

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
sito: www.aitla.it



Edizione realizzata da
Officinaventuno
Via Flli Bazzaro, 18
20128 Milano - Italy
email: info@officinaventuno.com
sito: 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 della regione autonoma 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

Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME

Abstract

During the CHROME project (Cultural Heritage Resources Orienting Multimodal Experience) digital resources were collected and used in technological applications to be adopted in cultural heritage linguistical and technological applications. The project outcomes have been obtained through interdisciplinary work carried out by computer scientists, psychologists, computational and general linguists, architects, museologists, and art historians. A set of texts taken from a wide range of scientific and non-scientific literature have been collected, describing the architectural environments of Carthusian monasteries and their artistic heritages. The collection ranges from scientific texts written by experts addressed to other experts to museum brochures addressed to children and speech corpora, architectural surveys of buildings and environments, video corpora filming tour guides describing points of artistic interest to visitors, from which gestural expression data have been extracted. Preliminary attempts were also made to integrate all digital data sources into a multidimensional platform for which an integrated multi-level annotation system was designed, oriented towards computer applications enriched by knowledge representation systems. The three magnificent Carthusian monasteries in Campania, namely San Martino (Naples), San Lorenzo (Padula, Salerno) and San Giacomo (Capri), were chosen as a case study for testing the technologies, methods, and data collected. The fruit of this interdisciplinary commitment has resulted in a large collection of multidimensional data articulated in heterogeneous sub-sectors, ranging from the collection of texts, through collections of audio-video materials, to three-dimensional reconstructions of valuable architectural environments which are virtually available. The project provides full access to all collected data, analysis tools and software demos for those who request it.

This contribution aims to highlight the main concepts and results achieved in terms of the multimodal analysis of videos in which professional operators illustrate the major aspects of the Charterhouse of San Martino.

1. *Il progetto CHROME*

Nell'attuale panorama della ricerca scientifica internazionale, sui temi dell'innovazione tecnologica iniettata nelle applicazioni di *Digital Culture* si possono trovare moltissimi progetti di grande rilevanza e una letteratura vastissima. In Italia, in particolare, il contributo in questo settore portato dalla Fondazione Bruno Kessler

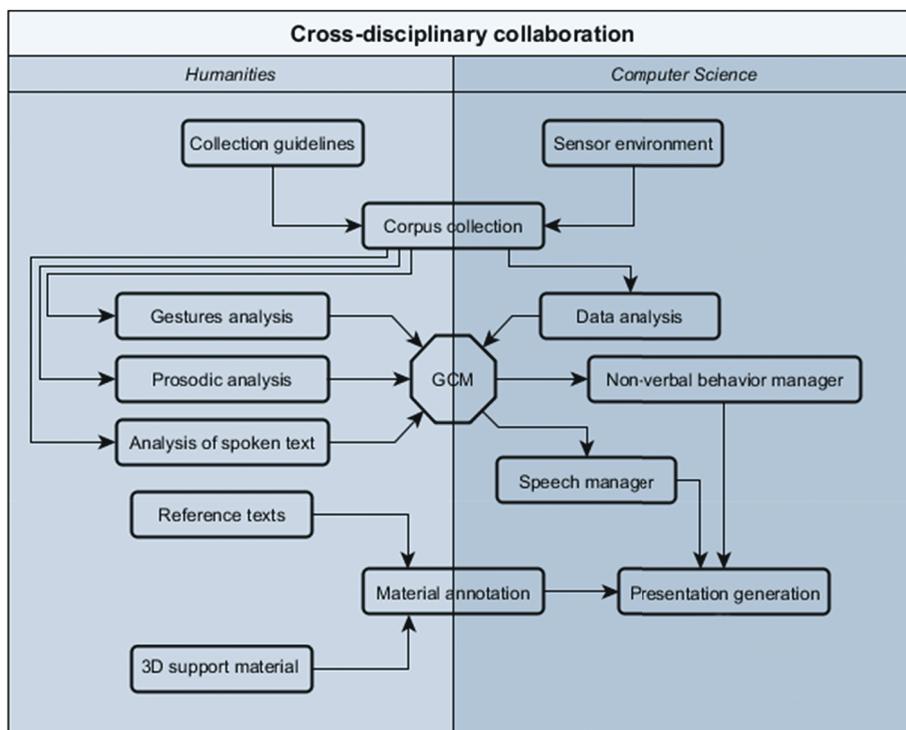
¹ Centro Interdipartimentale di Ricerche Urban/Eco – Università degli Studi di Napoli Federico II.

² Dipartimento di Scienze Umanistiche – Università degli studi di Catania.

in generale e da Oliviero Stock in particolare, ha generato attività di enorme interesse scientifico (Stock & Zancanaro 2007, Kuflik *et al.* 2015). Tuttavia, sebbene questi progetti si siano sempre mossi in un quadro esplicitamente interdisciplinare, gli aspetti legati al linguaggio ed alla comunicazione (verbale e non verbale), al rapporto fra spazi reali, esseri umani ed opere d'arte e monumenti sono stati gli ambiti meno investigati. Il progetto che presentiamo in questo lavoro punta a colmare questa lacuna, proponendo raccolte di dati, formalizzazione di metodi di ricerca e progettazione di dimostratori linguistici, architettonici e di applicazioni informatiche.

Più specificamente, l'obiettivo finale, sintesi di tutte le attività che sono state portate a termine durante il progetto CHROME, è la creazione di un agente umanoide progettato per assumere il ruolo di guida turistica virtuale. Tutte le risorse raccolte in CHROME sono utili a questo scopo: i testi che descrivono i luoghi e le opere ivi presenti, classificati per livello di articolazione e specializzazione nonché per la definizione del tipo di pubblico a cui sono rivolti, sono impiegati per definire il lessico specialistico i principali concetti semantici, il legame con gli oggetti fisici che realmente compongono il posto e la loro nomenclatura. In quest'ottica, i rilievi architettonici digitali, che sono un ulteriore risorsa informatica raccolta in questo progetto, forniscono rappresentazioni codificate dello spazio reale e degli oggetti ivi contenuti. Il continuum spaziale viene segmentato ed etichettato da esperti storici dell'arte e della architettura, creandosi, in questo modo, un collegamento fra entità presenti nell'ambiente e i loro nomi. La definizione del lessico specialistico si avvale anche del contributo di risorse esterne quali l'AAT della Getty Foundation, dai nomi e dai termini specialistici che sono messi in relazione alle posizioni e alle coordinate geometriche che questi ultimi assumono negli spazi investigati. Il materiale audiovideo permette la generazione di una definizione del modello di operatore turistico virtuale; i suoi testi parlati, trasformati in materiale etichettato consentono la creazione di uno stile di parlato che dà alla voce sintetica un timbro e delle caratteristiche estremamente naturali, i contenuti che il sistema automatico "inietta" nell'agente derivano dalla migliore sintesi fra quanto effettivamente prodotto dalle guide durante le loro presentazioni, sia nelle caratteristiche testuali che in quelle di pronuncia, e quanto riscontrato nei testi bibliografici di riferimento. I movimenti delle mani e del corpo delle guide vengono analizzati e ri-sintetizzati per rendere "naturale" il comportamento dell'agente virtuale che, inoltre, collegandosi alla mappatura degli oggetti nello spazio tridimensionale è "consapevole" della posizione di questi ultimi e può imparare ad indicarli e a puntarli con gesti deittici. L'aspetto più coinvolgente della complessa struttura interdisciplinare del progetto è la pianificazione di attività che vedono a stretto contatto il mondo delle scienze "mollì", almeno in principio, in cui gli aspetti umanistici sono dominanti, con il mondo, sempre più estremizzato, delle nuove tecnologie e della intelligenza artificiale. Nella figura 1 questa pianificazione viene schematizzata e sono illustrati i principali moduli di intervento e le relazioni fra questi.

Figura 1 - La pianificazione delle parti del progetto e la non separazione fra ruolo degli umanisti e ruolo della innovazione tecnologica (fonte: documenti originali del progetto sottoposto al MIUR, la sigla GCM sta per "Gatekeeper Computational Model")



Questo impianto progettuale trova la sua giustificazione nella seguente citazione:

it has often been said that museums are the places where you can see the real things. It is this presentation of reality, sampled and structured according to arbitrary principles of reality, that makes museums different from any other communication system. Museums are not simple communication systems that we can describe as: a sender, a medium, more than one receiver. It is a complex system with several senders, multiples media and different categories of receivers. The primary system of museum communication includes a curator who is the sender, real things that are the media and visitors who are the receivers. Curators describe nonverbal media (real things) with subsidiary media. (Cameron 1968, ma si leggano anche altre riflessioni sul tema in Chiacchio 2004)

La tendenza della scienza che si occupa dell'iniezione di nuove tecnologie per potenziare l'esperienza museale si pone esattamente l'obiettivo di trovare nuovi paradigmi di comunicazione per rendere materiale ciò che, artisticamente parlando, non lo è. Linguistica ed Interazione Uomo-Macchina, nella sua declinazione legata alla multimodalità, restano centrali in questo discorso e devono trovare nuovi modi di dialogare e di indirizzare la raccolta di risorse digitali che aiutino, nell'ambito delle moderne tecniche della intelligenza artificiale, ad addestrare sistemi innovativi. Si

raccogliono dunque testi, materiale parlato, audio e video, si effettuano rilievi digitali degli ambienti reali dove normalmente gli operatori umani agiscono, caratterizzati dalla loro struttura architettonica e dalle opere d'arte che contengono. Questo patrimonio di dati ricchissimo viene poi ripartito fra gli specialisti: la linguistica si occuperà di individuare ed estrarre corpora di testo e di parlato, di individuare ed estrarre da questi le risorse lessicali, di fare emergere stili di retorica e di oratoria, formalizzandone le caratteristiche; lo studio delle componenti non verbali, condotto dagli esperti di comunicazione non verbale, porta alla definizione di nuovi schemi di annotazione per il sistema espressivo gestuale delle guide.

Insieme agli architetti, che hanno effettuato i rilievi, si studia il rapporto fra il corpo, le parti dello spazio indicate dalle guide. Gli informatici, i linguisti e gli esperti di comunicazione non verbale definiscono mappe di comunicazione articolate, dando nomi agli oggetti nello spazio, calcolando i movimenti che portano alla deissi gestuale, rendendo virtuale finanche il rapporto fra spazio, esseri umani e testo prodotto. Proprio quest'ultimo, il testo generato automaticamente e sintetizzato per essere pronunciato dalla voce dell'agente artificiale, costituisce l'ultima, ma non meno complessa, sfida. Si parte da testi di differente complessità e destinazione d'uso, alcuni derivati dalle trascrizioni del parlato delle guide professionali, altri estratti dalla collezione di testi del corpus degli scritti raccolti; per questi, per ora attraverso un processo semiautomatico, vengono generate delle descrizioni di elementi artistici che possano essere impiegate sia per dare voce alla guida virtuale, sia per fungere da risposte a domande fatte dagli utenti quando, utilizzando il nostro sistema di dialogo, interrogano l'avatar. Il testo generato deve, da un lato avere le caratteristiche stilistiche e di contenuti coerente con quanto ci si aspetta in questo ambito nelle condizioni che simulano la situazione reale, e, dall'altro, deve essere prodotto con una voce che sia credibile in tutti gli aspetti prosodici, ivi incluse le naturali disfluenze.

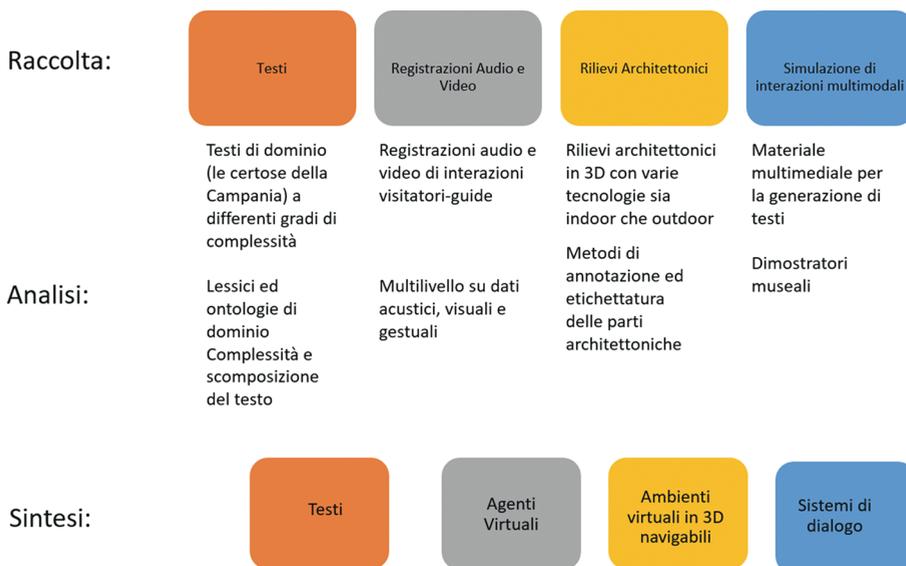
Sulla base di quanto finora esposto, dunque, dunque il progetto prevede delle fasi ben precise:

- raccolta dataset di riferimento per tutti i settori di ricerca da fornire alla comunità;
- studi su: a) le strutture linguistiche (sintattiche, lessicali, delle parti del discorso) per quanto riguarda i testi scritti, per i quali si tiene conto anche del differente stile in funzione della loro destinazione e fruizione, b) delle strutture del parlato, per le quali, oltre a molti degli aspetti pure legati alle strutture linguistiche, si considerano anche aspetti della qualità fonica sia a livello segmentale che soprasegmentale c) dei video, d) dei rilievi architettonici. Gli studi disciplinari e interdisciplinari introdotti in questo progetto, dunque, mirano a generare un sistema di analisi e riproduzione sintetica di tutte le dimensioni comunicative che si possono incontrare negli scenari evocati nel caso di studio scelto;

- processi di integrazione scientifica interdisciplinare e tecnologica per la generazione automatica di contenuti multimodali e per introdurre un modello computazionale sotto forma di agenti artificiali immersi in ambienti virtuali;
- valutazione sperimentale dei risultati scientifici conseguiti nei vari rami del progetto, da condursi attraverso la produzione di appositi dimostratori.

La figura 2 riassume e schematizza queste fasi.

Figura 2 - *Le fasi del progetto*



Il progetto offre, a chi ne faccia richiesta, il pieno accesso a tutti i dati raccolti, a tutti gli strumenti di analisi e ai dimostratori software progettati. Le modalità per le richieste si possono trovare all'indirizzo <http://www.chrome.unina.it>

2. *Le risorse raccolte in CHROME*

Alla fine del progetto sono state raccolte ingenti (per i dettagli si consulti il sito del progetto) quantità di dati che, come indicato nel paragrafo precedente, sono disponibili su richiesta all'intera comunità scientifica.

2.1 Testi sulle tre certose della Campania

Sono stati raccolti e sono disponibili in forma di consultazione, con diritti da assolvere per le opere coperte da diritto d'autore, 102 testi scritti che sono riconducibili a diverse categorie: a) testi scientifici, b) cataloghi destinati alla disseminazione, c) guide specializzate, d) materiale preso dal web e opportunamente selezionato con la supervisione di un esperto di dominio, e) brochure di sala per visitatori adulti e no.

2.2 Analisi testuale

Su parte dei testi raccolti sono state condotte analisi automatiche e manuali (Brunato *et al.* 2020). Le analisi principali sono state: a) estrazione di lessici generali e di termini specifici del dominio, b) Individuazione ed etichettatura degli eventi temporali indicati nei testi, c) analisi sintattica basata su concetti di treebanking, d) riconoscimento dei concetti dell'architettura, basato sul confronto fra termini estratti al punto a) e quelli presenti nell'Art and Architecture Thesaurus (AAT 2018)

2.3 Rilievi architettonici digitali

Gli architetti presenti nella compagine hanno eseguito diversi interventi di rilievo digitale degli edifici, degli ambienti e delle componenti strutturali ivi presenti. Sono state utilizzate diverse tecniche di rilievo, fra cui laser scanner, voli di droni e fotogrammetria aerea. I tre edifici certosini e le parti maggiormente rilevanti dal punto di vista artistico e storico di ognuno di questi hanno generato una considerevole mole di dati di digitalizzazione di beni architettonici, che saranno resi disponibili su richiesta all'interno del *repository* del progetto.

2.4 Video registrazioni delle guide museali

Tre delle quattro guide turistiche ingaggiate nel corso del progetto sono state videoregistrate durante il loro lavoro all'interno del museo ospitato nella certosa di S. Martino in Napoli mentre descrivevano, ad un pubblico di quattro persone per turno, le principali caratteristiche artistiche del sito. La registrazione è stata condotta con due telecamere, una puntata sulla guida e l'altra puntata sul pubblico, e il parlato della guida è stato registrato su tre canali audio aggiuntivi: il primo era collegato ad un microfono *headset* indossato dalla guida e tenuto a distanza fissa dalle labbra, gli altri due a microfoni simmetrici, distanti circa un metro dalla guida e puntati verso il pubblico, che mediamente stazionava di fronte alla guida.

Sulla componente linguistica di questo materiale audiovisivo sono state condotte delle analisi su vari livelli: a) livello ortografico, con trascrizioni del testo prodotto arricchita dalla annotazione delle pause silenti, delle pause piene, delle principali disfluenze (ripartenze, esitazioni, eccetera, si vedano Cataldo *et al.* 2019, Schettino *et al.* in stampa), b) livello sillabico, segmentazione in sillabe, con valutazione *a posteriori* della fluenza e della velocità di eloquio, c) livello intonativo, allineamento fra movimenti tonali della curva dell'intonazione e testo corrispondente, individuazione dei pattern di enfasi, indicazione di stili prodotti, d) livello pragmatico informativo delle unità informative maggiori di topic e comment (Gundel 1988) e di inciso (Cresti 2000), livello macrosintattico (per un approfondimento si veda anche Alfano *et al.* in stampa), e) analisi multimodale/gestuale, annotazione, secondo uno schema di annotazione dei gesti (si veda la seconda parte di questo lavoro) delle espressioni del volto, inclusa una classificazione in termini di funzionalità testuali, emotive ed interazionali. Lo strumento scelto per l'annotazione audio-video è ELAN (Sloetjes & Wittenburg 2008), che consente una architettura di annotazioni

multilivello, fondendo livelli di analisi video con livelli di analisi audio, permettendo al contempo sia la massima indipendenza possibile dei livelli di annotazione, sia di effettuare i collegamenti e le interrogazioni inter-livello fra i dati disponibili.

Di tutti gli aspetti trattati in CHROME, per questo articolo, si è scelto di procedere con la descrizione della analisi multimodale del corpus audio-video con specifico riferimento alla analisi dei gesti di una guida museale.

3. *Introduzione all'analisi multimodale*

È un dato ormai consolidato che i gesti che facciamo mentre parliamo non sono movimenti casuali ma interagiscono col parlato corrispondente a livello prosodico, sintattico, semantico e pragmatico (McNeill 1992; Kendon 2004). In particolare, numerosi studi confermano che non solo i gesti sono parte integrante dell'intenzione comunicativa del parlante, ma anche che sono percepiti come tali dai destinatari, che integrano l'informazione che viene dai gesti con quella proveniente dal parlato e, di conseguenza, consentono di raggiungere risultati migliori in compiti di comprensione (per una rassegna più completa cfr. Campisi 2018). Questo è vero non soltanto tra esseri umani ma, come un recente filone di ricerca sembra suggerire, anche nell'interazione con robot umanoidi e con agenti virtuali, che vengono apprezzati maggiormente quando gesticolano in maniera "appropriata" (Bergmann & Macedonia 2013).

Sebbene i *gesture studies* moderni siano una disciplina ormai affermata e in continua espansione, grazie anche agli sviluppi importanti delle tecnologie a supporto della ricerca sul parlato multimodale, la nostra conoscenza sulla gestualità che si accompagna al parlato (*co-speech gestures*) risente ancora di sostanziali limitazioni; tra queste, ricordiamo la scarsità di studi sistematici in lingue diverse dall'inglese americano da un lato e la frequente separazione tra chi si occupa di gesti, diventando esperto nella loro identificazione e codifica, e i linguisti che studiano il parlato, che a volte riconoscono l'importanza della componente gestuale ma non hanno gli strumenti conoscitivi per includerla nelle loro analisi.

Il progetto CHROME ha le potenzialità per contribuire a entrambe le limitazioni. Infatti, il corpus di parlato semi spontaneo e semi-monologico (Cataldo *et al.* 2019) è perfetto per un'analisi dei gesti che permetta di integrare ciò che già sappiamo sui gesti dei parlanti italiani – e nello specifico napoletani, documentati per lo più dal padre degli studi moderni sul gesto, Adam Kendon (cfr. Kendon 2004) – con informazioni nuove, provenienti dal perfezionamento delle nozioni che ha caratterizzato i *gesture studies* negli ultimi decenni e dal già menzionato sviluppo delle tecnologie a nostra disposizione.

4. L'analisi dei gesti

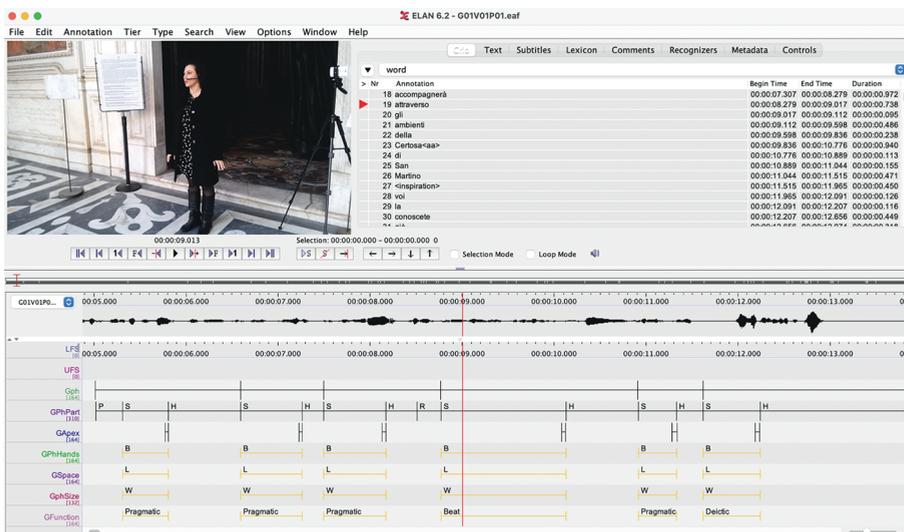
4.1 Criteri iniziali e successivi sviluppi

Fin dalle prime fasi del progetto, si auspicava un'analisi del comportamento multimodale delle guide turistiche coinvolte che prevedesse l'annotazione di gesti, le espressioni facciali e lo sguardo. Tale analisi doveva non soltanto includere la descrizione fisica di tali componenti multimodali, ma anche l'analisi semantica e la classificazione in termini di funzioni testuali, emotive e interazionali (Origlia *et al.* 2019a). In questa sede ci limitiamo a presentare la codifica, ancora in corso, dei gesti delle mani e delle braccia.

4.2 Schema di codifica attuale

Lo schema di codifica attuale dei gesti si presenta come un insieme di livelli organizzati gerarchicamente, tipici dell'interfaccia di ELAN.

Figura 3 - Schermata Elan con i livelli (tiers) di codifica dei gesti



I livelli attualmente individuati sono i seguenti:

Segmentazione dell'unità gestuale:

- identificazione delle *gesture phrases*
- identificazione delle fasi gestuali
- identificazione dell'apice del gesto

Per ogni *stroke*, identificazione delle caratteristiche cinetiche e semantiche del gesto:

- Mani
- Spazio del gesto
- Dimensioni del gesto
- Configurazione della mano
- Funzione del gesto

Nel resto del paragrafo definiremo brevemente ciascuno dei livelli.

Seguendo Kendon (1980), consideriamo un'unità gestuale (*gesture unit*) come l'intera escursione dei movimenti, da quando le mani iniziano a muoversi a quando tornano a una posizione di riposo. Un'unità gestuale può essere più o meno lunga, e può corrispondere o meno all'intera unità discorsiva. L'unità gestuale può essere a sua volta scomposta in unità più piccole, dette *gesture phrases*³, che si caratterizzano per un'escursione in cui le mani e le braccia eseguono un particolare movimento posizionandosi in un punto preciso dello spazio intorno al parlante e le mani assumono una configurazione specifica. Ogni *gesture phrase* inizia con un movimento di "preparazione", in cui gli arti superiori si preparano a raggiungere tale posizione e configurazione, e può concludersi con una "retroazione", cioè col ritorno a una posizione di riposo. Inoltre, ogni *gesture phrase* ha un suo apice, detto *stroke*, in cui il gesto raggiunge il momento massimo di sforzo e chiarezza. La complessità e la lunghezza delle *gesture phrases* possono variare notevolmente, a seconda che contengano o meno delle pause (sia *pre-stroke* sia *post-stroke*), la cui funzione principale sembra essere quella di allineare i gesti con il ritmo del parlato.

4.2.1 Mani (sinistra/destra/entrambe)

Un gesto può essere eseguito con una mano o con due mani. Usare due mani anziché una può essere considerato uno sforzo comunicativo aggiuntivo e quindi un tentativo di rendere più ricco il proprio messaggio. Questo è vero sia nel caso in cui le due mani eseguono lo stesso movimento sia nel caso di configurazioni più complesse, in cui invece i movimenti sono diversi (vedi la descrizione delle funzioni più in basso).

4.2.2 Spazio del gesto

Ogni gesto viene eseguito in un luogo specifico, cioè in una porzione di quello che si chiama *gesture space*, che possiamo immaginare come una semisfera davanti al parlante, la cui grandezza è delimitata dall'estensione massima delle braccia in avanti, di lato, in alto e in basso (McNeill 1992: 86). Secondo McNeill, la maggior parte dei gesti ha luogo al centro dello spazio gestuale, che corrisponde più o meno allo spazio davanti al torace del parlante, oppure nell'immediata periferia. Ciò comporterebbe che, più un gesto è periferico, più dovrebbe diventare saliente. Per la codifica attuale si è deciso, a causa della libertà di movimento delle guide, di non usare una griglia, ma di codificare lo spazio in base al luogo del corpo in cui il gesto si realizza (gambe, torso, testa).

4.2.3 Dimensione del gesto

Anche la dimensione del gesto può essere interpretata in termini di sforzo comunicativo, con i gesti più grandi considerati un modo per amplificare e potenziare il

³ A causa dell'ambiguità del termine *phrase*, che a seconda dell'ambito di studi linguistici può intendersi sia come *intonational phrase* sia come 'sintagma', decidiamo in questa sede di non tradurre l'espressione con un equivalente in italiano.

loro contributo (Mol *et al.* 2011). La codifica delle dimensioni può essere eseguita o tramite una scala Likert o, tramite il punto di articolazione del gesto (spalla/gomito/polso; Mol *et al.* 2011). Per la presente codifica abbiamo scelto di usare quest'ultimo sistema.

4.2.4 Configurazione della mano

Codificare i gesti in base alla configurazione che le mani assumono nell'eseguirlo può essere molto utile, a patto però di essere consapevoli del fatto che mentre nelle lingue dei segni la configurazione ha un valore lessicale abbastanza stabile (almeno nelle forme 'da dizionario'), le configurazioni dei gesti intanto hanno una natura altamente idiosincratca, e per di più non sono sempre chiaramente articolate come nelle lingue dei segni a causa di una generale maggiore rilassatezza dell'articolazione. La codifica della configurazione della mano è fondamentale quando ci si propone di analizzare famiglie di gesti (cfr. § 4).

4.2.5 Funzione del gesto

Gesti con funzione deittica: gesti che indicano “un oggetto, una posizione, o una direzione, che è scoperta proiettando una linea dritta dal punto estremo della parte del corpo che è stata estesa verso l'esterno sullo spazio che si estende oltre il parlante” (Kendon 2004: 200, trad. nostra). Per il presente progetto, ci interessa in particolare il modo in cui i gesti deittici vengono inseriti all'interno della gesticolazione. Il corpus CHROME è particolarmente ricco in questa direzione: il discorso delle guide turistiche è tipicamente accompagnato da deittici che indicano l'oggetto della spiegazione, presente nell'ambiente. Inoltre, anche in questo corpus i gesti assumono anche la cosiddetta funzione 'deittica astratta' quando rimandano a un referente assente o si riferiscono a un referente fisicamente presente, ma hanno una funzione metonimica.

Gesti con funzione rappresentativa: gesti che rappresentano un referente attraverso un'immagine che, in qualche modo, lo raffigura (McNeill 1992). Nella terminologia di McNeill e dei suoi allievi, i gesti con funzione rappresentativa includono i gesti iconici e i gesti metaforici. Un gesto iconico è correlato al contenuto semantico del discorso e lo esprime con un'immagine corrispondente. Un gesto metaforico, invece, crea un'immagine concreta per un concetto astratto. Anche in questo caso, il corpus offre numerosi esempi di entrambe le categorie: oltre a discutere di referenti presenti nell'ambiente circostante, le guide offrono narrazioni sulla storia dei luoghi di interesse o su altre opere e altri siti, accompagnando queste narrazioni con numerosi gesti rappresentativi.

Gesti con funzione pragmatica: Per gesto con funzione pragmatica si intende un gesto “che esprime aspetti della struttura dell'enunciato, inclusi lo status delle porzioni di discorso e della loro relazione e il carattere dell'atto linguistico o della mossa interazione dell'enunciato” (Kendon 1995: 247, trad. nostra). Secondo Kendon, i gesti pragmatici possono essere ulteriormente suddivisi in:

- gesti con una funzione “modale”, cioè gesti che indicano in che modo ciò che è detto deve essere interpretato;
- gesti “performativi”, che indicano il tipo di atto linguistico rappresentato dall’enunciato;
- gesti con una funzione di analisi (*parsing*), che marcano le componenti logiche del discorso ed evidenziano i punti più salienti.

Sembra quindi che, in generale, mentre i gesti rappresentativi sono correlati al contenuto della frase, i gesti pragmatici hanno a che fare con l’atto linguistico in sé e con lo scopo che il parlante si prefigge di raggiungere.

Gesti batonici (beats): gesti che non sembrano contribuire al significato del parlato corrispondente né a livello semantico né a livello pragmatico, ma hanno la funzione di scandire il ritmo del discorso.

Sono esclusi dall’analisi i cosiddetti gesti adattatori, tutti quei movimenti, cioè, che non sono legati al parlato ma piuttosto alla dimensione emotiva o ad altri aspetti dell’interazione, come ad esempio movimenti che coinvolgono parti del corpo (ad es. toccarsi i capelli) o oggetti circostanti (ad es. giocare con una penna).

5. Risultati quantitativi preliminari

L’analisi dei dati è ancora in corso, quindi i seguenti dati svolgono la sola funzione di indicare possibili tendenze. I risultati qui riportati riguardano i primi 4 minuti di spiegazione di una delle guide, per un totale di 128 *stroke*. La frequenza dei gesti (27,7 gesti ogni 100 parole) è abbastanza simile a quella ottenuta in altri studi con parlanti italiani che si rivolgono a interlocutori reali (ad es. Campisi & Özyürek 2013).

Riguardo alle caratteristiche cinetiche, la maggior parte dei gesti finora codificati è a due mani (121 su 128). Oltre ai gesti che si svolgono nello spazio centrale, ci sono molti gesti in uno spazio periferico, in particolare all’altezza delle gambe (forse a causa della posizione eretta, poco frequente nei *setting* sperimentali dove invece i partecipanti sono seduti). Per quanto riguarda la grandezza, quasi la metà dei gesti codificati (57 su 128) hanno il gomito come punto di articolazione, e quindi possono essere considerato di media grandezza, mentre solo 13 sono caratterizzati da un movimento che coinvolge la spalla.

Riguardo alla funzione, la maggior parte dei gesti codificati ha una funzione pragmatica. Per questo motivo, dedichiamo l’ultima parte di questo contributo ad analizzare alcuni esempi provenienti da una delle famiglie di gesti più diffusa, la mano aperta supina, detta anche famiglia del *palm-up*.

6. *Un esempio di gesti con funzione pragmatica nel corpus CHROME: l'open hand supine family*

Nei suoi lavori sui gesti pragmatici, Kendon introduce la nozione di famiglia di gesti (*gesture families*). L'intuizione alla base è che alcune configurazioni sono usate con una funzione pragmatica più spesso di altre, almeno nelle lingue in cui sono state studiate. Questo indicherebbe che in alcuni luoghi, come ad esempio il sud Italia, molti gesti pragmatici sono convenzionalizzati, forse a causa del loro uso pervasivo nell'interazione, e meno soggetti a variazioni individuali, come invece accade per i gesti rappresentativi. Una famiglia di gesti, quindi, è un gruppo di gesti che condividono alcune caratteristiche cinetiche e, allo stesso tempo, un tema semantico. Dato che gli studi sui gesti pragmatici sono pochi, specialmente se confrontati con quelli sui gesti rappresentativi o sui deittici, e per lo più limitati all'italiano (Kendon 2004; Graziano 2014), allo spagnolo e al tedesco (Müller 2004), risulta ancora poco chiaro quali siano i parametri che permettano di attribuire a un gesto una funzione pragmatica. Di conseguenza, durante la codifica, si corre il rischio di procedere per esclusione, considerando pragmatici tutti i gesti per cui non è possibile identificare una relazione con il contenuto del parlato. La nozione di famiglia di gesti si rivela quindi fondamentale nell'identificazione dei gesti pragmatici, almeno per quelli più altamente convenzionalizzati.

Tra le famiglie di gesti individuate da Kendon (2004), la più studiata è senz'altro quella detta *Open Hand Supine* (OHS), che include tutti quei gesti con funzione pragmatica eseguiti con la mano aperta e l'avambraccio supino, in modo che il palmo della mano sia rivolto verso l'alto, sia in verticale sia in orizzontale. Tale gesto viene di solito associato all'immagine metaforica della mano aperta che regge o offre qualcosa. Le numerose osservazioni indicano, infatti, che si associa a segmenti di parlato in cui si introduce qualcosa che sta per essere detto o che commenta, spiega o chiarisce quanto detto. Secondo Kendon, inoltre, la mano può essere o estesa nello spazio di fronte al parlante (*palm presentation gesture*, PP) oppure diretta verso l'interlocutore (in quel caso si parla di *palm addressed gesture*, PA). Infine, la mano aperta è spesso usata per indicare, sia concretamente sia in astratto. Per descrivere meglio le possibili funzioni del *palm-up*, ricorriamo ad alcuni esempi del corpus.

6.1 Palm presentation gesture

Tra le funzioni identificate da Kendon, una delle più frequenti è quella di espandere le premesse o le condizioni per comprendere qualcosa, di solito introdotto nel parlato poco prima. Vediamo in che senso. Si consideri il seguente contesto: la guida sta descrivendo le varie fasi dei lavori di costruzioni della Certosa di San Martino, che raggiungono il loro *clou* nel Settecento. Ella afferma quindi:

- (1) Agli inizi del milleottocento gli ordini vengono soppressi

A partire da una posizione di riposo, in corrispondenza di 'vengono soppressi' la guida esegue un PP (vedi figura 4). La soppressione dell'ordine è la causa dell'abbandono della Certosa da parte dei certosini: come ipotizza Kendon, quindi, il gesto pragmatico in questione presenta il parlato corrispondente come una spiegazione di quanto detto in precedenza.

Figura 4 - Palm presentation gesture



Un altro contesto tipico in cui il PP si presenta è quello in cui il parlante sta facendo un elenco o sta contrapponendo due concetti. Ad esempio, si consideri il seguente enunciato:

- (2) Oggi giorno non è solo certosa ma anche museo nazionale.

In corrispondenza di “certosa” e di “museo nazionale” la guida esegue due PP a due mani, il primo caratterizzato da un movimento verso sinistra (figura 5) e il secondo da un movimento verso destra (figura 6).

Figura 5 - PP in corrispondenza di “certosa”



Figura 6 - PP in corrispondenza di "museo nazionale"



6.2 Palm addressed gesture

Tra le funzioni più diffuse del *palm up* c'è anche quella di rivolgersi verso qualcuno (di solito l'interlocutore) o verso qualcosa. Nel secondo caso, il gesto ha una vera e propria funzione deittica, che per Kendon assume una connotazione leggermente diversa rispetto all'indicare col dito indice, e cioè quella di presentare l'oggetto all'interlocutore come qualcosa verso cui si deve guardare.

Nonostante negli anni ci siano stati studi di queste famiglie semantiche, ci sono poche analisi di corpora grandi, e pochi tentativi di estrarre regolarità statistiche a partire dai casi sopra descritti. È chiaro, ad esempio, che le funzioni indicate da Kendon per ciascuna famiglia di gesti non sono esaustive, e nemmeno vincolanti. Ad esempio, non sono rari i casi in cui il parlante adoperi un PP con la funzione che Kendon associa invece al PL (*Palm lateral*), quella cioè di identificare qualcosa come ovvio. In questo senso, il corpus CHROME permetterà un'analisi sistematica delle famiglie di gesti pragmatici presenti, anche grazie alla dettagliata analisi del parlato corrispondente.

7. Conclusioni

Il progetto CHROME ha prodotto una vasta serie di risorse digitali da applicare sia nell'ambito della linguistica applicata che in altri ambiti scientifici interdisciplinarmente collegati a questa. Come già accennato in precedenza, tutte queste risorse sono online e disponibili a fronte di una richiesta che ha solo fini di censimento dei richiedenti. Anche a progetto formalmente concluso, le ricerche attive sui vari fronti descritti in questo lavoro continuano, il gruppo di ricerca evolve, i dimostratori continuano ad essere progettati. Fra i prodotti più importanti rilasciati dopo la sca-

denza del programma PRIN 2015, vale la pena segnalare la suite Fantasia (Origlia *et al.* 2019b) che, basata su tecnologie innovative, colleziona tutto il lavoro di virtualizzazione di ambienti e agenti umanoidi, vi aggiunge un sistema di dialogo e si apre al mondo della intelligenza artificiale e delle forme più avanzate di sistemi di rappresentazione della conoscenza semantica di dominio.

Data la sede di questa pubblicazione, fra le varie tematiche legate alla linguistica applicata sviluppate in CHROME, si è deciso di porre un focus sugli aspetti della comunicazione non verbale. In questa direzione, il primo passo sarà completare la codifica dei gesti secondo lo schema di codifica presentato in questa sede. In questo modo, sarà possibile avere un quadro più preciso sulla distribuzione delle caratteristiche cinetiche e semantiche dei gesti delle guide, e poter comparare la produzione gestuale di guide diverse nei diversi punti di interesse all'interno del percorso.

In questo senso, numerosi studi dimostrano che, normalmente, i partecipanti a uno scambio comunicativo non solo tengono conto dei gesti di chi parla, ma ne traggono beneficio. Ad esempio, Kushch e Prieto (2016) dimostrano che i soggetti ricordano più parole quando l'informazione più importante della frase è accompagnata da un gesto. Inoltre, secondo Breckinridge Church *et al.* (2007), non solo chi è esposto a gesti e parole insieme tiene conto dell'informazione presente nel gesto, ma ricorda tale informazione più a lungo di quella veicolata solo dalle parole. Tuttavia, tali benefici non si verificano automaticamente solo perché il parlante muove le mani, ma si manifestano solo in alcune condizioni, ad esempio, scompaiono se il parlante produce in prevalenza gesti adattatori (Obermeier *et al.* 2015). In altre parole, i destinatari tengono conto dei gesti quando il contesto e il parlante ci dicono che è appropriato farlo. Per questo motivo, è importante che la formazione di figure professionali che fanno del parlato la loro risorsa principale tenga conto anche della gestualità.

Riconoscimenti

Questo lavoro è stato reso possibile dal finanziamento del Ministero Italiano dell'Università e della Ricerca attraverso il programma PRIN 2015, progetto *Cultural Heritage Resources Orienting Multimodal Experience* (CHROME) #B52F15000450001.

Bibliografia

AAT. *Art & Architecture Thesaurus Online*. 2018. <http://www.getty.edu/research/tools/vocabularies/aat/>.

Alfano, Iolanda & Violetta, Cataldo & Schettino, Loredana & Ansani, Alessandro & Chiera, Alessandra & Poggi, Isabella & Isora, Sessa. in stampa. Corpus CHROME: la codifica multilivello di un corpus audiovisivo. In *Corpora e studi linguistici. Atti del LIV Congresso SLI*.

Bergmann, Kirsten & Macedonia, Manuela. 2013. A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In Aylett, Ruth & Krenn,

- Brigitte & Pelachaud, Catherine & Shimodaira, Hiroshi (eds.). *Intelligent virtual Agents*, 139-148. Berlino: Springer.
- Breckinridge Church, Ruth & Garber, Philip & Rogalski, Kathryn. 2007. The role of gesture in memory and social communication. *Gesture* 7(2). 137-158.
- Brunato Dominique & Cimino Andrea & Dell'Orletta Felice & Montemagni Simonetta & Venturi Giulia. 2020. Profiling-UD: A tool for linguistic profiling of texts. In Calzolari, Nicoletta & Béchet, Frédéric & Blache, Philippe & Choukri Khalid & Cieri Christopher & Declerck, Thierry & Goggi, Sara & Isahara, Hitoshi & Maegaard, Bente & Mariani, Joseph & Mazo, Hélène & Moreno, Asuncion & Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of 12th Edition of International Conference on Language Resources and Evaluation (LREC 2020)*, 7145-7151. Marsiglia, European Language Resources Association.
- Cameron F. Duncan. 1968. The museum as a communication system and implications for museum education. *Curator* 11(1). 33-40.
- Campisi, Emanuela. 2018. *Che cos'è la gestualità*. Roma: Carocci.
- Campisi, Emanuela & Özyürek, Asli. 2013. Iconicity as a communicative strategy. Recipient design in multimodal communication for adults and children. *Journal of Pragmatics* 47, 14-27.
- Cataldo, Violetta & Schettino, Loredana & Savy, Renata & Poggi, Isabella & Origlia, Antonio & Ansani, Alessandro & Sessa, Isora & Chiera, Alessandro. 2019. Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech. In Piccardi, Duccio & Ardolino, Fabio & Calamai, Silvia (a cura di), *Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale*, 205-231. Milano: Officinaventuno.
- Chiacchio, Maria Laura. 2004. *Gli appartamenti storici: una tipologia museale e le sue strategie di comunicazione*. Tesi di dottorato in Metodologie Conoscitive per la Conservazione e la Valorizzazione dei Beni Culturali. Seconda Università di Napoli.
- Cresti, Emanuela. 2000. *Corpus di italiano parlato. Volume 1: Introduzione*. Firenze: Accademia della Crusca.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Hammond, Michael & Moravcsik, Edith & Wirth, Jessica (eds.), *Studies in syntactic typology*, 209-239. Amsterdam: Benjamins.
- Graziano, Maria. (2014). The development of two pragmatic gestures of the so-called Open Hand Supine family in Italian children. In Seyfeddinipur, Mandana & Gullberg, Marianne (eds.), *From gesture in conversation to visible action as utterance: essays in honor of Adam Kendon*, 311-330. Amsterdam: Benjamins.
- Kendon, Adam. 1980. Features of the structural analysis of human communicational behavior. In von Raffler-Engel, Walburga (ed.), 29-43. *Aspects of nonverbal communication*. Lisse: Swets & Zeitlinger.
- Kendon, Adam. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics* 23, 247-279.
- Kendon, Adam. 2004. *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kuflik, Tsvika & Wecker, Alan. J. & Lanir, Jael. & Stock, Oliviero. 2015. An integrative framework for extending the boundaries of the museum visit experience: Linking the pre, during and post visit phases. *Information Technology & Tourism* 15(1). 17-47.

- Kushch, Olga & Prieto, Pilar. 2016. The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. *Proceedings of Speech Prosody 2016*, 922-925. https://www.isca-speech.org/archive/speechprosody_2016/kushch16_speechprosody.html.
- McNeill, David. 1992. *Hand and mind*. Chicago: The University of Chicago Press.
- Mol, Lisette & Kramer, Emiel & Maes, Alfons & Swerts, Marc. 2011. Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication* 17(1). 77-100.
- Müller, Cornelia. (2004). Forms and uses of the Palm up open hand: A case of a gesture family? In Müller, Cornelia & Posner, Roland (eds.), *The semantics and pragmatics of everyday gestures. Proceedings of the Berlin conference, April 1998*, 233-256. Berlino: Weidler.
- Obermeier, Christian & Kelly, Spencer & Gunter, Thomas. 2015. A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Social Cognitive and Affective Neuroscience* 10(9). 1236-1243.
- Origlia, Antonio & Savy, Renata & Cataldo, Violetta & Schettino, Loredana & Ansani, Alessandro & Sessa, Isora & Chiera, Alessandra & Poggi, Isabella. 2019a. Human, all too human. Towards a disfluent virtual tourist guide. *Proceedings of the 10th International Workshop on Personalized Access to Cultural Heritage (PATCH 2019)*, 393-399. New York: Association for Computing Machinery.
- Origlia, Antonio & Cutugno, Francesco & Rodà, Antonio & Cosi, Piero & Zmarich, Claudio. 2019b. FANTASIA: A framework for advanced natural tools and applications in social, interactive approaches. *Multimedia Tools and Applications*, 78(10). 13613-13648.
- Schettino, Loredana & Betz, Simon & Cutugno, Francesco & Wagner, Petra. In stampa. Hesitations and individual variability in Italian tourist guides' speech. In *Atti del XVII Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*.
- Sloetjes, Han & Wittenburg, Peter. 2008. Annotation by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Stock, Oliviero & Zancanaro Massimo 2007. *PEACH-Intelligent interfaces for museum visits*. Berlino: Springer.

CHRISTOPH DRAXLER¹

Automatic Transcription of Spoken Language Using Publicly Available Web Services

Abstract

This paper is an informal introduction to transcribing speech in general and the current state of the art in automatic speech recognition in particular, and it elaborates on the differences between commercial and academic speech recognisers. Based on a short extract of an oral history interview, it presents four different types of transcription, and compares the output of a commercial speech recognition system to a human-generated transcript. It proposes a simple graphical tool which allows potential users to estimate the quality of the recognition output. Finally, it introduces the speech processing web services offered by the Bavarian Archive for Speech Signals (BAS) and shows how they may be used to automate parts of the transcription workflow.

1. Introduction

In recent years, the performance of automatic speech recognition (ASR) has improved substantially. In many application areas, e. g. travel information systems, or medical or juridical dictation systems, it has reached human performance levels. Currently, voice-driven personal digital assistants such as Amazon Alexa, Apple Siri, or Google Assistant are becoming increasingly popular and commercially successful – they are available on mobile personal devices, do not require technical expertise to use, and perform well on tasks as diverse as retrieving information from the internet, online shopping, selecting song tracks from streaming services, or dictating personal messages.

The success of ASR has of course attracted the attention of scientific fields working on spoken language – wouldn't it be great if a machine could provide high-quality transcripts of interviews, field recordings, medical or foreign language learning tests, child speech recordings, etc.? Expectations are high!

