

# FARE LINGUISTICA APPLICATA CON LE *DIGITAL HUMANITIES*

a cura di

Jacopo Saturno - Lorenzo Spreafico

studi AltLA **14**

AltLA

**studi AltLA 14**

FARE LINGUISTICA APPLICATA  
CON LE *DIGITAL HUMANITIES*

a cura di  
JACOPO SATURNO - LORENZO SPREAFICO

Milano 2022

L'AItLA pubblica una collana di monografie e di collettanee sui diversi temi della linguistica applicata. I manoscritti vengono valutati con i consueti processi di revisione di pari per assicurarne la conformità ai migliori standard qualitativi del settore. I volumi sono pubblicati nel sito dell'associazione con accesso libero a tutti gli interessati.

*Comitato scientifico*

Cecilia Andorno, Giuliano Bernini, Simone Ciccolone, Silvia Dal Negro, Maria Elena Favilla, Anna De Meo, Laura Gavioli, Antonietta Marra, Natacha S.A. Niemants, Elena Nuzzo, Jacopo Saturno, Lorenzo Spreafico, Marilisa Vitale.

© 2022 AItLA - Associazione Italiana di Linguistica Applicata  
Via Cartoleria, 5  
40100 Bologna - Italy  
email: [info@aitla.it](mailto:info@aitla.it)  
sito: [www.aitla.it](http://www.aitla.it)



Edizione realizzata da  
Officinaventuno  
Via Flli Bazzaro, 18  
20128 Milano - Italy  
email: [info@officinaventuno.com](mailto:info@officinaventuno.com)  
sito: [www.officinaventuno.com](http://www.officinaventuno.com)

ISSN: 2724-4830

ISBN edizione cartacea: 978-88-97657-50-7

ISBN edizione digitale: 978-88-97657-51-4

edizione digitale distribuita secondo la licenza Creative Commons CC BY NC ND



# Indice

Introduzione	5
FRANCESCO CUTUGNO - EMANUELA CAMPISI Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME	9
CHRISTOPH DRAXLER Automatic Transcription of Spoken Language Using Publicly Available Web Services	27
VALERIA CARUSO - ROBERTA PRESTA Approcci digitali alla lessicografia: sviluppo e valutazione di un dizionario per smartphone	49
SIMONE CICCOLONE - GIULIA GROSSO Macchine parlanti o macchine apprendenti? Osservazioni preliminari sull'uso dell'etero-ripetizione in produzioni di chatbot e nell'interlingua	69
VALENTINA DE IACOVO - MARCO PALENA La variazione prosodica in italiano: l'utilizzo di un chatbot Telegram per la didattica assistita per apprendenti di italiano L2 e nella valutazione linguistica delle conoscenze disciplinari	85
GIOVINA ANGELA DEL ROSSO - SILVIA BRAMBILLA L'accuratezza della trascrizione ASR sul parlato non-standard. L'italiano nell'OH Portal	99
MARTA MAFFIA - MASSIMO PETTORINO - ANNA DE MEO - ROSA DE MICCO - GIOACCHINO TEDESCHI - ALESSANDRO TESSITORE Ritmo e malattia di Parkinson. Prime riflessioni sull'annotazione automatica del parlato di pazienti in stadi iniziali	117
LUISA REVELLI - ANTONIO MASTROPAOLO - DANIELE PAOLO RADICIONI La <i>sottodeterminazione</i> nei testi giuridici: verso un'analisi linguistico- computazionale	131
FLAVIA SCIOLETTE - EMILIANO GIOVANNETTI Un modello per domarli tutti: verso una rappresentazione del testo come esplicitazione di documento, lingua e contenuto	145
LORENZO SPREAFICO Fare linguistica applicata con le <i>digital humanities</i> : interfaccia di metodo o di comodo uso?	159
Indice autori	173

