descrivere sistematicamente la struttura di queste costruzioni e il contributo funzionale dei loro elementi costitutivi. L'analisi, di tipo qualitativo, si basa su dati tratti da corpora diacronici e sincronici dell'italiano scritto e parlato (DiaCORIS, CODIT, ItWaC, itTenTen, KIParla), interrogati mediante stringhe volte a estrarre domande retoriche positive e negative, seguite o meno da clitici di prima/seconda persona (*mi/ti*) o da forme di *andare* + *a* + infinito.

I dati diacronici mostrano anzitutto che il sottotipo in (1) è storicamente anteriore rispetto a quello in (2)-(4), che presenta un grado maggiore di cristallizzazione formale. Questo sviluppo può essere interpretato come un processo di costruzionalizzazione (Traugott & Trousdale 2013), innescato dalla progressiva riduzione di una sequenza dialogica esplicita (domanda + risposta) a una forma monologica compatta che incorpora la dialogicità nella sola interrogativa retorica. Si tratta, più precisamente, di un caso riconducibile a quei processi descritti come *monologizzazione* o *grammaticalizzazione polifonica* (Calaresu 2018; Sansò 2022, 2025 per fenomeni analoghi nei marcatori di topic): in queste costruzioni la miratività, intesa come gestione delle aspettative, viene costruita attraverso una base dialogica simulata, ormai integrata nella grammatica della narrazione monologica, che consente al parlante di negoziare e (ri)strutturare il *Common Ground*.

Riferimenti bibliografici

- Calaresu, Emilia 2018. Grammaticalizzazioni polifoniche o “verticali” e sintassi dialogica. Dagli enuncianti-eco ai temi sospesi: l'infinito anteposto in strutture del tipo “mangiare, mangio”. In Greco, Paolo, Vecchia, Cesarina & Sornicola, Rosanna (a cura di), *Strutture e dinamismi della variazione e del cambiamento linguistico, Atti del convegno DIA III, Napoli, 24-27 novembre 2014*, 505-521. Napoli: Giannini.
- Celle, Agnès, Jugnet, Anne, Lansari, Laure & L'Hôte, Emilie. 2017. Expressing and describing surprise. In Celle, Agnès & Lansari, Laure (a cura di), *Expressing and describing surprise*, 215–244. Amsterdam: John Benjamins. <https://doi.org/10.1075/bct.92>
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111386683-010>.

- Di Caro, Vincenzo Nicolò, Massaro, Angelapia & Molinari, Luca. 2025. Ethical dative, pseudo-expletive negation, and mirative GO in Italian. A preliminary analysis of stackable markers of mirativity. *Revue roumaine de linguistique* 70(1-2). 89–123.
- Masini, Francesca 2012. Costruzioni verbo-pronominali “intensive” in italiano. In Bambini, Valentina, Ricci Irene & Bertinetto, Pier Massimo (a cura di), *Linguaggio e cervello – Semantica. Language and the brain – Semantics, Atti del XLII congresso internazionale di Studi della Società di Linguistica Italiana (SLI), Pisa, 25-27 settembre 2008*, 1–22. Roma: Bulzoni.
- Sansò, Andrea 2022. Discourse markers from processes of monologization: Two case studies from Italian. In Voghera, Miriam (a cura di), *From speaking to grammar*, 201-225. Bern: Peter Lang.
- Sansò Andrea 2025. From intersubjective to textual through monologization. A case study from Italian. In Scivoletto, Giulio & Takamura, Ryo (eds.), *Semantic-pragmatic change from intersubjective to textual meanings*, 221–243. Amsterdam: John Benjamins
<https://doi.org/10.1075/pbns.353.08san>
- Traugott, E.C. & G. Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: OUP.
- Trotzke, Andreas. 2024. *Non-canonical questions*. Oxford: OUP.

Non io che: uno studio della miratività tra esperienti e negazione

Questo contributo indaga la struttura *non X (che) V* nell'italiano digitato, inquadrandola tra quelle costruzioni negative che possono esprimere sorpresa in italiano. Tale struttura, che deriva dall'inglese *not X V*, diffusosi online dal 2018 (Fiorentini 2026, 59), serve principalmente a «svelare in prima persona qualcosa che [chi scrive] ha fatto, detto o pensato: una figuraccia, un'abitudine o pensiero ricorrente che viene “confessato” in maniera autoironica» (Albertini 2023). In questo senso, può mostrare anche usi mirativi (1):

(1) *non mio padre che dopo vent'anni si ricorda il nome di Zac Efron (unexpected)* [Instagram]

La miratività è stata definita come codifica linguistica dell'inaspettatezza o della violazione delle aspettative (Aikhenvald 2012; Cruschina & Bianchi 2021; DeLancey 1997): sembra quindi collocarsi all'intersezione tra evidenzialità e marcatura dell'atteggiamento epistemico. Come è noto, le lingue romanze non dispongono di marche grammaticali apposite per questo significato pragmatico (Dessi Schmid, Momma & Wiesinger 2025), ma sfruttano strategie alternative (cfr. de Nazaré Ferreira-Silva 2011; Rodríguez Rosique 2024; Squartini 2018). In italiano, troviamo perifrasi verbali (Dessi Schmid, Momma & Wiesinger 2025; Masini et al. 2019; Pisciotta 2025), marche lessicali (D'Achille & Thornton 2020; Mauri & Russo 2025), strutture sintattiche marcate (Cruschina 2020).