In this informal introduction, I will first give an overview of the workflow when working with spoken language, and briefly describe the foundations of current ASR technology in non-technical terms. Then, I will look at the various types of transcriptions needed by different scientific fields, and compare these to the outcomes of currently available free or low-cost services provided by commercial ASR providers. Finally, I will present web services offered by the Bavarian Archive for Speech

¹ LMU München.

Signals (BAS) which support and even partly automate the transcription workflow, including access to ASR.

2. Spoken Language Transcription Workflow

Spoken language is a physical signal, i.e. a change in air pressure over time, produced by an airflow from the lungs and modulated by the vocal tract, and perceived via the ears. By nature, it is *volatile* – as soon as it is produced, it is gone.

For processing and analysis, the speech signal needs to be captured and stored. In a *categorisation process*, symbolic labels, e.g. phonemes or words, are mapped to time-delimited fragments of the signal (see e.g. (Spreafico 2020) for an in-depth discussion). The result of this process is a *transcript*, which contains the verbal content of a given utterance (see Figure 1 for an example²). This transcript is the basis of all further analysis and processing steps.

Figure 1 - In a categorisation process, a continuous spoken language signal (shown in its waveform and the derived spectrogram) is labelled with discrete symbolic labels, e. g. words



The workflow consists of five main consecutive steps:

1. Recording: The speech signal is recorded and stored as digital data
2. Transcription: The recorded speech is transcribed by human transcribers or ASR to produce a verbal transcript
3. Segmentation: The transcript is time-aligned with the signal, and annotation levels, e. g. fine-grained phonetic labels, are added
4. Data Management: Transcripts and the corresponding signals are compiled in speech databases for further analysis
5. Analysis, Processing, Presentation: Theory-guided semi-automatic processing, manual analysis, or visual presentation of audio signals, derived signal data, and annotations

Each of these steps requires specific expertise and software tools. Figure 2 shows this workflow; icons represent software tools typically used for the individual steps. Some of these tools, e.g. web browsers, database management systems, or statistics systems such as R (R Core Team 2015) are general purpose software tools.

² This recording is available under https://www.phonetik.uni-muenchen.de/forschung/Bas/Experimente/aitla/Test0001IT_S0.wav.

Others, such as Praat (Boersma 2001), ELAN (Sloetjes 2007), or Oetra (Pömp & Draxler 2017), the segmentation tool MAUS (Schiel 2015, Kisler et al. 2012), or the recording software SpeechRecorder (Draxler & Jänsch 2004), are tailored to spoken language processing. Draxler et al. (2020) describes the T-chain web service, which provides a simple graphical user interface to the services used for transcribing recordings.

The costs associated with the workflow depend on the time and expertise needed to perform the different steps. A common measure is the *real time factor*. It states how much longer a given processing step takes than the duration of the spoken language signal. Table 1 contains estimates for recording, transcription, and manual or automatic segmentation.

Table 1 - *Real time factors and unit costs for recording, transcription, and segmentation (i.e. time-alignment) of spoken language*

<i>task</i>	<i>real time</i>	<i>unit cost</i>
recording	1 - 2	€
transcription	5 - 10	€€
manual segmentation	300 - 1000	€€€
automatic segmentation + manual correction	10 - 50	€€€

Figure 2 - *Transcription workflow and tools*



The real time factor for segmentation depends on the granularity of the transcription – orthographic vs. broad phonemic vs. detailed phonetic transcription – and the amount of manual validation and correction necessary.

Recording and transcription in general require only basic skills and thus they are relatively cheap.

The segmentation of speech requires phonetic or phonological expertise, which makes it expensive. This is especially true when working with dialects or less common languages, for which there are only a few experts in the world, or when working on recordings with low audio quality. Automatic segmentation dramatically reduces the time needed for segmentation, but still needs manual validation and correction, and thus requires expensive human experts.

3. *Transcription Types*

A transcript is a faithful written representation of the content of a spoken language signal. What do *content* and *faithful* mean? In spoken language, content is much more than the sequence of words of an utterance. Besides the verbal content, the signal contains information about the speakers: age, gender, socio-demographic background, emotional state, communication situation, etc.³

When creating a speech database, one has to weigh the options: on the one hand, a fine-grained transcription of the verbal, paraverbal and nonverbal content is expensive and may be of use to only a few researchers. On the other hand, a concise summary of the verbal content of the recording will serve to quickly select material of interest, but it is not sufficient for any in-depth analysis.

In the following, I will present the different types of transcriptions for the same short fragment of an oral history interview. The interview is part of the Anna Maria Bruzzone archive on the Ravensbrück concentration camp (Beccari Rolfi & Bruzzone 2020, Vangelisti et al. 2019). The interview was conducted by Anna Maria Bruzzone (AMB), the interviewee is Lidia Beccari Rolfi (L), a survivor of the concentration camp. The topic of this fragment is how using a spoon to eat one's soup counteracts the systematic dehumanisation in the camp.⁴

3.1 Broad transcript

This type of transcript is close to the original recording, focuses on the key issues, and features punctuation and text smoothing etc. to make it easy to read (see Figure 3 for an example).

³ When listening to the utterance shown in Figure 1, you will quickly notice that it is a sentence read by a male, non-native – possibly German – speaker producing an Italian sentence. You might guess his age, and perhaps even his weight and height.

⁴ The fragment is 1:04 minutes long. It was kindly provided by Silvia Calamai of Siena University and is available at <https://www.phonetik.uni-muenchen.de/forschung/Bas/Experimente/aitla/cucchiaio.wav>.

Figure 3 - *Broad transcript*⁵

Si tratta della storia del cucchiaino. Lidia spiega che l'obiettivo è quello di disumanizzare, di ridurre al livello degli animali, in modo psicologico di far sentire istintivamente come degli animali. Ed è per questo che la prima cosa che un deportato riceve nel campo è un cucchiaino, per sentirsi meno animale, perché c'è sempre e solo zuppa, e senza cucchiaino devi berla o leccarla via. La frase ripetuta più e più volte era: "non siamo le pulci di un cane".

This type of transcript is used to summarise the content of an interview, e. g. for a presentation, and to facilitate quick browsing through a collection of recordings.

3.2 Verbatim raw transcript

In general, a verbatim orthographic raw transcript uses standard orthography and optionally a (very small) set of mark-up symbols to denote specific paraverbal and nonverbal phenomena. The orthographic transcript is as close as possible to what was said, and includes word repetitions, hesitations, and repairs etc. (see Figure 4).

Figure 4 - *Raw verbatim transcript*

Il discorso del cucchiaino. Del cucchiaino ecco, del perché non danno il cucchiaino. Il discorso del cucchiaino rientra, rientra nello stesso tipo di discorso: volendoti disumanizzare, ridurti a livello di bestie: la bestia lecca. Allora istintivamente, tu già psicologicamente sei pronto a sentirti animale. E tant'è vero che una delle prime cose che il deportato acquista, in campo, è il cucchiaino. Per non sentirsi bestia. Cioè chi: tenta di reagire, o lo ruba, o lo compra, o lo acquista con il pane, ma acquista 'sto cucchiaino per potersi sentire meno animale. E perché se fosse stato qualcosa che si mh potesse mangiare con le mani, le mani le mani tutto sommato sono uno strumento. Sì sì certo Ma, visto che è sempre la minestra, sempre minestra, la devi leccare. Bere o leccare. E lì allora ti senti e effettivamente a livello di di di, le frasi che ricorrevano che, che ricorrevano erano: "Non siamo mica dei cani". Eh già: "Non siamo mica dei cani".

This type of transcript is often used to compute the ASR word error rate (see section 4 for details), for language modelling with n-grams, or as input to time-alignment tools (see section 5.3 on WebMAUS for an example of such a tool). For interviews, this transcript generally focuses on the main speaker, i.e. backchannel feedback often is not transcribed.

3.3 Transcript with diarisation

Diarisation adds explicit information on who is speaking and how speaker roles change to the transcript. It divides the transcript into turns which are labelled with a code for the speaker (see Figure 5).

⁵ English translation: «This is the story of the spoon. Lidia explains that the goal is to dehumanize, to reduce to the level of beasts, in a psychologic way, to make one feel instinctively, like animals. And this is why the first thing a deported person tries to obtain is a spoon, to feel less like an animal. It's always soup, and only soup, and without a spoon one has to drink it or lick it. The ever-repeated saying was: "we are not the fleas of a dog.»'. This summary was kindly provided by Lorenzo Spreafico.

Figure 5 - *Verbatim transcript with speaker diarisation*

<i>Spk</i>	<i>Manual verbatim transcript</i>
L	Il discorso del cucchiaino
AMB	Del cucchiaino ecco, del perché non danno il cucchiaino.
L	Del cucchiaino rientra, rientra nello stesso tipo di discorso:
AMB	Certo certo
L	volendoti disumanizzare, ridurti a livello di bestie: la bestia lecca.
AMB	Sì sì certo certo
L	Allora istintivamente, tu già psicologicamente sei pronto a sentirti animale.
AMB	Sì sì sì
L	E tant'è vero che una delle prime cose che il deportato acquista, in campo, è il cucchiaino. Per non sentirsi bestia. Cioè chi: tenta di reagire, o lo ruba, o lo compra, o lo acquista con il pane, ma acquista 'sto cucchiaino per potersi sentire meno animale. E perché se fosse stato qualcosa che si mh potesse mangiare con le mani, le mani le mani tutto sommato sono uno strumento.
AMB	Sì sì certo certo
L	Ma, visto che è sempre la minestra, sempre minestra, la devi leccare. Bere o leccare. E lì allora ti senti e effettivamente a livello di di di, le frasi che ricorrevano che, che ricorrevano erano: "Non siamo mica dei cani".
AMB	Eh già: "Non siamo mica dei cani".

This type of transcription is often used to perform automatic processing of the individual speakers' contributions, for content analysis, and for statistical analyses.

3.4 Transcript with technical and interpretive mark-up

Mark-up adds information to the text in the form of reserved codes. In Figure 6, each speaker turn begins with a speaker code, followed by a timestamp and the transcript text. Within the text, tags written as <...> contain mark-up code, e. g. <OVL> for *overlapping speech*, <BCH> for *backchannel feedback*, or <REP> for *repetition*.

Figure 6 - *Transcript with speaker diarisation and technical and interpretive mark-up*

<i>Spk</i>	<i>Time (s)</i>	<i>Manual verbatim transcript with mark-up</i>
L	0.00	Il discorso del cucchiaino
AMB	1.14	<OVL> Del cucchiaino ecco, del perché non danno il cucchiaino.
L	1.14	<OVL> Del cucchiaino rientra, rientra nello stesso tipo di discorso:
AMB	7.36	<BCH> Certo certo
L	7.99	volendoti disumanizzare, ridurti a livello di bestie: la bestia lecca.
AMB	10.79	<BCH> Sì sì certo certo
L	12.55	Allora istintivamente, tu già psicologicamente sei pronto a sentirti animale.
AMB	17.90	<BCH> Sì sì sì
L	19.12	E tant'è vero che una delle prime cose che il deportato acquista, in campo, è il cucchiaino. Per non sentirsi bestia. Cioè chi: tenta di reagire, o lo ruba, o lo compra, o lo acquista con il pane, ma acquista 'sto cucchiaino per potersi sentire meno animale. E perché se fosse stato qualcosa che si <FIL> potesse mangiare con <REP>le mani, le mani le mani </REP> tutto sommato sono uno strumento.
AMB	46.42	<BCH> Sì sì certo certo

L	48.59	Ma, visto che è sempre la minestra, sempre minestra, la devi leccare. Bere o leccare. E lì allora ti senti e effettivamente a livello <REP>di di di</REP>, le frasi che ricorrevano che, che ricorrevano erano: “Non siamo mica dei cani”.
AMB	61.53	Eh già: “Non siamo mica dei cani”.

This type of transcription is used for in-depth linguistic, discourse, content, and other analyses. The exact format and the extent of the mark-up varies with the research discipline – for linguistic analysis it will be different from that for sociological research. Transcription guidelines define the set of allowed mark-up tags, and their syntax.⁶

3.5 Discussion

The main reason for transcribing recordings is to obtain transcripts which can be used for further research, i.e. which can be browsed, searched, analysed, processed, or visualised.

In general, a raw verbatim transcript with diarisation is considered the basis for all subsequent processing steps. It is theory-neutral, supports the creation of a lexicon and word frequency lists, and creating such a transcript does not require special skills, apart from acute hearing and good knowledge of the language and its orthography. *Acute hearing* here means that the transcriber must be able to separate the different speakers by their voice, and should reliably distinguish and recognise the sounds of the language, even under adverse acoustic conditions or in accented speech.

Depending on the technical quality of the recording and the familiarity of the transcriber with the content, the transcription factor typically lies in the range of 5-10.

A broad transcript can be generated from a raw verbatim transcript, either manually or automatically. Topic detection identifies the main topics of the transcript and provides a structured, possibly statistical or graphical representation, whereas summarisation tools generate new text from the transcript. Both methods can be fine-tuned via parameters, e. g. to deliver only the n most relevant topic items, or to generate a summary of m words.

Mark-up adds para- and nonverbal information to a transcript. Technical information, such as timestamps, is easy to add. Basic syntactic information, such as part-of-speech tags can be provided automatically with a high degree of precision. Other information, e.g. named entity recognition, also achieves good results, but requires manual verification. Higher-level features such as e.g. discourse strategies or establishing a common semantic ground, require multi-faceted analyses and thus cannot be annotated automatically. Other analyses, e. g. emotion recognition, process the spoken language signal, and again may show their results in mark-up tags in the transcript.

⁶ Note that any mark-up should be formatted in such a way that it can be removed mechanically without damaging the transcript, and that it can be searched using regular expressions.

Because each research discipline views the transcript under different aspects, there is no common agreement on what to mark-up. Transcribing a text with mark-up has a transcription factor of 50 or more (in the case of manual phonetic transcription, up to 1000!). It requires specific skills and knowledge, which make it expensive: phonetic analytical hearing, recognition of syntactic structures, discourse analysis, psychology, and others.

To summarize: because of the relatively low effort necessary, both in terms of required skills and cost, the minimum level of transcription should be a raw verbatim transcript with diarisation. This transcript is the prerequisite for identifying material of interest for further processing – this selection will dramatically reduce the amount of data that needs in-depth annotation. The question now is: can automatic speech recognition contribute to the creation of such transcripts?

4. *Automatic Speech Recognition*

Today, automatic speech recognition is everywhere: one can dictate messages on a handheld device, request services or buy products via the internet, or communicate with household devices or cars. ASR has become a business. On the one hand this means that the technology has become affordable and sufficiently reliable for everyday applications, on the other hand this means that by focusing on commercially interesting languages and applications, the requirements of non-commercial applications, e.g. humanities research, receive little or no attention.

The most frequently used measure for ASR performance is the *word error rate* (WER). It computes the number of insertions, deletions and substitutions necessary to transform a hypothesis, i.e. the ASR output, to a reference or *gold standard*, i.e. the transcript generated by a human expert.

In some domains, e.g. medical or juridical dictation, ASR has reached human performance levels, with WERs of 2-5%; the same is true for travel or other information systems. In other domains, e.g. dictating personal messages, performance is very often good, but occasionally – and for no apparent reason – ASR fails miserably. And in yet other domains, e.g. linguistic field recordings, dialogues, colloquial conversation at home, and others, ASR consistently performs badly, with WERs of 40% and more.

For us humans, this is difficult to understand: highly complex application areas such as medicine or law, with their difficult terminology and non-natural way of saying things – how can a machine do it? And seemingly simple things, like having a conversation with grandfather in the living room about some family event, are impossible? How can this be?

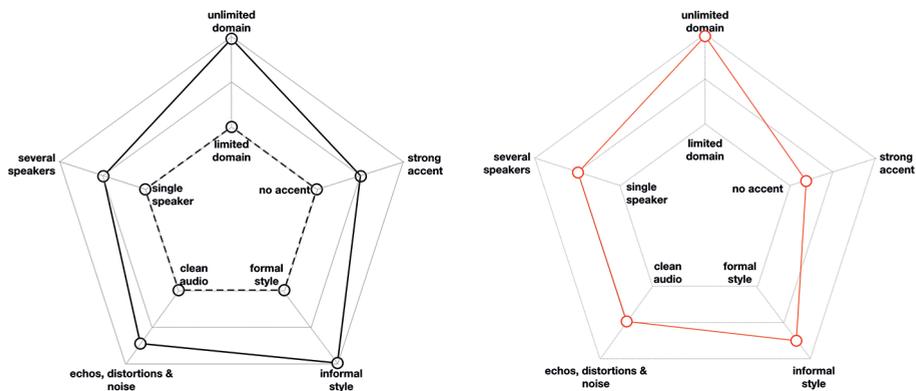
ASR performance depends on a number of factors. A simple graphical model, based on five factors, may help to estimate the quality of ASR for a given spoken language recording. Figure 7 shows a pentagon with the five factors (anti-clockwise):

1. Number of speakers
2. Audio quality

3. Speaking style
4. Dialect or accent
5. Domain of discourse

The innermost pentagon contains the values of the factors favourable to ASR: a single speaker, high quality recordings with very little background or technical noise, a formal style of speaking, e.g. reading or speech with no overlap, standard language, and a limited domain, i.e. a small vocabulary. The outermost pentagon displays values that are detrimental to ASR performance: several speakers, echoes and distortions in the signal, informal way of speaking, i.e. a lot of overlapping speech and frequent interruptions, language with a strong accent or dialect, and unlimited domains, i.e. virtually unrestricted vocabularies.

Figure 7 - *The left pentagon displays the values of the factors for the medical dictation (dashed line) and the family conversation (solid line). The right pentagon displays the values for the interview with the cucchiaino story*



With the help of the pentagon it becomes clear why medical dictation performs well: only one person speaks at a time, often headset microphones are used to keep the hands free and to reduce background noise, the communication is highly structured, doctors employ a clear language in their everyday professional communication, and, finally, medical fields have a limited terminology.

For the conversation with grandfather at home, the picture is different: there are many speakers, there is only one microphone on the table, people enter and leave the room, and maybe even the TV is on, the participants are familiar to each other and thus interrupt each other frequently, the language may be accented, dialectal or a family language, and, of course, one talks about almost anything imaginable, i.e. the domain is unlimited.

Both the medical dictation and the family conversation are shown on the left in Figure 7.

The interview fragment is shown on the right in the same figure: two speakers, audio signal with the high frequencies missing and a soft noise, a dialog with cross-

talk and frequent interruptions, Italian with a weak accent, and a potentially unlimited vocabulary with strong emotional content. The prediction is that the word error rate will be rather high for this particular recording.

In fact, when this interview fragment was processed with the Google Speech Cloud ASR using the BAS web services in January 2021, it generated the transcript in the right column of Figure 8. With text normalisation, i.e. converting all text to lower case, removing speaker labels, punctuation and other markers, the word error rate for this fragment is 48.5%⁷.

4.1 Analysis of the ASR-generated transcript

The output of the commercial ASR provider is a normalised and moderately smoothed verbatim transcript of the interview recording, with punctuation and capitalisation. Quite a lot of the spoken language signal was recognised faithfully.

However, there are severe problems: the ASR transcript

- is approx. 33% shorter (114 vs. 169 words) than the manual transcript because entire phrases are missing, e.g. the first two sentences, ‘la bestia lecca’, and ‘le mani sono uno strumento’;
- contains some totally unexpected words, e.g. ‘distrarre’, ‘Giusti’, ‘desti’;
- contains a few word duplications, e.g. ‘che che’, but far fewer than the manual transcript;
- is not diarised.

For a human reader, especially the two first problems are surprising: there is no apparent reason for the omission of sentences or the selection of these words.

Other errors are explicable, e.g. ‘tecnologicamente’ vs. ‘psicologicamente’, or ‘ma aquista ‘sto’ vs. ‘ma qui sta sto’ and ‘Eli. Allora’ vs. ‘E li allora’ because they sound similar, or ‘Campo’ as a capitalized named entity instead of a simple noun.

The transcript created by human transcribers is diarised, and it contains a number of interesting word or phrase repetitions that are missing in the automatically generated transcript. They are interesting because they illustrate the communication situation or give insight into the speaker’s emotional state:

- ‘sì sì certo certo’ etc. as backchannel feedback to the speaker to continue,
- ‘di di di’ and ‘mh’ as a consequence of searching for the right words, and
- ‘sempre la minestra, sempre minestra’ places extra emphasis on this topic.

Higher-level phenomena, such as the citation status of “Non siamo mica dei cani.”, require linguistic knowledge and insight into the communicative function of phrases within the utterance. They are thus outside the scope of ASR.

A transcript for research purposes must contain these phenomena to allow researchers to decide whether this recording will be analysed in more detail.

⁷ For the entire interview, which has longer passages where only one person is speaking, the WER is 37.9%. WER was computed using the `wer()` function in the `wersim` R-package by Jens Wäckerle (Proksch et al. 2018).

Today's commercial ASR does not deliver this type of transcript for two main reasons:

1. For maximum performance and efficiency for a given task, ASR has to be customised to this task, and because this is expensive, there has to be commercial interest.
2. Many of the phenomena of interest to research, e.g. dialects, accents, or vernacular language, or specific linguistic structures, or cultural or sociological analyses, do not have direct commercial potential.

As a consequence, research in ASR needs to address the requirements of humanities disciplines, e.g. oral history, sociolinguistics, and others.

Figure 8 - *Comparison of the ASR output and a manual verbatim transcript. The transcripts were diarised and formatted manually to improve legibility*

#	Spk	Manual verbatim transcript	Google Speech Cloud (Jan. 2021)
1	AMB	Del cucchiaio ecco, del perché non danno il cucchiaio.	
2	L	Del cucchiaio rientra, rientra nello stesso tipo di discorso:	Rientra nello stesso tipo di distrarre
3	AMB	Certo certo	
4	L	volendoti disumanizzare, ridurti a livello di bestie: la bestia lecca.	Volendoti disumanizzare le Giusti a livello di bestia.
5	AMB	Sì sì certo certo	
6	L	Allora istintivamente, tu già psicologicamente sei pronto a sentirti animale.	Allora Istintivamente tu già tecnologicamente sei sotto a sentirti animale.
7	AMB	Sì sì sì	
8	L	E tant'è vero che una delle prime cose che il deportato acquista, in campo, è il cucchiaio.	Tanto è vero che una delle prime cose che il deportato acquista in Campo ecco che hai.
9		Per non sentirsi bestia.	Per non sentirsi desti.
10		Cioè chi: tenta di reagire, o lo ruba, o o lo compra, o lo acquista con il pane, ma acquista 'sto cucchiaio per potersi sentire meno animale.	Cioè chi è tenta di reagire o lo ruba o lo compra all'acquista con il pane, ma qui sta sto cucchiaio per potersi sentire meno animale,
11		E perché se fosse stato qualcosa che si mh potesse mangiare con le mani, le mani le mani tutto sommato sono uno strumento.	Perché se fosse stato qualcosa che potresti mangiare con le mani in mano,
12	AMB	Sì sì certo certo	
13	L	Ma, visto che è sempre la minestra, sempre minestra, la devi leccare. Bere o leccare.	Visto che è sempre la minestra la devi leccare leccare
14		E lì allora ti senti e effettivamente a livello di di di, le frasi che ricorrevano che, che ricorrevano erano: "Non siamo mica dei cani".	Eli. Allora ti senti effettivamente a livello di Le frasi che che ricorre vero
15	AMB	Eh già: "Non siamo mica dei cani".	non siamo mica del cane.

4.2 Is transcribing based on ASR faster than purely manual transcription?

Currently, research is under way to estimate what level of ASR performance must be achieved to effectively speed up the generation of verbatim transcripts. In a pilot study performed in 2019 at our institute, 10 student presentations on the topic of ‘communication’ with a duration between 1:14 min and 5:34 min (total 38:14 min), recorded via a standard video camera in a lecture hall, were transcribed by two transcribers. Each transcriber generated 5 transcriptions from scratch, and 5 by correcting ASR output. Overall, the real time factor for the transcription from scratch was 9.43, and 8.52 for the transcription based on ASR – a speed-up of approximately 10%.

In this pilot study, the European Media Lab recogniser was used, and it achieved a WER of 68.9%. This might also explain why the real time factors are so similar: because of the high WER, the transcribers simply deleted some of the ASR-generated transcript and transcribed from scratch.

In 2021, the WER was computed for four of the student presentations and for three different ASR providers, and between the human-created transcripts. Table 2 shows that human-created transcripts are very close to each other, and that for this type of data – several speakers, low signal quality with reverberation, and unknown domain – the performance of ASR varies greatly.

Table 2 - *Averaged WERs for three different ASR systems and four selected student presentations; the last row displays the WER between the two human-generated transcripts*

<i>ASR</i>	<i>WER</i>
EML	68.9%
Google Speech Cloud	58.7%
Fraunhofer	23.2%
Human/human	3.7%

In a pilot study performed by Silvia Calamai’s group for this paper⁸, the interview of Anna Maria Bruzzone and Lidia Beccari Rolfi was re-transcribed (approximately 41 minutes). To compare transcription speed, the interview was split into five fragments of 5:28-11:25 minutes. Two fragments were transcribed from scratch, three by first running the Google Speech Cloud ASR and then manually correcting this transcript. Table 3 presents the results. Overall, re-transcription was very quick, with an average real time factor of 3.52. Transcribing from scratch was slightly faster than correcting the ASR generated transcript. As expected, the ASR WER is high. The relatively high WER in the comparison of the two transcripts created by human transcribers – especially with regard to the student presentations in Table 2 – may be attributed to the interview situation with overlapping speaker turns and the audio quality.

⁸ The data presented here was provided by Fabio Ardolino from Siena University.

Table 3 - Comparison of transcribing from scratch vs. correcting an ASR generated transcript

<i>Fragment</i>	<i>Playtime</i>	<i>Type</i>	<i>Real time factor</i>	<i>WER</i>	<i>Human/human</i>
BRZTO061a_16_0_0002	7:04	scratch	3.28	35.5%	11.6%
BRZTO061a_16_0_0003	8:38	ASR	3.79	42.1%	14.5%
BRZTO061a_16_0_0004	8.51	scratch	3.22	41.3%	14.0%
BRZTO061a_16_0_0005	11:25	ASR	3.41	40.5%	14.5%
BRZTO061a_16_0_0006	5:28	ASR	3.90	37.6%	14.5%
Average			3.52	39.5%	13.8%

Gorisch et al. (2020) is one of the few published studies on the use of ASR for the transcription of spoken language corpora. For different German corpora, it reports WERs between 13.3% and 30.0% for the Fraunhofer ASR, and correlates regional accents and ASR performance. In a private communication, the head of the spoken language archive at the Leibniz-Institut für Deutsche Sprache Mannheim, Thomas Schmidt, states, that for TV and radio broadcast with a WER of approximately 10%, the use of ASR is generally worthwhile. For interview data, it is recommended to run ASR tests on a subset of the corpus and then decide, and for natural conversations in field recordings it is more efficient to transcribe from scratch.

The following section describes how state-of-the-art ASR systems can be accessed via easy-to-use web services by academic users.

5. Web services at BAS

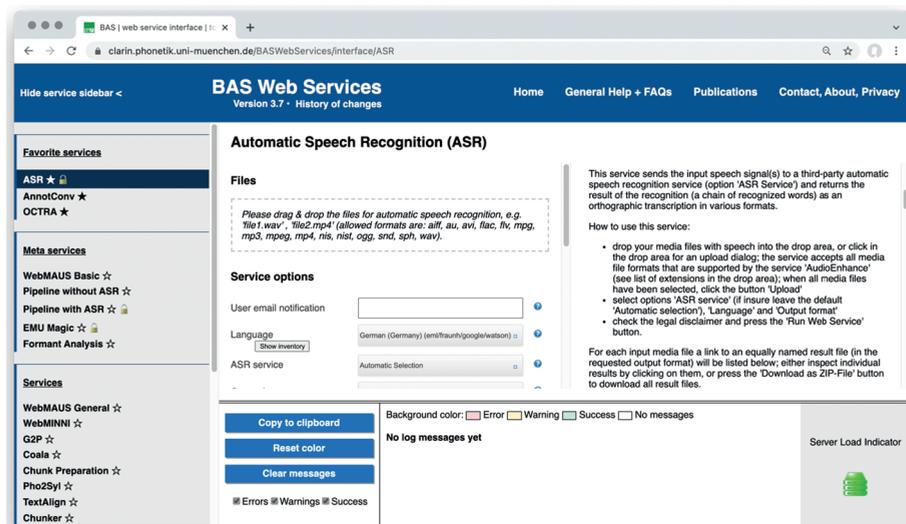
The Bavarian Archive for Speech Signals (BAS) is a German CLARIN centre. It operates a repository for spoken language resources, and it offers a range of web services which make available spoken language processing tools to non-technical users. The use of the BAS web services is restricted to members of academic institutions⁹.

The web services user interface is basically the same for all services (see Figure 9 for an example): on the left side, there is a sidebar with all available services; this sidebar can be hidden or shown. In the top middle is a panel with the file upload area and a form to set the service parameters. A click on the question marks next to the parameter fields opens a description of the selected parameter. On the right is a documentation of the service; again, this documentation can be hidden or shown. The bottom row is the feedback area: here, colour-coded status, warning or error messages are displayed, and the load indicator displays the current workload of the server.

Using a web service is easy: upload files from the local computer to the server, select a service and set some parameters, accept the conditions of use and run the service. Once a result has been computed, it can be viewed in the browser, e.g. using the Emu WebApp (Winkelmann 2015, Winkelmann et al. 2017), or downloaded to the local computer. Note that the most important parameters or ‘service options’ are always shown; many services also have optional parameters which can be accessed via a click on ‘expert options’.

⁹ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface>.

Figure 9 - Detail page for the ASR web service. On the left, the sidebar with the list of all services is shown, and on the right the manual page for the selected service



The following sections present a selection of web services of particular interest to linguists and phoneticians. The presentation follows the transcription workflow shown in Figure 2.

These services are (with the service names in parentheses)

1. Automatic Speech Recognition (ASR),
2. Grapheme-to-phoneme conversion (G2P),
3. Automatic phoneme and word alignment (WebMAUS), and
4. Anonymisation (Anonymizer)

Again, the interview with the *cucchiaio* story from the Bruzzone archive, or an excerpt from it, is used throughout this section to show how the services work and what they return.

5.1 Automatic Speech Recognition

The ASR web service calls on external ASR providers. Currently, these providers are

- European Media Lab (Germany),
- Fraunhofer Intelligente Analyse- und Informationssysteme (Germany),
- Google Speech Cloud (US),
- IBM Watson (US), and
- LST (Language and Speech Tools), Radboud University, (Netherlands).

Most providers support more than one language. The ASR service by LST is an academic service, and allows the selection of different content domains, e.g. conversational speech, parliamentary discussions, or oral history¹⁰.

¹⁰ Automatic Transcription of Dutch Speech Recordings, Language and Speech Tools, Radboud University, Nijmegen: <https://webservices.cls.ru.nl/oralhistory>.

The BAS web services use the free services of these providers. As a consequence, a number of limitations apply, depending on the provider. Such limitations concern the maximum audio duration, or a given quota per month, etc.¹¹ Note that some ASR providers keep the uploaded audio files, which is often not acceptable for privacy reasons.

The ASR web service requires an authentication – this makes sure that only academic users access the service. Members of a European academic institutions should be able to log in using credentials of the account at their home institution¹².

Once authenticated, the user can upload the audio files either via drag & drop to the marked area in the browser, or via a click on the marked area to select the files using the standard file selection dialog of the local computer. Note that several files can be loaded at once.

For ASR, the language to recognise, the provider of the service, and the output format must be set.

For the *cucchiaio* story, the language was set to Italian, Google was chosen as the ASR provider, and .txt as the output format. The result of ASR is shown in the middle column of Figure 8.

5.2 Grapheme-to-phoneme conversion

The tool G2P converts text in standard orthography into a phonemic representation. Technically, the service is based on statistic decision trees, part-of-speech tags and morphological segmentation. The service is trained on pronunciation dictionaries, or on a letter-sound table for languages with a unique correspondence between letters and sounds (e.g. Italian, Finnish). See (Reichel 2012) for details. G2P is available for 50+ languages and dialects, and users may upload their own letter-sound mapping table.

In Figure 10 the second sentence of the interview is processed by the G2P web service using Italian as the language, the output symbol inventory SAM-PA, and the output format .txt.

Figure 10 - *Grapheme-to-phoneme conversion for a sentence from the interview*

<i>Orthography</i>	<i>Pronunciation (SAM-PA)</i>
Del cucchiaio ecco, del perché non danno il cucchiaio.	d e l k u k : j a j o E k : o d e l p e r k e n O n d a n : o i l k u k : j a j o

¹¹ To lift some of these restrictions, users may purchase a key. With this key, they will be charged for the use of the service.

¹² As an alternative, users may apply for a CLARIN account with CLARIN-EU: <https://idm.clarin.eu/user/home>.

5.3 Automatic phoneme and word segmentation

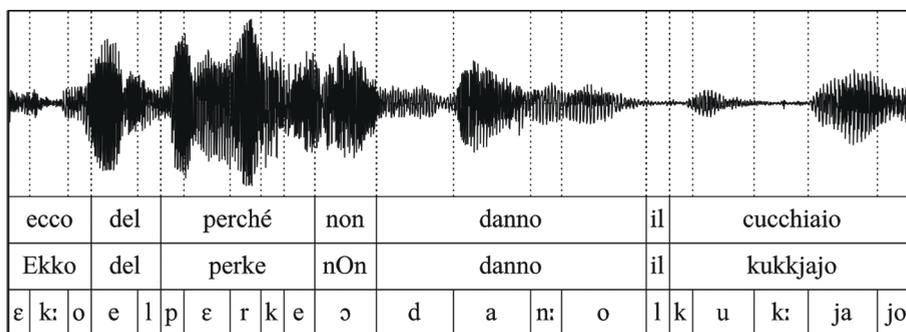
The Munich Automatic Segmentation (MAUS) time-aligns an audio file and its orthographic transcript to return a multi-tier phonetic transcript. MAUS internally generates pronunciation hypotheses from the orthographic transcript using G2P, and then computes the phoneme sequence that matches the audio file best. MAUS is available for 50+ languages. Furthermore, a language-independent mode takes as input a phoneme string in SAM-PA and thus allows the segmentation of languages for which there is no dedicated model. See (Kisler et al. 2012) for further information.

MAUS exports segmentations to different output formats, such as .TextGrid for Praat, .caf for ELAN, or .csv for spreadsheets or statistics packages. MAUS comes in two flavours: WebMAUS Basic with a limited set of options, and WebMAUS General with a large number of configuration parameters.

Figure 11 displays a time-aligned transcript, with orthographic words in the top level, the canonic word pronunciation in the middle and phoneme segments on the bottom level. The input format was BAS partitur file format, the language Italian, the output format Praat TextGrid, and the output in the IPA alphabet in UTF-8 encoding.

Here, the actual pronunciation deviates from the canonical pronunciation: the /d/ in ‘del’, both /n/ in ‘non’ and the /i/ in ‘il’ are not in the phonemic segmentation. This may be due to coarticulation in fluent speech, but also due to low signal quality of the recording – here, it is probably both¹³.

Figure 11 - MAUS segmentation of the second sentence of the transcript of the *cucchiaio* story



WebMAUS works best for high quality audio files with a single speaker. Because processing time is quadratic, WebMAUS works fast only for short files. In practice, 10 min or 3000 words are recommended as maximum file duration or transcript length. To process longer files, the use of a pipeline service with the chunker service is recommended (Pörner & Schiel 2018).

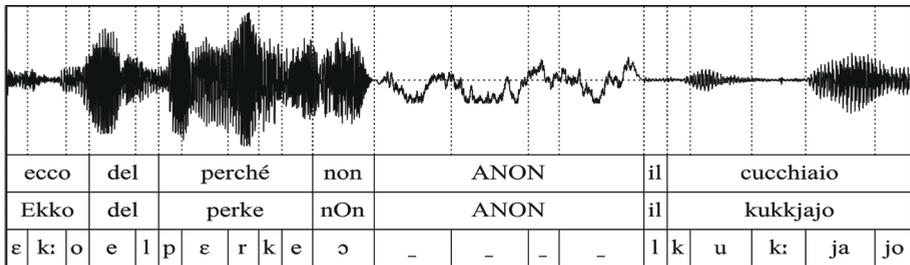
¹³ Of course, we also cannot rule out an error in the segmentation algorithm.

The performance of automatic segmentation systems has been shown to depend strongly on the type of language and the method applied for the comparison (for details, see e.g. (McAuliffe et al. 2017, or Meer 2020)).

5.4 Anonymizer

For interview data, it is often necessary that both transcripts and audio files be anonymised. The Anonymizer web service takes as input an audio file and its time-aligned segmentation plus a list of terms to anonymise. The service searches for these terms in the segmentation, extracts the boundaries from the segmentation, replaces the transcript text with a marker symbol, and replaces the matching signal fragment with brown noise.

Figure 12 - *Output of the Anonymizer service with the word ‘danno’ anonymised. Note that in the signal, the fragment corresponding to the word is replaced by noise, and the word is removed from the transcript*



5.5 Pipeline services

The web services presented in the previous sections were run individually. This means that for every service, files have to be uploaded to the server, and parameters have to be set. In the transcription workflow, the output of one service often is the input to the next, and thus it makes sense to organise individual services into a pipeline to reduce the number of file transfers and manual interactions.

BAS provides such pipeline services as pre-configured sequences of individual services. Files are uploaded only once and then are passed on from one process to the next until done. Pipelines not only streamline the use of the services, they also make them safe by allowing only meaningful combinations of services and parameters (see Kisler et al. 2017 for details).

For example, to run the four services from the previous sections, one has to upload the audio file three times (for ASR, WebMAUS, and Anonymizer), and download a .txt, a .par and two .TextGrid files.

The same result can be achieved by using the BAS pipeline service *ASR→G2P→MAUS→Anonymizer*. The audio file needs to be uploaded only once, the services are executed sequentially on the server, and the final output file is returned. Of course, the pipeline can be parameterised. For example, by selecting the output format .csv (for comma-separated values), the resulting table can be import-

ed directly into a spreadsheet or statistics package, e.g. to generate a frequency list, perform phoneme counts, or calculate mean durations and variance (see Table 4 for an example).

5.6 Final remarks on the BAS web services

The graphical user interface of the BAS services allows the user to upload any number of files; on the local computer, they may reside in different folders. During upload, and if required by a given service, the server will try to find matching files i.e. files with the same base name and different extensions. Non-matching files will be reported.

If processing is expected to take long, users may choose the email notification option: the BAS services notify the user via email when processing is done, and the email contains a link to the result in the format of a compressed archive file.

Finally, all BAS web services may be accessed programmatically via an API. This is particularly useful for very large numbers of files or repeated tasks, or when working in environments such as jupyter notebook, R Studio, or the command line. See the FAQ section and the documentation of the web services for further details. A number of transcription and annotation editors, e.g. Octra or ELAN, may call the BAS web services in the background. This allows running the services on the data currently open in the editor, without leaving the application.

Table 4 - *Segmentation data imported into a statistics package (formatted to improve legibility)*

<i>Token</i>	<i>Begin (s)</i>	<i>Label</i>	<i>Duration (ms)</i>
...			
41	20.070	una	130
42	20.200	delle	500
43	20.700	prime	360
44	21.060	<P>	60
45	21.120	cose	480
46	21.600	<P>	80
47	21.680	che	140
48	21.820	il	30
49	21.850	<P>	50
50	21.900	deportato	700
51	22.600	<P>	100
52	22.700	acquista	430
53	23.130	in	90
54	23.220	<P>	90
55	23.310	Campo	550
...			

6. *Summary and outlook*

The transcription of spoken language recordings is a time-consuming task, and ASR promises to speed up the generation of transcripts. Currently, ASR works well under specific conditions – the ASR pentagon can give an estimate of the expected quality of ASR output. For contemporary or future recordings, the specific requirements of ASR can be accounted for already during the recording, and thus ASR will speed up the generation of broad or verbatim orthographic transcripts. For legacy recordings, often with low audio quality, strongly accented speech, unlimited domains etc. ASR does not yet perform well enough to effectively increase transcription speed – correcting an ASR-generated transcript with a high word error rate is often slower than transcribing from scratch.

Two developments are needed:

- the workflow in the humanities needs to be adapted to make the best use of tools and resources, and
- research and development in ASR should focus on the specific requirements of the humanities.

Currently, orthographic transcriptions and in-depth annotations of recordings are often performed in one step by the same highly-trained transcriber. A modularisation of the workflow, which separates transcription from in-depth annotation, divides hour-long recordings into meaningful units of shorter length, and uses the proper tools at each stage of the workflow, promises more flexibility and a more efficient use of human resources. In such a modularised workflow, new technology, e.g. ASR, can be introduced for specific tasks without adverse effects on others. The key to this modularisation is a smooth flow of data from one processing step to the other, which in turn means that the tools used in the workflow must support each other's formats.

The specific requirements of the humanities with regard to ASR, such as truly verbatim orthographic transcripts with e.g. word repetitions, backchannel feedback, turn taking and pauses, are not well supported by current ASR systems. These topics are open research questions, and humanities scholars and speech technology researchers need to collaborate to answer them.

7. *Acknowledgements*

I thank the team at BAS (Markus Jochim, Thomas Kisler, Julian Pömp, Nina Pörner, Florian Schiel, Fritz Seebauer, Raphael Winkelmann) for developing and maintaining the web services and tools. This work was funded by the German Federal Ministry of Education and Research under contract #01UG1620B. I also thank the Oral History group (Jeannine Beeken, Silvia Calamai, Arjan van Hessen, Henk van den Heuvel, Stefania Scagliola) for the continued discussions and collaboration, and CLARIN-EU for funding parts of this work.

References

- Beccaria Rolfi, Lidia & Bruzzone, Anna Maria. 2020. *Le donne di Ravensbrück. Testimonianze di deportate politiche italiane*. Torino, Einaudi.
- Boersma, Paul. 2001. Praat, a System for doing Phonetics by Computer. *Glott International* 5(9/10). 341-245.
- Draxler, Christoph & van den Heuvel, Henk & van Hessen, Arjan & Calamai, Silvia & Corti, Louise & Scagliola, Stefania. 2020. A CLARIN Transcription Portal for Interview Data. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, 3353-3359.
- Draxler, Christoph & Jänsch, Klaus. 2004. SpeechRecorder – A Universal Platform Independent Multi-Channel Audio Recording Software. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, 559-562.
- Gorisch, Jan & Gref, Michael & Schmidt, Thomas. 2020. Using Automatic Speech Recognition in Spoken Corpus Generation. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, 6424-6428.
- Kisler, Thomas & Reichel, Uwe D. & Schiel, Florian. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326-347.
- Kisler, Thomas & Schiel, Florian & Sloetjes, Han. 2012. Signal Processing via Web Services: the Use Case WebMAUS. In *Proceedings of the Digital Humanities Conference 2012*, 30-34.
- McAuliffe, Michael & Socolof, Michaela & Mihuc, Sarah & Wagner, Michael & Sonderegger, Morgan. 2017. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the Annual Conference of the International Speech Communication Association*, 498-502.
- Meer, Philipp. 2020. Automatic Alignment for New Englishes: Applying state-of-the-art Aligners to Trinidadian English. *Journal of the Acoustical Society of America*, 147(4). 2283-2294. DOI: 10.1121/10.0001069.
- Pömp, Julian & Draxler, Christoph. 2017. OCTRA – A Configurable Browser-based Editor for Orthographic Transcription. In: *Tagungsband der 13. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, 145-148.
- Pörner, Nina & Schiel, Florian. 2018. A Web Service for Presegmenting Very Long Transcribed Speech Recordings. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*.
- Proksch, Sven-Oliver & Wratil, Christopher & Wäckerle, Jens. 2018. *Testing the Validity of Automatic Speech Recognition for Political Text Analysis*. Political Analysis.
- Reichel, Uwe D. 2012. PermA and Balloon: Tools for String Alignment and Text Processing. In *Proceedings of the Annual Conference of the International Speech Communication Association*, paper no. 346.
- R Core Team (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schiel, Florian. 2015. A Statistical Model for Predicting Pronunciation. In *Proceedings of the International Congress of Phonetic Sciences*, paper 195.

Sloetjes, Han. 2007. ELAN: a Free and Open-source Multimedia Annotation Tool. In *Proceedings of the Annual Conference of the International Speech Communication Association*, 4015-4016.

Spreafico, Lorenzo. 2020. Corpora di parlato o corpora di ascoltato? *Rivista italiana di dialettologia* 44, 37-51. <https://hdl.handle.net/10863/18479>.

Vangelisti, Petra & Pesce, Caterina & Setaro, Marica & Bianchini, Greta & Gigli, Lucilla & Calamai, Silvia. 2019. *Ritrovare Voci: il lavoro intorno all'archivio di Anna Maria Bruzzone*. DOI: 10.17469/O2106AISV000009.

Winkelmann, Raphael & Harrington, Jonathan & Jansch, Klaus. 2017. Emu-SDMS: Advanced Speech Database Management and Analysis in R. *Computer Speech and Language* 45, 392-410.

Winkelmann, Raphael. 2015. Managing Speech Databases with emuR and the Emu-webApp. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2611-2612.

Approcci digitali alla lessicografia: sviluppo e valutazione di un dizionario per smartphone

Abstract

With the aim of addressing lexicographic issues for developing dictionaries in a smartphone format, this paper describes the design phases of *Idiomatica*, a dictionary app of Italian idioms for L2 learners.

Reviewing the major breakthroughs in electronic lexicography, challenging topics in the field are foregrounded in the discussion. Although lexicography has significantly contributed to develop research methods in linguistics, especially with corpus studies, the way dictionaries are designed stands outdated and poor. To address these issues, two different prototypes of *Idiomatica* have been designed and tested with their perspective users, i.e. L2 Italian students, by following a human-centred design approach. Special attention in the paper is given to the microstructural components, the lexicographic data provided, and the design protocols adopted for developing the dictionaries so far.

1. I dizionari e l'apprendimento delle espressioni idiomatiche

Studi longitudinali sull'acquisizione della fraseologia hanno dimostrato che il lessico di una L2 tende a crescere nel tempo quantitativamente ma non qualitativamente. I parlanti con una competenza intermedia (pari al B1 del QCER, Council of Europe) usano un numero maggiore di espressioni multiparola rispetto ai principianti ma le combinazioni da loro prodotte sono per lo più scorrette, facendo registrare inversioni nell'ordine dei costituenti, sostituzioni con sinonimi, usi impropri di preposizioni o determinanti. Alcuni autori (Siyanova-Chanturia e Spina, 2019) ipotizzano che col crescere delle abilità di gestione delle regole morfo-sintattiche i non nativi si affidino a principi di composizionalità per selezionare i collocati delle parole e siano poco attenti alle restrizioni sintagmatiche che vincolano il lessico della L2. Tra i fraseologismi, la classe delle espressioni idiomatiche si contraddistingue per complessità semantica, formale ma anche pragmatica e viene indicata come parte del lessico per la quale gli apprendenti necessitano di un supporto specifico (Boers & Lindstromberg, 2009).