## Introduzione

È forse una curiosa coincidenza che questo volume sia stato elaborato in un momento storico nel quale, a causa dell'incertezza legata alla perdurante pandemia di Covid-19, le attività di ricerca e didattiche si svolgevano pressoché unicamente in modalità telematica. Il recente esponenziale sviluppo degli strumenti informatici ha infatti permesso di mantenere una sostanziale continuità nelle attività universitarie pur in presenza di restrizioni rigide agli spostamenti e alle possibilità di aggregazione, replicandone per quanto possibile struttura e modalità di svolgimento. Nei casi di maggior successo, tuttavia, non si è assistito alla semplice trasposizione di pratiche preesistenti alla modalità telematica. Al contrario, gli utenti più ricettivi hanno colto l'opportunità di sviluppare pratiche innovative che sfruttassero appieno il potenziale degli strumenti a disposizione, contribuendo all'elaborazione non solo di nuove soluzioni, ma di nuovi approcci ai problemi esistenti. A causa della magnitudine del fenomeno, in tempi recentissimi si è anche sviluppata una meta-scienza avente per oggetto la riflessione sul ruolo degli strumenti informatici nel contesto odierno, specie in relazione all'attività didattica. È evidente infatti che per quanto essi abbiano acquistato enorme popolarità in relazione a un contesto emergenziale, dunque eccezionale, molti di essi verosimilmente entreranno nel patrimonio comune delle pratiche accademiche, naturalmente in seguito a un opportuno accomodamento alle mutate circostanze.

Si può cercare un parallelo con la traiettoria appena descritta anche nell'ambito della ricerca scientifica di matrice umanistica. La recente comparsa di possibilità tecniche prima proibitive ha spianato la strada per nuovi approcci – basti pensare alla crescente popolarità delle metodologie statistiche di stampo bayesiano, fino a tempi recenti di scarsa diffusione proprio a causa delle limitazioni tecniche degli elaboratori. Si assiste oggi a interpretazioni diverse delle possibilità offerte dai nuovi strumenti, spazianti dal ricorso all'elaboratore in quanto utile sostituto dell'utente umano in compiti meccanici e ripetitivi, essenzialmente riconducibili al contare, fino ad approcci innovativi in cui l'informatica diventa strumento euristico in grado di suggerire decisioni anche di notevole portata, quando non addirittura di prenderle autonomamente sulla base di un algoritmo di intelligenza artificiale.

Tale diversità di approcci si rispecchia in una terminologia non del tutto univoca. L'etichetta *digital humanities* in particolare sembra riunire sotto di sé concezioni anche molto diverse del ruolo degli strumenti informatici nella pratica scientifica, accomunate semplicemente dal ricorso agli strumenti elettronici. Alla luce dell'ubiquità di questi ultimi, è evidente come tale concetto si presti poten-

zialmente ad una varietà estremamente ampia di interpretazioni. Ad esso si affianca poi anche il precedente termine “linguistica computazionale”, che a rigore si può forse interpretare come uno specifico ambito di applicazione delle *digital humanities*. Ai poli estremi si possono identificare da un lato l’applicazione di strumenti informatici a problemi di ordine linguistico, dall’altro un approccio metodologico volto a impostare tutte le fasi della ricerca – a partire dalla costruzione e categorizzazione della base di dati – esplicitamente alla luce di procedure computazionali. Così è anche nel caso dei contributi raccolti in questo volume, i quali propongono uno spaccato della ricca riflessione che si è recentemente sviluppata e si sviluppa tuttora intorno a questi temi.

Dopo aver introdotto il tema della trascrizione di dati orali in senso ampio, Christoph Draxler presenta le tecnologie attualmente a disposizione dello studioso per la trascrizione automatica del parlato, mettendone a confronto le prestazioni coi risultati ottenibili da un utente umano esperto.

Francesco Cutugno e Emanuela Campisi mostrano alcuni risultati del progetto interdisciplinare CHROME (*Cultural Heritage Resources Orienting Multimodal Experience*), tra i cui obiettivi rientra lo sviluppo di risorse digitali per la fruizione dei beni culturali. In particolare, si descrive l’articolazione di un sistema di annotazione multilivello capace di gestire una grande varietà di dati di diversa natura, spaziando dalla trascrizione del parlato ai rilievi architettonici. Dopo aver introdotto gli obiettivi e l’architettura generale del progetto, il contributo esemplifica le potenzialità dello strumento sulla base di un video in cui un operatore professionista illustra ai visitatori la certosa di S. Martino di Napoli, con particolare attenzione al rapporto tra parlato e gestualità.

Valeria Caruso e Roberta Presta descrivono le fasi progettuali di *Idiomatica*, una app dizionario per smartphone dedicata alle espressioni idiomatiche italiane e rivolta specificamente agli apprendenti di questa lingua come L2. Dopo una rassegna degli approcci attualmente esistenti alla progettazione di app dizionario, le autrici descrivono nel dettaglio le soluzioni adottate per far fronte alle mancanze individuate e aggiornare la struttura di questi prodotti. La discussione si sofferma in particolare su problemi legati alla microstruttura delle voci, alla raccolta dei dati e al formato finale di presentazione.