Inoltre, è stato osservato un legame tra alcune costruzioni negative e miratività (Cifuentes Honrubia 2023; Di Caro, Massaro & Molinari 2025; García Negroni 2022; Sawada & Sawada 2022; Voicu 2024;) o espressione della sorpresa (Godard & Marandin 2006, 190). Tra le costruzioni di questo tipo abbiamo in italiano p. es. *no, no vabbè* (Fiorentini 2024), *non è vero, non ci credo, non è possibile* (cfr. Vincent, Laforest & Nicole 1995), interiezioni come *incredibile* (cfr. García Negroni & Libenson 2022), o anche marche come *mica* (Magistro 2022; Mari 2025), oltre a strutture come le *Surprise Negation Sentences* (Greco & Moro 2015; cfr. anche Di Caro, Massaro & Molinari 2025) (2).

(2) *E non mi è scesa dal treno Maria?* (Greco 2019, 776)

La nostra indagine propone, da un lato, delle riflessioni teoriche, basate sulla letteratura esistente, sullo statuto della negazione nell'effetto di sorpresa; dall'altro, un approfondimento degli usi mirativi della costruzione *non X (che) V* sulla base di dati di italiano digitato tratti da social network (3-4):

- (3) *non io che in realtà vivo vicino al vero palazzo dove si svolge tutto quel dramma* [Reddit/anime]
- (4) *non io facendo le prove per quando finalmente smetterò di pensare al mio m@l3ss3r3* [TikTok]

A questi dati, di taglio qualitativo, affiancheremo i risultati di un questionario sperimentale mirato, che non si concentri tanto sull'identificazione, quanto sul *grado* di sorpresa (Filippi-Deswelle 2015) evocato da diversi tipi di soggetto in posizione *X*. Il questionario sarà somministrato a circa 50 parlanti, cui verrà chiesto di valutare il livello di sorpresa percepito, indicarne l'esperienza e motivarne l'interpretazione.

L'analisi indagherà così due aspetti: (i) quale tipo di relazione si instaura tra negazione e miratività; (ii) in che misura il soggetto (s)favorisce l'interpretazione della sorpresa, contribuendo alla prototipicità dell'effetto mirativo.

In definitiva, questo lavoro mira a contribuire al dibattito sulla miratività, approfondendo alcune caratteristiche specifiche di costruzioni mirative, e che emergono anche in marche esogene e innovative, come quella qui approfondita.

Riferimenti bibliografici

- Albertini, Agnese. 2023. Non io che, il significato del meme social. *Cultweb*. (13 novembre 2023) <https://cultweb.it/web/non-io-che-il-significato-del-meme-social/>
- Aikhenvald, Alexandra Y. 2012. The essence of mirativity. *Linguistic Typology* 16(3). 435–485.
- Cifuentes Honrubia, José Luis. 2023. (No) faltaba/faltaría más: Covert negation, mirativity and epistemic modality. In Rodríguez Rosique, Susana & Antolí Martínez, Jordi M. (a cura di), *Verb and context: The impact of shared knowledge on TAME categories*, 99–130. Amsterdam: John Benjamins.
- Cruschina, Silvio & Bianchi, Valentina. 2021. A surprising association and an unexpected move: Mirative implicatures at the syntax-semantics interface. In Trotzke, Andreas & Villalba, Xavier (a cura di), *Expressive meaning across linguistic levels and frameworks*, 86–107. Oxford: Oxford University Press.
- Cruschina, Silvio. 2020. Quello che le parole non dicono. L'anteposizione focale e l'espressione della sorpresa. In Garavelli, Enrico, Monticelli, Daniele, Ploom, Ülar & Suomela-Härmä, Elina (a cura di), *Italianistica 2.0. Tradizione e innovazione: Atti del XII Congresso degli Italianisti della Scandinavia*, Helsinki-Tallinn, 13-14 giugno 2019, 51–66. Helsinki: Uusfilologinen yhdistys.