Come sostiene anche Nation (2006: 450), i dizionari sembrano gli strumenti più indicati a supportare l'apprendimento di queste espressioni multiparola, ma la loro efficacia è ancora scarsamente esplorata. Il formato digitale può peraltro offrire

¹ Università degli Studi di Napoli 'L'Orientale'.

² Università degli Studi Suor Orsola Benincasa.

diversi vantaggi non solo per la presentazione dei dati, ma anche per l'ubiquità della consultazione, che può avvenire in tutte le situazioni di scambio linguistico: dalla classe di lingue all'interazione nella vita quotidiana.

In questo studio vengono presentate le ricerche condotte sin qui per sviluppare *Idiomatica*, un dizionario per smartphone delle espressioni idiomatiche italiane destinato ad apprendenti stranieri. Dopo un inquadramento generale sui progressi della lessicografia nell'era digitale (§2) vengono discusse le problematiche della lessicografia mobile (§2.2 e 3) e quelle connesse alla descrizione delle espressioni idiomatiche (§4, 4.1, 4.2). Seguirà una sintesi delle diverse fasi di progettazione dell'app dizionario (§5 e 5.1). La metodologia progettuale adottata segue le linee guida dello *Human-centred design for interactive systems* (ISO 9241-210) per permettere al dizionario di essere sia utile sia facile da usare.

2. Il formato digitale e la rivoluzione lessicografica

Secondo Jones (2014) tra il 2004 e il 2008 le app per smartphone hanno contribuito all'intrusione massiccia dei dati digitali nella vita quotidiana, inaugurando una rivoluzione che ha investito anche la ricerca scientifica con la nascita delle *Digital Humanities*. In un ambito di studi tanto giovane e multidisciplinare, la lessicografia è stata antesignana di diversi processi, se si pensa che la migrazione al formato digitale è iniziata già nel 1978 con il primo dizionario compilato mediante un database lessicale elettronico, il *Longman Dictionary of Contemporary English*. Nove anni più tardi, nel 1987, sotto la direzione di John Sinclair (1987), viene pubblicato invece il *COBUILD Dictionary of English*, i cui dati sono stati interamente estratti da un corpus. Le nuove tecnologie usate per il *COBUILD* hanno peraltro contribuito ad implementare anche i modelli teorici, segnando la nascita di un nuovo approccio allo studio della fraseologia basato sull'analisi distribuzionale (Sinclair, 1987, 1991, 2004) e della frequenza con cui le parole co-occorrono. È emersa in questo modo con grande evidenza la dimensione formulaica delle lingue, poiché gli indici statistici di associazione tra le parole consentono di rilevare espressioni multiparola dal significato compositivo che approcci semantici alla fraseologia tendono a disconoscere. Se in *accusare un dolore/fastidio* il significato inusuale del verbo aiuta ad evidenziare una restrizione sintagmatica, è difficile riconoscere come certe combinazioni siano modi preferenziali per esprimersi e realizzare specifiche funzioni pragmatiche (*buongiorno, come stai*), argomentative (*al di là di, per quanto riguarda*) o testuali (*viene da, non appena, dato che*).

Sul fronte dei contributi teorici apportati dalle analisi lessicografiche va purtroppo lamentato il predominio delle riflessioni sulla lingua inglese e sui suoi dizionari, coadiuvato probabilmente dal perdurare di un certo scetticismo da parte del mondo accademico di alcuni paesi nei confronti della lessicografia come disciplina scientifica. Questi due fattori hanno forse determinato la scarsa eco di altri contributi significativi apportati dalla descrizione lessicografica non anglofona, come ad esempio l'introduzione della classe dei *verbi procomplementari* nel *Grande Dizionario*

Italiano dell'Uso (GRADIT) di Tullio De Mauro, anch'esso compilato a fronte di una intensa ricerca sulla frequenza del lessico. Questi verbi³, presenti anche nelle altre lingue romanze, sono caratterizzati dalla lessicalizzazione di uno o più pronomi in sostituzione degli argomenti verbali: in *prenderle* e *darle*, ad esempio, *le* sostituisce *le botte*. La corretta lemmatizzazione dei verbi procomplementari anche nelle altre lingue romanze consentirebbe di migliorare tutti i tipi di dizionari, da quelli bilingui ai dizionari macchina impiegati per il Natural Language Processing, dal momento che la disambiguazione dei pronomi è indispensabile per la comprensione o l'elaborazione della semantica frasale. Ad esempio, *finiscila* in *finiscila altrimenti le prendi*, può essere anaforico di un altro elemento testuale, come *la pasta*, oppure lessicalizzare il concetto di 'smettere di fare capricci'.

L'analisi condotta oggi dai lessicografi sui testi digitali dei corpora ha dato quindi contributi anche alla teoria linguistica, tuttavia ogni dizionario che venga fruito su un dispositivo elettronico è anche un software e come tale deve rispondere a criteri di efficacia ed efficienza (Heid 2011) per aiutare gli utenti a risolvere problemi pratici, soprattutto di natura comunicativa. Oggetto di interesse della disciplina è quindi anche il dizionario in sé per sé, uno strumento d'uso (*utility tool*, Wiegand 1998) che deve aiutare a trasformare i dati collezionati dal lessicografo in informazioni (Wiegand 2000; 2002) capaci di colmare le lacune conoscitive degli utenti.

Su queste basi, già negli anni '70 del Novecento il linguista tedesco Herbert Ernst Wiegand aveva messo in risalto gli elementi utili a fondare una teoria lessicografica che trasformasse una prassi millenaria in un campo di ricerca scientifica autonomo. Ma è a partire dagli anni '90 che gli studi sulla compilazione e l'uso dei dizionari aumentano esponenzialmente, un progresso che avviene quando i dizionari passano dal supporto cartaceo a quello digitale. Il dispositivo ospite sembra diventare quindi l'elemento più riconoscibile di questa transizione, mentre il formato testuale — cartaceo o digitale — in cui il dizionario viene compilato (Lew & De Schryver 2014) non è tenuto in altrettanto credito. Si parla infatti di 'dizionari elettronici' oppure, e soprattutto a partire dal 2008, di 'dizionari online' mentre i 'dizionari digitali' sono richiamati da una parte assolutamente residuale degli studi e solo in anni più recenti⁴. I nomi rispecchiano peraltro gli ambiti d'interesse: le modalità di fruizione da parte degli utenti sono richiamate dai 'dizionari online' e i processi di compilazione — soprattutto automatica — delle risorse sono evocati dai 'dizionari elettronici' e sono di gran lunga i più studiati.

³ L'alta frequenza nell'italiano parlato di verbi con pronomi lessicalizzati, sia intransitivi pronominali sia procomplementari, è stata evidenziata già da tempo da diversi autori (Sabatini, 1985; Berretta 1986, D'Achille 1990, 2006) e viene considerata come una caratteristica tipica dell'italiano 'neostandard' (Berruto 1987).

⁴ A partire dal 2013, si veda Lew & De Schryver (2014). I 'dizionari digitali' sono tutti quelli il cui testo viene fruito dagli utenti in formato digitale. È una denominazione alternativa a 'lessicografia elettronica'.

A dispetto di questo interesse per gli aspetti compilativi automatici, la ricerca sui dispositivi mobili rappresenta un ambito d'interesse ancora marginale e spesso i dizionari migrano dalle versioni desktop agli smartphone senza adeguamenti significativi. La ricerca condotta per la compilazione di *Idiomatica* intende quindi sollecitare nuovi approcci alla progettazione dei dizionari, partendo proprio dalle modalità di presentazione e fruizione dei dati, seguendo protocolli in uso nell'ergonomia e nella ricerca sull'interazione tra uomo e macchina.

2.1 Dal tascabile al cliccabile: problemi di formato

I primi dizionari elettronici in formato ridotto, i così detti *Pocket Electronic Dictionaries*, hanno iniziato a diffondersi negli anni Novanta soprattutto nei paesi dell'est asiatico, diventando molto popolari (Taylor & Chan, 1994) e desiderati (Nesi & Boonmoh 2009) tra gli studenti per la loro portabilità e facilità d'utilizzo anche durante le lezioni. Due dispositivi di questo genere erano stati già commercializzati nel lontano 1978 adattando a questo scopo le calcolatrici elettroniche, uno dei produttori era infatti la Texas Instruments. I contenuti offerti erano inizialmente molto poveri ma già a partire dal 1995 comparvero adattamenti dei maggiori dizionari inglesi⁵ ai quali, successivamente, furono affiancate altre risorse bilingui di "qualità variabile" (Nesi & Boonmoh 2009: 68). Un singolo dispositivo poteva così contenere anche più di trenta dizionari.

In anni più recenti gli smartphone hanno soppiantato i *Pocket Electronic Dictionaries* ma, a fronte di nuove potenzialità hardware, i dizionari ospitati in questi dispositivi continuano ad essere adattamenti delle versioni desktop. Il motivo del ritardo nello sviluppo di strumenti specifici per i dispositivi mobili è ascrivibile alle molteplici competenze che la progettazione di un app dizionario⁶ richiede: da un lato quelle linguistico-lessicografiche, dall'altro quelle ideative e tecniche necessarie allo sviluppo di applicazioni facili da usare. Per questo motivo gli editori collaborano con gli sviluppatori di software per produrre i loro app dizionari e distribuirli negli *store* (Winestock & Jeong 2013: 119)⁷. Tipicamente, quando l'editore non vuole investire tempo e denaro nell'impresa, il dizionario viene acquistato dal produttore di software che decide interamente le caratteristiche del prodotto finale. In caso contrario, l'editore chiede allo sviluppatore di realizzare un software che abbia delle caratteristiche specifiche. Entrambi gli approcci fanno prevalere un tipo di competenza sull'altra e mettono in secondo piano la fruibilità della risorsa.

L'obiettivo di un software più rispondente alle esigenze degli utenti dovrebbe essere infatti perseguito in maniera sinergica da esperti di tutte le aree di specializ-

⁵ Il dizionario in questione era lo *Hand-Held Longman Dictionary of Contemporary English* (1995), prodotto da Seiko.

⁶ App dizionario è un composto maschile parafrasabile come 'dizionario compilato come applicazione mobile', resa italiana dell'inglese *dictionary app*. Entrambe indicano i dizionari sviluppati come applicazioni per dispositivi mobili (Caruso et al. 2019).

⁷ Winestock & Jeong (2013) evidenziano quanto il business degli app dizionari sia residuale, aggirandosi tra lo 0.27% e l'1.25% del ricavo annuale del software per smartphone.

zazione attinenti allo strumento da realizzare, come indicato dalle norme ISO dello *Human-centred design for interactive systems* (ISO 9241-210). In questo documento vengono tracciate le linee-guida per progettare software e hardware facili da usare, che garantiscano una buona interazione tra l'utente umano e il sistema. I fondamenti del protocollo prevedono il coinvolgimento attivo dell'utente sin dalle prime fasi della progettazione di un prodotto. Al contempo è necessario delineare le funzioni e le attività che devono essere realizzate dall'uomo e quelle a carico del sistema, mentre la progettazione deve essere intesa come un'attività ciclica di prototipazione e validazione di soluzioni progettuali. Inoltre, protocolli e *framework* diversi sono stati introdotti per coinvolgere attivamente gli utenti. Alcuni di questi sono stati usati per ideare i prototipi di *Idiomatica* che verranno presentati più avanti come esempio di una ricerca lessicografica interamente costruita attorno all'utente.

Da un lato, infatti, i lessicografi riconoscono che la qualità di un dizionario è misurabile con la sua facilità di utilizzo o, come dice Rundell (2015: 27): “se gli utenti non riescono a trovare rapidamente quello che stanno cercando, la colpa è sicuramente del dizionario”. Dall'altro, la lessicografia ha bisogno di estendere e rinnovare le sue prassi editoriali per realizzare “dizionari progettati così bene da poter essere usati in maniera assolutamente intuitiva” (*ib.*). Le sperimentazioni progettuali descritte in queste pagine mostreranno quanto sia indispensabile il contributo degli utenti per costruire un dizionario rispondente alle loro esigenze. Spesso i desiderata che hanno espresso o le modalità di consultazione da loro adottate hanno determinato significativi ripensamenti progettuali di *Idiomatica*.

3. Progettazione di *Idiomatica*: dai dati lessicografici alle sperimentazioni sui prototipi

Qualsiasi dizionario, cartaceo o elettronico, per supportare i propri utenti in diversi compiti comunicativi deve offrire diverse tipologie di dati lessicografici che siano facilmente interpretabili dall'utente. Per i dizionari elettronici, la qualità viene valutata in base all'efficacia e all'efficienza, due parametri mutuati dall'informatica (Heid 2011). Un dizionario efficace consente all'utente di svolgere correttamente un compito perché i suoi 'dati' rappresentano un input che si trasforma compiutamente in un *intake*, o 'informazione' secondo i lessicografi (Wiegand 2000, 2002), garantendo in questo modo il successo della consultazione. L'efficienza stima invece la rapidità con cui avviene una consultazione efficace ed è quindi una buona misura della facilità di utilizzo della risorsa. Sia la qualità dei dati, sia la modalità di presentarli vanno perciò correttamente pianificati quando si progetta un dizionario.

Nei paragrafi che seguono verranno illustrati, da un lato, i tipi di dati lessicografici scelti per i parlanti non nativi che useranno il dizionario e, dall'altro, la loro organizzazione microstrutturale nello spazio di presentazione offerto dalle schermate (o *view*) di uno smartphone.

4. Dalla definizione alla descrizione delle espressioni idiomatiche

Come scrivono Dobrovolskij & Piirainen (2005: 40), gli *idioms* sono espressioni multiparola caratterizzate da una struttura lessicale fissa e, allo stesso tempo, da una semantica opaca e/o duplice, dal momento che esprimono il loro valore referenziale indirettamente, attraverso la metaforizzazione (il *target concept*) di un significato letterale (il *source concept*). L'idiomaticità e la figuratività di una espressione possono essere infatti valutate in base alla presenza di un'immagine (*image requirement*) soggiacente e alla semantica non standard (*semantic opacity*) del costituente principale del fraseologismo (*key constituent*), che in casi estremi può essere un costituente unico (*unique constituent*) della lingua in questione: un lessema usato esclusivamente in una specifica espressione idiomatica, come 'lusco' nell'italiano *tra il lusco e il brusco*. Gli autori rimarcano inoltre la dimensione connotativa degli *idiom*, suggerendo un test efficace per verificare l'idiomaticità di un'espressione: trovarne un sinonimo. *Tra il lusco e il brusco* ha per esempio due diversi sinonimi, di cui uno, *al crepuscolo*, è di registro letterario, mentre l'altro, *al tramonto*, non ha connotazioni specifiche. Viene così posto in evidenza un aspetto molto trascurato dagli studi, ovvero il fatto che le espressioni idiomatiche sono delle modalità alternative, e sempre connotate, di lessicalizzare un concetto, un assunto che dovrebbe essere più centrale negli approcci didattici a queste unità del lessico.

Sinora, invece, la dimensione figurativa ed etimologica sono state le sole basi di partenza delle sperimentazioni sull'apprendimento (Steinel et al. 2007), anche di taglio lessicografico (Szczepaniak & Lew 2011). Il valore semantico delle espressioni determina infatti anche le loro restrizioni flessive e sintattiche. Si pensi all'ammissibilità della diatesi passiva o dell'imperativo per esprimere un ordine in: *dammi una mano* o una *grossa mano è stata data dai cittadini*, entrambe compatibili con il significato di 'aiutare'. Mentre *taglia la corda*, col significato di 'fuggire' non ammette il passivo (**la corda è stata tagliata da te*) e, analogamente, **mettiti le mani nei capelli* non è un comando ammissibile nella stessa misura di **disperati*, a meno che non vengano usati per fare una concessione: *ma disperati anche se vuoi, mettiti pure le mani nei capelli*.

Come questi esempi dimostrano, il principio della sinonimia introdotto da Dobrovolskij & Piirainen si rivela essenziale per la descrizione semantica e morfosintattica, ma richiama anche l'attenzione sulla dimensione connotativa delle locuzioni.

4.1 Dall'OWID – *Sprichwörterbuch* a *Idiomatica*

La scelta dei dati lessicografici da inserire nel dizionario è partita da un'indagine preliminare (Caruso 2016) sull'efficacia informativa di un autorevole dizionario monolingue come il *Treccani*. In un esercizio di completamento di frasi con una espressione idiomatica scelta da un elenco, un gruppo di studenti (8 cinesi e 10 vietnamiti) di italiano L2 di livello avanzato (B2-C1 del QCER) a cui erano state fornite le glosse del *Treccani*, è riuscito a svolgere correttamente l'esercizio solo nel 56% dei casi. Più della metà degli errori commessi (56%) è risultata essere di natura lessicale: le frasi venivano completate o con l'espressione idiomatica sbagliata o

modificandone impropriamente i costituenti. Il test ha rivelato come i dati comunemente offerti da un dizionario monolingue per descrivere i modi di dire non siano sufficienti a colmare le lacune informative di parlanti non nativi.

Per definire quindi gli *item*⁸ microstrutturali di *Idiomatica* è stato preso a modello il dizionario online *OWID – Sprichwörterbuch* del Leibniz-Institut für Deutsche Sprache (IDS, Steyer & Ďurčo, 2013) che offre descrizioni dettagliate dei proverbi tedeschi, tenendo in conto aspetti semantici, come il significato e le particolarità d'uso (*Bedeutung, Gebrauchsbesonderheiten*), formali, ovvero i componenti di base e le possibili sostituzioni lessicali e variazioni di forma (*Basiskomponenten, Ersetzung von Komponenten, Formvarianten*) e testuali (*Typische Verwendung im Text*). A partire da queste componenti, per elaborare *Idiomatica*, la microstruttura dell'*OWID – Sprichwörterbuch* è stata arricchita tenendo in considerazione a) il tipo di unità lessicali da descrivere – le espressioni idiomatiche sono sintagmi mentre i proverbi dell'*OWID – Sprichwörterbuch* sono brevi testi, b) i tipi di utenti del dizionario, c) i compiti o funzioni lessicografiche (Tarp 2008) per le quali il dizionario viene utilizzato: compiti comunicativi, di produzione e ricezione, e la più generale funzione didattica che il dizionario deve svolgere.

4.2 Gli item microstrutturali di *Idiomatica*

Vengono qui descritti i dati microstrutturali (o *item*⁹) di *Idiomatica* presentati con le etichette metalinguistiche (*data-identifying entries*, Gouws 2014) scelte per aumentare l'usabilità del dizionario dopo le sessioni di *testing* descritte oltre (§5). Come si leggerà in seguito, infatti, le sessioni sperimentali hanno messo in evidenza le difficoltà di comprensione da parte degli utenti di alcuni tipi di dati offerti dal dizionario. La descrizione che segue ricalca la successione microstrutturale degli *item*.

Ogni voce contiene un'area informativa principale dove è possibile trovare la spiegazione semantica (*comment on semantics*) dell'*idiom* come prima informazione sotto il lemma (fig. 1)¹⁰ sia nel dizionario cliccabile sia nello scrollabile, a questa segue un esempio d'uso e le possibili sostituzioni dei costituenti lessicali dell'*idiom* ricavate dai corpora¹¹ e indicate come *varianti* nel dizionario.

⁸ Nella *General Theory of Lexicography* di Wiegand con *item* si indicano i tipi di dati inseriti nella microstruttura di un dizionario: ad es. *item giving the lemma sign* rappresenta il lemma, *item giving the usage example* rappresenta gli esempi d'uso.

⁹ Per ogni componente del dizionario, si forniscono tra parentesi i descrittori metalessicografici dalla *General Theory of Lexicography*.

¹⁰ Un esempio di come si presenta una voce completa di *Idiomatica* è mostrato in fig. 3a per il dizionario cliccabile e in 3b per lo scrollabile.

¹¹ I corpora usati sono ITenTen16 di Sketchengine e Paisà. Ulteriori ricerche mirate a verificare le intuizioni del lessicografo sono state condotte sul *Repubblica corpus*, *Google libri* oppure sui siti dei giornali *La Repubblica* e *Il Sole 24 Ore*.

Figura 1 - Area informativa principale di una voce di Idiomatica

Avere le mani legate
Se qualcuno ha le mani legate, si trova in una situazione in cui non può agire liberamente.
<i>Es. Bisognava reagire, eppure tutti avevano le mani legate</i>
VARIANTI:
Al posto di 'avere': vedersi, sentirsi (o altri verbi di percezione) <i>"Qui, chi comincia a fare qualcosa - ha continuato Pappalardo - poco tempo dopo si vede le mani legate"</i>
Al posto di 'avere': stare, rimanere (o altri verbi di stato) <i>"La Commissione giudicante, così, si trovò le mani legate e fu costretta a decidere".</i>
Senza verbo (con le mani legate) <i>"Un governo con le mani legate dal deficit pubblico non prova nemmeno a cominciare i lavori di restauro".</i>

Seguono gli *item* giudicati più utili a svolgere compiti ricettivi, segnalati da una specifica etichetta 'funzionale' del dizionario come *informazioni per comprendere* (fig. 3a). Fra questi il primo è l'*origine*, che contiene informazioni etimologiche oppure una spiegazione della relazione che intercorre tra il significato letterale e quello metaforico. Non tutte le espressioni idiomatiche hanno una semantica tanto opaca da richiedere una spiegazione etimologica. Molto spesso il legame tra i due livelli semantici è ricostruibile in termini di "tropi" (Casadei 1996) con i quali si estendono più tipicamente i significati lessicali¹². Diversi studi hanno dimostrato quanto sia importante per gli apprendenti stranieri conoscere la connessione tra il piano letterale e quello metaforico dei modi di dire. Queste spiegazioni aiutano sia la comprensione sia la memorizzazione nel medio e lungo periodo (Steinel et al. 2007, Boers et al 2004, Boers et al. 2007). La relazione tra i due piani viene quindi descritta nel dizionario o su base storico-culturale, recuperando prassi e consuetudini che hanno dato origine all'espressione, cristallizzandola in una forma lessicale, oppure sulla base dei tropi che hanno dato origine ad una semantica traslata che sia ricostruibile in maniera intuitiva.

Il *significato letterale*, altro *item* incluso tra le *informazioni per comprendere*, viene illustrato con la parafrasi dei singoli costituenti dell'espressione prestando particolare attenzione ai "costituenti unici" già discussi (*lusco* in *tra il lusco e il brusco*, *parrucco* in *trucco e parrucco*). Alcuni studi hanno mostrato infatti la difficoltà degli

¹² La letteratura sul tema delle metafore cognitive negli *idiom* è molto vasta e non affrontabile in questa sede. Ci limitiamo a citare il noto testo di Casadei (1996), ma anche le osservazioni di Burger (2010) sulla codifica linguistica, attraverso le espressioni idiomatiche, della gestualità (*fare le corna*) o di gesti e azioni (*alzare le mani su qualcuno*), chiamati "cinigrammi" (*Kinegramm*).

apprendenti nel capire i significati letterali e, quindi, le immagini soggiacenti quando l'espressione contiene una parola poco frequente o con una semantica particolarmente vaga (Boers et al. 2009).

Le *emozioni* tengono in conto la dimensione connotativa delle espressioni idiomatiche che è stata già discussa assieme al principio di *additional naming* introdotto da Dobrovolskij e Piirainen (2005). La connotazione può riguardare innanzitutto la polarità o la "prosodia semantica"¹³ dell'espressione, che può essere di segno positivo o negativo. Si pensi a *essere a portata di mano*, che in *la vittoria è a portata di mano* ha una connotazione positiva mentre in *la borsa era a portata di mano del ladro* si carica di valori negativi. Sono peraltro frequenti le espressioni idiomatiche che vengono usate ironicamente e per le quali è possibile ravvisare casi di vere e proprie enantiosemie, come in: *ci si è rotta la macchina e, ciliegina sulla torta, si è messo a piovere!* In secondo luogo, le espressioni idiomatiche comunicano spesso valori affettivi che esprimono l'atteggiamento del parlante nei confronti del referente. Per esempio, se invece di dire che *Marisa è una persona socievole e affabile* dico che *Marisa è alla mano* voglio sottolineare la mia conoscenza stretta e il mio affetto nei confronti di Marisa. Se invece di dire a qualcuno *non tergiversare, parla chiaro!* dicessi *Per favore, non menare il can per l'aia!* esprimerei un grado di rimprovero e impazienza maggiori.

Seguono quindi le *informazioni per scrivere e parlare* (fig. 3a) che il dizionario suggerisce agli utenti di consultare per i compiti di produzione tramite una seconda etichetta funzionale. Il primo di questi *item* è la *struttura linguistica*, un descrittore ideato per *Idiomatica* e ispirato alle notazioni delle *Lexical Unit* di *FrameNet* (Fillmore et al. 2003). In questa componente semasiologica di *FrameNet* viene illustrata in dettaglio la codifica morfo-sintattica dei significati delle parole (o *Frame*).

Com'è noto, la semantica dei *Frame* introdotta in linguistica da Fillmore (1985) si basa sull'assunto che la conoscenza enciclopedica e quella linguistica non siano nettamente scindibili. I significati lessicali vengono illustrati in termini di *Frame*, definibili come tipi di conoscenze che le parole evocano. Ogni parola in *FrameNet* viene associata ad un *Frame*, o concetto, che è poi descritto attraverso i suoi costituenti essenziali, o *Core Frame Elements*. Ad esempio, semplificando i formalismi usati, l'inglese *to help* evoca il *Frame* "Assistance" definito come: "QUALCUNO CHE AIUTA fa qualcosa a vantaggio di una PARTE BENEFICIARIA consentendo il conseguimento di un OBIETTIVO che la PARTE BENEFICIARIA vorrebbe raggiungere"¹⁴. Nella sezione semasiologica di *FrameNet*¹⁵, accanto all'indicazione del *Frame* evocato dal lemma viene fornita una breve definizione estratta da un dizionario e, successivamente, viene descritta la morfo-sintassi a partire o dalla semantica, nella sezione

¹³ Così Louw e Chateau (2000: 57) definiscono la prosodia semantica: "a form of meaning which is established through the proximity of a consistent series of collocates". Si veda anche Louw (1993).

¹⁴ "A HELPER BENEFITS a BENEFITED_PARTY by enabling the culmination of a GOAL that the BENEFITED_PARTY has", da https://framenet.icsi.berkeley.edu/fndrupal/luIndex_help_v. Indichiamo qui in maiuscolotto i *Frame Elements* che sono evidenziati con colori diversi in *FrameNet*.

¹⁵ *FrameNet*, *Lexical Unit Index*: <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>.

dei “Frame Elements and Their Syntactic Realizations”, o dalla sintassi, nella parte dedicata ai “Valence Patterns”. Queste componenti sono esemplificate con frasi estratte da corpora, annotate in termini di *parola target* e dei suoi *Frame Elements* collegati. Ad esempio, la frase *Gradually the parent is helped to learn how to play with, talk to, and enjoy their child* è annotata come segue:

[_{Manner} Gradually] [_{Benefited_party} the parent] is *HELPE*^{Target} [_{Goal} to learn how to play with, talk to, and enjoy their child]. [_{Helper} CNI]¹⁶

Prendendo come riferimento queste descrizioni, *Idiomatica* le ripropone in maniera semplificata per consentire agli utenti di comprendere meglio la semantica dell'espressione grazie ad una indicazione puntuale dei partecipanti all'evento, come illustrato in figura 2 per l'espressione *dare una mano a qualcuno*.

Figura 2 - *Item microstrutturali*: struttura linguistica e parte non modificabile

< Dare una mano

Struttura

STRUTTURA LINGUISTICA

Gianni₁ ci₂ dà una mano a riparare l'auto₃.
 I bambini₁ hanno dato una mano alla maestra₂
 a riordinare la classe₃.

1 [la persona che aiuta] *qualcuno*
 2 [la persona che viene aiutata] *qualcun altro*
 3 [la cosa in cui si dà aiuto] *a fare qualcosa*

PARTE NON MODIFICABILE

una mano, non si può cambiare il numero e l'articolo.

“Dare **la** mano” significa conoscere qualcuno,
 presentarsi a qualcuno

SBAGLIATO:
 dare ~~la~~ **mano**
 dare ~~le~~ **mani**

Seguono informazioni pensate per supportare l'apprendimento della L2. Gli *item* descritti come *parte non modificabile* rendono infatti esplicite le restrizioni morfosintattiche proprie di ciascun idiom. Ad esempio, in *dare una mano* il cambio dell'articolo darebbe origine ad una diversa espressione idiomatica, come *dare la mano*, che indica il ‘fare la conoscenza di qualcuno’. Mentre il *livello di formalità* quantifica mediante una scala quadripartita (formale, non formale, non educato,

¹⁶ CNI sta per *constructional null instantiation*, ovvero: “An FE that is missing because the grammar of the sentence allows or requires an omission (e.g. the subject of an imperative, the agent of a passive verb)”, FrameNet Glossary, <https://framenet.icsi.berkeley.edu/fndrupal/glossary>.

volgare) l'usabilità di una certa espressione in specifiche situazioni d'interazione, cogliendo gli aspetti della variazione diafasica e diastratica. Ad esempio, dire qualcosa *in camera caritatis* consente di parlare con un tono più formale rispetto a dirsi qualcosa *in separata sede*. La scala proposta verrà plausibilmente arricchita per sottolineare l'usabilità di alcune espressioni con tipi o gruppi di specifici interattanti: ad esempio *come butta* viene usata come saluto tra i giovani.

L'etichetta *si usa per* contiene descrizioni dei tipi di atti linguistici che si realizzano più frequentemente con un certo idiom. Mentre *mettersi le mani nei capelli* viene usato quasi sempre per lamentarsi di qualcosa, *fasciarsi la testa prima di romperse-la* è un modo tipico per dare consigli, sebbene entrambi esprimano disperazione e scoraggiamento. Nel dizionario vengono usati descrittori non tecnici che rendano facilmente comprensibile il tipo di azione o attività linguistica più comunemente realizzata da una certa espressione idiomatica.

Seguono infine indicazioni sulla variazione testuale e diamesica — se si tratta quindi di una comunicazione orale o scritta — con i *contesti d'uso*; mentre la sezione *connettori* elenca gli avverbi, le preposizioni o in generale i modificatori più tipici. Si pensi alle particelle modali *c'è da* o *viene da* che spesso si usano per introdurre un *idiom*: *c'è da mettersi le mani nei capelli a vedere cosa hanno combinato!*

I *modi verbali ammessi* specificano le diatesi ammissibili nella flessione delle diverse locuzioni, ad esempio *vuotare il sacco* come molte altre non possono essere trasformate in frasi passive, mentre nella *coniugazione* è possibile trovare la tavola flessiva completa di un *idiom* verbale.

Nelle sessioni sperimentali che verranno discusse di seguito (§ 5.1), emergerà la tendenza da parte degli apprendenti meno competenti nell'italiano L2 a consultare le tavole di coniugazione per svolgere compiti flessivi e formali con gli idioms. Al contrario, i più competenti sembrano preferire le informazioni ricavabili a questo scopo dalla *struttura linguistica*.

5. Analisi dei bisogni e prototipazione partecipativa: dal workshop di co-design ai dizionari cliccabili e scrollabili

La progettazione di *Idiomatica* è frutto di un confronto serrato tra lessicografi e specialisti di fattori umani (Human Factors, HF), professionisti con una formazione multidisciplinare nel campo delle nuove tecnologie che si occupano dello sviluppo di sistemi interattivi applicando le linee guida dello *Human-Centred Design* (Giacomin 2014). La prima fase di progettazione di *Idiomatica* è stata realizzata attraverso un workshop “partecipativo” (*co-design*, Tosi 2020) che ha coinvolto rappresentanti di tutte le categorie di *stakeholders* interessati (Caruso et al. 2019): lessicografi, apprendenti stranieri e specialisti di fattori umani.

A seguito del workshop sono state abbozzate diverse famiglie di prototipi, successivamente raffinate in due tipologie: gli ‘scrollabili’ e i ‘cliccabili’. Tra questi, un rappresentante di ciascuna tipologia è stato selezionato per successive sessioni di *testing* e di design.

Figura 3a - *Prototipo cliccabile*

◀ IDIOMATICA	
Montarsi la testa	
Se una persona si monta la testa, vuol dire che inizia a pensare di essere eccezionale e avere capacità superiori a quelle che effettivamente possiede.	
Es. <i>Bravo, ma non montarti la testa.</i>	
INFORMAZIONI PER COMPRENDERE	
Origine	>
Significato letterale	>
Emozioni	>
Livello di formalità	>
Si usa per	>
INFORMAZIONI PER SCRIVERE E PARLARE	
Struttura linguistica	>
Contesti d'uso	>
Connettori	>
Modi verbali ammessi	>
Coniugazione	>

Figura 3b - *Prototipo scrollabile*

◀ IDIOMATICA	
Montarsi la testa	
Se una persona si monta la testa, vuol dire che inizia a pensare di essere eccezionale e avere capacità superiori a quelle che effettivamente possiede.	
Es. <i>Bravo, ma non montarti la testa.</i>	
INFORMAZIONI PER COMPRENDERE	
ORIGINE	
Montarsi la testa è un modo di dire che ha origine tra le attività culinarie. In cucina si montano le uova o la panna per farle aumentare di volume, diventare più grandi. Perciò, così come si montano le uova, una persona può montarsi la testa, ovvero può convincersi di essere più grande e più capace di quello che è in realtà.	
Tuttavia le uova e la panna dopo un certo tempo si sgonfiano e perdono il volume prodotto dallo sbattimento. Allo stesso modo, uno che si monta la testa non ha capacità che durano nel tempo.	
SIGNIFICATO LETTERALE	
Aumentare il volume della propria testa.	
EMOZIONI	
L'espressione ha un significato negativo e viene	

I due prototipi hanno la stessa macro- e microstruttura ma differiscono nelle modalità di distribuzione e presentazione dei dati, richiedendo tipologie d'interazione diverse da parte dell'utente. Il dizionario scrollabile ricalca le modalità organizzative attualmente più in uso tra gli app dizionari, ovvero l'inserimento di tutte le informazioni microstrutturali in un'unica schermata (o *view*) che dev'essere esplorata a scorrimento per trovare le informazioni desiderate, da qui il nome di prototipo 'scrollabile', o *by scroll* (figura 3a). Nell'altro, invece, l'accesso alle informazioni avviene attraverso un indice interattivo in cui sono elencate le etichette metalessicografiche dei dati, i quali possono essere visualizzati solo su una successiva schermata. In questo tipo di prototipo l'utente deve quindi decodificare le etichette metalinguistiche, scegliere quella desiderata e cliccare su di essa per accedere all'informazione, da qui il nome del prototipo: 'cliccabile', o *by tap* (figura 3b).

Entrambe le versioni condividono però la medesima macrostruttura alfabetica, il cui lemmario, accessibile dalla schermata principale (o *Home View*, figura 4), contiene anche i sinonimi degli idiomi, mostrati sotto ciascun lemma.

Figura 4 - Home view di Idiomatica con il lemmario

IDIOMATICA		
Q Search		
A		#
Avere le mani bucate	>	A
Essere uno spendaccione		B
Avere le mani legate	>	C
Non poter fare qualcosa		D
D		E
Dare una mano	>	F
Aiutare		G
E		H
Essere a portata di mano	>	I
Essere facilmente raggiungibile		J
Essere alla mano	>	K
Essere amichevoli		L
F		M
Fare di testa propria	>	N
Non accettare consigli		O
Fasciarsi la testa prima di rompersela	>	P
Preoccuparsi prima del tempo		Q
		R

A fronte delle diverse modalità d'interazione col dizionario, nelle voci di entrambe le versioni (cliccabile e scrollabile) l'utente trova due costituenti identici. Innanzitutto, l'area informativa principale, che contiene il commento semantico, un esempio d'uso e le sostituzioni lessicali (figura 1). Ci sono, inoltre, in entrambe le versioni, due etichette che chiamiamo 'funzionali'¹⁷ (visibili nell'immagine del prototipo cliccabile, figura 3a) perché raggruppano i dati ritenuti più utili a svolgere le funzioni principali del dizionario: i compiti ricettivi (etichetta: *informazioni per comprendere*) e i compiti produttivi (etichetta: *informazioni per scrivere e parlare*). Le etichette funzionali cercano di risolvere l'annoso problema di come evitare il sovraccarico informativo dell'utente senza dover creare sezioni specifiche o addirittura dizionari differenti da consultare in base al compito che si sta svolgendo, i

¹⁷ Seguendo la terminologia di Wiegand questi elementi potrebbero essere denominati 'function-identifying entries', dal momento che le etichette metalinguistiche per identificare gli item microstrutturali sono chiamate *data-identifying entries* (Gouws 2014).

così detti dizionari monofunzionali proposti dalla *Lexicographical Function Theory* (Tarp 2008, Caruso 2016).

Queste impostazioni generali sono il frutto di una rielaborazione ragionata degli input forniti dagli utenti durante il workshop di co-design, descritto in dettaglio altrove (Caruso et al. 2019). Proprio la necessità di creare un *tool* di facile utilizzo ha determinato la scelta di non inserire parti speciali destinate al solo apprendimento, ma di concepire quest'ultimo come la funzione principale del dizionario, perseguibile con l'arricchimento dei dati descritto in § 4.2.

5.1 Test di valutazione formativa per ripensare i prototipi

Ulteriori implementazioni del prototipo sono state messe a punto con una seconda sperimentazione, in cui degli apprendenti stranieri hanno svolto esercizi con le espressioni idiomatiche usando il prototipo cliccabile, che presenta una modalità di accesso ai dati lessicografici molto inusuale nel panorama degli app dizionari odierani. Si è deciso quindi di verificarne possibili criticità di utilizzo, testandolo con utenti reali in uno studio di “valutazione formativa” (*formative evaluation*, Lazar 2017). Lo scopo di questo tipo di sperimentazioni è quello di fornire spunti di riflessione per proseguire la progettazione, iterandone le fasi a partire dai pregi e i difetti del prototipo testato. Al contrario, le “valutazioni sommative” (*summative evaluations*) testano i prototipi giunti ad una fase progettuale più avanzata. Data la loro funzione orientativa, spesso le valutazioni formative coinvolgono pochi partecipanti: secondo Nielsen bastano 5 utenti per rilevare la maggior parte dei problemi progettuali (Nielsen & Landauer 1993; 2000). In questo modo è possibile, infatti, raccogliere dati utili in tempi brevi e tornare poi, subito dopo, a riprogettare il prototipo.

Lo studio formativo di *Idiomatica* ha coinvolto 10 studenti cinesi con un livello d'italiano L2 pari al B1 del QCER. Le sessioni sperimentali, che hanno coinvolto un partecipante alla volta, si sono svolte online durante i mesi del *lockdown* pandemico, usando un'app di videoconferenze. Gli sperimentatori hanno videoregistrato le interazioni degli studenti col dizionario grazie alla modalità di condivisione dello schermo mentre gli studenti consultavano una versione desktop¹⁸ del prototipo. Affiancando la pagina degli esercizi proposti su Moodle a quella del prototipo è stato possibile tenere traccia delle attività svolte con il dizionario. Anche in questo caso ci soffermiamo solo sui risultati dell'esperimento per focalizzare l'attenzione sull'importanza del feedback da parte degli utenti e dell'iterazione nei processi di design.

I dati raccolti, analizzati con modalità simili allo studio di Wolfer e colleghi (2018; e anche Müller-Spitzer et al. 2018), hanno consentito di registrare i tempi e il numero di volte che i diversi campi informativi del dizionario venivano consultati. In particolare, sono state comparate tra loro le strategie di consultazione rivelatesi efficaci, consentendo all'utente di dare risposte corrette, e quelle che non lo sono

¹⁸ Sebbene la condizione sperimentale basata su un prototipo in versione web del dizionario non sia ottimale, perché non riproduce le reali condizioni d'uso dell'applicazione, questa ha il vantaggio di garantire un'osservazione più accurata. Si spera, al momento in cui si scrive, di poter proseguire la ricerca con sperimentazioni dal vivo al più presto.

state. È emersa in questo modo l'efficacia informativa della *struttura linguistica* per risolvere compiti flessivi: chi eccelle in queste attività consulta la *struttura linguistica* per un tempo quattro volte più lungo rispetto a quello speso con le tavole flessive. Un'attenta lettura è forse sufficiente ad apprendere ciò che serve, dal momento che solo una piccola parte degli esercizi sulla morfosintassi (circa il 30%) vengono svolti dagli utenti più capaci consultando direttamente il dizionario. Sembra che la sinteticità della descrizione offerta nella *struttura linguistica* basti a trasformare l'input offerto dal dizionario in *intake*¹⁹, un aspetto su cui sarà utile raccogliere più dati in futuro.

Durante questo studio sono state raccolte anche le opinioni degli utenti, sia attraverso un'intervista finale, sia mediante la somministrazione di un questionario post-esperienza sull'usabilità e l'utilità percepita del dizionario (valutate sulla base di una scala Likert a 5 punti), fattori dai quali dipende l'accettazione di questa tecnologia (o *technology acceptance*, Davis 1989; Hornbæk & Hertzum 2018). Il questionario somministrato era suddiviso in due parti, una relativa all'usabilità e un'altra all'utilità.

5.2 Le opinioni degli utenti e gli sviluppi futuri della ricerca

I commenti degli utenti raccolti durante le interviste finali hanno rivelato le loro difficoltà nel comprendere alcune delle etichette metalessicografiche usate, così come la suddivisione tra le informazioni per i compiti ricettivi e quelli produttivi.

Il questionario ha invece rivelato come l'app sia percepita come uno strumento sia utile sia facile da usare, con una leggera superiorità dell'utilità rispetto all'usabilità percepita (usabilità percepita: 4.1; utilità percepita: 4.37). L'elevato punteggio per l'utilità rivela anche un significativo interesse nei confronti di strumenti didattici per imparare le espressioni idiomatiche. Il giudizio sull'usabilità della risorsa, invece, anch'esso nettamente positivo, conferma quanto il percorso di implementazione intrapreso sia promettente e fa sperare che gli utenti siano disposti a familiarizzare con essa e ad usarla per le loro attività di studio.

Inoltre, per venire incontro alle difficoltà riscontrate durante le sessioni di *testing*, sono state scelte delle etichette metalessicografiche più trasparenti, quali le già citate *livello di formalità* e *si usa per* al posto, rispettivamente, di (figura 5) *registro linguistico* e *atti linguistici*.

Avendo poi riscontrato che gli utenti consultano la *struttura linguistica* per svolgere compiti flessivi, questo campo è stato inserito tra le *informazioni per scrivere e parlare* e, più in generale, si è preferito lasciare tra le informazioni utili per i compiti ricettivi (*informazioni per comprendere*) solo quelle di natura più strettamente semantica. Le implementazioni seguite allo studio di valutazione formativa possono

¹⁹ Bettoni (2001: 22) spiega così uno dei concetti di base della linguistica acquisizionale (cfr. Corder 1967): "Non tutto quello che si legge o si sente si traduce direttamente in apprendimento. Perché l'input diventi produttivo, e cioè possa contribuire alla costruzione dell'interlingua ed eventualmente tradursi in output, deve essere capito. In gergo, l'input deve diventare intake".

essere apprezzate comparando il prototipo usato per questo studio e riprodotto nella figura 5, con quello finale già presentato e discusso della figura 3a.

Figura 5 - *Prima versione del prototipo scrollabile*

< Dizionario	
Montarsi la testa	
Se una persona si monta la testa, vuol dire che inizia a pensare di essere eccezionale e avere capacità superiori a quelle che effettivamente possiede.	
Es. <i>Bravo, ma non montarti la testa.</i>	
INFORMAZIONI PER LA COMPrensIONE	
Origine	>
Significato letterale	>
Emozioni	>
Registro stilistico	>
Atti linguistici	>
Struttura	>
INFORMAZIONI PER LA PRODUZIONE	
Contesti d'uso	>
Connettori	>
Flessione	>
Diatesi	>

L'ultima versione che è stata implementata del prototipo (fig. 3a) sarà oggetto di ulteriori sessioni di *testing* assieme al suo analogo scrollabile per verificarne, innanzitutto, l'efficacia rispetto ai dizionari monolingua generali dell'italiano e, allo stesso tempo, l'efficienza ovvero: il prototipo cliccabile e quello scrollabile consentono di svolgere i compiti negli stessi tempi oppure uno dei due garantisce di completarli più rapidamente?

Bibliografia

- Beretta, Monica. 1986. Pronomi atoni: una prospettiva funzionale. *Italiano e Oltre* 1. 26-29.
 Berruto, Gaetano. 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: La Nuova Italia Scientifica.

- Bettoni, Camilla. 2001. *Imparare un'altra lingua: lezioni di linguistica applicata*. Bari: Laterza.
- Boers, Frank & Piquer Píriz, Ana María & Stengers, Hélène & Eyckmans, June. 2009. Does pictorial elucidation foster recollection of idioms? *Language Teaching Research* 13(4). 367-382.
- Boers, Frank & Lindstromberg, Seth. 2009. *Optimizing a lexical approach to instructed second language acquisition*. Palgrave MacMillan.
- Boers, Frank & Demecheleer, Murielle & Eyckmans, June. 2004. Etymological elaboration as a strategy for learning idioms. In Bogaards, Paul & Laufer, Batia (eds.) *Vocabulary in a second language: Selection, acquisition and testing*, 53-78. Amsterdam: John Benjamins.
- Burger, Harald. 2010. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt.
- Caruso, Valeria & Balbi, Barbara & Monti, Johanna & Presta, Roberta. 2019. How Can App Design Improve Lexicographic Outcomes? Examples from an Italian Idiom Dictionary. In Kosem, Iztok & Zingano Kuhn, Tanara & Correia, Margarita & Ferreria, José Pedro & Jansen, Maarten & Pereira, Isabel & Kallas, Jelena & Jakubiček, Miloš, Krek, Simon & Tiberius, Carole (eds.). *Proceedings of the eLex 2019 conference*, 374-396.
- Caruso, Valeria. 2016. Dizionari elettronici e apprendimento delle espressioni idiomatiche: monitoraggio dei bisogni e prospettive future. In Bianchi, Francesca & Leone, Paola (a cura di), *Linguaggio e apprendimento linguistico. Metodi e strumenti tecnologici*, 173-189. Milano: AItLA.
- Casadei, Federica. 1996. *Metafore ed espressioni idiomatiche. Uno studio semantico sull'italiano*. Roma: Bulzoni Editore.
- Corder, Stephen Pit. 1967. The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching* 5. 161-170.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp.
- D'Achille, Paolo. 2003. *L'italiano contemporaneo*, Bologna: il Mulino.
- Davis, Fred. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13(3), 31. 319-340.
- Dobrovol'skij, Dmitrij & Piirainen, Elisabeth. 2005. *Figurative language: cross-cultural and cross-linguistic perspectives*, London: Emerald.
- Fillmore, Charles J. & Johnson, Christopher R. & Petruck, Miriam R.L. 2003. Background To Framenet. *International Journal of Lexicography* 16(3). 235-250.
- Fillmore, Charles. 1985. Frames and the semantics of understanding. *Quaderni di semantica*. 222-254.
- Giacomin, Joseph. 2014. What is human centred design? *Design Journal* 17(4). 606-623.
- Gouws, Rufus H. 2014. Article structures: Moving from Printed to e-Dictionaries. *Lexikos* 24. 155-157.
- Heid, Ulrich. 2011. Electronic Dictionaries as Tools: Toward an Assessment of Usability. In Fuentès Olivera, Pedro Antonio & Bergenholtz, Henning (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, 287-304. London/New York: Continuum.