Lo studio di Simone Ciccolone e Giulia Grosso è dedicato all’interazione tra utenti umani e chatbot progettati per simulare una conversazione informale, con particolare attenzione all’uso strategico della ripetizione di frammenti del turno dell’utente da parte della macchina. Come è noto, la strategia della eteroripetizione è frequentemente utilizzata nelle varietà di apprendimento per segnalare la propria partecipazione all’interazione, da un lato, e per negoziare forma e significato, dall’altro. Da qui l’ipotesi di un parallelo tra gli obiettivi di un apprendente di L2 e di una macchina che dovrebbe simularne il comportamento.

Di chatbot si occupano anche Valentina De Iacovo e Marco Palena, seppur da una prospettiva eminentemente didattica. Il loro studio descrive infatti un chatbot destinato agli apprendenti di italiano L2 e progettato per facilitare l’ap-

prendimento delle strutture ritmiche e intonative della lingua. Se da un lato sempre nuovi strumenti digitali facilitano al ricercatore il compito di confrontare i profili prosodici di un apprendente con la loro realizzazione attesa, individuando potenziali criticità diffuse, da un punto di vista didattico rimane il problema di come rendere gli apprendenti consapevoli delle proprie deviazioni dal modello. Dopo aver descritto la struttura del chatbot, la sua efficacia didattica è verificata mediante uno studio pilota.

Il lavoro di Giovina Angela del Rosso e Silvia Brambilla affronta il problema della trascrizione automatica, con particolare attenzione alle prestazioni del servizio Google-IT's ASR, il cui comportamento è stato finora descritto unicamente sulla base di dati audio di buona qualità. Lo studio si concentra invece sul caso contrario, evidenziando come il parlato non-standard e la presenza di rumore di fondo possano notevolmente peggiorare la precisione della trascrizione. Obiettivo di tale analisi è anche l'identificazione di errori ricorrenti, la consapevolezza dei quali può facilitare il processo di correzione manuale della trascrizione da parte dell'utente.

Marta Maffia e colleghi descrivono un esempio di applicazione degli strumenti della linguistica alla diagnosi precoce del Parkinson, la malattia neurodegenerativa più diffusa dopo l'Alzheimer, tra i cui primi sintomi si trovano diverse alterazioni del ritmo del parlato rilevabili con gli strumenti della fonetica sperimentale. Lo studio si propone dunque di verificare l'efficacia di uno strumento di annotazione automatica progettato per misurare due noti correlati della malattia. Si presentano i primi promettenti risultati di tale metodologia, seguiti da una discussione di alcuni punti critici emersi dalla valutazione.

Luisa Revelli, Antonio Mastropaolo e Daniele Paolo Radicioni si pongono all'intersezione di linguistica computazionale e studi giuridici, proponendosi nel loro lavoro di definire le manifestazioni linguistiche della sottodeterminazione, cioè la proprietà di un enunciato di veicolare diverse interpretazioni, spesso specificamente ricercata in ambito legislativo. L'analisi è applicata in particolare al corpus di leggi delle regioni autonome della Val d'Aosta. Il testo presenta l'impianto teorico e metodologico di un progetto di ampio respiro, anticipandone alcuni risultati preliminari.

Flavia Sciolette e Emiliano Giovannetti descrivono le prime fasi della complessa elaborazione concettuale di un modello olistico di rappresentazione del testo come diasistema. Il modello si articola in un insieme di componenti di diversa natura (grafica, linguistica, concettuale ecc.), i cui elementi formano insiemi separati ma strettamente interconnessi e capaci di influenzare l'intero diasistema. Le potenzialità del modello sono esemplificate sulla base di due testi, cioè il Talmud babilonese e il dizionario della terminologia botanica dell'occitano antico.

Lorenzo Spreafico, da ultimo, in un lavoro di natura speculativa riflette sulla relazione tra linguistica applicata e *digital humanities* alla luce tanto delle dinamiche interne alla due discipline – entrambe sono alla ricerca di una identità, seppur per opposte ragioni avendo la prima perso quella che aveva, e la seconda non

ancora trovato la sua propria – quanto delle posizioni più o meno esplicitamente prese dagli autori degli altri contributi inclusi del volume.

Come si vede, i contributi raccolti nel volume spaziano tra una grande varietà di temi e ambiti di ricerca. Comune a molti è il fatto che si discutano in primo luogo i problemi connessi all'elaborazione concettuale del modello di riferimento per la successiva analisi linguistica, a testimoniare il profondo cambio di prospettiva che le potenzialità offerte degli strumenti informatici richiedono inevitabilmente per essere sfruttate appieno.

Milano, febbraio 2022

*Jacopo Saturno, Lorenzo Spreafico*