- D'Achille, Paolo, Thornton, Anna Maria. 2020. Storia di un imperativo divenuto interiezione: *ammazza!*. In Faraoni, Vincenzo & Loporcaro, Michele (a cura di), *'E parole de Roma. Studi di etimologia e lessicologia romanesche*, 163-194. Berlin/Boston: De Gruyter.
- De Nazaré Ferreira-Silva, Marília. 2011. A mirative construction description: observations about an occurrence in Brazilian Portuguese spoken in Belém, *Acta Scientiarum. Language and Culture* 33(2). 207-210. <https://doi.org/10.4025/actascilangcult.v33i2.12195>
- DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33-52.
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111386683-010>.
- Di Caro, Vincenzo Nicolò, Massaro, Angelapia & Molinari, Luca. 2025. Ethical dative, pseudo-expletive negation, and mirative GO in Italian. A preliminary analysis of stackable markers of mirativity. *Revue roumaine de linguistique* 70(1-2). 89–123.
- Filippi-Deswelle, Catherine. 2015. Les degrés de la surprise. In Depraz, Natalie & Serban, Claudia (a cura di), *La Surprise, à l'épreuve des langues*, 1–13. Paris: Hermann.
- Fiorentini, Ilaria. 2024. *No vabbé, anche no*. No come segnale discorsivo in italiano contemporaneo. *Quaderni Borromaici* 11. 39–56.
- Fiorentini, Ilaria. 2026. *L'italiano dei social network*. Roma: Carocci.
- García Negroni, María Marta & Libenson, Manuel. 2022. On the dialogic frames of mirative enunciations. The Argentine Spanish discourse marker *mirá* and the expression of surprise. *Pragmatics* 32(3). 329–353. <https://doi.org/10.1075/prag.20015.gar>
- Magistro, Giuseppe. 2022. *Mica* preposing as focus fronting. *Glossa: a journal of general linguistics* 7(1). 1–39. <https://doi.org/10.16995/glossa.8494>
- Greco, Matteo. 2019. On the syntax of surprise negation sentences: A case study on expletive negation. *Natural Language & Linguistic Theory* 38. 775–825. <https://doi.org/10.1007/s11049-019-09459-6>
- Greco, Matteo & Moro, Andrea. 2015. Surprise negation (Sneg) sentences. Poster presentato alla *Göttingen Summer School on Negation*. University of Göttingen.
- Godard, Danièle & Jean-Marie Marandin. 2006. Reinforcing negation: The case of Italian. In Stefan Müller (a cura di), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, Varna, 24-27 luglio, 2006, 174–194. Stanford: CSLI Publications.

- Mari, Alda. 2025. Disconfirmational MICA in surprise questions (and some notes on concessivity). In Rodriguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 39-61. Berlin/Boston: De Gruyter Mouton.
- Masini, Francesca, Mattiola, Simone & Vecchi, Greta. 2019. La costruzione “prendere e V” nell’italiano contemporaneo. In Moretti, Bruno, Kunz, Aline, Natale, Silvia & Krakenberger, Etna (a cura di), *Le tendenze dell’italiano contemporaneo rivisitate*, 137–152. Milano: OfficinaVentuno.
- Mauri, Caterina & Russo, Antonia. 2025. The mirative values of Italian *altro che*. In Rodriguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 143–178. Berlin/Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111386683-007>
- Rodríguez Rosique, Susana, 2024. What is a morphological future doing in a *si*-clause? Traces of mirativity in Spanish, *Lingua* 304. 103710. <https://doi.org/10.1016/j.lingua.2024.103710>
- Pisciotta, Flavio. 2025. Ai confini tra evidenzialità e miratività: usi d’apparenza dei verbi sintagmatici di movimento con fuori. In Cerruti, Massimo, Onesti, Cristina Regis, Riccardo, Ricca, Davide (a cura di), *Continuo e discreto nelle scienze del linguaggio. Atti del LVI Congresso Internazionale di Studi della Società di Linguistica Italiana, Torino, 14-16 settembre 2023*, 355–382. Milano: OfficinaVentuno
- Sawada, Osamu & Sawada, Jun. 2022. The mirative demonstrative in Japanese. *Journal of East Asian Linguistics* 34. 1–57. <https://doi.org/10.1007/s10831-024-09292-y>
- Squartini, Mario. 2018. Mirative extensions in Romance: evidential or epistemic?. In Guentchéva, Zlatka (a cura di), *Epistemic Modalities and Evidentiality in Cross-Linguistic Perspective*, 196–216. Berlin/Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110572261-009>
- Vincent, Diane, Laforest, Marty & Nicole, Julie. 1995. L’étonnement et l’étonnant dans le discours oral spontané. *L’exclamation - Faits de langues* 6. 121–130.
- Voicu, Roxana. 2024. Contraintes discursives sur l’émergence du sens procédural de la construction mirative : [il n’est pas jusqu’à SN + relative négative]. In Neveu, Franck, Prévost, Sophie, Montébran, Aurore, Steuckardt, Agnès & G. Bergounioux, Gabriel (a cura di) *Congrès Mondial de Linguistique Française - CMLF 2024*, Lausanne, 1-5 luglio 2024. *SHS Web of Conferences* 191. 12006. <https://doi.org/10.1051/shsconf/202419112006>

Risposte polari e miratività: uno studio sull'italiano

Questo contributo esamina un insieme di strategie responsive dell'italiano che combinano marcatura della polarità e miratività. La letteratura sulle *polarity particles* ha analizzato le risposte in termini di conferma e inversione di polarità (*reversal*, Farkas & Bruce 2010; Roelofsen & Farkas 2015; per l'italiano cf. Servidio, Bocci e Bianchi 2018; Andorno & Crocco 2018; Andorno & Rosi 2018). Parallelamente, la miratività è stata descritta come categoria funzionale relativa all'espressione della sorpresa (Aikhenvald 2012; Dessì-Schmidt et al. 2025; Squartini 2017;). Rimane tuttavia poco esplorata l'intersezione tra inversione di polarità e codifica mirativa (cf. Martinez 2023).