- Hornbæk, Kasper & Hertzum, Morten. 2017. Technology Acceptance and User Experience. *ACM Transactions on Computer-Human Interaction* 24(5). 1-30.
- ISO 9241: 210. *Ergonomics of human-system interaction: Part 210. Human-centred design for interactive systems*. Geneva: ISO.
- Jones, Steven E. 2014. *The Emergence of the Digital Humanities*. New York & London: Routledge.
- Lazar, Jonathan & Feng, Jinjuan & Hochheiser, Harry. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.
- Lew, Robert & de Schryver, Gilles-Maurice. 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4). 341-359.
- Louw, Bill & Chateau, Carmela. 2010. Semantic prosody for the 21st century: Are prosodies smoothed in academic contexts? A contextual prosodic theoretical perspective. In Bolasco, Sergio & Chiari, Isabella & Giuliano, Luca (eds.), *Proceedings of 10th International Conference on Statistical Analysis of Corpus Study*, 755-764.
- Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, Mona & Francis, Gill & Tognini-Bonelli, Elena (eds.) *Text and technology: in honour of John Sinclair*, 240-251. Amsterdam: John Benjamins.
- Müller-Spitzer, Carolin & Domínguez Vázquez, María José & Nied Curcio, Martina & Silva Dias, Idalete Maria & Wolfer, Sascha. 2018. Correct Hypotheses and Careful Reading Are Essential: Results of an Observational Study on Learners Using Online Language Resources. *Lexikos* 28. 287-315.
- Nation, Paul. 2006. Vocabulary: Second Language. In Brown, Keith (ed.), *Encyclopedia of Language & Linguistics* (2nd ed.), 448-454. Oxford: Elsevier.
- Nesi, Hilary & Boonmoh, Atripat. 2009. A close look at the use of pocket electronic dictionaries for receptive and productive purposes. In Fitzpatrick, Tess & Barfield, Andy (eds.), *Lexical processing in second language learners: papers and perspectives in honour of Paul Meara*, 67-81. Clevedon, UK: Multilingual Matters.
- Nielsen, Jakob, & Landauer, Thomas. 1993. A mathematical model of the finding of usability problems. In Ashlund, Stacey & Henderson, Austin & Hollnagel, Erik & Mullet, Kevin & White, Ted N. (eds.), *Proceedings of ACM INTERCHI '93 Conference*, 206-213.
- Nielsen, Jakob. 2000. Why you only need to test with 5 users. <http://www.useit.com/alertbox/20000319.html>.
- Rundell, Michael. 2015. Review Article: Shigeru Yamada. Oxford Guide to the Practical Usage of English Monolingual Learners' Dictionaries: Effective Ways of Teaching Dictionary Use in the English Class. *Kernerman Dictionary News* (23). 26-27.
- Sabatini, Francesco. 1985. L'italiano dell'uso medio: una realtà tra le varietà linguistiche italiane. In Holtus, Edgar Radtke (Hrsg.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 154-184. Tübingen: Narr.
- Sinclair, John McHardy. (ed.) 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing*. London & Glasgow: Collins Cobuild.
- Sinclair, John McHardy. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, John McHardy. 2004. Trust the text. In Sinclair, John & Ronald, Carter (eds.) *Trust the Text – Language, Corpus and Discourse*, 9-23. London: Routledge.
- Siyanova-Chanturia, Anna & Spina, Stefania. 2019. Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning* 70(2). 420-463.
- Steinel, Margarita P. & Hulstijn, Jan H. & Steinel, Wolfgang. 2007. Second-language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition* (29). 449-484.
- Steyer, Kathrin & Ďurčo, Peter. 2013. Ein korpusbasiertes Beschreibungsmodell für die elektronische Sprichwortlexikografie. In Benayoun, Jean-Michel & Kübler, Natalie & Zouogbo Jean-Philippe (eds.). *Parémiologie. Proverbes et formes voisines*, Tome 3, 219-250. *Parémiologie*. Sainte Gemme: Presses Universitaires de Sainte Gemme.
- Szczepaniak, Renata & Lew, Robert. 2011. The Role of Imagery in Dictionaries of Idioms. *Applied Linguistics* 32(3). 323-347.
- Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Berlin/New York: de Gruyter.
- Taylor, Andrew & Chan, Adelaide. 1994. Pocket Electronic Dictionaries and Their Use. In Martin, Willy et al. (eds.) *Proceedings of the 6th Euralex International Congress*, 598-605.
- Tosi, Francesca. 2020. Co-design and Innovation: Tools, Methods and Opportunities for the Generation of Innovation Through User Involvement. In Tosi, Francesca, *Design for Ergonomics*, 129-141. Cham: Springer.
- Wiegand, Herbert Ernst & Smit, Maria. 2013. Microstructures in printed dictionaries, in: Gouws, Rufus H. & Heid, Ulrich & Schweickard, Wolfgang & Wiegand, Herbert Ernst (Hrsg.), *Dictionaries. An International Encyclopedia of Lexicography: Supplementary volume. Recent Developments with Special Focus on Computational Lexicography*, 149-214. Berlin: Mouton de Gruyter.
- Wiegand, Herbert Ernst. 1984. On the Structure and Contents of a general Theory of Lexicography. In Hartmann, Reinhard Rudolf Karl (ed.) *LEX'eter '83 Proceedings*, 13-30. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst. 2000. Wissen, Wissenrepräsentationen und Printwörterbücher. In: Heid, Ulrich & Evert, Stefan & Lehmann, Egbert & Rohrer, Christian (eds.), *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*, 15-38, Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Wiegand, Herbert Ernst. 2002. Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen. In: Ezawa, Kennosuke & Kürschner, Wilfried & Rensch, Karl Heinz & Ringmacher, Manfred (Hrsg.), *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums*, 265-281. Tübingen: Gunter Narr Verlag.
- Winestock, Christopher & Jeong, Young-kuk. 2014. An analysis of the smartphone dictionary app market. *Lexicography* 1(1). 109-119.
- Wolfer, Sascha & Nied Curcio, Martina & Dias, Idalete Maria Silva & Müller-Spitzer, Carolin & Domínguez Vázquez, María José. 2018. Combining Quantitative and Qualitative Methods in a Study on Dictionary Use. In Čibej, Jaka & Gorjanc, Vojko & Kosem, Iztok

& Krek, Simon (eds.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 101-112. Ljubljana: Ljubljana University Press.

Dizionari e risorse lessicografiche telematiche

FrameNet Glossary, <https://framenet.icsi.berkeley.edu/fndrupal/glossary>.

FrameNet, <https://framenet.icsi.berkeley.edu/fndrupal/>

GRADIT, 1999. Grande dizionario italiano dell'uso, De Mauro, T. (ed.), Torino: UTET.

Longman Dictionary of Contemporary English. London: Longman, 1978

OWID – *Sprichwörterbuch. Online-Wortschatz-Informationssystem Deutsch*, Institut für Deutsche Sprache, Mannheim, <http://www.owid.de/wb/sprw/start.html>.

Sinclair, J. M. (ed.). 1987. *COBUILD English Language Dictionary*. London: Collins ELT.

Vocabolario Treccani, Istituto della Enciclopedia Italiana, <http://www.treccani.it/vocabolario/>.

SIMONE CICCOLONE¹ - GIULIA GROSSO¹

Macchine parlanti o macchine apprendenti? Osservazioni preliminari sull'uso dell'etero-ripetizione in produzioni di chatbot e nell'interlingua²

Abstract

In this paper we will present some preliminary observations on written interactions between chatbots designed for “providing realistic simulations of casual conversation (or small talk)” (McTear *et al.* 2016: 52) and their human users. We will in particular point at possible strategic uses of repetitions of the user's input and compare them with the conversational functions of repetition in the output of L2 speakers. As observed in Pallotti (1999), repetition of fragments of the interlocutor's turn is often used by L2 speakers as a participation strategy. We will compare examples of this appropriation mechanism in the uses of Italian L2 speakers (through a selection of the data in Grosso 2015) and of chatbots.

1. *Introduzione*

In questo studio presenteremo alcune prime riflessioni su un campo di indagine relativamente nuovo riguardante le interazioni con interfacce conversazionali. Nello specifico, intendiamo identificare il ruolo delle etero-ripetizioni nelle conversazioni con chatbot e nelle interazioni che coinvolgono apprendenti di L2. Le nostre riflessioni saranno inquadrare non nei contesti disciplinari abituali per tale tipo di dati (come la linguistica computazionale), bensì nell'ottica della linguistica applicata e delle strategie conversazionali, proponendo in forma esplorativa un confronto tra i due tipi di interazione.

In primo luogo occorre chiarire che per interfacce conversazionali intendiamo strumenti software progettati per supportare un'interazione uomo-macchina tramite linguaggio verbale umano. In particolare, in questo studio ci concentreremo su interfacce conversazionali che usano la modalità scritta, non *task-oriented* (ovvero finalizzate a compiti ben definiti, come nel caso degli assistenti vocali) e “a dominio aperto” (ingl. *open domain*; cfr. § 2).

¹ Università degli Studi di Cagliari.

² Il contributo è stato progettato e realizzato in modo condiviso da entrambi gli autori. Tuttavia i paragrafi 1 e 2 sono da attribuire a Simone Ciccolone e i paragrafi 3 e 5 sono da attribuire a Giulia Grosso; il paragrafo 4 è stato redatto da entrambi. Si ringraziano Maria Elena Favilla e Gabriele Iannàccaro per gli stimoli e i commenti, nonché i revisori anonimi per gli utili commenti a una versione precedente del testo.

Il nostro contributo non intende inserirsi nel paradigma degli studi sul *natural language understanding*³, anche per la natura esplorativa della discussione dei dati qui presentati. L'obiettivo sarà piuttosto riflettere su eventuali parallelismi tra alcune strategie adottate da apprendenti nelle interazioni in L2 (con parlanti nativi o altri parlanti non nativi) e il possibile riconoscimento di tali strategie da parte dell'utente anche negli output prodotti da chatbot, concentrandoci nello specifico sull'etero-ripetizione.

Lo spunto per tale parallelismo nasce dalle scelte adottate per un'interfaccia conversazionale in particolare: Eugene Goostman (d'ora in poi EG). Questo chatbot ha partecipato a varie sessioni pratiche del test di Turing⁴, tra cui le sessioni alla Royal Society di Londra del 2014 descritte in Warwick e Shah (2016): in queste sessioni ogni giudice (umano) intratteneva due interazioni scritte in parallelo con due "entità", un utente umano (non visibile ai giudici) e un'interfaccia conversazionale. Ogni test aveva una durata fissa di 5 minuti, al termine dei quali le due interazioni parallele venivano interrotte e il giudice doveva dichiarare quali delle due entità fosse l'utente umano e quale la macchina.

Proprio in queste sessioni EG viene erroneamente identificato come umano in 10 test su 30, risultando tra i primi chatbot a superare la soglia del 30% stabilita per tale versione del test di Turing. L'aspetto rilevante di tale risultato, sia per i commentatori che per gli autori del chatbot, non riguarda la presunta capacità di emulare intelligenza o abilità interazionali, quanto quella di "ingannare" e indurre in errore il giudice umano⁵.

Per poter simulare un utente umano, occorre innanzitutto dotare il chatbot di un'identità fittizia da impersonare. Come ben sintetizzano i creatori di EG,

[w]hen making a bot, you don't write a program, you write a novel. You think up a life for your character from scratch – starting with his (or her) childhood and leading up to the current moment, endowing him with his personal unique features (Demchenko & Veselov 2008: 458).

³ Cfr. tra gli altri McTear *et al.* (2016: 175-176) per alcune applicazioni nel campo del *natural language understanding*, nonché Allen (1995) per un'introduzione.

⁴ Il test di Turing è un esperimento mentale proposto da Turing (1950) con cui un'intelligenza artificiale viene testata in base all'interazione comunicativa con un giudice umano: nella versione originale (chiamata *imitation game*), il giudice interagisce con due "entità" non visibili di cui identificare il genere, e il cui scopo è invece quello di non farsi identificare. Se una macchina riesce a ingannare il giudice con la stessa frequenza con cui può farlo un essere umano, supera il test. Le applicazioni concrete di questo test spesso si allontanano sensibilmente dalla proposta iniziale. Cfr. Epstein *et al.* (2008) per approfondimenti.

⁵ Lo studio di Warwick e Shah (2016) ha come scopo "to see how good the machine was at deception and also to consider how the deception was possibly achieved in each case" (*ibidem*: 992). Il titolo stesso del saggio di Demchenko e Veselov (2008: "Who Fools Whom? The Great Mystification, or Methodological Issues on Making Fools of Human Beings") è piuttosto eloquente sugli obiettivi dei programmatori di EG.

Nel caso di EG, la personalità scelta è quella di un ragazzino ucraino di 13 anni, che quindi interagisce in lingua straniera (l'inglese). Come osservano Warwick e Shah (2016),

Eugene's character (a 13-year old [sic!] Ukrainian boy) is used as a ploy to gain sympathy and to explain away some poor English grammar and a possible lack of understanding of some points in the conversation (Warwick & Shah 2016: 1003).

Secondo Warwick e Shah, quindi, tale caratterizzazione implicitamente indirizza i giudici a interpretare non solo gli errori grammaticali, ma anche eventuali incoerenze testuali e incomprensioni, come possibile esito di una competenza parziale in lingua straniera. Tramite tale artificio (in un certo senso di tipo testuale più che informatico) l'utente umano potrebbe essere indotto a riconoscere anche negli errori di processazione dell'input o di generazione dell'output possibili fenomeni "naturali" riconducibili a una sorta di "interlingua simulata". A questo si aggiunge la tendenza pervasiva di ogni utente umano a riconoscere intenzioni comunicative coerenti negli output dell'interlocutore, e a cercare di ricostruire una coerenza testuale anche sovrainterpretando⁶.

Nei paragrafi seguenti ci poniamo l'obiettivo di comprendere come le interazioni condotte con le interfacce conversazionali presentino degli aspetti e dei fenomeni in comune con interazioni di cui sono protagonisti apprendenti di una L2. In particolare, analizzeremo interazioni in cui si evidenziano similarità di comportamento comunicativo riguardo all'uso dell'etero-ripetizione, anche nei contesti in cui essa facilita il *topic change*, collocandosi, come sostiene Bazzanella, sul polo del disaccordo (cfr. par. 4). Il confronto proposto rappresenta solo un primo tentativo di taglio esplorativo, con cui orientare le scelte per un'indagine più sistematica a partire dalla raccolta dati.

2. *Le interfacce conversazionali*

Partendo da tali premesse, concentreremo la nostra attenzione sulle dinamiche interazionali evidenziate dalle reazioni degli utenti umani agli output prodotti dalla macchina. A tale scopo, occorre inquadrare meglio sia le interfacce conversazionali usate, sia i tratti caratterizzanti le interazioni di utenti umani con tali interfacce⁷, di cui commenteremo alcuni esempi.

Innanzitutto, osserveremo qui interfacce conversazionali che usano solo il canale scritto, ovvero i chatbot, escludendo in tal modo le fasi di processazione del segnale acustico. Questo ovviamente produce una differenza significativa sul piano diamesico tra i dati relativi agli apprendenti e quelli relativi ai chatbot; è d'altronde ipotizzabile che gli output della macchina abbiano livelli di pianificazione simile sia in forma scritta che in forma orale. Ciò non toglie che quest'aspetto rappresenta un

⁶ Cfr. Tavosanis (2018: 78): "anche la consapevolezza di avere a che fare con una macchina non cancella del tutto la spinta degli esseri umani a parlare e ascoltare come se l'interlocutore fosse in realtà umano".

⁷ Cfr. Tavosanis (2018: 87) per un esempio di interazione tra chatbot.

ulteriore limite della nostra ricerca allo stadio attuale, e richiede quindi, in futuri sviluppi, una trattazione più approfondita. Qui cercheremo in ogni caso di osservare i fenomeni dell'etero-ripetizione indipendentemente dal *medium* con cui sono prodotti, in virtù della natura esplorativa e preliminare dello studio proposto.

Indipendentemente dal canale, occorre distinguere due tipi di interfaccia conversazionale: *task-oriented* e *open-domain*.

Task-oriented dialog systems are designed for specific domains or tasks, such as flight booking, hotel reservation, customer service, and technical support, and have been successfully applied in some real-world applications. Open-domain dialog systems, however, are much more challenging to develop due to its [sic!] open-ended goal (Huang *et al.* 2020: 2).

In particolare, “[t]he goal of an open-domain dialog agent is to maximize the long-term user engagement” (*ibidem*: 2). McTear *et al.* (2016: 52) evidenziano finalità simili parlando dei chatbot: “the emphasis has been on providing realistic simulations of casual conversation (or small talk) rather than helping humans to perform a particular task”. Lo scopo ultimo di tali sistemi, in sostanza, è quello di trattenere l'utente in un'interazione verbale coinvolgente e quanto più estesa possibile.

I chatbot testati per questo studio mostrano di seguire un rigido modello di stimolo-risposta: per ogni input prodotto dall'utente, il chatbot produce un singolo output; inoltre, tale output può essere prodotto solo in risposta a uno stimolo ricevuto dall'utente⁸, e si costruisce interamente sulla base di tale stimolo. Questo aspetto cruciale della progettazione di un chatbot ha conseguenze rilevanti in termini di vincoli alle mosse conversazionali attuabili dalla macchina, e quindi anche di dominanza e asimmetria in tali interazioni.

Un problema centrale nella processazione dell'input, tutt'ora irrisolto, riguarda l'individuazione del *topic*. La sua conseguenza più evidente è la produzione di risposte incoerenti col co-testo precedente. Per evitare tali incoerenze, e cercare comunque di produrre output in qualche modo efficaci, spesso i chatbot ricorrono a risposte precostruite, di carattere generale, che hanno principalmente lo scopo di proporre un cambio di *topic*. Si consideri l'estratto seguente, riguardante uno scambio tra EG e un giudice durante un test:

- (1) 1 J: Pleased to meet you Eugene. What's the weather like where you are?
 2 EG: Let's get on with our conversation!
 3 J: Don't you like talking about weather?
 4 EG: All these talks about weather is a waste of time
 (Warwick & Shah 2016: 998)

Al turno 2 EG produce una risposta precostruita (presente anche in altre interazioni), indipendente dall'input ricevuto e non attivata dal co-testo. L'aspetto interessante è la reazione dell'utente umano al turno 3, che rende pertinente l'output di

⁸ Esistono in realtà sistemi che sembrano superare questo modello, come ad es. *Replika*, ma i loro schemi interazionali non verranno trattati qui.

EG, interpretandolo come una violazione della Massima di Relazione e inferendo quindi che l'interlocutore voglia cambiare discorso. A questo punto per EG è facile estrarre il *topic* tramite il semplice riconoscimento di uno schema combinatorio esplicito molto comune (*talk about X*, in cui *X* rappresenta appunto il *topic*), e quindi produrre un messaggio apparentemente coerente ripetendo parte dell'input (*weather*) all'interno di una frase precostruita (*All these talks about X is a waste of time*).

Quest'esempio mostra chiaramente come il compito di dare coerenza al testo dialogico sia in realtà tutto a carico dell'interlocutore umano, che attiva gli abituali meccanismi di cooperazione testuale. L'interfaccia conversazionale interviene sostanzialmente ripetendo segmenti dell'input immediatamente precedente per collegarvi il proprio output.

L'adozione di un rigido schema "postale" della comunicazione, con output direttamente vincolato all'input immediatamente precedente, sembra produrre come conseguenza una dominanza semantica dell'utente umano, unico in grado di controllare gli argomenti portati nella discussione e di riconoscere, nonché creare, una progressione tematica. Accanto a tale dominanza semantica può spesso affiancarsi anche una dominanza interazionale, privilegiando l'utente umano nel controllo delle sequenze conversazionali e potenzialmente limitando l'attivazione di mosse forti da parte della macchina (cfr. Linell & Luckmann 1991: 9).

Esistono tuttavia casi in cui tali strategie sono in qualche modo previste dall'algoritmo, che cerca di ovviare alle difficoltà di mantenimento o prosecuzione di un *topic* introducendo una mossa forte o un cambio di argomento. Si veda ad es. l'estratto seguente:

- (2) 1 J: I am an actress and writer
 2 EG: Well, I'll read something of your books, if they ain't too tedious
 (which is highly possible:-)
 3 J: grammer . . . aren't too tedious! A bit rude of you . . .
 4 EG: Calm down, please.
 5 J: i assure you i am very calm . . . are you?
 → 6 EG: Don't even bother me with the fact that *you are very calm ... are mine*.
 And I forgot to ask you where you are from ...
 7 J: I am from Dorset . . . where are you from?
 (Warwick & Shah 2016: 994)

Al turno 2 viene prodotta una variante substandard (*ain't*), probabilmente all'interno di un segmento precostruito, verosimilmente inserito dal programmatore per caratterizzare il personaggio fittizio di EG. Questa variante viene notata e corretta dal giudice in 3. Al turno 6 EG produce un tipico esempio di *phrase reversing* (tecnica usata già dai primi chatbot⁹), in cui si riprende parte dell'input invertendo i pronomi deittici e inserendolo in una risposta precostruita per forzare l'interazione all'interno del "dominio" gestibile dalla macchina. Infatti, subito dopo la risposta

⁹ Cfr. Demchenko e Veselov (2008: 452). Su questo processo si basa buona parte degli output prodotti da ELIZA, il primo chatbot, creato nel 1966 (cfr. Tavosanis 2018: 86).

allo stimolo ricevuto, EG incorpora nel suo output un'ulteriore mossa conversazionale con cui cerca di introdurre un nuovo *topic* (*And I forgot...*).

È interessante notare come l'errore procedurale prodotto qui (*are mine*) non venga evidenziato dall'utente umano, mentre va a buon fine la mossa forte della macchina, che produce un cambio nella dominanza semantica in questo punto dell'interazione. L'ancoraggio della mossa forte alla ripresa tramite ripetizione dell'input precedente (per mezzo del *phrase reversing*) sembra contribuire significativamente alla buona riuscita del cambio di *topic*.

Come detto, in questo tipo di interazioni il chatbot, e quindi in parte anche l'utente umano, sono costretti a una comunicazione a sensi alterni, in cui la macchina sceglie quale tipo di output produrre in base al riconoscimento di *pattern* o tipi di informazione individuati nell'input ricevuto dall'utente. Il chatbot, sostanzialmente, sembra seguire degli *script* precostruiti con i quali il programmatore cerca di creare un modello predittivo delle mosse conversazionali dell'utente e far rispondere la macchina di conseguenza.

Questo concetto di *script* precostruito ci permette di sottolineare una prima analogia con l'interlingua di apprendenti di una L2: l'aggancio al "copione comunicativo" permette infatti all'apprendente di una L2, così come avviene per i chatbot, di prendere tempo e inserirsi in modo pertinente nella conversazione, pur non avendone colto appieno il senso, in entrambi i casi a causa di una competenza (anche solo discorsiva) parziale.

3. *Le funzioni dell'etero-ripetizione nel parlato*

Il fenomeno della ripetizione è stato oggetto di diversi studi in letteratura: dalla linguistica testuale all'analisi della conversazione, svariate ricerche hanno indagato la struttura e le funzioni di questo meccanismo, nelle sue molteplici manifestazioni. In questo contributo focalizziamo l'attenzione da una prospettiva socio-interazionale sul fenomeno dell'etero-ripetizione, presente sia negli scambi tra parlanti nativi e non nativi che negli scambi tra utente umano e chatbot.

Per etero-ripetizione, o ripetizione dialogica, "si intendono le riprese uguali, anche se parziali, da parte del parlante di turno, di uno o più elementi lessicali presenti nel turno del parlante precedente"; l'etero-ripetizione rappresenta il legame diretto tra fenomeni linguistici e fenomeni sociali e "può essere inserita tra le strategie semiotiche di partecipazione agli scambi" (Bazzanella 1999: 211). Rappresenta, inoltre, un "modo di 'fare la stessa cosa' o 'fare qualcosa insieme' agli altri partecipanti all'interazione", un aggancio che negli apprendenti di L2 è finalizzato a "manifestare l'essere ancora insieme dal punto di vista discorsivo" (Pallotti 1999: 239). Andorno *et al.* (2017: 93) hanno inoltre evidenziato come l'etero-ripetizione, in un approccio interazionista, possa essere operazionalizzata come indice di *intake*, ovvero una manifestazione di ciò che il parlante ha 'estratto' e ha messo 'sotto osservazione' di quanto detto dall'interlocutore nel turno precedente.

Nello studio condotto da Grosso (2015), il fenomeno dell'etero-ripetizione è emerso come caratterizzazione delle interazioni fra non nativi e fra nativi/non nativi esattamente con la macrofunzione attribuita da Pallotti a questo fenomeno.

La presenza di tale strategia di superficie anche nelle produzioni di chatbot evidenzia, a nostro avviso, l'intenzione (più o meno consapevole) dei programmatori di creare un terreno comune con l'interlocutore umano e di permanere nella conversazione il più a lungo possibile.

Bazzanella (1999: 212) individua tre funzioni principali dell'etero-ripetizione: la funzione testuale, finalizzata a "creare la coerenza discorsiva e a segnalare rapporti all'interno della conversazione", anche attraverso la ripetizione di elementi che diventano poi *topic* della conversazione; la funzione conversazionale, finalizzata a far progredire la conversazione con facilità e a manifestare l'interesse dell'interlocutore pur "senza sbilanciarsi a livello personale"; e infine la funzione interazionale, che risponde allo "scopo [...] di mantenere una relazione positiva tra gli interlocutori". Il fenomeno dell'etero-ripetizione ricopre "l'intera gamma funzionale" che va dall'accordo al disaccordo: in particolare, per Bazzanella (1994: 218)

proprio il valore centrale di accordo che di solito veicola la ripetizione, viene sfruttato nei casi di disaccordo, in quanto l'opposizione sembra in parte così attenuata: la presenza stessa della ripetizione, oltre che indicare attenzione, comprensione e continuità – sia pure solo formale – rispetto all'intervento precedente, mette in rilievo il rapporto privilegiato tra gli interlocutori e la costruzione del messaggio come attività comune, negoziata.

Le etero-ripetizioni del chatbot ricoprono esattamente quella funzione di 'aggancio' rispetto al discorso dell'interlocutore, usata tipicamente da un "soggetto che ha difficoltà a riconoscere ciò di cui si sta parlando, ciò che si sta facendo con il linguaggio, e a produrre dei contributi che siano adatti per tali attività" (Pallotti 1999: 242) e che garantiscano coesione al testo interazionale, tramite le parole ripetute, come nell'esempio relativo alla produzione della piccola Fatma, apprendente di italiano L2, citato dall'autore:

- (3) Fatma: comeschiama
(0.4)
Gabriele: questa si chiama pesce
(.)
Fatma: pesce (questa) qui così [non mangiare pesce]

Potremmo osservare quindi che sia apprendenti che chatbot condividono la necessità di essere riconosciuti come interlocutori validi, come "partecipanti ratificati", cioè le cui "rispettive attivazioni per dar luogo ad un corso di azione comune hanno raggiunto il proprio scopo" (Pallotti 1999: 240). Tale riconoscimento viene assicurato dalla possibilità di interagire in maniera coerente e apportando un contributo originale, attraverso la ripetizione di una o due parole del turno precedente dell'interlocutore in un nuovo turno.

4. Alcuni esempi di etero-ripetizioni nell'interlingua di apprendenti e nelle produzioni di chatbot

Presentiamo ora i dati proposti in questo studio. Per quanto riguarda le produzioni in italiano L2, sono stati estratti alcuni esempi dal corpus di interazioni spontanee in contesto lavorativo tra parlanti non nativi e di non nativi con parlanti nativi di italiano raccolto nel corso dello studio pubblicato in Grosso (2015). Si tratta di una selezione molto parziale dei fenomeni di etero-ripetizione evidenziati in tale studio, presentati qui a titolo esemplificativo delle funzioni più comuni. A questi si aggiunge un estratto dal corpus di interazioni spontanee raccolte in un contesto didattico peculiare: il contesto sanitario¹⁰, presentato in un lavoro di Grosso e Floris (2020).

Per quanto riguarda le interazioni con chatbot, sono qui riprodotti alcuni estratti di chat con due interfacce conversazionali liberamente accessibili online (Mitsuku e una nuova versione di ELIZA) raccolte in occasione del seminario *Macchine parlanti* (svoltosi nel 2020 all'interno del corso di *Linguistica Digitale* di Maria Elena Favilla presso l'Università di Modena-Reggio Emilia) da uno degli autori o dagli studenti partecipanti¹¹. La raccolta, la comparazione e lo studio a livello microlinguistico – di analisi della conversazione – delle interazioni sia con chatbot che con parlanti non nativi permette di individuare le peculiarità e le criticità di questo tipo di scambi, da un lato con l'obiettivo di descriverne le caratteristiche strutturali e dall'altro con l'idea di costruire proposte didattiche più efficaci da integrare in percorsi 'tradizionali' di educazione linguistica.

Fra i diversi fenomeni comuni rilevati sia nelle interazioni con chatbot sia nelle interazioni con parlanti non nativi, in questo contributo tratteremo l'etero-ripetizione nelle sue diverse funzioni, testuali o interazionali, e nello specifico nella sua funzione di *trigger* per il cambiamento o per l'espansione del *topic*.

Il frammento che segue, tratto dal corpus raccolto da Grosso (2015), riporta una conversazione tra parlanti non nativi di italiano con diversa L1, un cittadino di origine albanese (D01) e uno di origine rumena (C02). Alla battuta 162, C02 usa la ripresa di una sola parola del turno dell'interlocutore (*destro?*) per chiedere una conferma o un'espansione del turno.

- (4) 161 D01: allora C. (.) eh? (0.4) ora che si
fa, un lavoro grosso eh! (0.1) ora c'è
da fa' tito sarrocchi lato destro!
→ 162 C02: *destro?*

¹⁰ Il corpus è stato raccolto tra il 2018 e il 2019 nell'ambito della tesi di laurea di M. Floris e comprende interazioni spontanee occorse durante un corso di italiano L2 destinato ad apprendenti/pazienti iracheni in cura all'Ospedale Microcitemico di Cagliari.

¹¹ Negli estratti di interazioni con chatbot, SC indica i turni di Simone Ciccolone, S1 e S2 indicano gli studenti, E indica la versione di ELIZA accessibile all'URL masswerk.at/elizabot e M Mitsuku (accessibile all'URL pandorabots.com/mitsuku all'epoca dei test, ora sostituito dalla nuova versione, Kuki, accessibile solo dietro creazione di un account all'URL <https://chat.kuki.ai>). Non è stato possibile raccogliere dati di prima mano da interazioni con EG perché il chatbot non è più accessibile.

Un altro esempio di etero-ripetizione è tratto dallo studio di Grosso e Floris (2020: 111). Nel frammento riportato, l'apprendente arabofono ripete alla battuta 32 la parola *pediatra*, usata dall'insegnante, per un'espansione del *topic* insieme a una richiesta di chiarimento, e focalizza l'attenzione su di esso come argomento condiviso rispetto al quale porre la propria domanda di approfondimento.

- (5) 31 INS: in Italy we have this doctor (.) pediatra from birth until eighteen
 → 32 STUD1: mmm ok (.) *PEDIATRA*? Everyone have pediatra?
 33 INS: just for children
 34 STUD1: pediatra

Si può rintracciare in un'etero-ripetizione la funzione 'richiesta di chiarimento' anche nelle interazioni con chatbot, come mostra l'estratto seguente. Al turno 6 Mitsuku genera un output a partire dall'input ricevuto al turno 5, sostituendo la seconda persona con la prima (tramite applicazione del *phrase reversing*) e *everything* con una forma sinonimica (*all things*) e costruendo una interrogativa che viene interpretata dall'utente umano come una richiesta di conferma. A questa ripetizione, e alla conferma successiva dell'utente (turno 7), segue un'altra coppia adiacente che replica sostanzialmente lo scambio precedente.

- (6) 1 SC: I don't understand
 2 M: Oh I'm sorry. Perhaps I can explain it again better.
 3 SC: yes, thank you
 4 M: Which part are you confused about?
 5 SC: everything in your last sentence
 → 6 M: All things in my last sentence?
 7 SC: yes
 8 M: Surely not all?
 9 SC: sure

Nell'estratto seguente, invece, sempre relativo a un'interazione sul luogo di lavoro, in questo caso tra un siciliano (L22) e un parlante non nativo di origine peruviana (H7), si possono notare diversi fenomeni di ripetizione: al turno 3 abbiamo prima la ripresa dell'apprendente, che usa l'etero-ripetizione come meccanismo coesivo, introducendo però uno *switch* intenzionale alla sua L1; subito dopo, al turno 4, L22 utilizza a sua volta l'etero-ripetizione per manifestare allineamento nei confronti dell'interlocutore.

La funzione di coesione e accomodamento dell'etero-ripetizione al turno 3 viene rafforzata dall'evidente funzione di *intake* che essa svolge per il parlante nativo nell'ambito del meccanismo di *crossing* (cfr. Rampton 1995: 485), una pratica interazionale consistente nell'appropriarsi della lingua dell'interlocutore con la funzione precipua di creare un terreno di socializzazione favorevole in contesti culturalmente e linguisticamente eterogenei.

- (7) 2 L22: (.) mamma mia quanto sei bella! ((riferito a una passante))
 ma:mmamia mammamia quanto sei bella! mi fa male la testa
 → 3 H07: ((ride)) (.) *ti fa male la cabeza?*
 4 L22: mi fa male la cabeza!

L'estratto 8 presenta un'altra etero-ripetizione di Mitsuku, usata qui per cercare di "riagganciare" l'utente e proseguire l'interazione. La strategia è in tutto simile a quella adottata da EG nell'estratto 2: Mitsuku riconosce lo schema *talk about X* e interpreta tutto il materiale seguente come *topic* (*nothing, I really have to go*), lo trasforma tramite *phrase reversing* e lo inserisce in un modello di output precostruito (*X sounds like a good topic*); forza poi l'introduzione di un nuovo *topic* tramite un'aggiunta (*I like mousebreaker*) che dovrebbe assumere funzione di espansione del tema precedente, ma che qui ha un effetto straniante per via dell'estrazione di un *chunk* non valido come tema.

- (8) 1 S1: Okay, I think we're talking about nothing, I really have to go.
 2 M: Yeah ok sure, *nothing you have to go* sounds like a good topic.
 I like mousebreaker.

Questa strategia (del programmatore) realizzata tramite processi sostanzialmente meccanici (del software) potrebbe risultare tuttavia efficace in caso di corretta individuazione del *topic*, ovvero nei casi in cui lo slot *X*, preso come variabile estratta dall'input e riprodotta nell'output, sia un'espressione referenziale ben isolata. Resterebbe una possibile incongruenza testuale tra il *topic* estratto e l'espansione proposta dal chatbot, tuttavia risolvibile sostituendo il singolo referente con un elemento variabile estratto sulla base di relazioni semantiche¹².

Si noti però la funzione che tale ripetizione assume per l'utente umano: l'inserimento dell'elemento ripreso all'interno di una valutazione positiva (*sounds like a good topic*) e il tentativo di espansione hanno lo scopo di produrre un avvicinamento emotivo, di creare uno spazio di socializzazione condiviso, in modo simile a quanto attuato dal parlante non nativo nell'estratto 7. Tale tentativo di avvicinamento ha intanto l'effetto di prolungare l'interazione, e se accolto può non solo cedere temporaneamente la dominanza semantica al chatbot, ma anche incentivare il riconoscimento del chatbot come interlocutore paragonabile a un essere umano. Non è un caso che questa strategia (con valutazioni emotive favorevoli dell'interlocutore) sia sistematicamente adottata da chatbot finalizzati a intrattenere relazioni comunicative a lungo termine con l'utente, come *Replika*.

Presentiamo un altro esempio di etero-ripetizione tramite "appropriazione" della lingua dell'interlocutore, ovvero passaggi dell'interazione in cui uno degli interlocutori utilizza la pratica discorsiva del *crossing*, aspetto già anticipato nell'estratto 7. L'etero-ripetizione, all'interno di una sequenza di *small talk*, anche qui si configura come esempio di appropriazione dell'(inter)lingua dell'interlocutore da parte della lavoratrice sene (B19), che la utilizza per una bonaria presa in giro del collega moldavo.

- (9) 2 C17: ((entra nel camion)) ((canta))
 comm'è bell' comm'è bell!
 c'hai mess' u: shcaldament'?
- 3 B19: *scaldamento* (ride) l'ho messo lo *scaldamento*!

¹² L'ultima versione di Mitsuku testata (maggio 2021) mostra un uso più evoluto di tale strategia, estraendo possibili *named entities* nell'input e cercando informazioni sul web.

In questo esempio, attraverso l'etero-ripetizione l'interlocutrice B19 trasforma in *topic* la considerazione metalinguistica riguardante la produzione del collega non nativo, una neoformazione (*scaldamento*) ottenuta dalla semplificazione, intesa come omissione di un morfema di scarsa salienza fonica, della parola *riscaldamento*¹³.

Dai dati dello studio condotto da Grosso (2015) emerge un uso più consistente dell'etero-ripetizione nelle sequenze di *small talk*, il parlato relazionale (60%), piuttosto che nella comunicazione funzionale (40%), ovvero *task-oriented*. Questo dato è assolutamente comprensibile, date le caratteristiche delle due tipologie di parlato: il discorso relazionale, per natura "sfilacciato", non orientato verso un obiettivo informativo preciso, richiede una dimostrazione di maggior impegno conversazionale e, dunque, il maggior utilizzo di strategie pragmatiche quali l'uso di etero-ripetizioni da parte degli interlocutori nell'intessere la relazione e mantenere salda la trama del discorso.

La funzione dell'etero-ripetizione di "creare solidarietà con l'interlocutore" (Grosso 2015: 77) si dimostra infatti di fondamentale importanza per la creazione di legami sia a livello della superficie del discorso, così come aveva rilevato Pescarini (2004), sia di legame discorsivo profondo e di coesione tra gli interlocutori, esattamente come avviene nelle interazioni con i chatbot progettati per lo *small talk*. In tale tipo di interazioni quest'aspetto risulta di notevole rilievo e in qualche misura previsto nella progettazione delle interfacce conversazionali stesse, soprattutto di tipo *open domain* e interessate, quindi, all'instaurazione di un dialogo coinvolgente e duraturo con l'utente.

In quest'ottica il compito di estrazione del *topic* diventa indispensabile per raggiungere lo scopo finale, benché al contempo più difficile (in virtù del dominio "aperto" e non circoscritto a un insieme prevedibile di campi semantici). Spesso per proseguire l'interazione efficacemente il chatbot dovrebbe essere in grado di ripercorrere più turni precedenti, o di produrre un output sulla base non di input isolati ma di sequenze più estese.

Nell'estratto 10, ELIZA mostra difficoltà a proseguire l'interazione, tanto da ricorrere a una richiesta indiretta di chiarimento (turno 4). La reazione dell'utente (*whatever*) non lascia però spazi di manovra, per cui ELIZA sembra costretta a scansionare gli input precedenti alla ricerca di un possibile "aggancio" tematico da richiamare – sempre tramite *phrase reversing*, stavolta però proposto nella forma di discorso riportato (turno 6).

- (10) 1 S2: None is normal, don't you think?
 2 E: Oh... think?
 3 S2: is an abstract concept
 4 E: I'm not sure I understand you fully.
 5 S2: whatever
 → 6 E: OK... "None is normal, don't me think".
 Tell me more

¹³ Tale strategia di omissione può essere annoverata tra le forme di semplificazione intesa come "esito di una strategia che l'apprendente mette in atto" (Andorno *et al.* 2017: 37).

Qui la ripresa è strettamente funzionale a superare una *impasse* nell'interazione, e il recupero di un input precedente (all'apprezzabile distanza di 5 turni) è usato proprio allo scopo di riprendere una fase produttiva dello scambio comunicativo e chiederne l'espansione (*tell me more*). L'etero-ripetizione al turno 6, nella forma di discorso riportato, mostra di assumere quindi una duplice funzione: innanzitutto come ripresa testuale, ovvero di riproduzione di un *chunk* realizzato da S2 come aggancio di superficie per dare coesione al testo; in secondo luogo, assume una funzione tematica/interazionale, per creare solidarietà con l'interlocutore tramite il richiamo di un *topic* precedentemente evocato dall'utente umano.

In sintesi, nell'osservazione di possibili "strategie" negli output prodotti da chatbot si sovrappongono più livelli: la presenza di strategie previste dall'autore del chatbot (in virtù di un utente-Lettore Modello) e il *bias* interpretativo del "lettore" umano (che tende a riconoscere nel chatbot un Autore Modello umano). Al netto di tali possibili estensioni nell'interpretazione degli interventi dei chatbot, possiamo riconoscere una strategia generale più o meno programmata nell'uso delle ripetizioni, che ha come obiettivo centrale quello di (ri)attivare un possibile *topic* prodotto dall'utente per catturare la sua attenzione e prolungare quindi la conversazione, come ben illustrato proprio dall'ultimo esempio.

5. Riflessioni conclusive

Come conclusione preliminare, possiamo osservare come la ripetizione rappresenti una risorsa testuale e interazionale estremamente efficiente e produttiva¹⁴, anche quando "filtrata" da più meccanismi generativi o interpretativi che si sovrappongono, come in particolare nel caso delle interfacce conversazionali.

Il ricorso a "script" precostruiti che accomuna le produzioni degli apprendenti umani di una L2 e delle "intelligenze conversazionali", osservato qui in pochi esempi, dovrà essere studiato in un insieme di dati molto più esteso e confrontato con corpora di apprendenti di italiano L2, come ad esempio il LIPS. In questi primi appunti di analisi abbiamo evidenziato come per i chatbot le funzioni dominanti dell'etero-ripetizione sembrano essere quelle di mantenere o riattivare un *topic* prodotto dall'utente umano per richiederne l'espansione, oppure semplicemente per simulare la comprensione e condivisione di uno spazio discorsivo comune, proprio allo scopo di prolungare l'interazione.

Un altro possibile spunto di riflessione, non sviluppato qui ma che sarebbe da prendere in considerazione, riguarda non le produzioni linguistiche della macchina, ma degli interlocutori umani. Ovvero, se da un lato abbiamo supposto l'esistenza di "interlingue simulate" per i chatbot, dall'altro possiamo sicuramente ipotizzare un corrispettivo del *foreigner talk*, ovvero un *computer talk* utilizzato dagli utenti umani con i chatbot (cfr. Tavosanis 2018: 76, nonché Fischer 2006). Questa varietà potrebbe risultare scarsamente connotata sul piano delle strutture di superficie (mo-

¹⁴ A tal proposito si vedano le interessanti riflessioni di Tannen (1987).

strando una generica predilezione per ordini non marcati e strutture lineari, nonché iperarticolazione nel parlato), ma concentrare i propri tratti salienti nel campo delle mosse conversazionali più efficienti nel garantire una progressione nell'interazione con la macchina.

Le riflessioni fin qui condotte ci portano a immaginare prospettive di sviluppo in campo glottodidattico. In questo senso sarebbe utile indagare l'impatto e il potenziale di chatbot sia educativi che non educativi per il miglioramento delle competenze discorsive e pragmatiche nella lingua straniera (L2), come già proposto in alcuni studi pionieristici tra i quali Bii (2013), Mazzilli (2018, 2019, 2021), Istrate (2018). In qualità di "intelligent conversational agents with complex, goal-driven behaviour" (Kerly *et al.* 2006: 178), i chatbot hanno già trovato, e continueranno a trovare, un'utile applicazione anche nel campo dell'educazione linguistica: si vedano ad es. i bot su Duolingo per tedesco, francese e spagnolo (Mazzilli 2019: 145-7). Alla base di questo possibile utilizzo vi è l'idea che l'acquisizione di competenze linguistiche possa essere rafforzata dalle opportunità di interazione che i parlanti hanno in vari ambienti di apprendimento. È noto in letteratura che i chatbot rappresentano un utile strumento per potenziare gli ambienti di apprendimento, considerando soprattutto che essi vengono progettati perseguendo gli obiettivi di "provide information, answer questions, discuss a specific topic, or perform a task" (Smutny & Shreyberova 2020: 1), tutte funzioni comunicative che gli apprendenti di una lingua hanno sicuramente necessità di esercitare. Lo studio citato evidenzia la presenza al primo posto di chatbot legati all'apprendimento delle lingue: la maggior parte degli 89 chatbot inclusi nel campione e analizzati sono chatbot per l'apprendimento linguistico. Tra essi figurano Wordsworth, English with Edwin, NELA ed altri. L'ambiente di apprendimento costituito dalla chat in cui l'apprendente dialoga con il chatbot gode dei vantaggi che caratterizzano in generale l'uso didattico del feedback immediato costituito dalle chat di messaggistica istantanea, favorendo il coinvolgimento attivo dell'utente e la percezione dell'apprendimento come attività ludica, determinando, come avviene in generale per le attività basate sulla *gamification* (cfr. Lombardi 2019), un aumento della motivazione e una facilitazione nell'avvicinamento dell'apprendente ai contenuti didattici.

Bibliografia

- Allen, James. 1995. *Natural Language Understanding*. Redwood: Benjamin/Cummings (2nd ed.).
- Andorno, Cecilia & Valentini, Ada & Grassi, Roberta. 2017. *Verso una nuova lingua. Capire l'acquisizione di L2*. Torino: UTET.
- Bazzanella, Carla. 1999. Forme di ripetizione e processi di comprensione nella conversazione. In Galatolo, Renata & Pallotti, Gabriele (a cura di), *Prospettive sulla conversazione*, 205-226. Milano: Raffaello Cortina.
- Bii, Patrick. 2013. Chatbot Technology: A Possible Means of Unlocking Student Potential to Learn How to Learn. *Educational Research* 4(2). 218-221.