L'obiettivo è verificare se le forme considerate costituiscano una classe funzionale coerente definibile come “risposte polari mirative”, e quali proprietà ne delimitino il dominio rispetto a categorie contigue quali controaspettatività ed esclamatività (cf. Unger 2019). L'analisi viene condotta sia in prospettiva sincronica, con dati dal corpus KIParla (Mauri et al. 2019) e da ItTenTen20, che diacronica, con dati dal CODIT (Micheli 2021).

La Tabella 1 organizza le strategie in esame, che sono state analizzate secondo tre parametri:

- (i) FAMIGLIA DI COSTRUZIONI: Altroché, Interrogative in *-wh*, Attribuzione di una condizione/opinione all'interlocutore, Ma + imperativo, Menzione esplicita dell'atto linguistico precedente;
- (ii) POLARITÀ DELLA RISPOSTA E DELL'ASPETTATIVA: positiva o negativa;
- (iii) TIPO DI MIRATIVITÀ: può essere di *primo livello* (sorpresa per uno stato di cose diverso da quanto atteso) o di *secondo livello* (sorpresa per un'aspettativa o credenza rispetto a un determinato stato di cose, Mauri & Russo 2025; cf. anche Cruschina 2019).

Queste strategie si distinguono dalle risposte polari semplici perché esprimono una polarità opposta a quella resa saliente o presupposta nel turno precedente (*reversal*) in combinazione con la codifica della sorpresa.

- *altroché!*, *come no?!*, *eccome!*, *e me lo chiedi!?* realizzano affermazioni con miratività di secondo livello, a fronte di aspettative negative;
- *macché!*, *(ma) quando mai!?* codificano tipicamente un rifiuto con miratività di primo livello, a fronte di un'aspettativa positiva;

- Le strategie che codificano *attribuzione di una condizione/opinione all'interlocutore* possono comunicare sia un rifiuto che conferma con miratività di secondo livello, a fronte di un'aspettativa inattesa.

L'indagine diacronica mostra che l'attuale convergenza funzionale deriva da percorsi evolutivi distinti. *Altroché* sviluppa la funzione mirativa da usi correttivi; *macché* la acquisisce da *ma che* in interrogative metalinguistiche di riparazione. Le restanti strategie, meno costruzionalizzate, sviluppano il valore mirativo parallelamente al loro uso originario: per *eccome!* e *come no!?* esso si innesta sulla funzione di messa in discussione di quanto affermato in precedenza; per *(ma) scherzi?/secondo te?* e forme analoghe, sull'attribuzione di un'opinione o condizione all'interlocutore. *Ma va' là* deriva da un "movimento metaforico discorsivo di divergenza" (Molinelli 2019, 241); *e me lo chiedi?* codifica esplicitamente la sorpresa per l'atto interrogativo in sé.

Il contributo del lavoro è duplice: propone una delimitazione empiricamente fondata della nozione di risposta polare mirativa in italiano, collocandola nello spazio funzionale delle risposte di inversione di polarità, e mostra che nei contesti responsivi la miratività spesso comporta la messa in discussione di uno stato epistemico attribuito nel *common ground*, permettendo di distinguerla più chiaramente da una più generica controaspettatività e dall'esclamatività.

Famiglia	Forma	Esempio	Funzione
Altroché	• <i>Altroché!</i>	(1) "Allo spirito ci pensa mai?" "Altroché. Ma per quello leggo Platone." (ItTenTen20)	Si (<i>sono sorpreso che tu ne dubiti!</i>) → miratività di 2° livello
Interrogative in -wh	• <i>Macché!</i> , • <i>(Ma) quando mai?</i>	(2) KP017: ah ci vai a piedi da qua dal ca~ dai campi? PKP013: macché PKP013: prendiamo la macchina [...] (KPN004)	No (<i>sono sorpreso, perché mi aspettavo – e ti aspettavi – di sì!</i>) → miratività di 1° livello
	• <i>Come no!?</i> • <i>Eccome!</i>	(3) "E dei capelli non ti disse nulla?" "Come no. E non solo la prima volta." (ItTenTen20)	Si (<i>sono sorpreso che tu ne dubiti!</i>) → miratività di 2° livello
Attribuzione di condizione/opinione all'interlocutore	• <i>(Ma) scherzi!?</i> • <i>(Ma) secondo te?</i> • <i>(Ma) ti pare?</i> • <i>(Ma) sei matta?</i>	(4) PKP021: è buono stefano? PKP019: beh ma scherzi? PKP020: eh col basilico poi (KPN006) (5) "[...] Dimmi che non hai perso il ristorante e che non ti stai mettendo a fare il pescivendolo!!" [...] " Ma ti pare? .. Il ristorante è nostro! [...]" (ItTenTen20)	Si (<i>sono sorpreso che tu ne dubiti!</i>) → miratività di 2° livello No (<i>sono sorpreso del fatto che ti aspettavi di sì!</i>) → miratività di 2° livello
Ma + imperativo	• <i>Ma va' (là)</i>	(6) g: ma belin ste donne ti usano e basta? girimi: ma va là son il loro confidente (ItTenTen20)	No (<i>sono sorpreso del fatto che ti aspettavi di sì!</i>) → miratività di 2° livello
Menzione esplicita dell'atto linguistico precedente	• <i>E me lo chiedi?</i>	(7) "Volete assaggiarne una fetta?" 3 commenti: " E me lo chiedi? Certo che assaggerei volentieri (ItTenTen20)	Si (<i>sono sorpreso che tu ne dubiti!</i>) → miratività di 2° livello

Tabella 1: Risposte polari mirative: famiglie, forme, esempi e funzioni.

Riferimenti bibliografici

- Andorno, Cecilia & Crocco, Claudia 2018. A contribution from echo replies: In search for polarity contrast marking in Italian. In Dimroth, Christine & Sudhoff, Stefan (a cura di), *The Grammatical Realization of Polarity Contrast: Theoretical, empirical, and typological approaches*, 225–288. Amsterdam: John Benjamins Publishing Company.
<https://dx.doi.org/10.1075/la.249.09and>
- Andorno, Cecilia & Rosi, Fabiana. 2018. Italian sì/no in replies between polarity and agreement: A first inquiry on corpus data. In Gille, Johan & Norén, Coco (a cura di), *Self and other in dialogue: Romance studies on discourse and interaction*, 37–58. Bern: Peter Lang.
<https://doi.org/10.3726/978-3-653-06605-0>
- Aikhenvald, Y. Alexandra. 2012. The essence of mirativity. *Linguistic Typology* 16(3). 435–485.
- Bazzanella, Carla. 1995. *I segnali discorsivi*. In Renzi, Lorenzo, Salvi, Giampaolo & Cardinaletti, Anna, *Grande grammatica italiana di consultazione*. Bologna: Il Mulino.
- Bernini, Giuliano. 1995. *Le profrasi*. In Renzi, Lorenzo, Salvi, Giampaolo & Cardinaletti, Anna, *Grande grammatica italiana di consultazione*. Bologna: Il Mulino.
- Cruschina, Silvio. 2019. L’anteposizione focale. *Neuphilologische Mitteilungen* 120(2). 247–268.
<https://www.jstor.org/stable/26950770>
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (a cura di), *Expressing Surprise at the Crossroads: Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783111386683-010>.
- Farkas, Donka F. & Bruce, Kim B. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118. <https://doi.org/10.1093/jos/ffp010>
- Martinez, Randi. 2023. Mirativity in English response particles: An analysis from the syntax/semantics interface. *Yale Working Papers in Grammatical Diversity* 5(3). 1–25.
- Mauri, Caterina, Ballarè, Silvia, Gorla, Eugenio, Cerruti, Massimo & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Bernardi, Raffaella, Navigli, Roberto & Semeraro, Giovanni (a cura di), *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), Bari, 13-15 novembre. 2019*, 243–249. Aachen: CEUR Workshop Proceedings. <https://aclanthology.org/2019.clicit-1.37/>
- Mauri, Caterina & Antonia Russo. 2025. The mirative values of Italian *altro che*. In Rodríguez Rosique, Susana (a cura di), *Expressing surprise at the crossroads: Mirativity, exclamativity*

- and cooptation in Romance Languages*, 143–178. Berlin/Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9783111386683-007>
- Micheli, Silvia M. 2021. *CODIT: Corpus diacronico dell'italiano*. Ústav Českého národního korpusu FF UK, Prague. <http://www.korpus.cz/>.
- Molinelli, Piera, 2019. Congiunzioni e segnali funzionali: la produttività di ‘ma’ ed ‘e’ in diadi pragmatiche in italiano. *Storie e Linguaggi*, 229–254.
- Roelofsen, Floris & Farkas, Donka F. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414.
<https://doi.org/10.1353/lan.2015.0017>
- Servidio, Emilio; Bocci, Giuliano & Bianchi, Valentina. 2018. (Dis)agreement, polarity, and focus: Answering negative polar questions in Italian. *Glossa: a journal of general linguistics* 3(1). 1–28. <https://doi.org/10.5334/gjgl.360>
- Squartini, Mario. 2017. Italian non-canonical negations as modal particles: Information state, polarity and mirativity. In Fedriani, Chiara e Sansò, Andrea (a cura di), *Pragmatic Markers, Discourse Markers and Modal Particles: New Perspectives*, 203– 228. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.186.08squ>
- Unger, Christoph. 2019. Exclamatives, exclamations, miratives and speaker’s meaning. *International Review of Pragmatics* 11(2). 272–300.