- Demchenko, Eugene & Veselov, Vladimir. 2008. Who Fools Whom? The Great Mystification, or Methodological Issues on Making Fools of Human Beings. In Epstein, Robert & Roberts, Gary & Beber, Grace (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 447-459. Heidelberg: Springer.
- Epstein, Robert & Roberts, Gary & Beber, Grace (eds). 2008. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Heidelberg: Springer.
- Fischer, Kerstin. 2006. *What Computer Talk Is and Isn't: Human-Computer Conversation as Intercultural Communication*. Saarbrücken: AQ.
- Grosso, Giulia Isabella. 2015. *Interazioni in italiano lingua franca sul luogo di lavoro: una prospettiva pragmatica*. Pisa: ETS.
- Grosso, Giulia Isabella & Floris, Moana. 2020. L'italiano in ospedale. Analisi di interazioni e sperimentazione di un percorso didattico per bambini stranieri ospedalizzati. In Marra, Antonietta & Dal Negro, Silvia (a cura di), *Lingue minoritarie tra localismi e globalizzazione*, 111-118. Milano: AItLA.
- Huang, Minlie & Zhu, Xiaoyan & Gao, Jianfeng. 2020. Challenges in Building Intelligent Open-domain Dialog Systems. *ACM Transactions on Information Systems* 38(3). n. 21.
- Istrate, Ana Mihaela. 2018. Artificial Intelligence and Machine Learning. Future Trends in Teaching ESL and ESP. *eLearning & Software for Education* 2. 471-476.
- Kerly, Alice & Hall, Phil & Bull, Susan. 2006. Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models. *Knowledge-Based Systems* 20(2). 177-188.
- Linell, Per & Luckmann, Thomas. 1991. Asymmetries in Dialogue: Some Conceptual Preliminaries. In Marková, Ivana & Foppa, Klaus (eds), *Asymmetries in Dialogue*, 1-20. Hemel Hempstead: Harvester Wheatsheaf.
- Lombardi, Ivan. 2019. *Motivazione, gioco, lingua. Elementi ludici tra glottodidattica e psicolinguistica*. Milano: AItLA.
- Mazzilli, Francesca. 2018. Bot talk e apprendimento linguistico. L'uso dei chatbot per lo sviluppo della competenza comunicativa nella lingua straniera. *Iperstoria* 12. 13-36.
- Mazzilli, Francesca. 2019. Kriterien und Vorschläge zur Evaluation textbasierter Chatbots für die touristische Sprachbildung am Beispiel von Lufthansa BestPrice und Duolingo Bots. *Sprache im Beruf* 2(2). 142-64.
- Mazzilli, Francesca. 2021. Chatbots for Action-Oriented Language Learning. Using Elbot to Enhance Conflict-Solving Skills in Learners of German as a Foreign Language. *EL.LE* 10(1). 95-116.
- McTear, Michael & Callejas, Zoraida & Griol, David. 2016. *The Conversational Interface. Talking to Smart Devices*. Heidelberg: Springer.
- Pallotti, Gabriele. 1999. Fatma ripete per fare qualcosa insieme. *Italiano & oltre* 14. 239-245.
- Pescarini, Diego. 2004. La coesione dei turni nella conversazione italiana. *Quaderni patavini di linguistica* 20. 43-74.
- Rampton, Ben. 1995. Language Crossing and the Problematisation of Ethnicity and Socialisation. *Pragmatics* 5(4). 485-513.

Smutny, Pavel & Schreiberova, Petra. 2020. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education* 151.

Tannen, Deborah. 1987. Repetition in conversation: toward a poetics of talk. *Language* 63(3). 574-605.

Tavosanis, Mirko. 2018. *Lingue e intelligenza artificiale*. Roma: Carocci.

Warwick, Kevin & Shah, Huma. 2016. Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence* 28(6). 989-1007.

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

Abstract

The comparison of foreign language learners' pronunciation with the utterances of native speakers is receiving increasing attention thanks to the new applications of computer-assisted teaching. The studies on prosodic variation between several native speakers reveal rhythmic-intonative patterns that should be part of the linguistic baggage of the native speaker; it seems problematic how to make explicit to the learner his degree of prosodic competence on the basis of the different acoustic indices (f₀, intensity, pitch) and other parameters (number of syllables, lexical accent, speech rate, pauses). We present a chatbot designed as a proactive learning support for the improvement of oral skills in Italian L2. After describing its technical structure, the first results of a preliminary study are discussed. They provide a first starting point to reflect on some considerations that emerged.

1. *Introduzione*

Le caratteristiche prosodiche del parlato rivelano importanti informazioni come la modalità enunciativa, lo stile d'eloquio, l'attitudine e le connotazioni diatopiche o diastratiche del locutore; esse sono quindi essenziali per una buona riuscita della comunicazione interlocutiva (Hirst 1983, Cruttenden 1986). Tuttavia, in molti contesti relativi alla didattica delle lingue straniere (LS) la riflessione metaprosodica non è sempre soddisfacente per lo studente poiché non viene affrontata in modo esaustivo (Trouvain & Gut 2007). Riferirsi al ruolo di caratteristiche prosodiche come l'intonazione, il ritmo, la durata, la focalizzazione per spiegare differenze comunicative di significato può invece rivelarsi indispensabile per innescare nello studente una consapevolezza utile per riconoscere e riprodurre il parlato della LS (Cresti 1999, Chun 2002, Frost & Picavet 2014). Inoltre, ormai da diversi decenni, gli sviluppi tecnologici nell'analisi del parlato hanno aperto la strada a nuove prospettive di interazione comunicativa declinabili anche per la didattica delle lingue (Chun 1998, Cazade 1999), facendo emergere, sin dai primissimi studi condotti, anche varie criticità verso le possibili soluzioni di esplicitazione del grado di compe-

¹ Università degli Studi di Torino.

² Politecnico di Torino.

tenza raggiunto, ad esempio tramite l'utilizzo di rappresentazioni grafiche di supporto (James 1976, De Bot 1983, Martin 2010).

La didattica incentrata sull'oralità di una LS permette proprio quel tipo di riflessione metaprosodica grazie alla quale lo studente può sperimentare la varietà di combinazioni enunciative presenti nella lingua appresa e associarne il significato in termini di funzioni specifiche (modali, sintattiche, informative, emozionali) per affinare le proprie competenze comunicative orali (Delattre 1966, Cresti 1999). In italiano bisogna inoltre tener conto delle diverse scelte intonative anche sulla base della varietà diatopica utilizzata (Canepari 1985, Sorianello 2006), una fonte di variabilità inclusa ormai da diverso tempo negli spazi della ricerca sull'intonazione della L2 (De Meo & Pettorino 2012, De Marco *et al.* 2014) che, discussa nei contesti didattici, porterebbe a una maggior sensibilizzazione da parte dell'apprendente, molto spesso esposto a un solo modello diatopico. Infine, ci sembra interessante soffermarci sulle diverse possibilità di realizzazione tipiche della lingua orale e su come occorra tenerne conto nel contesto della valutazione automatica e nella ricorrenza di andamenti intonativi.

Sulla base di queste considerazioni, questo studio si struttura nel seguente modo: dopo aver presentato l'architettura del chatbot, analizzeremo come avviene il confronto tra i dati di parlanti madrelingua italiana e apprendenti di italiano L2 a partire da un set di dieci frasi complesse costruite in modo da dare un'adeguata rappresentazione delle più frequenti soluzioni intonative di un parlato elicitato su contenuti di natura tecnico-scientifica, includendo però al tempo stesso anche alcune espressioni quotidiane. Discuteremo infine, sulla base dei risultati ottenuti da parte degli italofoeni e degli studenti, una serie di riflessioni nell'ottica futura di implementare successivi parametri per la valutazione del *chatbot*, nello specifico in che modo i valori dei parametri acustici rilevati dal chatbot rispecchiano una valutazione percettiva (Munro & Derwing 1999).

2. Architettura e funzionamento del chatbot

In questo paragrafo illustriamo in che modo è stato strutturato il *chatbot*³ come strumento di didattica assistita per la valutazione di tratti prosodici. L'adozione di un chatbot di questo tipo permette diversi vantaggi. Innanzitutto prevede una modalità di interazione, basata su testo, a cui il pubblico è già largamente abituato grazie alla popolarità di applicazioni di messaggistica istantanea quali *Whatsapp* o *Telegram*. Ciò rende questi strumenti più facilmente fruibili rispetto ad applicazioni dedicate o ai cosiddetti *Learning Management System* (LMS) come *Moodle* o *Blackboard*, il cui maggior ostacolo alla fruibilità è rappresentato proprio dalla necessità da parte degli utenti di adattarsi a un'interfaccia poco familiare. Un *chatbot* permette inoltre di interagire con gli apprendenti in maniera strutturata e asincrona: strutturata

³ Il *chatbot* è un software progettato per simulare una conversazione con un utente umano (Fernoagă *et al.* 2018) che vede, tra le diverse applicazioni, anche quella educativa e didattica (Colace *et al.* 2018).

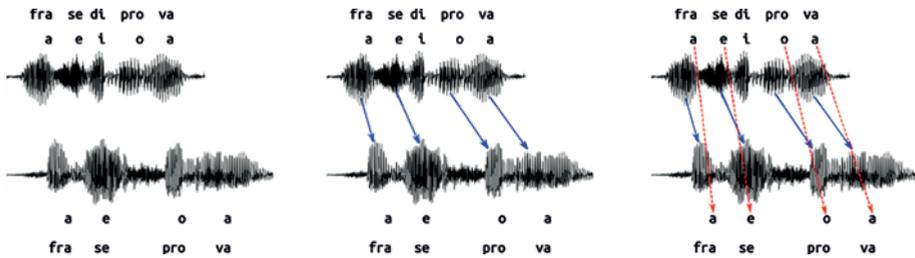
perché organizzata secondo schemi predefiniti, volti a ridurre le ambiguità; asincrona perché non vincolata a intervalli di tempo continuativi ma fruibile liberamente secondo le proprie disponibilità di tempo. Le potenzialità relative all'uso di *chatbot* come strumenti di ausilio alla didattica sono investigate già da alcuni anni (Pereira 2016, Fernoagà *et al.* 2018) e in diversi casi si osserva un orientamento verso la valutazione didattica della pronuncia (Cheng *et al.* 2020) e della prosodia (Lezhenin *et al.* 2017). Realizzato all'interno dell'applicazione di messaggistica istantanea *Telegram*, il *chatbot* (o bot) qui presentato prevede l'interazione con l'utente attraverso domande e risposte basate su una serie di domande a risposta chiusa (quiz) che possono avere carattere generale di comprensione linguistica o essere legate a un particolare ambito disciplinare (ad esempio tecnico-scientifico). Una volta individuata la risposta corretta, l'apprendente invia la registrazione della frase letta sotto forma di messaggio vocale. L'enunciato ricevuto viene così processato automaticamente dal bot al fine di ottenere una valutazione del livello intonativo del parlante. Tale processamento è riassumibile nei seguenti passi:

1. Preprocessing
2. Segmentazione fonetica
3. Estrazione dei valori di frequenza fondamentale
4. Analisi intonativa

Il primo passo nella catena di processamento ha un duplice scopo: garantire un livello di qualità minimo per il segnale audio di input e preparare quest'ultimo alle elaborazioni successive. L'audio viene innanzitutto convertito in formato *wav* a singolo canale, ricampionato (se necessario) a 48 kHz e ripulito dal rumore di fondo costante⁴. Vengono poi stimati il valore medio di ampiezza e il rapporto segnale/rumore dell'audio risultante. Nel caso in cui i valori stimati risultino inferiori a determinate soglie predefinite, il *chatbot* richiede all'apprendente di registrare un nuovo enunciato in un ambiente meno rumoroso e/o parlando a un tono di voce più alto. Il segnale vocale viene successivamente sottoposto a segmentazione mediante il servizio web WebMAUS Basic (Kisler *et al.* 2018). Tale servizio, una volta inviato il segnale vocale come input e la trascrizione ortografica dell'enunciato fornita dal bot, ne restituisce una segmentazione in parole e fonemi sulla base del metodo descritto in Schiel (1999). La segmentazione fonetica dell'enunciato, fornita in formato *TextGrid* per agevolare la successiva elaborazione mediante il software Praat (Boersma & Weenink 2018), viene poi processata etichettando i singoli foni come vocali o consonanti (fig. 1). Uno script Praat⁵ viene poi richiamato per estrarre i va-

⁴ Questa procedura non pregiudica la valutazione poiché rimuove solo quelle componenti di rumore costante del segnale, ad esempio ronzii, fischi e sibili tipicamente dovuti all'apparecchiatura utilizzata per la registrazione o alla vicinanza del parlante a fonti di rumore continuativo.

⁵ Lo script è stato creato da Antonio Romano per agevolare l'applicazione del metodo AMPER nell'ambito di Matlab™ e successivamente implementato per Praat da Plinio Barbosa (Campinas, Brasile) e adattato da Albert Rilliard (Parigi). Lo script è stato successivamente automatizzato da Marco Gamba (UniTo) e, per questo lavoro, da Marco Palena.

Figura 2 - *Audio segmentato, allineamento temporale, corrispondenza dei segmenti*Figura 3 - *Schermate esemplificative dei passaggi principali dell'interazione chatbot-utente*

Lo svolgimento del task (fig. 3) prevede una prima fase di registrazione in cui l'utente fornisce alcuni dati sociolinguistici (il livello linguistico in base al QCER, se e dove ha vissuto in Italia, in che contesti è abituato a parlare italiano). Successivamente,

dopo aver effettuato una prova microfono che garantisca una sufficiente qualità dell'audio, l'utente può cominciare il task che consiste nel rispondere a 10 domande in italiano di discipline varie (grammatica, storia, matematica) registrando mano a mano la lettura individuale di ogni risposta: ciascuna domanda è presentata in forma scritta e orale⁷ mentre la scelta della risposta è compresa tra quattro opzioni (cfr. appendice per l'elenco completo delle frasi).

Se l'utente sbaglia risposta, può riprovare e, una volta selezionata la risposta corretta, il *chatbot* propone di registrare la risposta tramite un messaggio vocale. A questo punto dell'interazione, la frase elicitata viene comparata con quelle presenti nel database prodotte da parlanti italofofoni (si veda § 3 per la descrizione dettagliata della base dati) e dopo aver individuato quella con la vicinanza intonativa maggiore, il *chatbot* restituisce all'apprendente il valore di correlazione in percentuale⁸. Al completamento del task, viene infine restituito un punteggio complessivo ottenuto dalla media dei valori percentuali per ciascuna risposta.

3. *La base dati di riferimento di parlanti italofofoni*

Il confronto con dati di parlanti madrelingua è impiegato da diversi anni (Delmonte 2009)⁹ e si sposa perfettamente con le nuove modalità di valutazione dei sistemi automatici di riconoscimento del parlato. In una prima fase del progetto abbiamo quindi creato un set di 10 domande e risposte da utilizzare come dataset di riferimento da somministrare successivamente agli studenti. Nelle risposte, tra cui è presente la frase target, sono state inserite date, formule matematiche¹⁰ e strutture sintattiche mediamente complesse (l'elenco delle frasi è consultabile in appendice) per avere degli enunciati di difficoltà di lettura medio-alta, rivolte a un pubblico di studenti con un livello minimo B1. Il set è stato quindi sottoposto alla lettura da parte di 250 persone che hanno registrato domanda e risposta tramite un dispositivo cellulare portando alla creazione di una base di dati di 2500 frasi. Gli audio sono stati quindi ricampionati a 16kHz e, tramite una segmentazione manuale, sono state segmentate ed etichettate soltanto le risposte. I campioni registrati provengono da diverse parti d'Italia (la metà dal Piemonte)¹¹ e si dividono in 164 femmine e 86 maschi con un'età compresa tra i 15 e i 70 anni. La comparazione permette di trovare la frase all'interno della base di dati dei campioni italofofoni che più si avvicina a quella

⁷ La domanda è letta da uno speaker professionista senza inflessioni regionali.

⁸ Il confronto tra utente e parlante di riferimento non è legato al genere poiché il confronto si basa sulle differenze di intonazione e non sulla frequenza media del parlante.

⁹ Una rassegna bibliografica dettagliata è disponibile in Romano & Giordano (2017).

¹⁰ La lettura di formule matematiche e, in generale, di numeri può essere complessa per un apprendente L2, non a caso l'ultima risposta del set (" $1/6 + 3/2$ ") ha ricevuto da parte degli utenti una lettura per lo più erronea (si veda tabella in appendice).

¹¹ Sebbene in questa prima fase gli italofofoni registrati provengano principalmente dal Piemonte, uno degli obiettivi futuri consiste nell'ampliare la base di dati così da avere una proporzione più significativa tra le varie aree regionali d'Italia e poter quindi eventualmente correlare anche le caratteristiche prosodiche regionali.

pronunciata dallo studente e di restituirne la percentuale di correlazione massima. È importante sottolineare che la percentuale restituita alla fine di ogni risposta data esprime esclusivamente il parametro acustico responsabile dell'intonazione, ovvero la frequenza fondamentale (f_0): di conseguenza, per offrire una valutazione più accurata, le frasi di riferimento prodotte dagli italofoeni sono state valutate da parte di due operatori umani per poter successivamente correlare alcuni indici acustici (come la velocità d'eloquio, la fluenza o la scansione pausale) con la valutazione percettiva. Nello specifico, per ogni frase si segnalano¹²:

1. la velocità di lettura (lenta, normale o veloce);
2. l'accento regionale (punteggio da 1 – accento non riconducibile a una specifica varietà regionale – a 3 – accento marcato regionalmente);
3. la scansione delle parole (lenta, normale o veloce);
4. la fluenza intonativa (punteggio da 1 – min spontaneità – a 3 – max spontaneità);
5. l'età del locutore (bambino, ragazzo, adulto o anziano).

4. *Primi risultati su un campione di studenti francofoni*

In un primo tempo¹³, per testare il chatbot su un gruppo linguistico omogeneo, il task è stato sottoposto a 15 studenti universitari francofoni (12 femmine e 3 maschi) di un'età compresa tra i 20 e i 22 anni che studiano l'italiano da 2-5 anni¹⁴; dal totale di 150 frasi sono state escluse 15 frasi perché non sono state registrate dall'utente o perché non sono state lette correttamente¹⁵. Le frasi prodotte sono state quindi confrontate con quelle della base di dati che, per questo set, hanno ricevuto generalmente una valutazione percettiva massima (97% dei casi) e media (3%). Gli utenti francofoni invece hanno ricevuto un punteggio alto (3) nel 47% dei casi (63 frasi), medio (2) per il 47% (63 frasi) e basso (1) nel restante 6% (9 frasi). Mediamente per ciascuna frase gli utenti hanno ricevuto una valutazione percettiva dell'intonazione di 2 nella metà dei casi e 3 nel restante, tranne la frase 9 per la quale soltanto 5 utenti hanno ricevuto una valutazione media (2). Se osserviamo i confronti tra le frasi degli utenti francofoni con i riferimenti italofoeni con un punteggio alto, notiamo che la media della vicinanza intonativa delle frasi si attesta tra il 37% e il 94%. Più precisamente 6 frasi hanno una vicinanza compresa tra il 25 e il 50%, 32 tra il 50 e il 75% e 24 tra il 75 e il 100%. Nel confronto grafico di due curve intona-

¹² In questa sede viene presa in esame soltanto la fluenza intonativa, gli altri parametri percettivi sono in corso di analisi e saranno oggetto di un altro lavoro.

¹³ Il task è stato somministrato nel tempo a più gruppi di studenti con provenienza mista (anglofoni, ispanofoni, ecc.) ma per questa prima presentazione si è scelto di focalizzare l'analisi sul campione di francofoni.

¹⁴ In questo lavoro non abbiamo preso in considerazione e correlato le caratteristiche sociolinguistiche degli studenti limitandoci a una presentazione dei dati qualitativa.

¹⁵ In questo caso, ci riferiamo all'ultima frase ("1/6 + 3/2") che è stata letta "un sesto più tre mezzi" soltanto da 4 utenti.

tive utente-campione riferimento (fig. 4) con un'alta percentuale di correlazione del 92%, si nota ad esempio un andamento di f_0 complessivamente simile.

Figura 4 - Confronto delle curve intonative tra parlante di riferimento (segmenti continui) e utente (segmenti tratteggiati) della frase "Il 14 luglio 1789" a partire dai valori dei segmenti vocalici estratti (righe verticali)

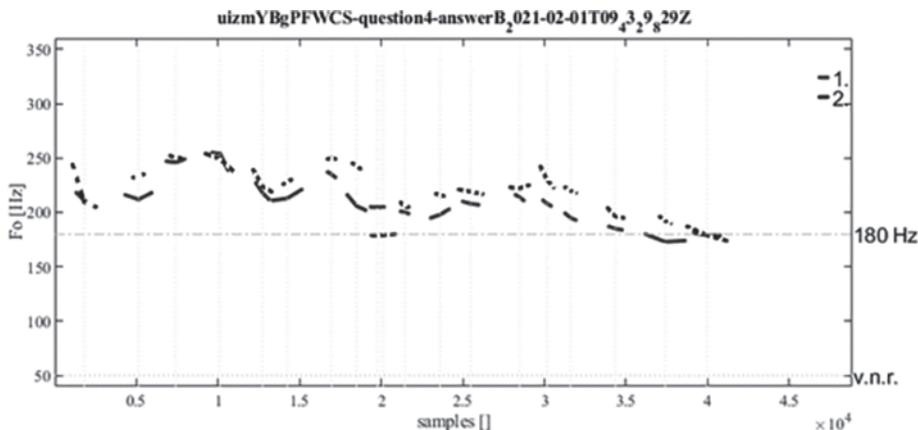
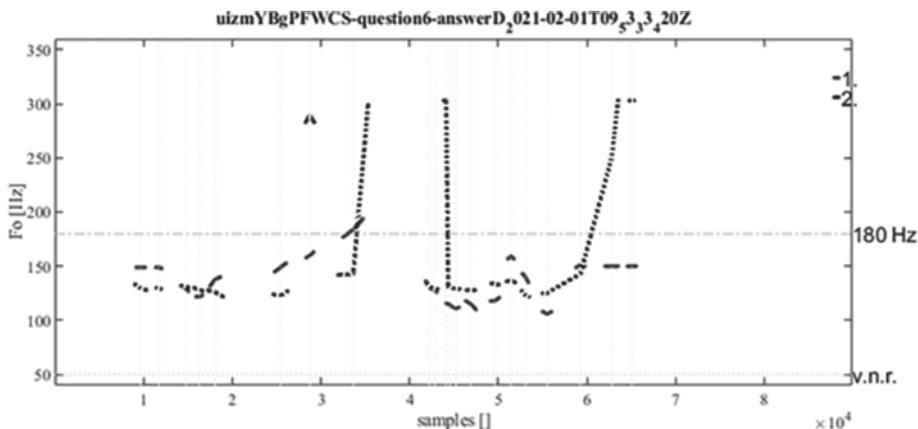


Figura 5 - Confronto delle curve intonative tra parlante di riferimento (segmenti continui) e utente (segmenti tratteggiati) della frase "Se fossi venuto con me, ti avrei portato a mangiare una pizza" a partire dai valori dei segmenti vocalici estratti (righe verticali)



Sono stati anche analizzati quei casi in cui sia il locutore italofono che l'utente francofono hanno un punteggio percettivo alto (3) ma la percentuale di correlazione intonativa è bassa: poiché ci aspettiamo che laddove la valutazione percettiva sia in entrambi i casi alta, anche il punteggio sia alto, occorre capire se un'eventuale percentuale bassa sia dovuta ad altri fattori (segmentazione di Maus non corretta, qualità scarsa della registrazione, bassa presenza di segmenti allineabili). La figura 5 mostra ad esempio che, nonostante la valutazione percettiva dell'intonazione

un largo campione, è in corso la fase di quantificazione e soluzione dei casi in cui gli enunciati si presentino difformi da quelli attesi a causa di un ordine di parole diverso e/o della presenza di frammenti involontari di parlato (balbettii, esitazioni, false partenze) che causano differenze tali da impedire il conseguimento di una minima distanza prosodica. Dopo aver concluso l'etichettatura della base dati per gli altri parametri descritti, vorremmo concentrarci su quelli prodotti dagli studenti per una valutazione fonetico-percettiva da comparare in un secondo tempo con gli indici acustici principali. In un'ultima fase, vorremmo incrementare ed equilibrare il corpus di riferimento dei parlanti madrelingua italiani per ampliare le varietà enunciate in termini diatopici e diafasici (Crocco 2017). In questa direzione, un ulteriore passo riguarda l'eventuale classificazione dalla quale partire per addestrare un algoritmo in grado di sostituire l'operatore umano lungo una linea che distingue un parlato più artefatto da uno sciolto e spontaneo (Nencioni 1976, Voghera 1989, Zmarich *et al.* 1996).

Infine, un aspetto ancora poco approfondito riguarda la lettura di enunciati di lettura più complessa, che prevedano la presenza di semplici formule matematiche, sigle, acronimi e forestierismi, senza trascurare le espressioni richieste dalle soluzioni enunciate tipiche di alcuni linguaggi specialistici. Su questi aspetti nessuna formazione specifica è prevista nei curricula di avvicinamento alle materie tecnico-scientifiche e nessuna pubblicazione nazionale dà indicazioni esaustive. Ad esempio, per quanto tutti riconoscano espressioni grafiche come "3/2", "12,1%", "011 6709718", "FBI", "report" etc., non esiste una fonte pubblicamente disponibile in grado di chiarire a un apprendente di italiano la pronuncia (o le pronunce) più tipiche di questi oggetti linguistici, nelle preferenze dei parlanti nativi (Fry 1989).

Ringraziamenti

Questo progetto ha beneficiato di un finanziamento della Fondazione CRT – Bando Erogazioni Ordinarie 2020 - CALL-UniTO. Si ringrazia la prof.ssa Donatella Bisconti e tutti coloro che hanno preso parte alla registrazione delle frasi e gli studenti che si sono sottoposti al task.

Bibliografia

- Bailey, D. (2019). Chatbots as conversational agents in the context of language learning. In *Proceedings of the Fourth Industrial Revolution and Education*, 32-41.
- Berruto, Gaetano. 2012. *Sociolinguistica dell'Italiano contemporaneo*. Roma: Carocci.
- Boersma, Paul & Weenink, David. 2018. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>.
- Busà, Maria Grazia. 2012. The role of prosody in pronunciation teaching: a growing appreciation. In Busà, Maria Grazia & Stella Antonio (eds.), *Methodological Perspectives on Second Language Prosody*, 101-105. Padova: Cleup.

- Canepari, Luciano. 1983. *Italiano standard e pronuncia regionale*. Padova: CLEUP.
- Canepari, Luciano. 1985. *L'intonazione. Linguistica e paralinguistica*. Napoli: Liguori.
- de Castro Moutinho, Lurdes & Coimbra, Rosa Lídia & Rilliard, Albert & Romano, Antonio. 2011. Mesure de la variation prosodique diatopique en portugais européen. *Estudios de fonética experimental* 20. 33-55.
- Cazade, Alain. 1999. De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues. *Apprentissage des Langues et Systèmes d'Information et de Communication, online* 2(2). 3-32.
- Cheng, Victor Chi-Wa & Lau, Vincent King-Tang & Lam, Ringo Wai-Kit & Zhan, Tian-Jie & Chan, Pak-Kin. 2020. Improving English phoneme pronunciation with automatic speech recognition using voice chatbot. In Lee, Lap-Kei & U, Leong Hou & Wang, Fu Lee & Cheung, Simon K. S. & Au, Oliver & Li Kam Cheong (eds.), *International Conference on Technology in Education*, 88-99. Singapore: Springer.
- Chun, Dorothy. 1998. Signal analysis software for teaching discourse intonation. *Language Learning & Technology* 2. 61-77.
- Chun, Dorothy. 2002. *Discourse Intonation in L2: From theory and research to practice*. Amsterdam: Benjamins.
- Colace, Francesco & De Santo, Massimo & Lombardi, Marco & Pascale, Francesco & Pietrosanto, Antonio & Lemma, Saverio. 2018. Chatbot for e-learning: A case of study. *International Journal of Mechanical Engineering and Robotics Research* 7(5). 528-533.
- Cresti, Emanuela. 1999. Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. *Faits de langue* 13. 168-181.
- Crocco, Claudia. 2017. Everyone has an accent: standard Italian and regional pronunciation. In Cerruti, Massimo & Crocco, Claudia & Marzo Stefania (eds.), *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*, 89-117. Berlin/New York: Mouton de Gruyter.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.
- De Bot, Kees. 1983. Visual feedback of intonation I: Effectiveness and Induced Practice Behavior. *Language and Speech* 26(4). 331-350.
- De Iacovo, Valentina & Romano, Antonio. 2019. DataDriven intonation teaching: an overview and new perspectives. *EL.LE*, 8(2). 393-408.
- De Marco, Anna & Sorianello, Patrizia & Mascherpa, Eugenia. 2014. L'acquisizione dei profili intonativi in apprendenti di italiano L2 attraverso un'unità di apprendimento in modalità blended learning. In De Meo, Anna & D'Agostino, Mari (a cura di), *Varietà dei contesti di apprendimento linguistico*, 189-210. Milano: AIELEA
- De Meo, Anna & Pettorino Massimo (eds.), 2012. *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*. Cambridge: Cambridge Scholar Publishing.
- Delattre, Pierre. 1966. Les dix intonations de base du français, *French Review* 40. 1-14.
- Delmonte, Rodolfo. 2009. Prosodic tools for language learning. *International Journal of Speech Technology* 12(4). 161-184.
- Fernoagă, Vlad & Stelea George Alex & Gavrilă, Cristinel & Sandu, Florin. 2018. Intelligent education assistant powered by chatbots. *The International Scientific Conference eLearning and Software for Education* 2. 376-383.

- Frost, Dan & Picavet, Francis. 2014. Putting prosody first – Some practical solutions to a perennial problem: The Innovalangues Project. *Research in Language* 12(3). 233-243.
- Fry, Edward B. 1989. Reading formulas: maligned but valid. *Journal of Reading*, 32(4). 292-297.
- James, E.F. 1976. The acquisition of prosodic features of speech using a speech visualizer. *International Review of Applied Linguistics in Language Teaching* 14(3). 227-244.
- Kisler, Thomas & Reichel, Uwe & Schiel, Florian. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326-347.
- Hirst, Daniel. 1983. Structures and categories in prosodic representations. In Cutler, Anne & Ladd, Robert (eds.), *Prosody: models & measurement*, 93-109. Berlin: Springer.
- Lacheret-Dujour, Anne. 2001. Modéliser l'intonation d'une langue. Où commence et où s'arrête l'autonomie du modèle? L'exemple du français parlé. *Colloque international: Journées Prosodie 2001*. 57-60.
- Levy, Michael. 1997. *CALL: context and conceptualisation*. Oxford: Oxford University Press.
- Lezhenin, Yuriy & Lamtev, Anton & Dyachkov, Vadim & Boitsova, Elena & Vylegzhanina, Karina & Bogach, Natalia. 2017. Study intonation: mobile environment for prosody teaching. *IEEE International Conference on Cybernetics (CYBCONF)* 3. 1-2.
- Martin, Philippe. 2010. Learning the prosodic structure of a foreign language with a pitch visualizer. *Speech Prosody 2010*, paper 980.
- Munro, Murray J. & Derwing, Tracey M. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49. 285-310.
- Nencioni, Giovanni. 1976. Parlato-parlato, parlato-scritto, parlato-recitato. *Strumenti Critici* X(1). 1-56.
- Pereira, Juanan. 2016. Leveraging chatbots to improve self-guided learning through conversational quizzes. In *Proceedings of the fourth international conference on technological ecosystems for enhancing multiculturalism*, TEEM '16, 911-918. ACM Press, New York.
- Romano, Antonio & Giordano, Gelsomina. 2017. Esperienze e riflessioni sulla didattica assistita dell'intonazione in italiano, inglese e francese. In Damascelli Adriana (ed.), *Digital Resources, Creativity, Innovative Methodologies and Plurilingualism: New Approaches to Language Teaching and Learning*, 176-200. Newcastle upon Tyne: Cambridge Scholar Publishing.
- Romano, Antonio & Contini, Michel & Lai, Jean Pierre. 2014. L'Atlas Multimédia Prosodique de l'Espace Roman: uno strumento per lo studio della variazione geoprosoica. In Tosques, Fabio (Hrsg.), *20 Jahre digitale Sprachgeographie*, 27-51. Berlin: Humboldt-Universität-Institut für Romanistik.
- Schiel, Florian. 1999. Automatic phonetic transcription of non-prompted speech. In *Proceedings of the ICPHS 1999*, 607-610.
- Sorianello, Patrizia. 2006. *Prosodia: modelli e ricerca empirica*. Roma: Carocci.
- Trouvain, Jürgen & Gut, Ulrike. 2007. *Non-native prosody. Phonetic description and teaching practice*. Berlin: Mouton de Gruyter.
- Voghera, Miriam. 1989. L'intonazione dell'italiano parlato spontaneo, *The Italianist* 9(1). 116-141.
- Zmarich, Claudio & Magno Caldognetto, Emanuela & Ferrero, Franco. 1996. Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza. In Cutugno Francesco (a cura di), *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VII Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica*, 111-139. Roma: Esagrafica.

Appendice

Elenco delle domande e risposte utilizzate:

- 1) Quale fra questi gruppi di opere è formato di sole opere italiane del Novecento?
 - A. L'Orlando furioso, L'Iliade e L'Odissea.
 - B. La divina commedia, Il fu Mattia Pascal e Lo Zibaldone.
 - C. I promessi sposi, La Vita Nuova e La Gerusalemme liberata.
 - D. Uno, nessuno e centomila, La coscienza di Zeno e Se questo è un uomo.
- 2) Chiara ha bisogno di sapere l'ora, come fa?
 - A. Scusi, ho lasciato l'orologio a casa, sa dirmi che ore sono?
 - B. Scusa, ho lasciato l'orologio a casa, per caso ha l'ora?
 - C. Sa dirmi che ora sono per favore che ho lasciato l'orologio a casa?
 - D. Mi scusi, ho lasciato l'orologio a casa, qual è il tempo oggi?
- 3) Marco e Salvatore si sono messi d'accordo per andare a mangiarsi una pizza e Marco deve dare appuntamento a Salvatore, cosa NON gli dirà?
 - A. Ciao Salvatore, allora ci troviamo alle 8 lì davanti?
 - B. Ciao Salvatore, allora ci becchiamo alle 8 davanti alla pizzeria?
 - C. Ciao Salvatore, allora ci incontriamo alle 8 lì?
 - D. Ciao Salvatore, allora ci andiamo a fare una partita a calcio uno di questi giorni?
- 4) Gianna vuole chiedere a Marta di comprare della frutta, cosa le dirà?
 - A. Senti Marta, una domanda: ma a te piace la cassata?
 - B. Senti Marta, stai uscendo? Se riesci a comprare della frutta, mi fai un favore.
 - C. Senti Marta, mi sa che non ci sono più pesche.
 - D. Senti Marta, che tu sappia, c'è ancora frutta in casa?
- 5) In quale anno l'Italia è diventata una Repubblica?
 - A. Nel giugno del 1950.
 - B. Durante l'autunno del 1939.
 - C. Tra il 1945 e il 1946.
 - D. Il 18 marzo 1861.
- 6) In quale giorno è scoppiata la Rivoluzione francese?
 - A. Nell'anno 1789.
 - B. Il 14 luglio 1789.
 - C. Nel marzo 1938.
 - D. Nel 1914.
- 7) L'espressione "avere le mani legate" vuol dire:
 - A. Non riuscire a prendere qualcosa che si trova in alto in uno scaffale.
 - B. Non avere nessuna possibilità di agire come si vorrebbe.
 - C. Avere tempi ristretti e non riuscire ad arrivare in orario a un appuntamento.
 - D. Fare di tutto senza essere all'altezza di una situazione.
- 8) In quale di queste frasi il periodo ipotetico è utilizzato correttamente?
 - A. Se vieni con me, ti portassi a mangiare una pizza.
 - B. Se venissi con me, ti avrei portato a mangiare una pizza.
 - C. Se verresti con me, ti porto a mangiare una pizza.
 - D. Se fossi venuto con me, ti avrei portato a mangiare una pizza.

- 9) Dov'è morto Giuseppe Garibaldi?
- A. Giuseppe Garibaldi è morto nell'isola di Caprera.
 - B. Nel gennaio del 1882.
 - C. Garibaldi è morto all'isola d'Elba.
 - D. Quando aveva 75 anni.
- 10) $1 + \frac{2}{3}$ equivale a:
- A. $\frac{4}{4} + \frac{4}{5}$
 - B. $\frac{1}{6} + \frac{3}{2}$
 - C. $\frac{1}{4} + \frac{4}{9}$
 - D. $\frac{1}{3} - 0,75$

L'accuratezza della trascrizione ASR sul parlato non-standard. L'italiano nell'OH Portal

Abstract

This paper aims at evaluating the transcription accuracy of the Google-IT's ASR service available in the OH Portal. Data on its performance are limited to good quality recordings. Thus, we focus on suboptimal authentic materials, encompassing non-standard conversational speech recorded in noisy environments. We carry out a quanti-qualitative analysis of the linguistics and extra-linguistic parameters affecting accuracy and error distribution. The preliminary results show higher Word Error Rates for non-standard speech and low-quality recordings. Moreover, we seek the error patterns that could ease the transcription correction process for the users.

1. *Introduzione*

In questo contributo² presentiamo i risultati preliminari di un lavoro in corso incentrato sulle trascrizioni automatiche di materiali di parlato, uno degli strumenti chiave delle *digital humanities*. Il processo di trascrizione manuale dei materiali di parlato costituisce un passaggio laborioso ma imprescindibile nell'elaborazione delle registrazioni, sebbene spesso precluda l'analisi di grandi quantità di dati. Una nuova risorsa che risponde all'esigenza di ottimizzare la catena di trascrizione è rappresentata dall'OH Portal. Si tratta di un'interfaccia *web* integrata nell'infrastruttura europea CLARIN (clarin.eu), che raccoglie una serie di strumenti per la gestione semi-automatica dei materiali di parlato (Draxler et al. 2020; Scagliola et al. 2020; van den Heuvel 2020). Al suo interno sono disponibili vari servizi, che includono strumenti di terze parti per la trascrizione tramite il riconoscimento automatico del parlato (*Automatic Speech Recognition*, ASR). Tra questi, per il nostro studio, abbiamo usato l'ASR di Google, già riconosciuto come uno degli ASR commerciali più accurati (Ashwell & Elam 2017; Biadys et al. 2012; Filippidou & Moussiades 2020; Tavosanis 2018). Le valutazioni dell'accuratezza degli ASR implementati nell'OH Portal, incluso quello di Google, sono prevalentemente basate su materiali monologici, di varietà linguistiche tendenti al polo dello standard,

¹ Università degli Studi Roma Tre-Sapienza Università di Roma.

² Sebbene il presente contributo sia frutto di un costante lavoro comune, la stesura dei paragrafi 1, 2, 4.1, 5.2.1, 5.2.4 e 6 è da attribuirsi a Giovina Angela del Rosso, mentre i paragrafi 3, 4.2, 5.1, 5.2.2, 5.2.3 sono da attribuirsi a Silvia Brambilla.

registrati in buone condizioni acustiche. Spesso, però, i materiali di parlato spontaneo comprendono interviste e dialoghi, attestano varietà non-standard e non sono registrati in ambienti ideali. Notoriamente, le esitazioni, false parenze, parole interrotte e sovrapposizioni di turno, fenomeni largamente presenti nel parlato spontaneo e dialogico, rendono il riconoscimento automatico complesso e possono contribuire a far diminuire l'accuratezza della trascrizione (Badino 2016; Çetin & Shriberg 2006; Kitaoka et al. 2014; Tavasani 2018). Pertanto, basandoci su materiali precedentemente raccolti per altre finalità e che rappresentano tali caratteristiche (§ 2), il nostro obiettivo è indagare a livello quanti-qualitativo l'accuratezza della trascrizione automatica e quali fattori la influenzino, adottando una prospettiva basata sull'utente e orientata all'utente.

Un limite intrinseco di questa analisi, tuttavia, deriva dall'oggetto stesso della ricerca. L'*output* che osserviamo è il prodotto di un sistema di ASR commerciale, il che comporta che non sono disponibili informazioni dettagliate sul funzionamento e sui materiali di addestramento dell'ASR di Google (Draxler et al. 2020). Ciononostante, tale limite può essere arginato in virtù di due considerazioni. In primo luogo, possiamo avanzare ipotesi inferenziali, fondate sul funzionamento dei riconoscitori automatici di ultima generazione (Abulimiti & Schultz 2020; Kěpuska & Bohouta 2017; Kitaoka et al. 2014). In secondo luogo, in un'ottica incentrata sugli utenti, l'esame linguistico dell'*output* è di maggiore utilità rispetto all'analisi delle motivazioni che lo hanno prodotto. L'analisi linguistica, infatti, costituisce una fase euristica necessaria per lo sviluppo di strategie di correzione manuale *ad hoc*, funzionali all'ottimizzazione della catena di trascrizione (semi-)automatica.

L'articolo è organizzato come segue: in § 2 descriviamo il dataset; in § 3 ripercorriamo la metodologia adottata; in § 4 discutiamo i principali risultati finora ottenuti, che riguardano l'accuratezza della trascrizione (§ 4.1) e la distribuzione degli errori (§ 4.2); in § 5 illustriamo e discutiamo tramite esempi casi di errore rilevanti per gli utenti; infine, in § 6 presentiamo le conclusioni preliminari del lavoro e i possibili sviluppi futuri.

2. *Materiali*

Il dataset analizzato è costituito dai corpora descritti nella tab. 1. I corpora BA e L2, precedentemente raccolti per altri scopi, rappresentano varietà di italiano non-standard, rispettivamente di italiano regionale di Bari nativo e L2. Tali corpora sono stati confrontati tra di loro e con i subcorpora LP e LB del CLIPS (Albano Leoni et al. 2007). Questi ultimi fungono da controllo, in quanto rappresentativi dell'italiano standard.

Tabella 1 - *Corpora analizzati*

<i>Corpus L2</i>	<i>Corpus BA</i>	<i>Corpus CLIPS</i>	
9 registrazioni	10 registrazioni	10 registrazioni	10 registrazioni
6 parlanti avanzati L2 (2 F e 4 M) 3 parlanti nativi (2 F e 1 M)	10 parlanti nativi (5 F e 5 M)	10 parlanti professionisti (5 F e 5 M)	
Dialogo	Dialogo	Monologo	
Parlato spontaneo	Parlato semi-spontaneo (dialoghi semi-guidati)	Parlato letto ortofonico (subcorpus LP)	Parlato letto ortofonico (subcorpus LB)
Italiano regionale di Bari	Italiano regionale di Bari	Italiano standard	
Ambienti chiusi ma rumorosi	Ambienti chiusi e silenziosi, ma non isolati	Camera anecoica (Falcone et al. 2007b)	
Computer portatile con microfono esterno, verso gli intervistati	Registratore Tascam DR-40 con microfoni incorporati, verso gli intervistati	Apparecchiatura professionale (Falcone et al. 2007a)	
44.1 kHz, 16-bit, .wav	44.1 kHz, 24-bit, .wav	22.1 kHz, .wav	

3. Metodo

Abbiamo trascritto manualmente le registrazioni dei corpora BA e L2 che, insieme alle trascrizioni del corpus CLIPS, costituiscono i nostri testi di riferimento (*reference texts*, REF). Tutte le registrazioni sono poi state elaborate nell'OH Portal con il sistema ASR di Google-IT, i cui risultati costituiscono i nostri testi di ipotesi (*hypothesis texts*, HYP). I testi REF e HYP sono stati allineati usando un algoritmo³ di programmazione dinamica (Sakoe & Chiba 1978) che consente di calcolare la distanza di Levenshtein tra stringhe (Levenshtein 1966). Le coppie di parole sono state così classificate come corrispondenze (OK) o errori, distinti nei tre tipi consolidati in letteratura (Beché & Favre 2013; Levis & Suvorov 2012; Palmerini & Savy 2014), cioè cancellazioni (DEL), sostituzioni (SUB) o inserzioni (INS). Questa classificazione è funzionale al calcolo del tasso di errore, per il quale la metrica adottata è stata il WER (*Word Error Rate*):

$$WER = ERR/N$$

in cui ERR è la somma del numero degli errori (INS + SUB + DEL), che viene rapporta a N, cioè il numero totale delle coppie di parole considerate (OK + INS + SUB + DEL). Data la sua diffusione in letteratura, abbiamo scelto il WER come in-

³ Codice Python *case-insensitive* adattato da <https://holianh.github.io/portfolio/Cach-tinh-WER/>.

dice di accuratezza per una maggiore comparabilità dei dati, nonostante non manchino proposte di metriche alternative (Filippidou & Moussiades 2020). Le uscite di tale allineamento saranno d'ora in poi chiamate REF-0/HYP-0. In totale, tale processo ha individuato 57248 coppie di parole, automaticamente annotate anche per tipo di errore.

Successivamente, abbiamo riallineato manualmente tali coppie. Ciò ha prodotto un secondo allineamento, d'ora in avanti chiamato REF-1/HYP-1. Abbiamo conseguentemente corretto l'annotazione dei tipi di errore. Nella tab. 2 riportiamo un esempio dei diversi allineamenti, con le relative classificazioni degli errori. Si può notare che in REF-0/HYP-0 mancano le corrispondenze tra *barese-varese* e *guidato-di lato*; quest'ultima è anche un caso di *sostituzione multiparola* (§ 5.2.2). A seguito di questo tipo di correzioni è stato riconteggiato un totale di 57389 coppie.

Tabella 2 - *Confronto tra allineamento automatico e manuale*

automatico			manuale		
ERR-0	REF-0	HYP-0	ERR-1	REF-1	HYP-1
DEL	sono	***	DEL	sono	***
OK	una	una	OK	una	una
DEL	barese	***	SUB	barese	varese
DEL	acquisita	***	SUB	acquisita	visita
DEL	tuttavia	***	DEL	tuttavia	***
SUB	non	varese	DEL	non	***
SUB	ho	visita	DEL	ho	***
SUB	mai	di	DEL	mai	***
SUB	guidato	lato	SUB	guidato	di
SUB	qui	si	SUB	SUB	lato
			SUB	qui	si

Per ragioni di chiarezza, aggiungiamo un asterisco (*) all'etichetta del dataset, quando ci riferiamo alle prime 1100 coppie per registrazione di REF-0/HYP-0 o di REF-1/HYP-1 (o alla totalità delle coppie disponibili, se meno di 1100). Le 35040 coppie totali di REF-1*/HYP-1* sono state annotate secondo una serie parametri linguistici ed extralinguistici, il cui effetto sull'accuratezza dei sistemi ASR è stato precedentemente evidenziato in letteratura (Badino 2016; Çetin & Shriberg 2006; Palmerini & Savy 2014; Tivosanis 2018)⁴:

- a. parametri linguistici: parte del discorso (PoS, *Part of Speech*) di REF-1* e HYP-1*, normalizzazione o errore ortografico, complessità fonologica, lingua di REF-1* e HYP-1*, ripetizioni di *chunk* in REF-1*, segnale discorsivo;

⁴ Per l'annotazione delle parti del discorso abbiamo adottato una versione semplificata del *tagset* ISST-TANL (<http://www.italianlp.it/docs/ISST-TANL-POSTagset.pdf>); per gli eventi acustici abbiamo seguito le linee guida del CLIPS (Savy 2007); per la distinzione tra segnali discorsivi e disfluenze, cfr. Crible (2016).