Esprimere la sorpresa per un pubblico di ascoltatori: l'esempio dell'incredulità ritualizzata nelle interazioni radiofoniche in italiano

Prendendo le mosse dai lavori goffmaniani sui “gridi di reazione” (*response cries*, Goffman 1978, 1981), diversi sono stati gli studiosi che, da una prospettiva di analisi conversazionale e di linguistica interazionale (Couper-Kuhlen & Selting 2018; Sidnell & Stivers 2013), si sono interrogati su come i parlanti manifestino la sorpresa, in particolare rispetto a informazioni fornite dall'interlocutore che risultano inattese e non conformi alle conoscenze e alle aspettative (personali, culturali) in un dato contesto interazionale (Wilkinson & Kitzinger 2006, 153).

Nel quadro di un crescente interesse per l'espressione delle emozioni (Robles & Weatherall 2021; Sorjonen & Peräkylä 2012), le manifestazioni della sorpresa in scambi dialogici – con forme quali gridi di reazione (cfr. inglese *wow, gosh, oh! oh my god*) ed espressioni di “incredulità ritualizzata” (*really? what? you're kidding*, ripetizione totale/parziale del turno precedente, “no” enfatico e turni di pseudo-riparazione con soluzione candidata, Heritage 1984, 339; Rossi 2020; Wilkinson & Kitzinger 2006) – sono state esaminate, in un'ottica emica e sequenziale, indagando il rapporto del turno del parlante che esprime sorpresa con il turno/i turni precedenti dell'interlocutore – in cui l'informazione fornita può essere o meno formulata come possibile oggetto di sorpresa –, come pure con i turni successivi dello scambio, che vedono la continuazione o la chiusura della sequenza a seconda del tipo di espressione della sorpresa utilizzato (cfr. Golato 2010; Golato & Betz 2008; Koivisto 2016; Thompson, Fox & Couper-Kuhlen 2015). Studi specifici sono stati inoltre dedicati agli aspetti prosodici e incorporati dell'espressione della sorpresa (Dix & Gross 2023; Goodwin 2000; Heath et al. 2012; Local 1996; Selting 1996; Persson 2020).

A partire da tali ricerche, con il presente contributo si offrono spunti per lo studio della miratività (Aikhenvald 2012; D'Achille & Thornton 2020; Delancey 1997; Rodriguez Rosique 2025), concentrandosi sull'italiano e, nello specifico, sull'espressione dell'“incredulità ritualizzata”. Lo studio, di taglio conversazionale, si basa su una collezione di interazioni telefoniche tratte dal programma radiofonico “Il Ruggito del Coniglio” (RAIRadio3, 50 telefonate, ca. 120'), che vertono sulla narrazione di vicende quotidiane in chiave umoristica. L'analisi esamina le manifestazioni di incredulità ritualizzata da parte dei conduttori (ad es. produzione di “eh?” e “no” enfatici, ripetizione totale/parziale del precedente turno del chiamante) considerandone il contesto sequenziale precedente e seguente e l'interazione con ulteriori elementi di espressione della sorpresa. Si osserva così come, sistematicamente, i conduttori da un lato replichino con espressioni di incredulità ritualizzata a turni

dei chiamanti che preannunciano il carattere non ordinario dell'evento narrato, e, dall'altro, sollecitano il parlante ad 'andare al punto', esprimendo incredulità quando questo viene raggiunto e producendo poi ulteriori espressioni di sorpresa.

Ragionando sul carattere co-costruito della manifestazione della sorpresa, l'analisi evidenzia come questa rappresenti al contempo una risorsa interazionale per la buona riuscita delle narrazioni – che si svolgono all'interno di scambi comunicativi radiofonici rivolti a un pubblico di ascoltatori a casa – inserendosi tra le strategie messe in campo dai conduttori per assolvere alle funzioni di intrattenimento, vicinanza e familiarità tipiche della comunicazione radiofonica (Atzori 2017; Hutchby 2006; Maraschio 2011; Rubino 2016; Sobrero 1997).

Riferimenti bibliografici

Aikhenvald, Y. Alexandra. 2012. The essence of mirativity. *Linguistic Typology* 16(3). 435–485.

Atzori, Enrica (2017), *La lingua della radio in onda e in rete*. Firenze: Cesati.

Couper-Kuhlen, Elizabeth & Selting, Margret 2018. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

D'Achille, Paolo & Thornton, Anna Maria 2020. La storia di un imperativo diventato interiezione: *ammazza!*. In Faraoni, Vincenzo & Loporcaro, Michele (a cura di), *'E parole de Roma. Studi di etimologia e lessicologia romanesche*, 163–194. Berlin/Boston: De Gruyter.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33–52.

Dix, Carolin & Groß, Alexandra 2024. Surprise about news or just receiving information? Moving and holding both eyebrows in co-present interaction. *Social Interaction: Video-Based Studies of Human Sociality* 6(3). 127.

Goffman, Erving 1978. Response cries. *Language* 54(4). 787–851.

Goffman, Erving 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press/Blackwell.