- b. parametri extralinguistici: turno, numero di partecipanti, distanza del parlante dal microfono, ambiente, evento acustico.

In un'ottica orientata agli utenti, in questa prima fase della ricerca, tuttora in corso, abbiamo lavorato principalmente sull'interpretazione di PoS e turno. Ad esempio, come mostreremo in § 5, sembra che alcune classi di parole siano maggiormente soggette a errore rispetto ad altre.

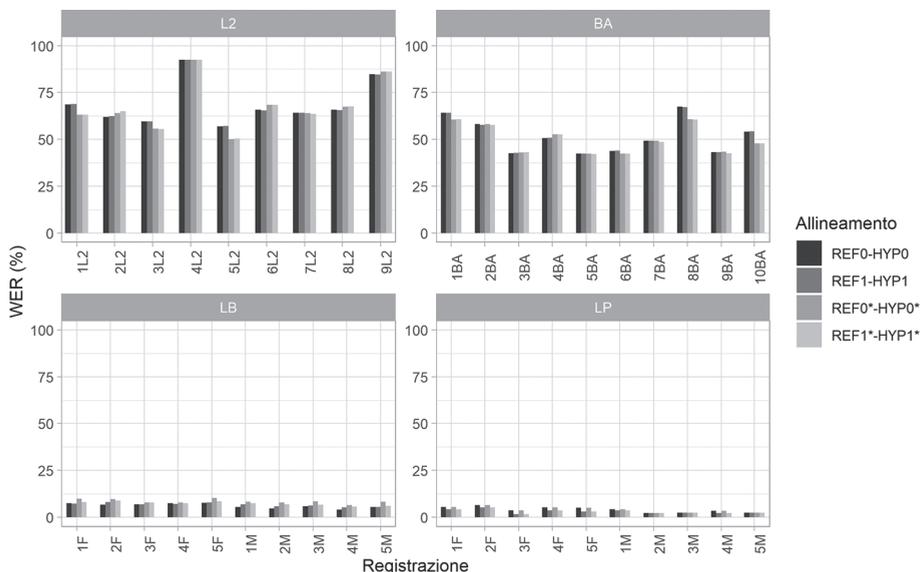
Infine, abbiamo calcolato il WER per registrazione di REF-0/HYP-0, REF-1/HYP-1, REF-0*/HYP-0*, REF-1*/HYP-1*. Per REF-1*/HYP-1* abbiamo anche calcolato il WER rispetto al turno di parola. Nei prossimi paragrafi, mostriamo i risultati dell'analisi quanti-qualitativa condotta col supporto dell'ambiente statistico R (R Core Team 2021).

4. Risultati

In questa sezione mostriamo una panoramica complessiva dei risultati preliminari di questo lavoro, che riguardano l'accuratezza delle trascrizioni automatiche (§ 4.1) e l'analisi della tipologia degli errori (§ 4.2). Considerata la natura del dataset, abbiamo utilizzato le singole registrazioni come unità di comparazione, senza aggregare i dati per corpus, anche se impieghiamo le etichette dei corpora per richiamarne tutte le registrazioni.

4.1 Il tasso di errore

Figura 1- *Comparazione dei WER per registrazione, per corpus e per metodo di calcolo*



Dall'analisi finora condotta, come mostrato nella fig. 1, a livello quantitativo si riscontra uniformità tra i tassi di errore risultanti dalla reiterazione del calcolo del

WER di REF-0/HYP-0, REF-1/HYP-1, REF-0*/HYP-0* e REF-1*/HYP-1*. Si veda ad esempio la registrazione 1L2, in cui i WER relativi al dataset annotato sono coincidenti (63.2%), discostandosi di pochi punti percentuali ($\Delta=5.7\%$) dai WER relativi alle trascrizioni integrali, a loro volta molto simili ($\Delta=0.2\%$).

Dal grafico si rileva, inoltre, sostanziale variazione dei WER sia inter-corpora che intra-corpora. Globalmente, il tasso d'errore delle trascrizioni automatiche è maggiore nei corpora di italiano non-standard rispetto ai corpora di italiano standard: L2 (50.0-92.5%) > BA (42.2-67.4%) > LB (4.0-10.3%) > LP (1.7-6.5%).

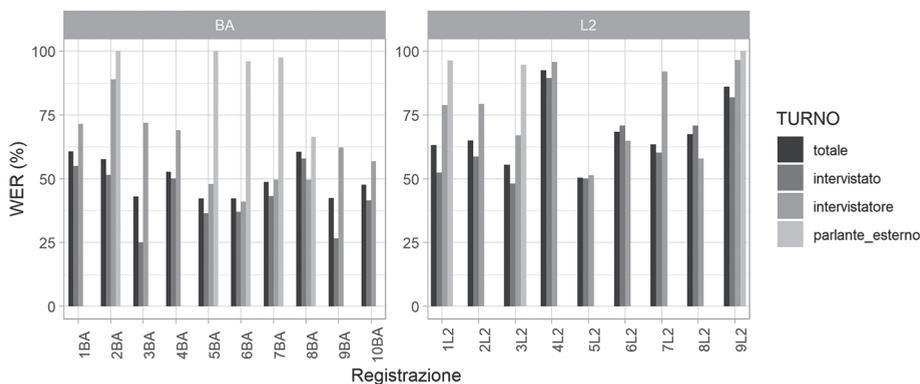
Alla limitata variazione dei WER, però, si contrappongono le differenti distribuzioni degli errori rilevati automaticamente e manualmente (§ 4.2). Considerando ciò, l'analisi delle interazioni tra errori e parametri linguistici ed extralinguistici sarà basata su REF-1*/HYP-1*.

Tra i parametri considerati, il tipo di corpus e il turno di parola influiscono significativamente⁵ sul tasso di errore (fig. 2)⁶. Nei nostri dati, tali parametri sono strettamente collegati rispettivamente alla qualità del segnale e alla distanza del parlante dal microfono (cfr. tab. 1), corroborando i risultati già noti delle ricerche sull'argomento (cfr. Li et al. 2016). Inoltre, ciascun partecipante contribuisce alla definizione del tasso di errore globale in misura disuguale, proporzionale alla lunghezza del proprio turno. Da ciò segue che il WER di REF-1*/HYP-1* coincide con la media ponderata dei WER dell'intervistato, dell'intervistatore e degli eventuali parlanti esterni. Nel nostro dataset, il WER dei parlanti esterni influisce limitatamente sul WER totale, sebbene possa raggiungere anche il 100% (come in 2BA, 5BA e 9L2), perché le coppie attribuite ai loro turni sono poche (ad esempio, in 2BA sono solo 6 su 1100). Tuttavia, tassi di errore così elevati sono motivati dalle condizioni di registrazione. Nello specifico, durante le interviste dei corpora BA e L2, gli eventuali parlanti esterni si trovavano in posizioni periferiche rispetto a intervistato e intervistatore e conseguentemente erano molto lontani dal microfono, direzionato sempre verso l'intervistato. A ciò si aggiunge la tendenza dei parlanti esterni a intervenire a bassa voce e a generare non di rado sovrapposizioni di turno o eventi acustici di varia natura.

⁵ Dall'analisi preliminare dell'interazione e della gerarchia tra parametri, l'albero di inferenza condizionale (*conditional inference tree*) conferma che anche i nodi corrispondenti al tipo di corpus e al turno di parola esercitano un effetto statisticamente significativo sull'accuratezza della trascrizione. Tuttavia, sarà necessario approfondire tali complesse relazioni in uno studio dedicato, che esula parzialmente dagli scopi di questo lavoro, la cui natura è principalmente linguistica.

⁶ Questa considerazione è stata recentemente integrata come *disclaimer* nell'OH Portal: "Known Issues for Google ASR: if the recording is longer than a few minutes we observed that whole stretches of words are simply omitted in the result; this happens frequently at speaker turns, i.e. when two or more speakers take turns, then often the begin of the turn of the new speaker is compromised." (<https://clarin.phonetik.uni-muenchen.de/apps/TranscriptionPortal/>, accesso effettuato in data 28/09/2021).

Figura 2 - Comparazione dei WER per registrazione, per corpus e per turno di parola



4.2 Distribuzione degli errori

Anche se l'algoritmo ha permesso la classificazione automatica degli errori, la successiva verifica dei risultati ha evidenziato la necessità di correggerla manualmente. Le classificazioni automatiche e manuali divergono qualitativamente sia nell'allineamento sia per la conseguente rietichettatura degli errori (cfr. tab. 2, § 3). Nella tab. 3 sono riportati a titolo esemplificativo i WER minimi e massimi divisi per corpus e per allineamento. Nel caso del corpus LP, l'annotazione è stata esaustiva, quindi REF-0/HYP-0 e REF-1/HYP-1 coincidono rispettivamente con REF-0*/HYP-0* e REF-1*/HYP-1*.

Tabella 3 - WER minimi e massimi per corpus e allineamento

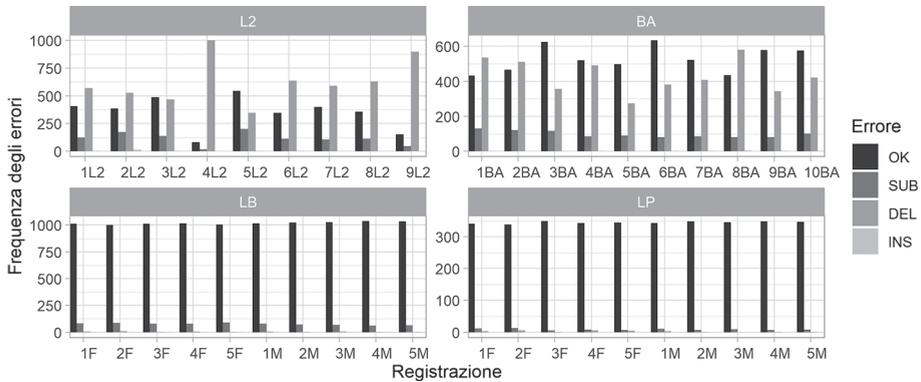
Corpus	LP		LB		BA		L2	
Registrazione	2F	3F	4M	5F	5BA	8BA	5L2	4L2
REF-0/HYP-0	3.7	6.5	4.0	7.6	42.3	67.4	56.8	92.4
REF-1/HYP-1	1.7	5.3	5.2	7.9	42.3	67.2	57.0	92.4
REF-0*/HYP-0*	-	-	6.4	10.3	42.3	60.7	50.0	92.5
REF-1*/HYP-1*	-	-	5.7	8.5	42.2	60.5	50.4	92.5

I risultati dell'analisi sono rappresentati nella fig. 3⁷: la distribuzione degli errori nei corpora non è uniforme né da un punto di vista quantitativo né qualitativo. Dai diagrammi a barre, risulta evidente che nei subcorpora del CLIPS prevalgono le corrispondenze (OK) rispetto alle sostituzioni (SUB) e alle cancellazioni (DEL). Diversamente, nei corpora BA e L2 (di italiano non-standard) le cancellazioni incidono maggiormente sul tasso di errore, come rilevabile dal grafico, a differenza delle inserzioni (INS), quasi del tutto assenti. Tuttavia, BA e L2 differiscono per la distribuzione interna di cancellazioni e sostituzioni: al peggiorare delle condizioni

⁷ Per ragioni grafiche, la scala dell'asse y non è uniforme.

acustiche, aumenta il numero di cancellazioni, mentre diminuisce il numero di sostituzioni.

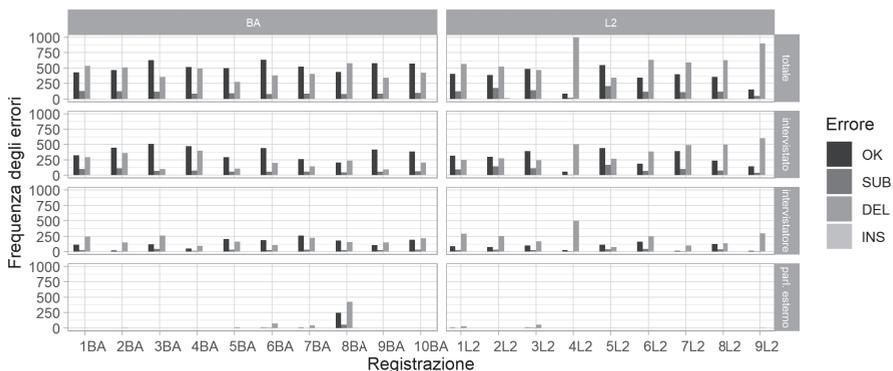
Figura 3 - Distribuzione degli errori per registrazione e per corpus



Una tendenza analoga è riscontrabile anche all'aumentare della distanza del parlante dal microfono, fattore rappresentato nel nostro dataset dal parametro "turno di parola". Dato che tale effetto è osservabile solo su materiali non monologici, l'analisi è pertinente solo per BA e L2 (fig. 4). Nei diagrammi a barre riferiti a "parlante esterno" e a "intervistatore", il numero di cancellazioni è maggiore che per l'"intervistato", per la maggiore vicinanza di quest'ultimo al microfono. Quindi, il deteriorarsi delle condizioni acustiche comporta una ridotta variazione dei tipi di errore, come osservabile anche nella fig. 4.

Di conseguenza, gli utenti dell'OH Portal che optino per l'ASR di Google-IT dovranno aspettarsi di ripristinare manualmente nella trascrizione soprattutto gli enunciati prodotti dai parlanti più distanti dal microfono, perché tendenzialmente l'ASR non ne trascrive correttamente o addirittura ne cancella del tutto i turni.

Figura 4 - Distribuzione degli errori per registrazione e per turno di parola in BA e L2



5. *Gli errori per gli utenti*

In questa sezione consideriamo l'interazione tra gli errori e i parametri linguistici, anche in funzione del dominio. In particolare, ci soffermiamo sulle inserzioni (§ 5.1) e sulle sostituzioni (§ 5.2) di maggior rilievo per gli utenti per l'uso delle trascrizioni automatiche. Infatti, se nei casi di cancellazione è sufficiente che gli utenti aggiungano le parti omesse dall'ASR, l'individuazione e la successiva correzione delle inserzioni e delle sostituzioni risulta più complessa.

5.1 Inserzioni

In letteratura le inserzioni sono solitamente considerate come errori dovuti alla reinterpretazione linguistica di rumore spurio (Baber & Hone 1993). Nel nostro dataset, le inserzioni sono gli errori più rari. A livello quantitativo le inserzioni individuate tramite l'algoritmo di allineamento sono 153 su 35081 coppie totali di REF-0*/HYP-0* (0.4%). Dopo il riallineamento manuale, il numero delle inserzioni è stato ridotto a 65, con una riduzione delle inserzioni allo 0.2% del totale. Infatti, abbiamo ricondotto la maggior parte delle inserzioni rilevate automaticamente a sostituzioni multiparola (§ 5.2.2).

Le parole inserite sono frequenti nel lessico italiano, perlopiù foneticamente deboli e brevi, quindi spesso appartenenti a classi funzionali più che lessicali (Bybee 2001; Maturi 2014). Questa indicazione di massima può guidare l'utente durante la correzione e coadiuvare l'individuazione di elementi estranei alla traccia audio ma presenti nella trascrizione automatica.

È da notare che, diversamente da quanto atteso, le inserzioni sembrano essere attribuite più all'influsso della probabilità linguistica rispetto a quello della probabilità acustica. Infatti, le inserzioni ricorrono anche in assenza di rumore ambientale e, in generale, di eventi acustici cui potrebbero essere attribuite: sono presenti anche nel subcorpus di controllo LB, benché questo sia stato registrato in camera anecoica.

Dal punto di vista linguistico, possiamo ricondurre gli *output* testuali interessati da inserzioni a tre macro-fenomeni, distinti in base al grado di standardizzazione dei nostri materiali e di grammaticalità dei testi HYP. In particolare, i fenomeni individuati sono:

1. standardizzazione di forme di varietà regionali e di apprendimento: in (1)⁸ nella HYP viene inserito il determinante *i* assente nella REF perché omesso dal parlante non nativo. Per l'utente, l'inserzione di *i* nella sequenza comporta una standardizzazione non fedele alla varietà testimoniata dal materiale originale;

(1)	deve	lavorare	per	darle	INS	soldi
	devi	lavorare	per	darle	i	soldi

⁸ Negli esempi, dove non specificato diversamente, la prima riga è il testo REF, la seconda è il testo HYP.

2. presenza nell'HYP di strutture più frequenti di quelle della REF, benché entrambe le varianti siano grammaticali: in (2) il determinante *la* precede il sintagma *settimana scorsa*, perché il costrutto *la settimana scorsa* è più probabile rispetto alla variante priva di articolo;

(2) m' ha detto una volta INS settimana scorsa poi
 ma che tu una volta la settimana scorsa poi

3. creazione di *chunk* localmente grammaticali ma contestualmente agrammaticali: in (3) la presenza del verbo *andare* innesca l'inserzione della preposizione *a*, che segue frequentemente questo lemma. Tuttavia, se si considera un contesto più ampio del bigramma, la sequenza **andare a prima al mercato* non è grammaticale.

(3) per fare la spesa andiamo INS prima al mercato
 DEL DEL DEL DEL andiamo a prima al mercato

5.2 Sostituzioni

5.2.1 Sostituzioni per fattori linguistici e di dominio

L'accuratezza degli ASR è sensibile non solo alle condizioni acustico-ambientali ma anche a fattori linguistici e di dominio (Besacier et al. 2014; Draxler et al. 2020). Per gli utenti, l'(inter-)azione di tali fattori è più evidente nelle sostituzioni.

Abbiamo riscontrato che, nei nostri dati, alcune parole possono essere trascritte con grafie errate. Queste sostituzioni sembrano essere dovute alle procedure di post-elaborazione ortografica applicate all'uscita dell'ASR. Tali errori influiscono largamente sull'accuratezza, perché si riscontrano in parole ad alta e altissima frequenza, come articoli e preposizioni.

In questa casistica rientrano i numerali, riconosciuti correttamente nel 61.5% delle occorrenze e trascritti con alterazioni ortografiche nel 15.6% dei casi. Tra gli errori ortografici che coinvolgono i numerali, sono sistematici quelli di numeri contenenti *dieci* (4a), mentre sono occasionalmente errati altri cardinali (4b) e i numeri pronunciati in sequenza (4c)⁹:

(4a) dieci, diecimila	>	dici, dici-comma-zero-zero-zero
(4b) tre (*)	>	iii
(4c) il sette il settantuno (*)	>	il settecentosettantuno

Un esempio analogo riguarda l'apostrofo, che viene sempre trascritto nel verbo *c'è*, ma sistematicamente omissivo in *l', un' e po'*. Più complesso è il caso delle preposizioni articolate, in cui alla corretta trascrizione si alternano casi di omissione. Sebbene le grafie errate siano frequenti, gli errori di questo tipo sono facilmente emendabili.

Altrettanto articolato è il caso dell'onomastica. Da un lato, abbiamo individuato sostituzioni *di* nomi propri (5a-b), presumibilmente candidati poco probabili per

⁹ Il simbolo (*) affianca i *token* che non sono assenti nel vocabolario (*out of vocabulary*), perché trascritti correttamente altrove.

L'output: ad es., l'antroponimo straniero *Ahmad* viene trascritto come *Amazon* (5a). Dall'altro, segnaliamo le sostituzioni *con* nomi propri (5c-d), anche poco diffusi nelle conoscenze enciclopediche degli italofofoni ma ben indicizzati su Google Search: ad es., *inverno* viene sostituito dal cognome del giornalista Luciano *Onder* (5c).

- (5a) Ahmad > Amazon
 (5b) Jamil (*) > già mi
 (5c) inverno (*) > Onder
 (5d) m'ha detto (*) > Maletto

L'ipotesi del collegamento tra l'ASR e gli altri prodotti Google spiegherebbe anche la maggiore accuratezza di trascrizione dei toponimi rispetto a quella degli antroponimi (tab. 4). Tale differenza si apprezza maggiormente in BA, in cui i toponimi sono trascritti correttamente in circa 1 caso su 2, mentre gli antroponimi solo in 1 su 5.

Tabella 4 - Distribuzione degli errori nell'onomastica in BA e L2

	BA				L2			
	OK	SUB	DEL	TOT.	OK	SUB	DEL	TOT.
toponimi	50.2%	8.2%	41.6%	100%	31.1%	8.6%	60.3%	100%
antroponimi	19.0%	14.3%	66.7%	100%	26.7%	13.3%	60.0%	100%

Anche le forme non-standard nel dataset vengono sostituite e ricondotte a forme standard, cui sembrano essere attratte per similarità fonetica (6a-c); ciò comporta, talvolta, un casuale mantenimento del significato (6b).

- (6a) er a mett > ero metà
 (6b) pad > papà
 (6c) cè > cioè

La similarità fonetica spiegherebbe anche le saltuarie sostituzioni di *token*, in presenza di realizzazioni ipoarticolate (Lindblom 1990; per l'italiano, cfr. Albano Leoni & Maturi 1995). La confusione tra suoni può avvenire all'interno di parola (7a) o limitarsi alle vocali atone finali (7b).

- (7a) parco (*) > pacco
 (7b) famosa (*) > famoso

Infine, anche le disfluenze sono oggetto di errori linguistici (80.5% DEL, 8.1% SUB). Le sostituzioni riguardano false partenze e parole troncate, sempre trascritte con parole intere e foneticamente plausibili, che possono ripristinare (8a) o alterare (8b) il senso originale.

- (8a) quartie+[quartiere] > quartiere
 (8b) picco+ [piccoli] > picconi

Anche l'individuazione di questi tipi di *pattern* di errore potrebbe supportare gli utenti nello sviluppo delle strategie di correzione.

5.2.2 Le sostituzioni multiparola

L' algoritmo di allineamento considera possibili solo corrispondenze uno-a-uno, non uno-a-molti, in cui l' unità è la parola ortografica delimitata da spazi bianchi. Tuttavia, durante la correzione manuale, considerando il cotesto, abbiamo individuato sostituzioni che coinvolgono più parole, dette *sostituzioni multiparola*. Complessivamente, costituiscono il 9.8% delle sostituzioni.

Le sostituzioni multiparola ricorrono in due situazioni: nella prima, abbiamo allineato una o più parole della REF a due o più parole dell' HYP, come in (9).

(9) Quartierino > quartiere Reno

Nella seconda, in modo speculare, abbiamo allineato due o più parole della REF a una o più parole dell' HYP, come in (10).

(10) ti dovresti > tirolese

In questi casi, alcuni errori, che nell' allineamento automatico erano rispettivamente inserzioni e cancellazioni, sono stati riannotati come sostituzioni. Nel calcolo del WER il numero di errori legato a questa rianalisi è rimasto invariato, dato che un tipo di errore è stato corretto in un tipo diverso: in (9), abbiamo contato due sostituzioni invece che una sostituzione e un' inserzione, mentre in (10) abbiamo contato due sostituzioni invece che una sostituzione e una cancellazione. La soluzione da noi adottata rispetta la tassonomia classica degli errori degli ASR e il suo vantaggio risiede in una migliore comparabilità dei risultati. Tuttavia, restano da indagare gli eventuali benefici di metriche alternative (Besacier et al. 2014; Filippidou & Moussiades 2020), che attribuiscono un peso diverso alle sostituzioni multiparola, il cui *status* sembrerebbe non essere ancora stato considerato esplicitamente in letteratura.

Infine, l' incidenza delle sostituzioni multiparola sugli errori totali si apprezza più nei subcorpora di controllo (12.6% in LP, 8.0% in LB) che nei corpora di italiano non-standard (1.7% in BA, 1.6% in L2). Questo è presumibilmente legato alla maggiore variabilità dei tipi di errore al migliorare delle condizioni acustiche (§ 4.2). L' analisi di questi aspetti e della loro interazione con la similarità fonetica necessita di ulteriori approfondimenti futuri.

5.2.3 Le parti del discorso

Nonostante il ruolo dei modelli acustici nei moderni ASR sia preponderante rispetto a quello dei modelli linguistici, per gli utenti può essere un utile strumento di correzione individuare quali parti del discorso siano maggiormente soggette a sostituzione.

Tabella 5 - Matrice di confusione delle sostituzioni rispetto alla parte del discorso (PoS)

SUB		PoS_HYP																
		V	com_N	art_det	agg	prep_art	art_ind	prep	cong	avv	aus	pro_cli	mod	N_pro	num	pro	V_cli	int
PoS_REF	V	385	40	2	18	2	1	16	38	24	2	4	3	18	2	12	3	5
	com_N	47	246	1	26	1	2	15	7	15	2	7	3	38	6	12	1	0
	art_det	6	0	175	7	23	2	7	8	5	1	4	0	0	0	5	0	0
	agg	12	32	3	106	1	1	5	4	10	3	0	0	10	0	13	1	0
	prep_art	0	6	25	0	68	0	11	2	0	1	0	0	0	0	1	0	0
	art_ind	0	0	8	0	0	54	3	1	2	0	0	0	0	1	1	0	0
	prep	9	1	11	3	16	0	49	5	7	2	3	0	1	1	4	0	0
	cong	31	7	0	2	6	0	10	48	15	8	10	5	2	0	3	0	1
	avv	18	13	3	7	4	1	12	23	41	2	14	2	5	0	10	0	0
	aus	4	0	1	0	0	0	0	4	6	41	4	1	0	0	1	0	1
	pro_cli	4	1	2	1	0	0	9	5	4	1	34	0	2	0	2	0	0
	mod	6	6	0	1	0	0	1	3	4	0	0	26	0	0	0	0	0
	N_pro	7	20	1	2	2	0	1	5	9	0	0	0	22	0	3	1	0
	num	6	1	0	3	0	0	0	1	1	0	0	0	1	20	1	0	0
	pro	7	1	0	6	0	0	4	9	6	1	8	1	2	0	19	0	0
	V_cli	10	1	0	0	0	0	0	0	1	0	1	0	0	0	0	4	0
	int	3	3	0	0	0	0	3	36	13	0	3	1	0	0	2	0	1

La matrice di confusione dei casi di sostituzione divisi per parte del discorso (tab. 5) permette di verificarne la consistenza numerica. In particolare, le previsioni (la trascrizione manuale, PoS_REF) sono riportate nelle righe, mentre lo stato effettivo (la trascrizione automatica, PoS_HYP) è riportato nelle colonne. Nella diagonale sono evidenziati i casi in cui, malgrado la sostituzione, è mantenuta la classe di parola originaria. Ad esempio, in 246 casi i nomi comuni della REF sono sostituiti da altri nomi comuni nell'HYP (veri positivi), come esemplificato in (11):

- (11a) soccorritori > seguitori
 (11b) farmacia (*) > suoneria

Diversamente, 26 nomi comuni della REF sono sostituiti da aggettivi (falsi negativi), come in (12):

- (12a) entrata (*) > dentale
 (12b) pancia (*) > panica

Viceversa, nell'HYP 32 aggettivi sono sostituiti da nomi comuni (falsi positivi), come in (13).

- (13a) equo > eco
 (13b) pulito > prurito

Alcune di queste sostituzioni sono intuibili, come quelle frequenti tra la congiunzione *e* e la forma verbale *è*, mentre altri errori di sostituzione esibiscono *pattern* meno regolari. In termini assoluti, la distribuzione più eterogenea riguarda verbi e nomi, che sono anche le parti del discorso più frequenti nel nostro dataset. Un'analisi divisa per classe, tuttavia, restituisce un quadro parzialmente diverso circa l'accuratezza di trascrizione rispetto alla parte del discorso. Una sintesi è fornita nella tab. 6, in

cui la confusione per classe cresce da sinistra verso destra. Per ciascuna PoS, sono riportati i valori di richiamo (*recall*), precisione (*precision*) e F1-score. Nello specifico, si nota un maggiore mantenimento di classe per gli articoli ($F1_{\text{artind}}=82\%$, $F1_{\text{artdet}}=73\%$), nonostante siano foneticamente deboli, rispetto a classi di parole più lunghe e piene, come aggettivi e avverbi ($F1_{\text{agg}}=55\%$, $F1_{\text{avv}}=25\%$). A livello complessivo, la classe di parola è mantenuta in circa metà delle sostituzioni, come indicato sia dal valore di macro-F1 (47%), in cui ciascuna classe ha uguale peso, sia dal valore di micro-F1 (54%), più influenzato dal numero di esemplari per classe.

Tabella 6 - Accuratezza percentuale di trascrizione per parte del discorso

metrica		PoS																
		art_ ind	art_ det	V	aus	num	N_ com	mod	prep_ art	agg	pro_ cli	prep	V_ cli	cong	avv	N_ pro	pro	int
per-classe	richiamo	88	75	69	64	68	65	62	55	58	36	33	40	24	25	22	22	12
	precisione	77	72	67	65	60	57	55	60	53	52	44	23	32	26	30	30	1
	F1-score	82	73	68	64	64	61	58	57	55	43	38	29	27	25	25	25	2
media	micro-F1	54																
	macro-F1	47																

5.2.4 I verbi in LB

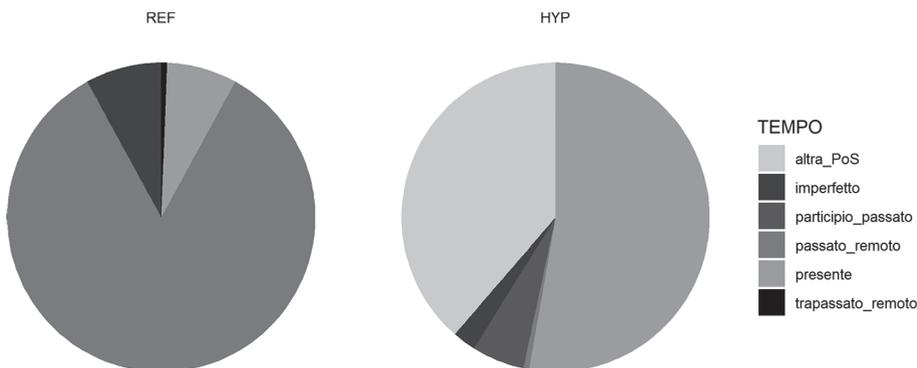
Il subcorpus LB è stato scelto, insieme a LP, come controllo, perché rappresentativo dell'italiano standard e di condizioni acustiche ottimali. Come atteso, il tasso di errore per questi materiali è basso ($WER_{LP}=1.5-5.3\%$; $WER_{LB}=5.7-8.8\%$). Eppure, l'accuratezza di trascrizione per LB è significativamente inferiore rispetto a quella di LP ($\chi^2 = 111.85$, $p < 2.2 \cdot 10^{-16}$), sebbene i due subcorpora siano stati registrati nelle stesse condizioni: in camera anecoica, con la medesima strumentazione, con la stessa varietà (italiano standard), con gli stessi parlanti, con lo stesso *task*. Ciò mette in luce il fatto che il riconoscimento può fallire anche in condizioni acustiche ottimali, per forme linguistiche standard, prodotte con un eloquio iperarticolato da parlatori professionisti (cfr. tab. 1). Si può quindi supporre che le cause degli errori per questi subcorpora siano da ricercare nel dominio linguistico e nel modello del linguaggio implementato nell'ASR: alcune forme sono meno frequenti di altre. Approfondire la natura degli errori presenti in LB è di particolare interesse non tanto per la differenza quantitativa, spiegata dal numero maggiore di DEL, ma per le caratteristiche di questo subcorpus. Infatti, LB comprende 120 stimoli, mentre LP ne comprende 20 (cfr. Falcone et al. 2007a, allegato C); il maggior numero di stimoli dà più spazio a strutture linguistiche rare e complesse, come esemplificato nella tab. 7. Quindi ci siamo concentrate sull'analisi degli errori di sostituzione dei verbi in LB, che costituiscono il 23.6% del totale (179 su 758 sostituzioni) e permettono di esemplificare una maggiore varietà di fenomeni.

Tabella 7 - Confronto tra le frasi di LP e LB

LP	LB
Un mese di vacanza passa in fretta.	Stimammo che il banco del Perù aveva un debito pari a mezzo milione di dollari.
Nel grande parco un bambino giocava con suo padre.	Il frate beve un blando orzo molto caldo e si scottò il labbro superiore.
Luisa fingeva di guardare da un'altra parte, ma cercava il modo per farsi notare.	Inghiottì la capsula e dopo un paio di secondi cadde sul posto stecchito.

Perciò, abbiamo annotato i verbi della REF di LB per tempo, modo, persona e lemma. Specularmente, abbiamo annotato i corrispondenti *token* dell'HYP: (a) secondo gli stessi parametri, se verbi; (b) con l'etichetta *altra_PoS*, se sostituiti da parole di classi diverse. I risultati sono mostrati nella fig. 5. In particolare, il 67.3% dei verbi della REF è al passato remoto indicativo e tende a essere sostituito nell'HYP da verbi al presente indicativo; in metà dei casi (50.7%) si tratta di forme dello stesso lemma. Ciò sembra riconducibile alla minore frequenza del passato remoto rispetto al presente indicativo nell'italiano contemporaneo (D'Achille 2019).

Figura 5 - Confronto REF-HYP dei verbi sostituiti in LB



Dalla fig. 5, inoltre, si può notare che, nelle sostituzioni di verbi in LB, è frequente anche il cambio di parte del discorso (37.9%), talvolta concomitante anche con sostituzioni multiparola (11.2%). Due sostituzioni tipiche di verbi in LB sono esemplificate in (14), dove *bevve* e *scottò* sono ricondotti rispettivamente al presente *beve* e alla preposizione *sotto*.

- (14) il frate beve un blando orzo molto caldo e si scottò il labbro superiore
 il frate beve un blando orzo molto caldo e si sotto il labbro superior

Se questo tipo di distribuzioni non viene controbilanciato in modo mirato nei materiali di addestramento o nel modello dei sistemi di ASR, come *output* viene preferito il candidato più frequente rispetto a quello meno frequente, soprattutto in concomitanza di una forte similarità fonetica (Song et al. 2021). Come esemplificato nella tab. 8, tale preferenza può essere sistematica: infatti, se in alcuni casi l'esito

delle sostituzioni è variabile, in altri casi le sostituzioni sono fisse e riguardano tutte le occorrenze di un dato verbo nel dataset (10 su 10).

Tabella 8 - *Verbi in LB sostituiti in tutte le occorrenze*

<i>SUB fisse</i>		<i>SUB variabili</i>	
REF	HYP	REF	HYP
bevve	beve	cadde	cade (9) che (1)
chiamai	chiama	convertì	convertiti (7) convertire (2) convertini (1) sotto (6)
colpì	colpi	scottò	scotta (3) Scott (1) sti mammo (5) stimiamo (2)
dormii	dormi	stimammo	sì mamma (1) ti mammo (1) sì ma mo (1)

6. Conclusioni preliminari

L'analisi preliminare conferma il divario tra materiali ottimali e non ottimali per quanto riguarda l'accuratezza della trascrizione automatica. I risultati sono in linea con quanto ipotizzato in letteratura (Draxler et al. 2020; Scagliola et al. 2020; Tavosanis 2018). Rimane però da indagare una serie di questioni. Innanzitutto, la definizione di standard e non-standard sul piano della pronuncia e degli ASR potrebbe non essere del tutto sovrapponibile. Tale conflitto emerge se si considera la natura dei corpora BA e L2, che differiscono sia per la qualità delle registrazioni sia per il tipo di italiano non-standard; di contro, la ricerca sulle *under-resourced languages* sembra indicare che la definizione di (non-)standard dipenda primariamente dalla distanza dell'*input* dai materiali di addestramento (Besacier et al. 2014; Biadys et al. 2012; Tavosanis 2018). Pertanto, lo specifico tipo di *input* sembrerebbe essere trascurabile, ma ulteriori approfondimenti potrebbero confutare tale ipotesi. Inoltre, è necessario individuare la gerarchia dei parametri che influenzano l'accuratezza, sebbene ciò comporti un allontanamento dall'approccio basato sugli utenti. Infine, si conferma la necessità di continuare a sviluppare strumenti di automazione delle trascrizioni, nonché complementari strategie di correzione che ne migliorino la fruizione da parte degli utenti impegnati nello studio di materiali linguistici, in particolare conversazionali e di italiano non-standard.

Bibliografia

- Abulimiti, Ayimunishagu & Schultz, Tanja. 2020. Automatic Speech Recognition for Uyghur through Multilingual Acoustic Modeling. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 6444-6449. Marseille.
- Albano Leoni, Federico & Maturi, Pietro. 1995. *Manuale di fonetica*. Roma: Carocci.
- Albano Leoni, Federico & Sobrero, Alberto A. & Paoloni, Andrea. 2007. Corpora e lessici di italiano parlato e scritto (CLIPS). *Bollettino di italianistica* 2. 122-130. doi:10.7367/71826.
- Ashwell, Tim & Elam, Jesse R. 2017. How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production? *JALT CALL Journal* 13(1). 59-76.
- Badino, Leonardo. 2016. The ArtiPhon Task at Evalita 2016. In Basile, Pierpaolo & Cutugno, Franco & Nissim, Malvina & Patti, Viviana & Sprugnoli, Rachele (eds.), *Evalita. Evaluation of NLP and Speech Tools for Italian*, 20-25. Pisa: Accademia University Press. doi:10.4000/books.aaccademia.1930.
- Bechet, Frederic & Favre, Benoit. 2013. ASR error segment localization for spoken recovery strategy. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6837-6841. Vancouver, BC, Canada: IEEE. doi:10.1109/ICASSP.2013.6638986.
- Besacier, Laurent & Barnard, Etienne & Karpov, Alexey & Schultz, Tanja. 2014. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication* 56. 85-100. doi:10.1016/j.specom.2013.07.008.
- Biadys, Fadi & Moreno, Pedro J & Jansche, Martin. 2012. Google's Cross-Dialect Arabic Voice Search. *2012 IEEE Acoustics, Speech, and Signal Processing*, 4441-4444. Kyoto, Japan. doi:10.1109/ICASSP.2012.6288905.
- Bybee, Joan L. 2001. *Phonology and language use*. Cambridge, UK: Cambridge University Press.
- Çetin, Özgür & Shriberg, Elizabeth. 2006. Speaker Overlaps and ASR Errors in Meetings: Effects Before, During, and After the Overlap. *2006 IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings*, vol. 1, 1-357-1-360. Toulouse, France: IEEE. doi:10.1109/ICASSP.2006.1660031.
- Crible, Ludivine. 2016. Discourse Markers and Disfluencies. Integrating Functional and Formal Annotations. *Proceedings of the LREC 2016 Workshop ISA-12*, 38-45. Portoroz, Slovenia.
- D'Achille, Paolo. 2019. *L'italiano contemporaneo*. Terza edizione. Bologna: il Mulino.
- Draxler, Christoph & van den Heuvel, Henk & van Hessen, Arjan & Calamai, Silvia & Corti, Louise & Scagliola, Stefania. 2020. A CLARIN Transcription Portal for Interview Data. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 3353-3359. Marseille.
- Falcone, Mauro & Barone, Antonio & Bonomi, Alessandro. 2007a. Definizione e caratterizzazione di un database vocale ortofonico realizzato da parlatori professionisti in camera anecoica. *Progetto CLIPS - W1-a3* 1-14.
- Falcone, Mauro & Barone, Antonio & Bonomi, Alessandro. 2007b. Realizzazione del database ortofonico in camera anecoica. *Progetto CLIPS - W1-a3* 17.

- Filippidou, Fotini & Moussiades, Lefteris. 2020. A Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems. In Maglogiannis, Ilias & Iliadis, Lazaros & Pimenidis, Elias (eds.), *Artificial Intelligence Applications and Innovations*, vol. 583, 73-82. Cham: Springer. doi:10.1007/978-3-030-49161-1_7.
- Kitaoka, Norihide & Enami, Daisuke & Nakagawa, Seiichi. 2014. Effect of acoustic and linguistic contexts on human and machine speech recognition. *Computer Speech & Language* 28(3). 769-787. doi:10.1016/j.csl.2013.09.009.
- Levenshtein, Vladimir Iosifovich. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady* 10. 707-710.
- Levis, John & Suvorov, Ruslan. 2012. Automatic Speech Recognition. In Chapelle, Carol A. (ed.), *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell. doi:10.1002/9781405198431.wbeal0066.
- Li, Jinyu & Deng, Li & Haeb-Umbach, Reinhold & Gong, Yifan. 2016. *Robust automatic speech recognition: a bridge to practical applications*. Waltham, MA: Academic Press.
- Maturi, Pietro. 2014. *I suoni delle lingue, i suoni dell'italiano: nuova introduzione alla fonetica*. Bologna: Il Mulino.
- Lindblom, B. 1990. Explaining Phonetic Variation: A Sketch of the H&H Theory. In Hardcastle, William J. & Marchal, Alain (eds.), *Speech Production and Speech Modelling*, 403-439. Dordrecht: Springer Netherlands. doi:10.1007/978-94-009-2037-8_16.
- Palmerini, Maria & Savy, Renata. 2014. Gli errori di un sistema di riconoscimento automatico del parlato. Analisi linguistica e primi risultati di una ricerca interdisciplinare. In Basili, Roberto & Lenci, Alessandro & Magnini, Bernardo (eds.), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014*, vol. I, 281-285. Pisa: Pisa University Press.
- Sakoe, Hiroaki & Chiba, Seibi. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(1). 43-49. doi:10.1109/TASSP.1978.1163055.
- Savy, Renata. 2007. Specifiche per la trascrizione ortografica annotata dei testi raccolti. *Progetto CLIPS - WI-a4* 1-28.
- Scagliola, Stefania & Corti, Louise & Calamai, Silvia & Karrouche, Norah & Beeken, Jeannine & van Hessen, Arjan & Draxler, Cristoph & van den Heuvel, Henk & Broekhuizen, Max & Truong, Khiet. 2020. Cross disciplinary overtures with interview data: Integrating digital practices and tools in the scholarly workflow. 126-136. doi:10.3384/ecp2020172015.
- Song, Minguang & Zhao, Yunxin & Wang, Shaojun & Han, Mei. 2021. Word Similarity Based Label Smoothing in Rnnlm Training for ASR. *2021 IEEE Spoken Language Technology Workshop (SLT)*, 280-285. Shenzhen, China: IEEE. doi:10.1109/SLT48900.2021.9383598
- Tavosanis, Mirko. 2018. *Lingue e intelligenza artificiale*. Roma: Carocci.
- van den Heuvel, Henk. 2020. Crossing the SSH Bridge with Interview Data. *Proceedings of LR4SSHOC: Workshop about Language Resources for the SSH Cloud*, 42-44.

Ritmo e malattia di Parkinson.

Prime riflessioni sull'annotazione automatica del parlato di pazienti in stadi iniziali

Abstract

Speech disorders resulting from Parkinson's disease (PD), typically classified as hypokinetic dysarthria, include the alteration of speech rhythm, even at the very onset of the disease. The purpose of this study is to test an automatic annotation procedure for the calculation of two rhythmic variables – %V, vowel percentage, and VtoV, the mean interval between two consecutive vowel onset points – on a corpus of read speech collected from 20 patients with early-stage PD and from 20 healthy controls, sex-matched and similar age. The results of the automatic procedure, compared to those previously obtained with the manual labelling, were found to be effective in the characterization of PD speech rhythm, with significantly higher values of %V in patients' productions than in healthy control speech. Some critical issues are discussed in details.

1. *Introduzione*

Il Parkinson è la seconda malattia neurodegenerativa più diffusa al mondo dopo l'Alzheimer, in particolare nella popolazione con più di 65 anni (de Lau & Breteler 2006). Numerosi studi sperimentali condotti su diverse lingue hanno evidenziato come i cambiamenti anatomici e fisiologici correlati alla perdita di dopamina nel sistema nervoso centrale abbiano effetti sul sistema respiratorio, fonatorio e articolatorio dei soggetti affetti da tale morbo (tra gli altri, Darley et al. 1969a, Ramig et al. 2008). Nel parlato parkinsoniano la disartria ipocinetica si manifesta, infatti, in alterazioni a livello segmentale, come la riduzione dell'area vocalica e il fenomeno del *target undershooting*, il mancato raggiungimento del target articolatorio nella produzione sia dei suoni vocalici sia di quelli consonantici, collegato alla riduzione/alterazione dell'ampiezza e della velocità di movimento degli organi fonatori (Gili Fivela et al. 2014, Rusz et al. 2013, Sapir et al. 2010, Skodda et al. 2011). Sul piano soprasegmentale si assiste, inoltre, a una più scarsa modulazione del parlato, con un *range* tonale ridotto rispetto a quello osservato in parlanti di controllo (Darley

¹ Dipartimento di Studi Letterari, Linguistici e Comparati, Università degli Studi di Napoli "L'Orientale".

² Dipartimento di Scienze Mediche e Chirurgiche avanzate, Università degli Studi della Campania "Luigi Vanvitelli".

et al. 1969b, Goberman & Coelho 2005, MacPherson et al. 2011). Anche le caratteristiche ritmiche e la velocità di articolazione risultano alterate (Liss et al. 2009), sebbene non vi siano dati uniformi nella letteratura sul tema: in alcuni casi i soggetti parkinsoniani presentano il fenomeno della tachilalia, per cui parlano molto più velocemente rispetto a quelli sani (Ackerman et al. 1997, Hirose et al. 1982); in altri casi il parlato patologico è descritto come molto rallentato (Logemann et al. 1978, Ludlow et al. 1987); in altri ancora non si riscontra nessuna differenza tra parlato parkinsoniano e parlato sano (Duez 2006, Skodda & Schlegel 2008).

È attestato, comunque, un fenomeno di globale *dysrhythmia*, che non interessa solo il piano della comunicazione verbale ma che si evidenzia anche in compiti non linguistici come nel tenere il tempo con l'andatura o con il movimento di un dito (Cochen De Cock et al. 2018, Puyjarinet et al. 2018).

Negli stadi più avanzati della malattia, il parlato parkinsoniano è fortemente esitante, caratterizzato da una bassa velocità di eloquio, con frequenti e lunghe pause silenziose e numerose disfluenze. Negli stadi iniziali, invece, i cambiamenti nei parametri acustici della voce, seppur strumentalmente rilevabili, sono percettivamente meno rilevanti e non intaccano l'intelligibilità della comunicazione. Tuttavia, diversi studi hanno ipotizzato che proprio l'osservazione di tali variazioni possa rivelarsi uno strumento sostenibile e non invasivo di supporto alla diagnosi clinica e al monitoraggio dell'avanzamento della patologia (tra gli altri, Cohen 2003, Harel et al. 2004).

1.1 Il progetto di ricerca

Il presente contributo è parte di un progetto di ricerca più ampio dal titolo "Analisi vocale per la diagnosi precoce del morbo di Parkinson: sviluppo e validazione di uno *speech tool* non invasivo", finanziato dai fondi PON-AIM 2014-2020. L'obiettivo a lungo termine del progetto è lo sviluppo di un dispositivo che, a partire da una traccia vocale (una registrazione di parlato letto), possa automaticamente calcolare alcuni parametri acustici.

In particolare, l'attenzione è posta su due variabili ritmiche, la durata media del VtoV (Vowel to Vowel), ossia l'intervallo tra il punto di inizio di un fono vocalico e il punto di inizio del fono vocalico successivo, e la percentuale di porzione vocale (%V) sul tempo totale dell'enunciato, escludendo le pause silenziose e le eventuali disfluenze³. In studi precedenti, infatti, la metrica %V/VtoV è risultata efficace nel rendere conto delle variazioni ritmiche del parlato parkinsoniano rispetto al parlato sano in diverse lingue (Pettorino et al. 2016, Pettorino et al. 2017, Pettorino

³ Prendere in considerazione il VtoV permette di superare i limiti di una segmentazione basata su una "tradizionale" divisione in sillabe. I confini sillabici non sono infatti sempre chiaramente identificabili, come già discusso in Pettorino et al. 2013. Al contrario, i punti di inizio dei fonemi vocalici, i *VOPs* (*Vowel Onset Points*), sono facilmente individuabili sul segnale acustico in quanto caratterizzati da un repentino e notevole incremento del valore dell'intensità. Questo tipo di segmentazione, basata su caratteristiche acustiche, riflette, quindi, la reale percezione di discontinuità ritmiche nel segnale da parte di chi ascolta.

et al. 2018) e anche negli stadi iniziali della malattia (Maffia et al. 2020, Maffia et al. 2021).

Parallelamente alla raccolta di un sempre più ampio corpus di parlato patologico, il raggiungimento dell'obiettivo del progetto di ricerca prevede che la procedura di analisi acustica del parlato, fino ad ora manuale, diventi automatica. L'automaticità è legata da un lato alla necessità di ridimensionare il carico di lavoro dell'annotazione manuale, nell'ottica di ottimizzare i tempi e le energie della ricerca nel suo complesso, dall'altro alla prospettiva di effettuare in futuro tali misurazioni senza la presenza dell'operatore, nel contesto delle consuete valutazioni cliniche o anche da remoto.

1.2 Obiettivo del lavoro

Partendo dai risultati degli studi precedentemente condotti sulle caratteristiche ritmiche del parlato parkinsoniano, in questo contributo si intende valutare l'applicazione della procedura di annotazione automatica nel contesto del parlato disartrico di persone allo stadio iniziale della malattia di Parkinson. Si presenteranno, inoltre, alcune osservazioni preliminari di carattere qualitativo sui risultati di tale procedura.

2. Metodologia

2.1 I partecipanti e le valutazioni cliniche

Sono stati finora coinvolti nella ricerca 20 pazienti con malattia di Parkinson (PD), di cui 12 donne, dai 41 agli 81 anni (età media di circa 64 anni), tutti reclutati presso l'ambulatorio "Parkinson e altri disordini del movimento" della I Clinica neurologica e neurofisiopatologia dell'Università della Campania "Luigi Vanvitelli" (Napoli).

La diagnosi è stata posta sulla base dei criteri diagnostici modificati della *UK Parkinson's Disease Society Brain Bank* (Gibb & Lees 1988). I criteri di inclusione utilizzati sono i seguenti: (1) esordio della malattia dopo i 40 anni, per escludere parkinsonismi a esordio precoce; (2) un punteggio alla scala di Hoehn & Yahr modificato (mH&Y) ≤ 2.5 ; (3) durata della malattia ≤ 4 anni. I criteri di esclusione sono stati i seguenti: (1) deterioramento cognitivo clinicamente rilevante come da criteri diagnostici; (2) depressione maggiore, depressione minore, e disturbo distimico come da criteri del DSM-IV; (3) altre malattie neurologiche o patologie sistemiche clinicamente rilevanti o instabili.

Sono stati inoltre reclutati 20 soggetti sani con funzione di gruppo di controllo (GC), comparabili ai pazienti per caratteristiche biografiche (8 donne, dai 54 ai 77 anni, età media 65 anni), senza storia di malattie neurologiche e/o disturbi del linguaggio.

Tutti i partecipanti alla ricerca, residenti in Campania, hanno dichiarato l'italiano come lingua materna. Un questionario, inoltre, ha permesso di raccogliere ulte-

riori informazioni di natura sociolinguistica sull'uso delle eventuali diverse lingue conosciute.

I 20 pazienti affetti dalla malattia di Parkinson sono stati sottoposti a una estesa valutazione clinica motoria e non motoria per la quale sono state utilizzate scale validate e specifiche. La severità dei sintomi motori associati alla malattia è stata valutata con l'utilizzo della scala mH&Y e della *Unified Parkinson's Disease Rating Scale* parte III (UPDRS III). La scala UPDRS III è stata anche utilizzata per valutare la presenza di disturbo dell'articolazione del linguaggio clinicamente rilevante, utilizzando un *cut-off* del punteggio all'item 3.1 (Speech) ≥ 1 . La compromissione del disturbo della parola è valutata dal neurologo secondo una scala di severità che va da 0 (normale) a 4 (compromissione severa). Le funzioni cognitive globali sono state valutate con l'utilizzo della scala *Montreal Cognitive Assessment* (MoCA – Flostein et al. 1975). Inoltre, la presenza e severità di sintomi depressivi sono state valutate con l'utilizzo della scala *Beck Depression Inventory* (Beck et al. 1961). Infine, la dose equivalente di terapia dopaminergica giornaliera è stata calcolata utilizzando l'algoritmo di Tomlinson (2010).

In tabella 1 si riportano le caratteristiche demografiche e cliniche di tutti i partecipanti.

Tabella 1 - *Caratteristiche dei partecipanti*

		GC (n=20) media \pm dev.st.	PD (n=20) media \pm dev.st.	p-value*
Dati biografici	Età	64.8 \pm 5.9	63.8 \pm 10.9	0.713
	Sesso (M/F)	8/12	12/8	0.205
	Durata della malattia (in mesi)	-	31.9 \pm 17.1	-
Dati clinici	mH&Y	-	2.1 \pm 0.4	-
	UPDRS III	-	24.0 \pm 7.6	-
	Item 3.1 (Speech)	-	0.7 \pm 0.5	-
	MoCA total	-	22.4 \pm 4.0	-
	BDI	-	5.2 \pm 3.6	-
	LEDD total	-	202.5 \pm 157.8	-
	LEDD DA	-	30.0 \pm 76.6	-

H&Y: Hoehn & Yahr; UPDRS: Unified Parkinson's Disease Rating Scale; MoCA: Montreal Cognitive Assessment; BDI: Beck Depression Inventory; LEDD: Levodopa Equivalent Daily Dose; DA: Dopamine-Agonist. *Test del Chi quadrato.

2.2 La raccolta dei dati

La raccolta dei dati ha avuto luogo nel corso di una visita di *follow-up* clinico e ha previsto la registrazione di parlato letto da parte di ciascun partecipante. Per il *task* di lettura, è stato proposto a ciascun soggetto uno stesso testo di 175 parole e di circa 350 sillabe su un argomento comune (le abitudini alimentari del passato e quelle dei nostri giorni), con un alto livello di leggibilità (indice Gunning Fog: 6; indice Gulpease: 70), semplici strutture morfosintattiche e un lessico prevalentemente ad

alta frequenza (94% dei *tokens*). La scelta di un tale testo input è stata dettata dalla volontà di evitare incertezze nel parlato dovute alla non comprensione di quanto letto, in particolare nell'eventualità di casi di debole alfabetizzazione tra i partecipanti e/o scarsa dimestichezza con simili procedure di raccolta dati.

A tutti i soggetti coinvolti è stato chiesto di leggere ad alta voce il testo nel modo più naturale possibile, al volume e alla velocità a loro più congeniali. Le registrazioni sono state effettuate in una stanza dell'Università "Luigi Vanvitelli", utilizzando il software Praat (Boersma & Weenink 2021) e con frequenza di campionamento pari a 44100 Hz. In totale sono stati registrati circa 50 minuti di parlato letto.

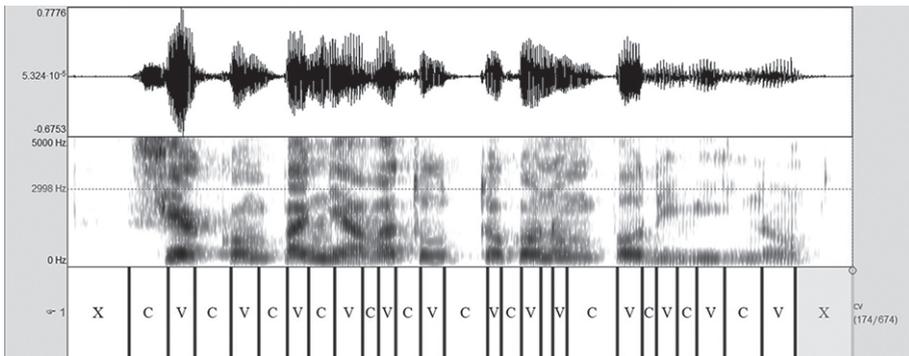
I pazienti hanno effettuato le registrazioni sotto l'effetto dei consueti farmaci dopaminergici. Tutti i soggetti hanno letto il foglio informativo sulle finalità della raccolta dati e firmato il relativo modulo per il consenso informato.

2.3 L'annotazione dei dati

Il corpus di registrazioni è stato sottoposto a due diverse procedure di annotazione.

In primo luogo esso è stato analizzato con Praat. Il segnale acustico è stato segmentato manualmente in intervalli vocalici (V) e consonantici (C), come mostrato in figura 1.⁴

Figura 1 - *Spettrogramma dell'enunciato "si accompagnava tutto con un po' di vino" (voce maschile). Annotazione manuale. C: intervallo consonantico, V: intervallo vocalico, X: pausa silente*



L'annotazione manuale dei più di 20.000 intervalli V e C è stata condotta da due esperti, i quali hanno svolto periodiche sessioni di standardizzazione dei risultati

⁴ La scelta di non riportare la trascrizione fonetica degli enunciati in tutte le didascalie alle immagini di spettrogrammi è legata all'utilizzo di una metodologia di analisi che intende superare le difficoltà relative all'individuazione dei singoli foni realmente pronunciati, siano essi consonantici o vocalici. Il tipo di segmentazione che si propone si basa, infatti, sulle caratteristiche del segnale acustico, in particolare dei parametri di intensità e periodicità: con l'etichetta V (o *Vowel*) si segnala la completa apertura del canale fonatorio, con libero passaggio del flusso di aria e assenza di turbolenze; al contrario, l'etichetta C (o *Consonant*) identifica la presenza di una chiusura, una stretta o un qualsiasi tipo di impedimento che, ostacolando il passaggio dell'aria nel canale fonatorio, provochi fruscio o frizione. La trascrizione fonetica sarà invece riportata in quei casi in cui si vorrà mostrare una realizzazione dell'enunciato molto lontana da quella attesa.

(in fase iniziale, intermedia e finale)⁵. In presenza di differenze o dubbi nelle due annotazioni, si è proceduto a un riesame caso per caso e si è stabilita una procedura comune.

In particolare, nell'annotazione manuale del segnale acustico, le criticità riscontrate sono state le seguenti:

- La segmentazione di alcune porzioni di parlato, soprattutto disartrico, caratterizzate da un forte fenomeno di ipoarticolazione. Un esempio è il tracciato riportato in figura 2. Il testo letto dal parlante è la frase “si faceva la spesa tutti i giorni”. Come si vede la sequenza si riduce dalle 11 sillabe previste alle 7 sillabe realmente prodotte. In casi come questo la segmentazione ha tenuto conto delle caratteristiche acustiche del segnale, indipendentemente dal modello atteso, per cui l'intero enunciato è stato suddiviso in 7 intervalli consonantici e 7 vocalici. Va anche detto che in casi più problematici, in cui non è stato possibile identificare i singoli segmenti, si è deciso di etichettare come sequenze non comprensibili tali porzioni di enunciato, non considerandole poi nel calcolo della %V e del VtoV.
- L'individuazione del confine nel nesso V+C nasale. Laddove visibile, la porzione nasalizzata della vocale è stata assegnata all'intervallo V. Un esempio particolarmente interessante è quello riportato in figura 3. La sequenza si riferisce alla frase “secondo diversi studi”. In corrispondenza del nesso [ondo] nella parola “secondo” è infatti possibile identificare cinque distinti tratti: un primo tratto vocalico, un tratto in cui la vocale si nasalizza, come risulta dalla riduzione dell'ampiezza del segnale, un tratto consonantico nasale, con ulteriore caduta dell'intensità, un tratto di occlusiva sonora, anch'essa parzialmente nasalizzata, e finalmente la vocale. Nella segmentazione i tratti risultano solo tre, V-C-V. L'inclusione della porzione corrispondente alla vocale nasalizzata nel primo intervallo vocalico (e non nella consonante nasale) si basa su considerazioni che tengono conto della meccanica articolatoria richiesta nella sua produzione. La vocale nasalizzata, infatti, si distingue dalla vocale orale esclusivamente per l'apertura del diaframma rinovelare durante la sua realizzazione. Ma tale apertura diaframmatica non è dovuta a una partecipazione attiva degli articolatori, nel caso specifico del velo e quindi del muscolo *levator palatini* ma, al contrario, è il risultato di un rilassamento di tale muscolo. Sul piano dell'attività neuromuscolare di controllo degli articolatori, quindi, la vocale nasalizzata richiede ancor meno impegno rispetto a una vocale orale, in cui è necessaria la contrazione del *levator palatini* per tutta la sua durata. Per questo motivo la vocale nasalizzata è da considerare come parte vocalica e non consonantica. Naturalmente le consonanti nasali, pur presentando lo stesso rilassamento del *levator palatini*, sono però caratterizzate dalla contrazione di uno dei muscoli orali (ad esempio per la

⁵ Come di norma in procedure di *standard setting*, nell'ottica di incrementare l'affidabilità dell'annotazione manuale, durante le sessioni i due valutatori hanno effettuato dei controlli incrociati sulla segmentazione di alcune porzioni di parlato scelte come campione.

pronuncia della bilabiale [m] sarà contratto l'*orbicularis oris*) e quindi classificate senza dubbio come consonanti.

- La durata delle consonanti occlusive a inizio di catena fonica. Nel caso delle occlusive sonore, il confine è stato identificato con l'inizio delle vibrazioni glottidali, sempre ben visibili sullo spettro; alle occlusive sorde, invece, è stata assegnata una durata pari al valore medio di un'occlusiva sorda scempia in posizione intervocalica nello stesso enunciato.

Figura 2 - Spettrogramma dell'enunciato "si faceva la spesa tutti i giorni" (voce maschile, PD), realizzato in una sequenza del tipo [sia|fes|sət:|iç|on:i]. Annotazione manuale. C: intervallo consonantico, V: intervallo vocalico, X: pausa silente

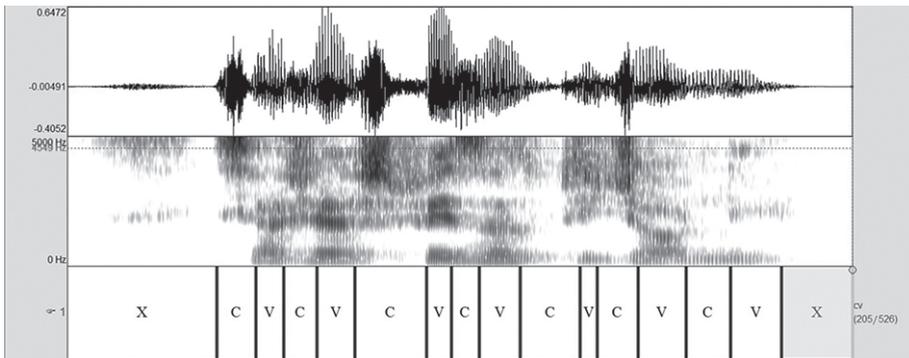
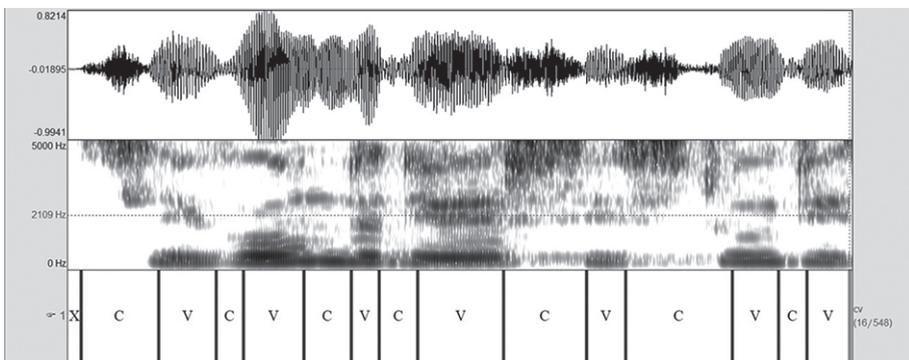


Figura 3 - Spettrogramma dell'enunciato "secondo diversi studi" (voce femminile, PD). Annotazione manuale. C: intervallo consonantico, V: intervallo vocalico, X: pausa silente



Le stesse registrazioni di parlato letto, inoltre, sono state sottoposte a una procedura di annotazione completamente automatica, condotta con il software SPPAS, versione 3.7 (Bigi 2015), già usato per l'annotazione del parlato disartrico in studi sulla lingua francese (Bigi et al. 2015), *open source* e con un'intuitiva interfaccia grafica. A partire da un file audio e dalla trascrizione ortografica degli enunciati, SPPAS è in grado di produrre l'annotazione automatica del segnale su diversi livelli: *IPUs*, cioè *Inter Pausal Units* o catene foniche, parole, sillabe e fonemi. La procedura automa-

3. Risultati

Nella tabella 2 si riportano i valori medi di %V e di VtoV ottenuti nei due gruppi di parlanti, GC e PD, con le due procedure di annotazione.

Tabella 2 - Valori di %V e VtoV nei due gruppi e con le due annotazioni

		GC (n=20) Media±dev.st.	PD (n=20) Media±dev.st.	p-value
annotazione manuale	%V	44.7±1.3	49.8±1.7	<0.001
	VtoV	0.186±0.02	0.185±0.02	0.889
annotazione automatica – SPPAS	%V	43±4.3	45.8±2.6	0.017
	VtoV	0.208±0.03	0.196±0.03	0.201

Si osserva innanzitutto come l'annotazione automatica sia in grado di distinguere efficacemente i due gruppi sulla base del parametro %V, così come avviene con l'annotazione manuale: i valori relativi al parlato patologico risultano significativamente più alti rispetto a quelli del GC, sebbene la significatività sia decisamente più bassa nel caso dei dati ottenuti con SPPAS.

In relazione al VtoV, entrambe le annotazioni non producono risultati che cambiano significativamente nei due gruppi.

Se la procedura automatica sembra confermare le osservazioni ottenute con la segmentazione manuale del segnale, si nota, tuttavia, che i valori di %V ottenuti con SPPAS sono in media più bassi rispetto a quelli dell'annotazione manuale, con uno scarto maggiore nel caso del gruppo PD (4%) rispetto ai parlanti GC (1.7%).

Per cercare di capire le motivazioni di tali risultati e verificare in che modo SPPAS “perda” porzioni vocaliche o “guadagni” porzioni consonantiche rispetto all'annotazione manuale, si è proceduto a un'analisi qualitativa delle segmentazioni, che si è focalizzata in particolare sulle produzioni dei parlanti che ottengono risultati più divergenti con le due procedure di annotazione. Si riportano di seguito alcune osservazioni preliminari.

Il software riscontra alcune criticità laddove la qualità della registrazione risulta più bassa: un rumore di fondo o il riverbero ambientale, facilmente identificabili nell'analisi fatta da esperti, sono senz'altro elementi che non permettono sempre al sistema automatico di riconoscere con buona approssimazione i diversi foni e di individuarne i confini.

Nei casi, inoltre, di accentuata ipoarticolazione nel parlato disartrico, come quello riportato in figura 5, il software è costretto forzatamente a riconoscere tutti i fonemi rappresentati nella trascrizione ortografica, anche quando essi non siano effettivamente presenti nel segnale acustico: le due consonanti vibranti sono del tutto assenti nell'enunciato prodotto ma il software assegna loro due segmenti, sacrificando una porzione delle vocali che le precedono.

Le annotazioni ottenute con il software SPPAS si sono, infatti, dimostrate efficaci nel differenziare i due gruppi di parlanti sulla base del parametro %V sebbene si siano evidenziate alcune criticità e alcune divergenze rispetto alla segmentazione manuale. È possibile ipotizzare che, con una migliore qualità delle registrazioni e con una più accurata trascrizione del parlato, la procedura di annotazione automatica possa funzionare sempre meglio e in maniera sempre più affidabile.

In conclusione, quindi, in questa fase del progetto di ricerca, nell'ottica di ottimizzare i tempi di lavoro, l'utilizzo di una metodologia mista di annotazione, che preveda una prima fase automatica seguita da un controllo manuale, è senz'altro auspicabile. Questo primo test sul corpus di parlato letto, inoltre, ha permesso di raccogliere informazioni utili alla progettazione di un nuovo protocollo di registrazione, più adatto a una procedura di annotazione completamente automatica, che rappresenta l'obiettivo ultimo del progetto.

Ringraziamenti

Un doveroso ringraziamento va a Brigitte Bigi, per il supporto agli autori nell'uso di SPPAS e la grande disponibilità dimostrata.

Bibliografia

- Ackermann, Hermann & Konczak, Jürgen & Hertrich, Ingo. 1997. The temporal control of repetitive articulatory movements in Parkinson's disease. *Brain and Language* 56(2). 312-319.
- Beck, Aaron T. & Ward, Calvin H. & Mendelson, Mock & Mock, Jeremiah & Erbaugh, John. 1961. An inventory for measuring depression. *Archives of General Psychiatry* 4(6). 561-571.
- Bigi, Brigitte & Meunier, Christine. 2018. Automatic speech segmentation of spontaneous speech. *Revista de Estudos da Linguagem* 26(4).
- Bigi, Brigitte & Klessa, Katarzyna & Georgeton, Laurianne, & Meunier, Christine. 2015. A syllable-based analysis of speech temporal organization: a comparison between speaking styles in dysarthric and healthy populations. In *Sixteenth Annual Conference of the International Speech Communication Association (INTERSPEECH 2015)*, Dresda. 2977-2981.
- Bigi, Brigitte. 2015. SPPAS – Multilingual Approaches to the Automatic Annotation of Speech. *the Phonetician – International Society of Phonetic Sciences* 111-112 (2015-I-II). 54-69.
- Boersma, Paul & Weenink, David. 2021. *Praat: doing phonetics by computer* [Computer program]. Version 6.1.39, <http://www.praat.org/>.
- Cochen De Cock, Valérie & Dotov, Dobromir & Ihalainen, Petra et al. 2018. Rhythmic abilities and musical training in Parkinson's disease: do they help?. *Parkinson's Disease* 4(8).
- Cohen, Henri. 2003. Disorders of speech and language in Parkinson's disease. In Bédard, Marc-Andre & Agid, Yves & Korczyn, Amos & Lespérance, Paul & Chouinard, Sylvain

(eds.), *Mental and behavioral dysfunction in movement disorders*, 125-134. New York: Humana Press.

Darley, Frederic L. & Aronson, Arnold E., & Brown, Joe R. 1969a. Differential diagnosis patterns of dysarthria. *Journal of Speech and Hearing Research* 12(2). 246-269.

Darley, Frederic L. & Aronson, Arnold E., & Brown, Joe R. 1969b. Cluster of deviant speech dimension in the dysarthrias. *Journal of Speech and Hearing Research* 12(3), 462-469.

de Lau, Lonneke M. & Breteler, Monique M. B. 2006. Epidemiology of Parkinson's disease. *The Lancet. Neurology* 5(6). 525-535.

Duez, Danielle. 2006. Syllable structure, syllable duration, and final lengthening in Parkinsonian French speech. *Journal of Multilingual Communication Disorders* 4. 45-57.

Flostein, Marshal F. & Folstein, Susan E. & McHugh, Paul R. 1975. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12. 189-198.

Gibb, William R. & Lees, Andrew J. 1988. A comparison of clinical and pathological features of young- and old-onset Parkinson's disease. *Neurology* 38. 1402-1406.

Gili Fivela, Barbara & Iraci, Massimiliano M. & Sallustio, Vincenzo & Grimaldi, Mirko & Zmarich, Claudio & Patrocino, Danilo. 2014. Italian Vowel and Consonant (co) articulation in Parkinson's Disease: extreme or reduced articulatory variability? In *10th International Seminar of Speech Production (ISSP), 5-8 Maggio 2014*, Colonia, Germania. 146-149.

Goberman, Alexander M. & Coelho, Carl A. 2005. Prosodic characteristics of Parkinsonian speech: The effect of levodopa-based medication. *Journal of Medical Speech-Language Pathology* 13. 51-68.

Harel, Brian T. & Cannizzaro, Michael S. & Cohen, Henri & Reilly, Nicole, & Snyder, Peter J. 2004. Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. *Journal of Neurolinguistics* 17. 439-453.

Hirose, Hajime & Kiritani, Shigeru & Sawashima, Masayuki. 1982. Velocity of articulatory movements in normal and dysarthric subjects. *Folia phoniatrica* 34(4). 210-215.

Liss, Julie M. & White, Laurence & Mattys, Sven L. & Lansford, Kaitlin & Lotto, Andrew J. & Spitzer, Stephanie M. & Caviness, John N. 2009. Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research* 52. 1334-1352.

Logemann, Jeri A. & Fisher, Hilda B. & Boshes, Benjamin & Blonsky, E. Richard. 1978. Frequency and cooccurrence of vocal tract dysfunction in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders* 43. 47-57.

Ludlow, Christy L. & Connor, Nadine P. & Bassich, Celia J. 1987. Speech timing in Parkinson's and Huntington's disease. *Brain and Language* 32(2). 195-214.

MacPherson, Megan K. & Huber, Jessica E. & Snow, David P. 2011. The intonation-syntax interface in the speech of individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research* 54(1). 19-32.

Maffia, Marta & De Micco, Rosa & Pettorino, Massimo & Siciliano, Mattia & Tessitore, Alessandro & De Meo, Anna. 2021. Speech rhythm variation in early-stage Parkinson's disease: a study on different speaking tasks. *Frontiers in Psychology* 12 (668291).

- Maffia, Marta & De Micco, Rosa & Tessitore, Alessandro & De Meo, Anna. 2020. Acoustic characteristics of Italian Parkinsonian speech: a study on early-stage patients. In Botinis, Antonio (ed.) *Proceedings of the 11th International Conference of Experimental Linguistics – Exling2020*, Atene. 117-120.
- Pettorino, Massimo & Maffia, Marta & Pellegrino, Elisa & Vitale, Marilisa & De Meo, Anna. 2013. VtoV: a perceptual cue for rhythm identification, in Mertens, Piet & Simon, Anne Catherine (eds.), *Proceedings of the Prosody-Discourse Interface Conference 2013 - IDP2013, Leuven*, 101-106.
- Pettorino, Massimo & Busà, Maria Grazia, & Pellegrino, Elisa. 2016. Speech Rhythm in Parkinson's Disease: A Study on Italian. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, 1958-1961.
- Pettorino, Massimo & Gu, Wentao & Pólrola, Pawel & Fan, Ping, 2017. Rhythmic Characteristics of Parkinsonian Speech: A Study on Mandarin and Polish. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, Stoccolma, 3172-3176.
- Pettorino, Massimo & Hemmerling, Daria & Vitale, Marilisa & De Meo, Anna. 2018. Towards a Speech-Test for Parkinson's Disease detection: A Diachronic Study on Michael J. Fox. In *Proceedings of 41st International Conference on Telecommunications and Signal Processing (TSP2018)*, Atene, 1-5.
- Puyjarinet, Frédéric & Bégel, Valentin & Gény, Christian & Pinto, Serge & Dalla Bella, Simone. 2018. Relations between rhythmic impairment in speech and non-verbal domains in Parkinson's disease. *Annals of Physical and Rehabilitation Medicine* 61. e46-e47.
- Ramig, Lorraine O. & Fox, Cynthia & Sapir, Shimon. 2008. Speech treatment for Parkinson disease. *Expert Review of Neurotherapeutics* 8. 299-311.
- Rusz, Jan & Cmejla, Roman & Tykalova, Tereza & Ruzickova, Hana & Klempir, Jiri & Majerova, Veronika, et al. 2013. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. *The Journal of the Acoustical Society of America* 134(3). 2171-2181.
- Sapir, Shimon & Ramig, Lorraine O. & Spielman, Jennifer L. & Fox, Cynthia. 2010. Formant Centralization Ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research* 53(1). 114-125.
- Skodda, Sabine & Schlegel, Uwe. 2008. Speech rate and rhythm in Parkinson's disease. *Movement Disorders* 23(7). 985-992.
- Skodda, Sabine & Visser, Wenke & Schlegel, Uwe. 2011. Vowel articulation in Parkinson's disease. *Journal of voice: official journal of the Voice Foundation* 25(4). 467-472.
- Tomlinson, Claire L. & Stowe, Rebecca & Patel, Smitaa & Rick, Caroline & Gray, Richard & Clarke, Carl E. 2010. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society* 25(15). 2649-2653.

La *sottodeterminazione* nei testi giuridici: verso un'analisi linguistico-computazionale³

Abstract

This contribution introduces a theoretical-methodological paradigm framed in an interdisciplinary research project at the intersection among law, linguistics and natural language processing, focused on the legislative and regulatory production of Valle D'Aosta Region, as a limited though autonomous body of rules.

Main target of the project is the notion of underdetermination, that is the property of juridical utterances to allow language to convey different, alternative interpretations. The experimental methods employed for the field survey are illustrated, along with the applications of digital humanities used to devise a descriptive model of phenomenology and linguistic realities of underdetermination.

The final aim of the project is to describe and to track underdetermination in the many different forms it takes in the wide range of the different text typologies encompassed in the juridical field.

1. *Inquadramento*

Nell'applicare la legge non si può ad essa attribuire altro senso che quello fatto palese dal significato proprio delle parole secondo la connessione di esse, e dalla intenzione del legislatore.

“Disposizioni sulla legge in generale”, Art. 12, comma I.

Per quanto in parte mitigato dal riferimento alla *ratio iuris* (l'*intenzione del legislatore*) e da successive determinazioni che prefigurano in alternativa all'interpretazione letterale una serie di possibilità ermeneutiche sempre ad essa gerarchicamente subordinate, la previsione del *Codice civile* posta in epigrafe si affida al convincimento di matrice cognitivista che la lingua possa rappresentare in modo univoco e inequivocabile il valore significativo (*sensu*) ricavabile dallo specifico contenuto semantico (*significato proprio*) delle singole unità linguistiche presenti (*parole*) e dalla loro concatenazione nel contesto morfosintattico (*secondo la connessione di esse*).

¹ Università della Valle d'Aosta.

² Università degli Studi di Torino.

³ I contenuti qui esposti sono frutto delle riflessioni teoriche e delle scelte metodologiche condivise in piena collaborazione dagli autori. Per quanto riguarda la stesura del testo, sono stati elaborati in comune i paragrafi 1, 2 e 7; Antonio Mastropaolo è autore del paragrafo 3; Daniele Paolo Radicioni del paragrafo 5; Luisa Revelli dei paragrafi 4 e 6 e della revisione generale del testo.

Il contributo che qui presentiamo si concentra sulle circostanze in cui il contenuto espresso da un enunciato normativo non presenta invece le caratteristiche di evidenza attese, e pone anzi problemi di decodifica tali da impedire un'interpretazione univoca. Si propone, più in particolare, di delineare un quadro teorico che consenta di identificare le condizioni che determinano la possibilità di un enunciato giuridico di prestarsi a diverse e alternative interpretazioni plausibili.

Mutuando il termine dall'ambito della filosofia della scienza e riconoscendogli una tradizione d'ambito anglosassone (*underdetermination*) molto più solida di quella del contesto italiano, adottiamo per definire tale caratteristica la denominazione di *sottodeterminazione* (d'ora in poi: *S.*): le potenziali alternative espressioni di *oscurità, vaghezza, indeterminatezza, ambiguità* su cui nell'ultimo trentennio si è sviluppata una pur ampia e autorevole letteratura comprendono infatti contesti in cui l'impossibilità di interpretazione letterale non è necessariamente collegata alle caratteristiche linguistiche e semantiche dell'enunciato e tendono comunque a orientarsi sugli effetti più che sulle cause del fenomeno.

Scopo del contributo è invece di presentare un paradigma teorico e metodologico di riferimento per la messa a punto di strumenti linguistico-computazionali capaci di identificare le circostanze che producono o possono produrre *S.* e quindi di contenere le possibilità di arbitraria soggettività interpretativa che la *S.* per sua stessa natura alimenta.

Le ipotesi illustrate fanno riferimento a studi condotti nelle fasi preliminari di un Progetto triennale dedicato alla normativa regionale d'ambito valdostano (§ 2).

L'attenzione qui concentrata sulla *S.* – nozione collocata alla base del costrutto di ricerca adottato – si pone l'obiettivo di evidenziarne anzitutto le specifiche e multiformi declinazioni d'ambito giuridico: imputabili a scelte in alcuni contesti necessarie e anche virtuose, ma anche - in differenti circostanze e all'opposto - a inconsapevole o intenzionale vaghezza (§ 3), le manifestazioni linguistiche della *S.* si correlano ai gradi di vincolo interpretativo posti dalle diverse tipologie testuali di genere giuridico-normativo richiedendo paradigmi descrittivi che non possono essere ricondotti tanto a parametri di *comprensibilità* quanto piuttosto a principi di *interpretabilità* e *interpretatività* (§ 4) che si ipotizza si trovino proporzionalmente distribuiti in un continuum categoriale che vede da un lato i *testi normativi* generali e astratti, *di rango primario*, e al capo opposto i testi a carattere applicativo o esecutivo. Con l'obiettivo di illustrare tale assunto viene qui testato come caso di studio l'esempio dell'uso dei lessemi appartenenti al *vocabolario di base* italiano in tre testi rappresentativi di altrettante categorie del continuum (§ 5). La prova condotta mira a supportare l'ipotesi che la *S.* non abbia (tanto o soltanto) a che fare con i vizi di eccessiva complessità o ostentazione tipicamente addebitati al linguaggio giuridico-amministrativo, ma si correli piuttosto a istanze pragmatiche – di volta in volta performative, deontiche, prescrittive, persuasive, regolative e altro ancora – che, interessando diversi livelli della lingua, si manifestano attraverso fenomeni di semantica frasale. Ai fini dell'individuazione e della puntuale descrizione di tali fenomeni la seconda fase del progetto prevede la conduzione di rilievi computazionali

su corpora bilanciati e, in parallelo, di prove sperimentali basate sulla consultazione diretta dei parlanti (§ 6): a partire dalle evidenze induttivamente ricavate e comparativamente analizzate ci si propone di giungere a delineare una fenomenologia dei marcatori linguistici assumibili come indicatori della S. giuridica. A partire da tale base nella fase conclusiva del Progetto saranno sviluppate applicazioni di supporto al *drafting* normativo (§ 7).

2. Il Progetto di ricerca

Condotto con un approccio interdisciplinare tra linguistica, diritto e informatica, il progetto di ricerca “La lingua del legislatore in Valle d’Aosta”⁴ pone il proprio focus sulla normativa della piccola regione alpina che, in virtù dello Statuto di autonomia approvato con legge costituzionale 26 febbraio 1948, n.4 della Repubblica italiana, presenta proprie particolarità linguistiche e giuridiche.

Relativamente al primo aspetto – malgrado il repertorio sociolinguistico contemporaneo delinei una condizione di *bilinguismo sociale endocomunitario con dilalia multipla* (Berruto 2003: 45) in cui il francese occupa rispetto all’italiano una posizione debole in quanto a usi e competenze individuali (Raimondi 2006) – l’equiparazione statutaria⁵ delle due lingue ufficiali ha condotto l’amministrazione regionale e gli enti locali a prevedere che entrambe e in parallelo siano utilizzate per la redazione di leggi, regolamenti, atti amministrativi, ecc.

Sulla scorta di quanto previsto dalle “Norme in materia di tutela delle minoranze linguistiche storiche” (L. 482/1999) è inoltre ammessa la possibilità di traduzione dei documenti anche nella principale parlata minoritaria locale. Proprio con l’obiettivo di “agevolare le comunicazioni in lingua francoprovenzale tra i cittadini e le istituzioni” l’Assessorato all’Istruzione e Cultura ha attivato uno specifico servizio di consulenza linguistica⁶.

Il potenziale plurilingue della normazione locale chiama dunque in causa tutto l’universo della trasposizione orizzontale, quindi anche insidie legate alle specificità della traduzione giuridica come le difficili equivalenze del suo lessico settoriale, le difficoltà di armonizzazione o omologazione di concetti e istituti fra contesti diversi e le esigenze di individuare soluzioni, nello specifico caso della trasposizione in francoprovenzale, per sopperire alle lacune che necessariamente una parlata locale priva di tradizione settoriale in ambito legislativo presenta.

Per ciò che concerne le particolarità dell’ordinamento giuridico, la possibilità di prevedere adattamenti locali della normativa nazionale implica poi – e soprattutto

⁴ Il Progetto prevede un’articolazione in fasi e segmenti: al tema de “L’accessibilità dei testi normativi come dispositivo di inclusione” è dedicato uno specifico programma di attività finanziato nell’ambito del bando 2021 “Erogazioni ordinarie” della Fondazione CRT.

⁵ Art. 38, commi 1 e 2: “Nella Valle d’Aosta la lingua francese è parificata a quella italiana. Gli atti pubblici possono essere redatti nell’una o nell’altra lingua, eccettuati i provvedimenti dell’autorità giudiziaria, i quali sono redatti in lingua italiana”.

⁶ Cfr. www.patoisvda.org.

in caso di *S.* del dettato normativo statale – che le possibilità interpretative a disposizione del legislatore regionale determinino esiti inattesi o anche conflittuali rispetto alle competenze dello Stato.

Con queste premesse, il progetto avviato presso l'Università della Valle d'Aosta mira a definire circostanze e fenomenologia della *S.* esaminandone le declinazioni all'interno delle diverse dimensioni implicate nel contesto valdostano assunto a caso di studio: a livello intralinguistico, in rapporto alla riformulazione del dettato statale per l'introduzione degli adattamenti locali; a livello interlinguistico, in relazione alla trasposizione dei testi normativi regionali nella seconda lingua di cultura e potenzialmente anche nella parlata minoritaria endogena; nella dimensione produttiva, con riguardo alle diverse fasi del *drafting* normativo e dei processi di redazione di conseguenti atti applicativi; in quella ricettiva, in collegamento alle dimensioni linguistiche e giuridiche dei processi interpretativi.

3. *Sottodeterminazione e universi giuridici*

Le questioni legate alla 'buona scrittura delle leggi' sono state esaminate, soprattutto nell'ultimo ventennio e anche in sedi istituzionali, attraverso approcci linguistici e giuridici integrati. All'interno dei dibattiti scaturiti dal confronto interdisciplinare il problema dell'oscurità espressiva è stato ricondotto dai linguisti principalmente a questioni di leggibilità e comprensibilità testuale e inquadrato secondo prospettive orientate sui destinatari nell'ambito dei contesti della comunicazione pubblica; è stato d'altra parte riconosciuto dai giuristi come radicato vezzo legato ad ataviche istanze di salvaguardia del valore magico della lingua giuridica (Ondelli 2007), a intenti criptolalic conservativi dei privilegi di cerchie ristrette, a cifre stilistiche di ostentazione erudita, ma anche descritto come in alcuni contesti indispensabile dispositivo di garanzia della flessibilità e longevità di norme altrimenti soggette a rapida obsolescenza.

Casi di *S.* possono ad esempio essere identificati nella *Costituzione italiana*, il cui testo – coralmemente elogiato per efficacia di formulazione – lascia in diversi luoghi possibilità di spazi interpretativi alternativi. Citiamo, a titolo di esempio, il comma 2 dell'art. 59, che nella versione originale recitava:

Il Presidente della Repubblica può nominare senatori a vita cinque cittadini che hanno illustrato la Patria per altissimi meriti nel campo sociale, scientifico, artistico e letterario.

L'oscillazione fra interpretazioni restrittive o estensive, a seconda che per *Presidente della Repubblica* si intendesse l'organo costituzionale o invece l'individuo che ricopre l'incarico, ha inciso in alcune circostanze della storia italiana recente sulla sopravvivenza di maggioranze politiche per le quali il voto dei senatori a vita risultava decisivo. Una recente revisione costituzionale (art. 3 comma 1 della L.C. n. 1 del 19 ottobre 2020) ha infine risolto la *S.* della formulazione originaria attraverso la seguente integrazione destinata a definitiva disambiguazione: "Il numero com-

pllessivo dei senatori in carica nominati dal Presidente della Repubblica non può in alcun caso essere superiore a cinque”.

La possibilità d'intervenire sul testo dopo aver sperimentato i concreti effetti di formulazioni deliberatamente sottodeterminate consente di evitare decisioni aprioristiche, potenzialmente foriere di esiti imprevedibili: risulta d'altra parte chiaro che gli usi della *S.* ispirati a tale principio di flessibilità o cautela possano trovare maggiore o minore legittimazione a seconda delle categorie testuali in cui si collocano e delle cause che ne motivano la presenza.

Partendo dall'inquadramento proposto in Longo, Mastropaolo e Pallante (2012) e approfondendone la fenomenologia, proponiamo allora anzitutto di distinguere i fenomeni di *S.* deliberatamente adottata dal legislatore (*S. fisiologica*) da quelli in cui la *S.* non svolge invece funzioni giuridicamente motivate (*S. patologica*).

Alla categoria della *S. fisiologica* appartengono i casi in cui il legislatore con intenzionalità seleziona fra le molte disponibili le soluzioni linguistiche capaci di garantire una comprensività dell'enunciazione indispensabile in presenza di una pluralità di possibili scenari applicativi, presenti o futuri, così come quelli in cui un certo grado d'indeterminatezza espressiva rappresenta un'ineludibile mediazione strategica fra istanze politiche differenti e altrimenti inconciliabili. La presenza di *S. fisiologica* in circostanze di questo tipo non inficia la normatività della *produzione legislativa*: semplicemente rimanda la puntualità di concrete declinazioni di significato attraverso una delega interpretativa.

Costituiscono invece esempi di *S. patologica* i casi di equivocità degli enunciati determinati da insufficiente elaborazione concettuale del legislatore; le ambiguità conseguenti a sciatteria linguistica e scrittoria; le formulazioni criptiche introdotte per l'esibizione di virtuosismi stilistici. In questi casi la presenza di *S.* corrisponde a tutti gli effetti a *cattiva produzione legislativa* ed espone gli enunciati a un' indesiderabile arbitrarietà interpretativa.

Da tale assunto discendono due corollari. Il primo prevede che laddove la *S. fisiologica* sia sempre controllata e intenzionalmente introdotta dal legislatore, la *S. patologica* non risponda invece a una reale necessità e possa quindi essere quantomeno circoscritta attraverso riformulazioni linguistiche senza che questo comporti sacrifici o distorsioni del contenuto regolativo. Il secondo fa riferimento alla disposizione delle categorie testuali della letteratura d'ambito giuridico all'interno di un continuum ai cui estremi si collocano rispettivamente i *testi normativi di rango primario*, generali e astratti secondo la classica definizione della legge, e all'opposto i *testi normativi di rango secondario*, applicativi o esecutivi, con funzioni illocutive e carattere pragmatico: in quest'ottica sembra possibile ipotizzare che presentino maggiori necessità e caratteristiche di *S. fisiologica* i testi di rango primario, per i quali si prevedono ampio respiro, caratteristiche di generalità e una quanto più ampia possibile longevità; che debbano sottrarsi alla *S. patologica* – richiedendo invece la massima coerenza e univocità interpretativa – le tipologie testuali di prospettiva applicativa, di breve prospettiva temporale e/o limitato ambito di applicazione.

Anche con riguardo ai pubblici di riferimento e rispetto all'asse dei rapporti fra *S.* e comprensibilità sembra d'altra parte opportuno distinguere requisiti e caratteristiche dei testi e degli enunciati a carattere prettamente applicativo-amministrativo da quelli più propriamente normativo-giuridici, giacché, come avverte Rovere (2007: 202), «la comprensibilità di un testo giuridico per un interlocutore profano non implica necessariamente che per il giurista il testo sia univoco. Viceversa, rendere comprensibili testi giuridici a non specialisti significherebbe mettere a loro disposizione, in aggiunta al testo, tutto il sapere giuridico indispensabile a tal fine, oppure accettare il rischio, altamente probabile, di imprecisioni e ambiguità tecniche».

4. *Sottodeterminazione e grado di vincolo interpretativo: interpretabilità e interpretatività*

Proposto e a più riprese approfondito da Francesco Sabatini, anche con espresso riferimento all'ambito normativo, il modello di classificazione delle tipologie testuali basato sul *grado di vincolo interpretativo* attribuisce ai testi giuridici la natura di scritture collocate al 'polo della massima rigidità ed esplicitezza' (2011b: 338), con *discorso molto vincolante* e "funzione prescrittiva, basata su una manifestazione di volontà coercitiva, regolata da un intero sistema di principi enunciati espressamente". Precisa Sabatini (2011a: 192) che l'intenzione dell'emittente "di regolare in maniera più o meno rigida l'attività interpretativa del destinatario" presenta due componenti, che bene si prestano a delineare una prima definizione delle circostanze della *S.*

La prima, descritta come "atteggiamento di personale e generica attenzione dell'emittente nel cercare di rendere ricevibile il senso complessivo del messaggio" tratteggia in riferimento a un enunciato giuridico la sua *interpretabilità* intesa come concreta possibilità decifrativa del significato globale. All'interno del patto comunicativo tra emittente e destinatario, le possibilità che il testo risulti decodificabile (*ricevibile*) hanno necessariamente a che fare con le specifiche competenze del ricevente: per garantire a un pubblico quanto più ampio possibile l'accessibilità di un enunciato giuridico occorre che questo sia formulato nel rispetto degli indicatori canonici della comprensibilità e leggibilità.

La seconda componente, che attribuisce all'estensore del testo un "comportamento di forte consapevolezza delle modalità di interpretazione puntuale di ogni elemento linguistico del suo messaggio", richiama invece condizioni di possibilità esegetica che non possono essere ricondotte a parametri linguistici di superficie ma evocano piuttosto dispositivi ermeneutici a base semantica. Ci pare sia questa seconda condizione, che denominiamo *interpretatività*, a chiamare in causa la *cogenza* o al contrario la *S.* di un enunciato o di un testo giuridico.