Golato, Andrea 2010. Marking understanding versus receipting information in talk: *Achso* and *ach* in German interaction. *Discourse Studies* 12(2). 147–176.
<https://doi.org/10.1177/1461445609356497>

- Golato, Andrea & Betz, Emma 2008. German *ach* and *achso* in repair uptake: Resources to sustain or remove epistemic asymmetry. *Zeitschrift für Sprachwissenschaft* 27. 7–37. <https://doi.org/10.1515/ZFSW.2008.002>
- Good, Jeffrey S. & Beach, Wayne A. 2005. Opening up gift-openings: Birthday parties as situated activity systems. *Texts* 25(5), 565–593. <https://doi.org/10.1515/text.2005.25.5.565>
- Goodwin, Charles 2000. Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics* 32, 1489–1522. [https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X)
- Heath, Christian, vom Lehn, Dirk, Cleverly, Jason & Luff, Paul 2012. Revealing surprise: The local ecology and the transposition of action. In Perakyla, Anssi & Sorjonen, Marja-Leena (a cura di), *Emotion in Interaction*, 212–234. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730735.003.0010>
- Heritage, John 1984. A change-of-state token and aspects of its sequential placement. In Atkinson, J. Maxwell & Heritage, John (a cura di), *Structures of Social Action. Studies in Conversation Analysis*, 299–345. Cambridge: Cambridge University Press.
- Hutchby Ian 2006. *Media Talk: Conversation Analysis and the Study of Broadcasting*. Maidenhead: Open University Press.
- Koivisto, Aino 2016. Receipting information as newsworthy vs. responding to redirection: Finnish news particles *aijaa* and *aha(a)*. *Journal of Pragmatics* 104. 163–179.
- Local, John 1996. Conversational Phonetics: Some Aspects of News Receipts in Everyday Talk. In Couper-Kuhlen, Elizabeth & Selting, Margret (a cura di), *Prosody in Conversation: Interactional Studies*, 177–230. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511597862.007>
- Maraschio, Nicoletta 2011. La lingua della radio. In Simone, Raffaele (a cura di), *Enciclopedia dell'Italiano*, 1217–1221. Roma: Istituto della Enciclopedia Italiana Treccani.
- Persson, Rasmus 2020. Prosody and grammar of other-repetitions in French: The interplay of position and composition. *Language in Society* 49. 585–618. <https://doi.org/10.1017/S0047404520000068>
- Robles, Jessica S. & Weatherall, Ann (a cura di). 2021. *How emotions are made in talk*. Amsterdam: John Benjamins Publishing Company.

- Rodriguez Rosique, Susana 2025. *Expressing Surprise at the Crossroads: Mirativity, Exclamativity and Cooptation in Romance Languages*. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783111386683>
- Rossi, Giovanni 2020. Other-repetition in conversation across languages: Bringing prosody into pragmatic typology. *Language in Society* 49(4). 495–520.
<https://doi.org/10.1017/S0047404520000251>
- Rubino, Antonia. 2016. Constructing pseudo-intimacy in an Italo-Australian phone-in radio program. *Journal of Pragmatics* 103. 33–48. <https://doi.org/10.1016/j.pragma.2016.07.009>
- Selting, Margret. 1996. Prosody As an Activity Type Distinctive Cue in Conversation: The Case of So-Called ‘Astonished’ Questions in Repair Initiation. In Couper Kuhlen, Elizabeth & Selting, Margret (a cura di). *Prosody in Conversation*, 231–270. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511597862.008>
- Sidnell, Jack & Stivers, Tanya (a cura di) 2013. *The Handbook of Conversation Analysis*. Wiley-Blackwell.
- Sobrero, Alberto A. 1997. Le telefonate in diretta: struttura, scelte linguistiche e organizzazione conversazionale. In AA.VV., *Gli italiani trasmessi. La radio*, 505–524. Firenze: Accademia della Crusca.
- Perakyla, Anssi & Sorjonen, Marja-Leena (a cura di) 2012. *Emotion in interaction*. Oxford: Oxford University Press.
- Thompson, Sandra A., Fox, Barbara A. & Couper-Kuhlen, Elizabeth. 2015. *Grammar in everyday talk: Responsive actions in conversation*, Cambridge: Cambridge University Press.
- Wilkinson, Sue & Kitzinger, Celia. 2006. Surprise as an interactional achievement: reaction tokens in conversation. *Social Psychology Quarterly* 69(2), 150–182.

L'espressione dell'inatteso nel portoghese europeo contemporaneo: strategie mirative nel parlato radiofonico

Il contributo indaga l'espressione dell'inatteso nel portoghese europeo contemporaneo, concentrandosi sulle strategie linguistiche attraverso cui i parlanti segnalano sorpresa, controaspettativa o improvvisa acquisizione di conoscenza. Si assume come punto di partenza la definizione funzionale di miratività proposta da DeLancey (1997) e ripresa da Aikhenvald (2012), trattando tuttavia la miratività non come categoria grammaticalizzata, bensì come effetto pragmatico emergente all'interfaccia tra struttura informativa e marcatori discorsivi.

L'analisi si basa su un corpus in costruzione di interviste radiofoniche e talk di intrattenimento nel portoghese europeo. Si tratta di contesti dialogici ad alta spontaneità, nei quali la sorpresa viene frequentemente negoziata nel corso dell'interazione (Dessi Schmid, Momma & Wiesinger 2025). L'indagine è prevalentemente qualitativa e mira a individuare configurazioni ricorrenti in sequenze di ridefinizione improvvisa dell'impegno epistemico.

L'attenzione si concentra su tre strategie principali:

1. Verba sentiendi e dicendi impiegati come marcatori pragmatici, verificando se forme come *olha* o *vê* possano funzionare come segnali di registrazione improvvisa o di controaspettativa nel flusso dialogico (cfr. Sánchez López 2017).
2. Anteposizione focale con interpretazione di controaspettativa, in linea con le analisi sul Mirative Fronted Focus (Cruschina & Bianchi 2021), come correlato sintattico di inattesa acquisizione di informazione.
3. Interiezioni e segnali discorsivi polifunzionali, con particolare attenzione al marcatore colloquiale *pá*. Si mostrerà che *pá* ricorre sistematicamente in contesti di ristrutturazione improvvisa del frame discorsivo e può assumere una funzione mirativa speaker-oriented, segnalando uno scarto tra aspettativa e nuova evidenza (cfr. Olbertz 2009).

L'ipotesi di lavoro è che nel portoghese europeo la miratività non sia codificata da marcatori dedicati, ma emerga come configurazione sistematica dall'interazione tra struttura informativa, orientamento discorsivo e gestione dell'allineamento epistemico tra interlocutori (Hengeveld & Olbertz 2012). Si mostrerà che determinate strategie – in particolare l'uso di marcatori discorsivi come *pá* in sequenze

di *surprise uptake* – ricorrono con regolarità in specifici contesti dialogici delle interviste radiofoniche.

Il contributo intende così mappare lo spazio funzionale occupato dall'inatteso nel portoghese europeo contemporaneo e contribuire al dibattito tipologico sulle strategie mirative nelle lingue romanze, offrendo al contempo una prima sistematizzazione descrittiva per questa varietà. Il caso del portoghese europeo offre un terreno comparativo particolarmente rilevante: da un lato, consente di mettere alla prova categorie funzionali discusse anche per l'italiano (struttura informativa, focalizzazione, marcatori discorsivi); dall'altro, contribuisce alla riflessione tipologica sulle lingue romanze prive di marcatori mirativi grammaticalizzati (Marques 2022; Marques & Pires de Oliveira 2016). L'analisi proposta intende quindi non solo colmare una lacuna descrittiva relativa al portoghese europeo, ma anche alimentare il dibattito teorico più ampio sull'espressione dell'inatteso nelle lingue romanze e sul rapporto tra sintassi, pragmatica e interazione.

Riferimenti bibliografici

- Aikhenvald, Y. Alexandra. 2012. The essence of mirativity. *Linguistic Typology* 16(3). 435–485.
- Cruschina, Silvio & Valentina Bianchi. 2021. Mirative implicatures at the syntax–semantics interface: A surprising association and an unexpected move. In Trotzke, Andreas & Villalba, Xavier (a cura di), *Expressive meaning across linguistic levels and frameworks*, 86–107. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198871217.003.0005>
- DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1(1). 33–52.
- Dessi Schmid, Sarah, Momma, Lydia & Wiesinger, Evelyn 2025. Mirativity in Romance: Speaker-oriented vs. heareroriented expression of unexpectedness. In Rodríguez Rosique, Susana (ed.), *Expressing Surprise at the Crossroads Mirativity, Exclamativity and Cooptation in Romance Languages*, 229–246. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111386683-010>.
- Hengeveld, Kees & Olbertz, Hella 2012. Didn't you know? Mirativity does exist! *Linguistic Typology* 16(3). 487–503. <https://doi.org/10.1515/lingty-2012-0018>
- Marques, Rui 2022. Conversation and mood in European Portuguese. In Gregoromichelaki, Eleni, Hough, Julian, Kelleher, John D. (a cura di), *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue, Dublin, 22-24 agosto 2022*. http://semdial.org/anthology/Z22-Marques_semdial_0022.pdf

- Marques, Rui & Pires de Oliveira, Roberta 2016. Mood and modality. In Wetzels, W. Leo, Costa, João & Menuzzi, Sergio (a cura di), *The Handbook of Portuguese Linguistics*, 408–424. Oxford: Wiley-Blackwell.
- Olbertz, Hella. 2009. Mirativity and exclamatives in Functional Discourse Grammar: Evidence from Spanish. *Web Papers in Functional Discourse Grammar* 82. 66–82.
- Sánchez López, Cristina 2017. Mirativity in Spanish. The case of the particle *mira*. *Review of Cognitive Linguistics* 15(2): 489–514. <https://doi.org/10.1075/rcl.15.2.08san>