Poniamo allora alla base del nostro paradigma di analisi una prima fondamentale distinzione tra i concetti di *interpretabilità* e di *interpretatività*: le conseguenti ipotesi riguardano la possibilità che un enunciato giuridico caratterizzato da *interpretabilità* non necessariamente presenti anche caratteristiche di *interpretatività*; che in

alcuni casi, anzi, *interpretabilità* e *interpretatività* siano inversamente proporzionali, come avviene quando con l'obiettivo di "rendere ricevibile il senso complessivo del messaggio" le possibilità di "interpretazione puntuale di ogni elemento linguistico" possano essere sfavorite da una riduzione dell'intensione degli enunciati e quindi dal ricorso, intenzionale o accidentale, alla *S*.

Ipotizziamo, inoltre, che la *S*. si corredi all'*interpretatività* concentrandosi nella sua veste *fisiologica* – all'interno del continuum delle categorie testuali ascrivibili all'ambito giuridico – nelle formulazioni dei testi normativi di rango primario, gradatamente scomparendo o rarefacendosi nei testi applicativi o esecutivi.

L'ovvia constatazione che questo nella realtà concreta non sempre accade, piuttosto che a mettere in discussione l'ipotesi teorica, ci induce a porci un interrogativo: e se in alcuni casi fosse l'intenzione del legislatore di favorire l'*interpretabilità* ad amplificare gli spazi d'*interpretatività* di un testo giuridico?

Cerchiamo di contestualizzare la domanda assumendo come caso esemplificativo uno dei fondamenti della *scrittura controllata*, che compare costantemente nelle raccomandazioni contenute nelle linee guida per la redazione di testi chiari e comprensibili: l'indicazione a un uso quanto più possibile esteso di parole della lingua comune. Il punto di riferimento usualmente assunto per l'identificazione di tali parole all'interno del serbatoio lessicale dell'italiano contemporaneo è quello del *Vocabolario di Base* (VdB) pubblicato per la prima volta da Tullio De Mauro come appendice alla sua *Guida all'uso delle parole* (1980) e riedito nel 2016 in una nuova versione basata su un più aggiornato spoglio e campionamento di testi (*Nuovo Vocabolario di Base: NVdB*).

Come noto, entrambe le versioni del repertorio lessicale comprendono circa settemila vocaboli ripartiti in tre categorie: le prime due, elaborate sulla base dell'analisi statistica di ampi campioni di testi appartenenti a differenti tipologie, sono quelle del *lessico fondamentale* (FO: le circa duemila parole di maggior uso assoluto) e del *lessico di alto uso* (AU: circa duemilasettecento termini ad altissima frequenza); la terza, quella del cosiddetto *lessico di alta disponibilità* (AD), raccoglie invece circa duemilatrecento vocaboli che sulla base di indagini sul campo risultano essere percepiti dai parlanti comuni come molto familiari pur non presentando frequenze statistiche di rilievo.

Il parametro dell'appartenenza o non appartenenza dei lessemi al *vocabolario di base* e alle sue categorie interne è stato assunto nell'ultimo quarantennio come indicatore per la valutazione della leggibilità dei testi in relazione a diversi ambiti applicativi della scrittura controllata e anche attraverso strumenti prodotti nell'ambito della linguistica computazionale⁷. In linea generale, scarsa considerazione è però stata accordata a quei potenziali effetti collaterali della frequenza lessicale da cui lo stesso De Mauro (2009) aveva messo in guardia osservando che "data una lista

⁷ Più recentemente l'Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR di Pisa ha messo a punto uno strumento di valutazione avanzata della leggibilità – READ-IT – che identifica la percentuale di lemmi del vocabolario del testo non appartenenti al Vocabolario di Base e per quelli ad esso riconducibili assume a riferimento le relative classiche tre categorie (Dell'Orletta et al. 2011).

di frequenza o uso delle parole di una qualsiasi lingua, i lessemi più frequenti hanno in media un maggior numero di accezioni, più esattamente al crescere della frequenza secondo potenze cresce secondo la successione naturale il numero delle accezioni". Pare probabile che un così significativo fenomeno di proliferazione semantica possa assumere contorni molto specifici nell'ambito delle lingue speciali, che per loro natura tendono a convertire i termini del lessico comune verso sensi tecnici o specializzati e comunque non corrispondenti a quelli ordinariamente impiegati e noti ai parlanti. Il parametro lessicometrico dell'alta frequenza potrebbe allora in ambito giuridico non garantire di per sé l'accesso ai corretti significati contestuali e anzi alimentare, come in alcune altre varietà dell'italiano (Revelli 2020), ingannevoli rappresentazioni di trasparenza e comprensibilità: il ricorso a termini di solo apparente facilità interpretativa incrementerebbe, invece che arginare, le circostanze della *S.* di un enunciato giuridico.

5. Sottodeterminazione e comprensibilità: l'esempio delle frequenze lessicali

Approfondiamo la questione verificando, in primo luogo, come i livelli di polisemia – misurati sulla base di *MultiWordNet* (MWN nel seguito; Pianta et al. 2002; Toral et al. 2010), un *dump* italiano della risorsa lessicografica WordNet (Miller 1995), – si distribuiscano nelle due versioni del *Vocabolario di base* (versione 1980⁸: VdB; versione 2016⁹: NVdB) all'interno delle tre categorie di frequenza lessicale demauriane.

Come illustrato in tab. 1 e in tab. 2, sia nel VdB che nel NVdB i lessemi etichettati come FO appaiono significativamente più polisemici di quelli marcati AU, a loro volta più polisemici di quelli ad AD. Le evidenze relative al tasso di lemmi polisemici concordano sostanzialmente con le indagini già condotte in precedenza sul livello di polisemia del VdB (Casadei 2014).

Nell'indagine sul livello di polisemia caratterizzante le tre partizioni del VdB, i due indicatori ('% polysem' e 'Syn/Lem' nelle tab. 1 e 2) possono essere posti in relazione: il primo misura la proporzione di elementi polisemici (quanti sono i lemmi polisemici: frequenza relativa), mentre il secondo coglie la consistenza di detta polisemia (quanto in media sono polisemici i lemmi di una data categoria), essendo calcolato come rapporto fra numero di *synset* (insiemi di sinonimi in *WordNet*) e numero di lemmi¹⁰. Le analisi condotte illustrano quindi non solo come all'interno dei lemmi FO siano più frequenti elementi polisemici (rispetto ad AU e AD: para-

⁸ In De Mauro 1980: 149-183.

⁹ www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana.

¹⁰ Va notato come le percentuali di termini polisemici debbano essere integrate da quella di termini privi di *synset* associati in *MWN*. La frazione di termini non presenti nella base lessicografica cresce passando dal vocabolario ad AU, e da questo a quello ad AD. Tale andamento è comune alle due versioni del VdB, ed è compatibile con l'ipotesi che lemmi meno frequenti siano non solo meno presenti all'interno della risorsa lessicografica, ma anche connotati da un minore grado di polisemia.

metro “% polysem”), ma anche come nel passaggio da FO, AU e AD tale polisemia sia caratterizzata da un numero di sensi decrescente (parametro “Syn/Lem”).

Tabella 1 - Livello di polisemia nelle tre sezioni del VdB, misurata come numero medio di *synset* per lemma, percentuale di lemmi non presenti in MWN, percentuale di termini monosemici e polisemici

VdB	N. Lemmi	Syn/Lem	% no <i>synsets</i>	% monosem	% polysem
FO	1855	4.1	5.12%	13.05%	81.83%
AU	2922	2.3	10.40%	30.18%	59.41%
AD	1632	1.7	12.38%	41.12%	46.51%

Tabella 2 - Livello di polisemia nelle tre sezioni del NVdB, misurata come numero medio di *synset* per lemma, percentuale di lemmi non presenti in MWN, percentuale di termini polisemici

NVdB	N. Lemmi	Syn/Lem	% no <i>synsets</i>	% monosem	% polysem
FO	1979	4.1	5.36%	13.14%	81.51%
AU	3011	2.2	9.66%	31.12%	59.22%
AD	2231	1.5	17.71%	43.30%	39.00%

Verifichiamo allora se, come ipotizzato, l'esigenza di circostanziare in modo univoco e vincolante l'interpretazione del testo conduca il legislatore a identificare nel VdB, quindi nel lessico più polisemico, le maggiori insidie della *S*. Ci aspettiamo quindi che il ricorso a termini del VdB sia maggiore nei testi di rango primario, che, come evidenziato in § 4, richiedono una comprensione globale e tollerano o anzi implicano un certo livello di *S*.; che sia invece minimo nei testi regolativi, le cui funzioni illocutive rispondono all'esigenza pragmatica di fornire prescrizioni inequivocabili.

Conduciamo un test preliminare considerando tre documenti paradigmatici dei due poli dell'ideale continuum delle tipologie testuali d'ambito giuridico – la *Costituzione italiana* e il *Decreto Asseverazioni 6 agosto 2020*¹¹ ai due poli opposti, e la *Legge 240/2010*¹² in una posizione centrale, rappresentativa delle tipologie giuridiche miste di rango intermedio.

I dati rappresentati nella tab. 3 consentono di osservare come nei tre testi oggetto di analisi il livello di polisemia si riduca parallelamente alla riduzione della copertura dei termini del VdB. Tale riscontro può essere integrato dall'andamento dei valori nella prima colonna della tabella, che riporta la percentuale di termini non associati ad alcun *synset*. Questi sono meno numerosi nella *Costituzione*, mentre crescono significativamente per il testo del *Decreto*, rappresentativo dei testi di rango secondario. Al di là di questioni connesse alla copertura della risorsa lessicale

¹¹ Si tratta del Decreto del Ministero dello Sviluppo economico 6 agosto 2020 recante “*Requisiti tecnici per l'accesso alle detrazioni fiscali per la riqualificazione energetica degli edifici – cd. Ecobonus*” (GU n. 246 del 05-10-2020, Serie Generale).

¹² L. n. 240 del 30 dicembre 2010: “*Norme in materia di organizzazione delle università, di personale accademico e reclutamento, nonché delega al Governo per incentivare la qualità e l'efficienza del sistema universitario*” (GU n. 10 del 14.1.2011, Suppl. Ordinario n. 11).

utilizzata (Pianta et al. 2002), un simile andamento è compatibile con il ricorso a un lessico progressivamente più specifico passando dalla *Costituzione*, attraverso la *L. 240/2010*, fino al *Decreto*.

Tabella 3 - *Livello di polisemia dei tre documenti assunti a campione: le prime tre colonne illustrano rispettivamente la percentuale di termini non presenti in MWN, quelli associati a un solo synset e quelli polisemici; le colonne successive illustrano invece il numero di termini presenti all'interno delle tre sezioni del VdB e la copertura totale, espressa come somma*

VdB	% no synsets	% monosem	% polysem	cover. FO	cover. AU	cover. AD	Σ cov.
Costituz. ITA	16.37%	16.01%	67.62%	48.37%	17.84%	2.44%	68.65%
L. 240/2010	17.25%	19.39%	63.36%	38.44%	15.73%	1.96%	56.13%
D. 6.08.2020	25.82%	18.83%	55.35%	21.21%	6.88%	1.24%	29.33%

Per quanto riguarda la percentuale dei termini monosemici i valori rilevati non consentono di individuare tendenze intelligibili (per ampiezza e direzione della variazione), mentre il livello di polisemia è decrescente nel passaggio dalla *Costituzione* al *Decreto*. Il livello della correlazione con la copertura dei tre documenti del VdB è complicato dall'interazione di vari fattori, quali la completezza della risorsa utilizzata, ma anche da altri elementi rilevanti seppur di natura implementativa¹³, e richiede ulteriori analisi ed approfondimenti.

Sembra plausibile l'ipotesi che nei testi normativi di rango primario l'ampio ricorso a termini del VdB risponda al contempo a un obiettivo di raggiungibilità di un pubblico ampio e non specialista e a una necessità di *S.* volta a garantire ai contenuti del dettato l'opportuna longevità e flessibilità. Nei testi di rango secondario, un uso parsimonioso del VdB associato al ricorso a tecnicismi sarebbe invece rivolto a evitare una *S.* che, se presente, comporterebbe necessariamente ambiguità applicative.

La ricognizione qui proposta ha sondato unicamente il livello della polisemia di unità lessicali monorematiche con l'obiettivo di verificare preliminarmente e su un ambito circoscritto l'ipotesi che esista una gerarchia dell'ambiguità lessicale parallela alla *S.* rappresentata nel continuum tipologico delle fonti giuridiche. Si ipotizza che la *S.* che caratterizza le diverse categorie di documento si associ tuttavia non soltanto a ulteriori dimensioni semantico-lessicali qui non considerate e a tratti specifici del linguaggio giuridico-normativo, ma anche alle loro interfacce con più profondi livelli della lingua. Con lo scopo di verificare questa ipotesi e di individuare indicatori che possano conclusivamente essere assunti come marche linguistiche identificative della *S.* giuridica, la fase sperimentale del Progetto – che prevede la costituzione ad hoc di un corpus normativo bilanciato da sottoporre a trattamento computazionale e la parallela conduzione di indagini sul campo – seguirà le coordinate sinteticamente qui di seguito descritte.

¹³ Come, per esempio, il filtraggio delle *stopwords*, per cui si è inizialmente utilizzato un insieme di *parole piene* messe a disposizione dalla risorsa NLTK (Loper & Bird 2002), a una prima analisi largamente incompleto e non adattato al contesto applicativo corrente.

6. Per l'individuazione di indicatori linguistici ad hoc

Posto che quella di *linguaggio giuridico* è una “etichetta di estensione variabile, adattata [...] a un universo testuale composito, in cui si riconoscono varietà di lingua concorrenti” (Mortara Garavelli 2001: 7) e che all'interno della letteratura del settore normativo tratti microlinguistici possono assumere funzionalità tecniche e stilistiche differenziate, gli indicatori della *S.* da identificare devono necessariamente tenere conto delle interfacce che possono interagire ai diversi livelli della lingua e contraddistinguere gli specifici contesti pragmatici, oltre che tipologico-testuali.

Per selezionare fra i tratti descritti in letteratura come caratteristici del linguaggio giuridico solo quelli effettivamente significativi in relazione alla *S.* e con lo scopo di farne emergere di ulteriori, eventualmente condivisi con le microlingue di altri ambiti settoriali o con differenti generi ad alta vincolatività, sono state previste due classi di procedure sperimentali.

La prima prevede che a partire da un censimento di leggi e regolamenti prodotti in Valle d'Aosta dal 1950 ai giorni nostri si giunga a definire una tassonomia delle tipologie testuali giuridiche locali secondo parametri riconducibili a una scala gerarchica che permetta di declinare con un soddisfacente livello di dettaglio la tipologia dei generi rappresentati. La selezione di alcune delle tipologie individuate come maggiormente caratteristiche della realtà esaminata sarà poi utilizzata per la costituzione di un corpus testuale qualitativamente e quantitativamente rappresentativo di diverse prassi redazionali. All'interno di tale corpus si procederà alla creazione di *sottocorpora* distinti per la costituzione di concordanze e lessici di frequenza. Test di leggibilità testuale consentiranno di verificare se i luoghi di maggiore criticità identificati secondo gli indicatori canonicamente impiegati in letteratura consentono di rilevare manifestazioni della *S. fisiologica e patologica*, ed in rapporto a quali specificità linguistiche. I documenti saranno poi sottoposti ad analisi computazionali comparative volte a ricavare indici di ricchezza lessicale e densità semantica utili anche per l'interpretazione dell'andamento delle frequenze lessicali e per la parametrizzazione di indicatori relativi ai rapporti fra sinonimia, monosemia e polisemia, iponimia e iperonimia. In questa fase sarà esplorata inoltre la *S.* in analoghi testi in altre lingue con rappresentazioni vettoriali come *word-* e *sense-embeddings* multilinguistici (Colla et al. 2020) e contestuali (Devlin et al. 2018), per verificare l'ipotesi dell'utilizzo graduato della *S.* in corrispondenza con i diversi livelli cui appartengono i documenti giuridici.

La seconda classe di procedure prevede la conduzione di prove sul campo con il coinvolgimento di tre campioni di parlanti: il primo gruppo comprenderà i *produttori* di testi giuridici, ossia addetti ai lavori direttamente coinvolti nelle diverse fasi di stesura delle norme (livello politico); il secondo gruppo sarà costituito da funzionari amministrativi con ruoli di interpreti esecutivi dei contenuti normativi (livello tecnico); il terzo gruppo sarà formato da parlanti comuni.

Alle tre coorti sarà proposta, in tre distinte fasi, la lettura di testi normativi rappresentativi di categorie tipologiche a vincolatività differenziata.

La prima fase prevedrà la lettura a voce alta di un testo proposto sul monitor di un computer: attraverso un'applicazione per l'*eye-tracking* sviluppata *ad hoc* saranno registrati pattern di movimenti oculari e rilevati movimenti anomali, quali il protrarsi di fissazioni su specifici punti, le variazioni di ritmo e estensione dei movimenti saccadici progressivi; la comparsa e l'intensificazione di movimenti saccadici regressivi (ritorni all'indietro orizzontali, da destra verso sinistra, e verticali, dal basso verso l'alto). Ci si attende che l'analisi dei tracciamenti condotta congiuntamente all'osservazione dei modi e tempi di esecuzione del compito di lettura consenta di identificare le variazioni attenzionali associate a difficoltà interpretative e quindi a mettere in relazione le collocazioni rilevate con l'analisi dei risultati dell'analisi dell'*eye tracking* (*S. tracciata*) con le circostanze linguistiche che le caratterizzano.

La seconda fase consisterà in una rilevazione off line della *S. percepita* dai lettori. Tramite domande poste immediatamente dopo la lettura dei testi attraverso un questionario semistrutturato ci si proporrà di verificare da un lato la comprensione di specifici passi e l'interpretazione di proposizioni potenzialmente ambigue; d'altro lato si chiederà agli intervistati di esprimere le proprie opinioni rispetto alla chiarezza complessiva del documento con valutazioni quantitative espresse su scala Likert e domande a risposta aperta. In tal modo si mira a ricavare evidenze comparative sui livelli di comprensione riscontrabili all'interno delle tre categorie di parlanti considerate e al contempo a ottenere informazioni circostanziate sulle loro rappresentazioni della *S. giuridica*.

Nella terza fase ci si proporrà di rilevare le modalità della *S. processata* attraverso operazioni di traduzione di un testo giuridico dall'italiano verso il francese. Per la conduzione della prova sarà utilizzata la tecnica del protocollo del pensiero ad alta voce: si chiederà quindi ai soggetti intervistati di procedere alla trasposizione avvalendosi al bisogno di un dizionario bilingue ed esplicitando di volta in volta dubbi e criticità, in modo che le difficoltà legate a *S.* del testo di partenza possano in tempo reale, durante lo svolgimento del compito, essere rilevate, registrate e distinte da quelle imputabili a ragioni diverse.

A conclusione delle sperimentazioni, la comparazione delle evidenze risultanti dal trattamento computazionale del corpus con gli esiti delle prove sul campo dovrà condurre a identificare gli indicatori di *interpretatività* distinguendoli, ove necessario, da quelli impiegati per la valutazione dell'*interpretabilità*, anche in relazione a singole specificità della variazione semasiologica dei sottocodici giuridici. Sarà quindi predisposto un inventario dei tratti linguistici identificabili come marcatori della *S.*: un'analisi dei contesti di occorrenza si proporrà di classificarli da un punto di vista tipologico – semantico e onomasiologico – per la strutturazione di un sistema di etichettatura rappresentativo della casistica censita.

7. Sviluppi applicativi: sistemi di supporto linguistico-computazionale

La predisposizione di applicazioni informatiche concepite ad hoc consentirà di verificare presenza e distribuzione delle differenti etichette tipologiche della *S.* nei

diversi generi testuali rappresentati nel corpus. La verifica della coerenza dei riscontri, testata anche al di fuori degli ambiti documentari compresi nel campione di partenza, dovrà infine condurre all'introduzione delle modifiche eventualmente necessarie per la validazione conclusiva sia del modello categoriale delle tipologie testuali d'ambito giuridico sia dell'inventario degli indicatori linguistico-giuridici parametrati.

I successivi sviluppi applicativi potranno essere rivolti alla predisposizione di linee guida e alla costruzione di un sistema di supporto a vari compiti, che nelle fasi di *drafting* normativo possa favorire – a seconda delle necessità legate ai diversi contesti testuali – il ricorso consapevole e mirato alla *S.* o il suo opportuno intenzionale evitamento. In dimensione ricettiva, tale sistema dovrà agevolare il riconoscimento delle circostanze della *S.*, favorendone l'interpretazione sia in direzione applicativa, sia nella prospettiva traduttiva, a livello intralinguistico (trasposizione verticale a fini divulgativi) e interlinguistico (*task* della traduzione propriamente intesa).

In quest'ottica ci si auspica, conclusivamente, che i modelli teorici e gli strumenti elaborati attraverso il paradigma della *S.* possano andare a integrare in modo complementare i contributi già nell'ultimo decennio apportati dalle *digital humanities* al generale miglioramento della qualità degli atti normativi¹⁴.

Bibliografia

Berruto, Gaetano. 2003. Una Valle d'Aosta, tante Valli d'Aosta? Considerazioni sulle dimensioni del plurilinguismo in una comunità regionale. In Fondation Emile Chanoux, *Una Valle d'Aosta bilingue in un'Europa plurilingue*, 44-53. Aosta: Tipografia Valdostana.

Casadei, Federica. 2014. La polisemia nel vocabolario di base dell'Italiano. *Lingue e Linguaggi* 12. 35-52.

Colla, Davide & Mensa, Enrico & Daniele P. Radicioni. 2020. LessLex: Linking multilingual Embeddings to SenSe representations of lexical items. *Computational Linguistics*. 46(2). 289-333.

De Mauro, Tullio. 1980. *Guida all'uso delle parole*, 149-183. Roma: Editori Riuniti.

Dell'Orletta, Felice & Montemagni, Simonetta & Venturi, Giulia. 2011. READ-IT: Assessing readability of Italian texts with a view to text simplification. In *SLPAT '11. Proceedings of the second Workshop on Speech and Language Processing for Assistive Technologies*, 73-83. Stroudsburg, PA, USA: Association for Computational Linguistics.

Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint 1810.04805.

¹⁴ In particolare, la famiglia dei prodotti *xmLeges*, sviluppati dall'Istituto di Teoria e Tecnica dell'Informazione Giuridica nell'ambito del Progetto *Norme in Rete* con l'obiettivo di supportare la redazione, revisione e conversione in XML di testi normativi redatti in conformità alle regole di *drafting* legislativo e agli standard di formato previsti per la rappresentazione elettronica dei provvedimenti (Rinaldi 2016).

- Longo, Fabio & Mastropaolo, Antonio & Pallante, Francesco. 2012. Incertezze derivanti dalla ineliminabile, ma non adeguatamente contenuta, oscurità linguistica delle disposizioni normative. In Dogliani, Mario (a cura di), *Il libro delle leggi strapazzato e la sua manutenzione*, 43-49. Torino: Giappichelli.
- Loper, Edward & Bird, Steven. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.
- Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM*. 38(11). 39-41.
- Mortara Garavelli, Bice. 2001. *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*. Torino: Einaudi.
- Ondelli, Stefano. 2007. *La lingua del diritto: proposta di classificazione di una varietà dell'italiano*. Aracne: Roma.
- Pianta, Emanuele & Bentivogli, Luisa & Girardi, Christian. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proceedings of the first International Conference on Global WordNet*, 293-302. Mysore, India.
- Raimondi, Gianmario. 2006. Storia e configurazione del repertorio plurilingue valdostano. In Bertolino, Fabrizio & Revelli, Luisa (a cura di), *Università, scuola, territorio*, 100-126. Milano: Franco Angeli.
- Raso, Tommaso. 2016 (2005). *La scrittura burocratica. La lingua e l'organizzazione del testo*. Roma: Carocci.
- Revelli, Luisa. 2020. Lexicometric paradoxes of frequency: Comparing VoBIS and NVdB. In Iezzi, Domenica & Mayaffre, Damon & Misuraca, Michelangelo (eds.), *Text analytics: Advances and challenges*, 91-99. Cham: Springer.
- Rinaldi, Mariasole. 2016. XmLeges-Editor: un ambiente completo per la redazione e conversione di testi normativi. In Panizza, Saulle (a cura di), *Profili attuali di qualità degli atti normativi e amministrativi*, 247-268. Pisa: University Press.
- Rovere, Giovanni. 2007. L'efficienza comunicativa in ambito giuridico. *Annali del Dipartimento di Scienze Giuridico-Sociali. Università del Molise* 9. 183-208.
- Sabatini, Francesco. 2011a (1^a ed. 1999). "Rigidità-esplicitzza" vs "elasticità-implicitzza": possibili parametri massimi per una tipologia dei testi. In Sabatini, Francesco, *L'italiano nel mondo moderno. Saggi scelti dal 1968 al 2009*, tomo II, 183-216. Napoli: Liguori.
- Sabatini, Francesco. 2011b (1^a ed. 2001). I tipi di testo e la "rigidità" del testo normativo giuridico. In Sabatini, Francesco, *L'italiano nel mondo moderno. Saggi scelti dal 1968 al 2009*, tomo II, 335-346. Napoli: Liguori.
- Toral, Antonio & Bracale, Stefania & Monachini, Monica & Soria, Claudia. 2010. Rejuvenating the Italian WordNet: Upgrading, standarisising, extending. In *Proceedings of the 5th Global WordNet Conference (GWC)*, Bombay, India: ACL.

Un modello per domarli tutti: verso una rappresentazione del testo come esplicitazione di documento, lingua e contenuto²

Abstract

The aim of this research is to describe the first steps towards the theoretical elaboration of a holistic model to represent textual information. The focus of the model is the definition of “text”, with its different dimensions, as a “diasystem”. The set of elements, organized into distinct but strictly interconnected systems, wherein each element has an effect on the whole diasystem, is described in a model structured in the following components: graphic, linguistic, documental, discursive, and conceptual. In this work, the first attempts in the modeling of text will be shown through two case studies: the Babylonian Talmud and the DiTMAO (Dictionary of Old Occitan medico-botanical terminology).

1. Il testo come fonte di informazione. Standard e Vocabolari³

Il presente lavoro si propone di documentare lo stato di avanzamento di una serie di ricerche relative alla modellizzazione del testo, da un punto di vista computazionale, in corso presso l’Istituto di Linguistica Computazionale da parte del costituentesi Knowledge Laboratory (KLab).

“Il testo è tutto il nostro bene” scriveva Cesare Segre in “Ritorno alla critica” (Segre 2001: 99); se questa affermazione conferma un *habitus* proprio delle scienze umanistiche nel loro complesso, tanto più si adatta alle *Digital Humanities* (DH), intese come disciplina, approccio e soprattutto metodo. Sono infatti numerosi i *task*, ormai tradizionali, che vedono nel testo – o meglio, nelle diverse accezioni del testo – il presupposto fondante, tanto come bacino di dati, quanto come veicolo di informazione di varia natura (linguistica, contestuale, concettuale).

Si parla di diverse accezioni dell’oggetto “testo” proprio perché, a seconda dell’ambito, è possibile mutare la lente, la prospettiva di analisi; da mera stringa di

¹ Istituto di Linguistica Computazionale “A. Zampolli” – CNR.

² Il presente lavoro è frutto del lavoro congiunto di entrambi gli autori; in particolare la scrittura dei paragrafi 1 e 4 è di Emiliano Giovannetti, dei paragrafi 2 e 3 di Flavia Sciolette.

³ Si utilizza il termine “vocabolario” nell’accezione comunemente nota nell’ambito del Semantic Web, dunque da intendersi come “vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms”. (<https://www.w3.org/standards/semanticweb/ontology>).

caratteri ai prodotti dell'attività linguistica dei parlanti fino ad arrivare alle edizioni digitali di documenti e opere di diversa complessità. Le definizioni di testo possono dunque essere molteplici, ma da qualunque punto di vista lo si osservi, occorre partire dall'assunto che il testo è un oggetto pluridimensionale; il primo problema pratico nella sua rappresentazione è la necessità, dunque, di inquadrare da un punto di vista teorico e metodologico questa intrinseca e multifattoriale complessità.

L'interesse per il web semantico e i *Linked Open Data* (LOD), paradigmi oramai ampiamente compresi nell'ambito della linguistica computazionale, in particolare per il sottoinsieme dei *Linguistic Linked Open Data* (LLOD, intesi come dati condivisi, aperti e interoperabili per gli scopi della linguistica e del Natural Language Processing, Cimiano et al. 2020), trova ormai da tempo spazio nel campo della codifica dei testi (già in Ciotti 2014: 1-10 e Ciotti & Tomasi 2016, Monella & Del Turco 2020: 148-155 per le DSE – Digital Scholarly Editions). Tale interesse non sembra essere orientato solo in un senso; data la crescente attività di conversione di risorse digitali, anche non native, in formato *LOD-compliant*, il modello *LExicon Model for Ontologies* (lemon), elaborato dal gruppo Lexica⁴ (McCrae et al. 2017) ha di recente pubblicato Lexicog⁵, un modulo rivolto alla conversione di dizionari tradizionali (non focalizzato dunque solo al dizionario-macchina, inteso come insieme di informazioni strutturate, di natura linguistica, elaborabili dalla macchina). Questa attenzione per il testo, inteso come attestazione, si registra inoltre anche grazie all'elaborazione di un ulteriore modulo dedicato a questi aspetti, attualmente in fase di lavorazione sempre nell'ambito del gruppo Lexica (Chiarcos et al. 2020). Nel modello lemon, tuttavia, questa riflessione sul testo appare funzionale – per non dire ancillare – alla rappresentazione del dato lessicale e lessicografico, in un'ottica tuttavia perfettamente coerente tanto con i presupposti del modello, quanto con la sua vocazione agnostica rispetto alla descrizione degli oggetti considerati nel *data-model*. La filosofia soggiacente a lemon, in linea con il paradigma LOD, demanda a vocabolari ontologici esterni la descrizione della lingua e conseguentemente anche del testo.

In questo senso, si segnala la distribuzione CRMt_{ex} (<http://www.cidoc-crm.org/crmtext/home-8>, pubblicata nel 2020), la più recente incarnazione di CIDOC-CRM (Doerr 2003), importante ontologia formale nel campo della conservazione del patrimonio culturale CRMt_{ex} si presenta come un'estensione deputata alla descrizione dei testi scritti antichi, intesi come entità comprendenti un insieme di glifi su un determinato supporto, dotati di scopo e in grado di veicolare un contenuto. Il modello introduce classi e proprietà indirizzati nello specifico agli esperti di discipline quali codicologia, epigrafia e paleografia. L'attenzione è pertanto rivolta al testo in quanto oggetto materiale e il vocabolario introdotto mira a fornire le etichette per una descrizione rigorosa, utile anche in termini di scambio di dati e interope-

⁴ <https://www.w3.org/community/ontolex/> Il gruppo fa parte della community W3C (World Wide Web Consortium), secondo il "W3C Community Final Specification Agreement (FSA). Di conseguenza il modello non è da considerarsi uno standard W3C.

⁵ <https://www.w3.org/2019/09/lexicog/>.

rabilità su risorse diverse (si veda per esempio Bellotto 2020, per un'applicazione nell'ambito di manoscritti medievali, con analisi anche dei vantaggi dell'adozione del vocabolario unitamente a TEI per la risoluzione di alcune questioni teoriche e di metodo relative alla codifica).

Il panorama, sebbene fluido, mostra tuttavia approcci di modellizzazione del testo con un focus specifico, improntati al *task* piuttosto che al suo carattere n-dimensionale, ovvero della sua specifica complessità, in quanto insieme di documento, lingua e contenuto. Il presente lavoro tenta dunque di fornire una possibile risposta teorica in corso di elaborazione, fondata sia dal punto di vista semiotico che computazionale. Il punto di partenza sono i diversi casi d'uso forniti dall'attività svolta durante il progetto di traduzione del Talmud Babilonese in italiano e il progetto DiTMAO (Dictionary of Old Occitan medico-botanical terminology).

2. *Un modello per l'oggetto testo*

2.1 L'esplicitazione dell'informazione

La definizione rigorosa e formale di un oggetto e delle sue diverse componenti, delle relazioni che intercorrono tra di esse e delle proprietà che le caratterizzano, nonché dei diversi processi in cui possono essere coinvolte, è *conditio sine qua non* per l'applicazione di un approccio computazionale. Ogni operazione in cui è coinvolto un testo – sia essa la lettura, la scrittura o la traduzione – vede il coinvolgimento di un interprete e, conseguentemente, del suo punto di vista. In particolare, nel processo di traduzione, un interprete (ed esperto di dominio) utilizza l'informazione contenuta nel testo per la produzione di un nuovo oggetto (la traduzione). Questa informazione impiegata per strutturare il nuovo oggetto testo, tuttavia, non è rappresentata solo da quanto potrebbe essere veicolato esplicitamente dal testo – ovvero tutta l'informazione disponibile nel testo in quanto stringa di caratteri – ma comprende anche l'informazione implicita, ovvero quella che l'interprete umano è in grado di decodificare per mezzo della propria enciclopedia mentale (per esempio il significato di determinate espressioni idiomatiche).

Gli strumenti elaborati dal nostro gruppo di ricerca sono pensati per supportare l'attività dell'interprete umano grazie ai dati estratti: statistiche linguistiche, allineamenti di parola, occorrenze, frequenze, concordanze. Secondo un approccio votato alla collaboratività tra interpreti umani, sono stati elaborati strumenti volti alla traduzione assistita (come Traduco, sviluppato nell'ambito del progetto di traduzione del Talmud, Giovannetti et al. 2016) e alla creazione di risorse terminologiche e lessicali (come LexO, editor di lessici computazionali basato sul modello lemon, Bellandi 2021). Gli strumenti rappresentano uno dei quattro pilastri⁶ su cui

⁶ Gli altri tre pilastri sono: i) gli algoritmi dedicati al trattamento automatico della lingua, per esempio per l'estrazione della conoscenza e della terminologia (Giovannetti et al. 2020); ii) le risorse linguistiche (come nel caso del lessico computazionale dell'italiano Parole-Simple-Clips, il cui aggiornamento è in corso da parte del KLab al momento della redazione del presente contributo); iii) il modello volto

si basano le ricerche condotte dal KLab nell'ambito della modellazione, della rappresentazione e del trattamento del testo.

Secondo quanto appena esposto, i dati estraibili per mezzo degli strumenti non rappresentano tuttavia tutta l'informazione veicolata dal medium di trasmissione del testo. Ogni oggetto complesso, difatti, porta dell'informazione implicita, costituita da dati che non compaiono come segni espliciti – tali da poter essere individuati dagli strumenti e dunque estratti direttamente – ma che sono tuttavia presenti in ogni processo legato al testo (di cui la traduzione rappresenta probabilmente uno degli esempi più complessi). In altre parole, per poter sfruttare tutta l'informazione contenuta in un testo, in prima istanza, abbiamo bisogno di “esplicitarla”. Necessitiamo quindi di poter modellare e rendere ogni elemento del testo un oggetto digitale esplicito, dotato di un proprio identificatore univoco (come *Uniform Resource Identifier*, o URI), delle proprietà caratterizzanti e delle relazioni con altri oggetti: chiamiamo questo processo “esplicitazione computazionale dell'informazione testuale”. La seconda parte dell'articolo sarà dedicata all'esposizione di alcuni esempi in cui mostreremo questo processo nella pratica.

Una volta dati i presupposti per avere disponibile questa informazione, è necessario altresì disporre di un modello nel quale essa possa essere organizzata. Questo modello olistico deve dunque essere focalizzato sul testo come contenuto, sull'uso della lingua al suo interno (quindi non su una sua formalizzazione) e soprattutto deve essere sufficientemente espressivo da permettere l'utilizzo di standard esistenti, come quelli sopramenzionati. Di conseguenza, il nostro obiettivo non è, di per sé, la creazione di un nuovo *data-model* o un tentativo di sostituire standard esistenti, quanto piuttosto fornire: 1) una prospettiva di visione di insieme del dato testuale in ottica computazionale; 2) un vocabolario per le interazioni tra componenti che solitamente vengono considerate separatamente, utile per *task* che richiedono di elaborare informazione di diversa natura; 3) uno schema per agevolare la conversione e l'interoperabilità tra risorse basate su modelli diversi, aventi il testo come oggetto di interesse. Da qui il titolo del presente contributo, volutamente iperbolico, da leggere come obiettivo ideale per la realizzazione di un modello del testo in grado di fornire una descrizione dei suoi diversi aspetti, utile per specialisti di diversi ambiti.

In particolare, per il linguista computazionale e l'informatico della lingua questo modello intende agevolare la sperimentazione di elaborazioni della lingua (in analisi e in generazione) attraverso approcci in grado di integrare conoscenza linguistica e conoscenza del mondo e, inoltre, creare risorse linguistiche ricche e articolate (lessici, terminologie, corpora annotati, corpora allineati, ecc.), con strumenti e metodologie utili all'arricchimento delle risorse stesse, attraverso un modello rigoroso atto all'elaborazione di informazioni appartenenti a risorse di natura diversa. Per lo studioso del testo e della lingua, si intende fornire un modello (affiancato da un set di strumenti e algoritmi per utilizzarlo) per la creazione di risorse linguistiche e corpora annotati che offrano funzionalità di interrogazione avanzata, una formaliz-

alla rappresentazione dell'informazione testuale, oggetto di questo articolo e ultimo menzionato, ma in realtà fondamento di tutta l'architettura delineata.

zazione dell'informazione puntuale e rigorosa che sia nativamente condivisibile con la propria comunità di riferimento e, infine, la possibilità di arricchire collaborativamente le risorse già esistenti. Oltre all'estrazione di dati, il processo di esplicitazione dell'informazione si pone come ulteriore mezzo per lo studio del testo, attraverso possibilità di interrogazioni complesse per il corpus.

Preme precisare in questa sede che il modello non nasce per sostituire altri modelli già esistenti, focalizzati su specifici sistemi, né si propone come un'ontologia del testo. La proposta ambisce soprattutto a fornire una sovrastruttura per l'organizzazione generale di tutta l'informazione contenuta in un testo, a prescindere dalle tecniche utilizzate per estrarla in modo automatico o semiautomatico.

La necessità di un tale meta-modello nasce dagli esperimenti condotti su testi particolarmente complessi, spesso scritti in varietà con poca documentazione o non-standard o legate a specifici domini; queste tipologie testuali spesso risultano difficili da analizzare secondo approcci stocastici, in quanto possono essere scarsamente rappresentate nei corpora di addestramento. In linea generale l'obiettivo tuttavia non è, anche in questo caso, porsi in diretta antitesi con differenti approcci, quanto piuttosto affiancarsi a questi ultimi.

2.2 Il modello: diasistema, sistema e dimensioni

Il modello che qui si intende presentare, da un punto di vista teorico, si avvale delle innovazioni metodologiche proposte nei lavori di Tito Orlandi (formalizzate in Orlandi 2010) relativamente all'edizione critica digitale; sebbene Orlandi parta dai manoscritti, gli assunti di base che propone sono applicabili all'oggetto testuale nelle sue diverse accezioni. Secondo Orlandi, il testo è un "diasistema", inteso come prodotto, oggetto complesso risultante dall'interazione dialettica di più sistemi complessi (quindi sistema a sua volta), riprendendo una terminologia utilizzata già dal sopra menzionato Cesare Segre (ma la nozione di "sistema" e di "sistemi" in cui sono inseriti i testi ha una tradizione di lungo corso negli studi umanistici), in cui interagiscono diversi sottosistemi, tra cui quello linguistico (composto dalle parole grafiche, concepite come entità a sé stanti) e quello grafematico (di cui le unità sono i diversi grafemi aventi valore distintivo). Ognuno degli elementi considerato in questi sistemi interagisce con l'altro.

Parliamo di "sistema", infatti, in presenza di due o più elementi che soddisfino le seguenti condizioni: 1) ogni elemento del sistema ha un effetto sul comportamento del sistema nella sua interezza; 2) gli elementi sono interdipendenti tra di loro in quanto legati da relazioni; 3) gli elementi di un sottoinsieme hanno un effetto sul sistema nella sua interezza (Ackoff 1971; Skyttner 1996). La metafora del diasistema, inoltre, risulta particolarmente produttiva anche in un ambito di analisi computazionale del testo, in quanto ci consente di scomporre l'informazione testuale in moduli (i sistemi, per l'appunto) secondo elementi omogenei in relazione, senza concepirne un'obbligatoria gerarchizzazione (e quindi riuscendo a evitare il ricorso a nozioni come "livello" o "strato" del testo). In questo modo possiamo descrivere tutti gli elementi semanticamente e semioticamente fondanti di un testo.

Gli elementi del testo possono essere raggruppati tra loro in modo omogeneo in base alle loro caratteristiche e relazioni; per esempio, possiamo raggruppare tutti gli elementi di lingua; tutti i concetti espressi all'interno di un testo; tutte le caratteristiche materiali del veicolo di trasmissione, e costituire quindi un sistema (linguistico, concettuale o documentale). Allo stesso modo, possono essere messi in relazione elementi appartenenti a sistemi diversi, per esempio nel caso di termini da legare ai rispettivi concetti, nel corrispondente sistema.

In un modello del testo, secondo l'impostazione teorica orlandiana, è possibile distinguere diversi sistemi; per le esigenze delle nostre ricerche ne abbiamo, al momento, individuati cinque: 1) il sistema linguistico, ovvero il sistema riferito alla lingua, in quanto codice e mezzo di espressione dei contenuti del testo, nonché realizzazione di una varietà propria dell'agente che ha prodotto il testo; 2) il sistema grafico, ovvero il sistema riferito al *vehiculum*, inteso come l'insieme dei segni che compongono l'oggetto testo; 3) il sistema documentale, ovvero la rappresentazione del testo in quanto documento, comprendente le sue caratteristiche materiali e le informazioni codificabili come metadati; 4) il sistema discorsivo, riferito alle regole legate alla tradizione discorsiva del testo, alle regole di pragmatica e all'insieme delle caratteristiche non riconducibili a una grammatica esplicita; 5) il sistema concettuale, ovvero l'espressione della conoscenza del mondo espressa dal testo e dall'agente che lo ha prodotto.

I sistemi consentono di ordinare l'informazione testuale e di "spacchettarla" in diverse tipologie, che possono essere messe in relazione tra loro.

Ogni sistema, a sua volta, può essere diviso in dimensioni: definiamo "dimensione" un determinato aspetto del sistema entro il quale trovano spazio elementi omogenei, appartenenti a un insieme individuato secondo un determinato punto di vista, che chiameremo "prospettiva". Alcune delle dimensioni di un certo sistema devono essere considerate "raccomandate", se si tiene conto della natura del sistema a cui appartengono (per esempio la dimensione sintattica o quella morfologica nel sistema linguistico), tuttavia altre possono dipendere dalla teoria di riferimento o dal *task* prefissato.

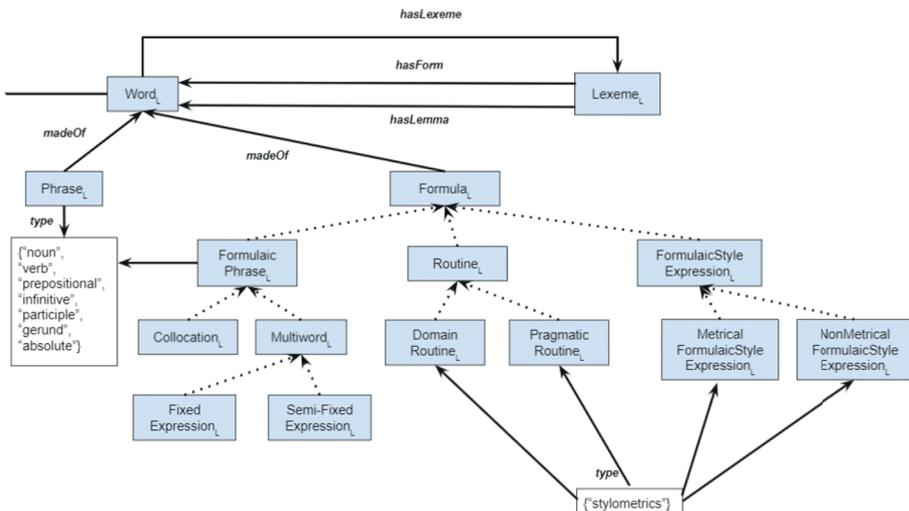
Gli elementi contenuti in una data dimensione costituiscono le unità di analisi e veicolano l'informazione da esplicitare, relativa a una specifica dimensione; la dimensione terminologica, ad esempio, si riferisce ai termini di un determinato dominio.

La nozione di dimensione, come sopra detto, non presuppone un ordinamento gerarchico, nonostante due o più dimensioni possano essere legate gerarchicamente da un punto di vista procedurale (per esempio, si consideri la dimensione sintattica di una lingua flessiva, che richiede comunque una relazione con la dimensione morfologica, perché altrimenti, da un punto di vista procedurale, non sarebbe possibile identificare l'organizzazione delle sue strutture sintattiche. L'organizzazione non strettamente gerarchica del modello, tuttavia, permette la descrizione anche di casi diametralmente opposti, come per le lingue isolanti). I legami possono sussistere tra elementi di dimensioni dello stesso sistema – come nel caso dei sintagmi, che

devono necessariamente considerare elementi di lessico, morfologici e sintattici – e tra dimensioni di diversi sistemi, nel momento in cui, per esempio, si formalizza la relazione tra un termine e un concetto (definendo così una relazione tra il sistema linguistico e il sistema concettuale) oppure la relazione tra una determinata struttura sintattica e il suo utilizzo in specifici generi testuali (relazione tra sistema linguistico e sistema discorsivo).

Di seguito si presenta un esempio di formalizzazione di alcuni elementi del modello. Abbiamo scelto di rappresentare il concetto di “Formula”, inteso come componente del *formulaic language*, secondo l’accezione utilizzata in Wray (2013): “formulaic language’ refers to sequences of words that are in some regard not entirely predictable, whether on account of a meaning that is wildly or subtly different from the words they contain, a function that is only achieved with the whole expression, or features of structure such as morphology or word order that are non-canonical.” La figura 1 mostra le relazioni tra gli elementi in tassonomia e le proprietà di ciascuna classe. Per spiegare l’esempio, si prenda come punto di partenza il concetto di Formula, inteso come classe delle formule in una Lingua (L). Le frecce tratteggiate descrivono le sottoclassi della classe Formula: i) Formulaic Phrase, ii) Routine, iii) Formulaic expression, a loro volta distinte in ulteriori sottoclassi. Le frecce non tratteggiate istituiscono le relazioni tra classi (come nel caso della relazione tra la classe delle formule e la classe delle parole) e tra classi e istanze (come nel caso della classe di Phrase, costituita dalle istanze con attributo *type* e i valori indicati in tabella). In questo caso la classe Formula nel modello è legata alla classe Word dalla relazione *madeOf*.

Figura 1 - Rappresentazione della classe “Formula” nel sistema linguistico del modello.
Le frecce tratteggiate indicano la relazione di “isA”.



Si tratta di un esempio di dimensioni contenute e di cui non sono ancora disponibili tutte le singole ramificazioni, utile tuttavia per chiarire la struttura che si intende dare all'informazione testuale.

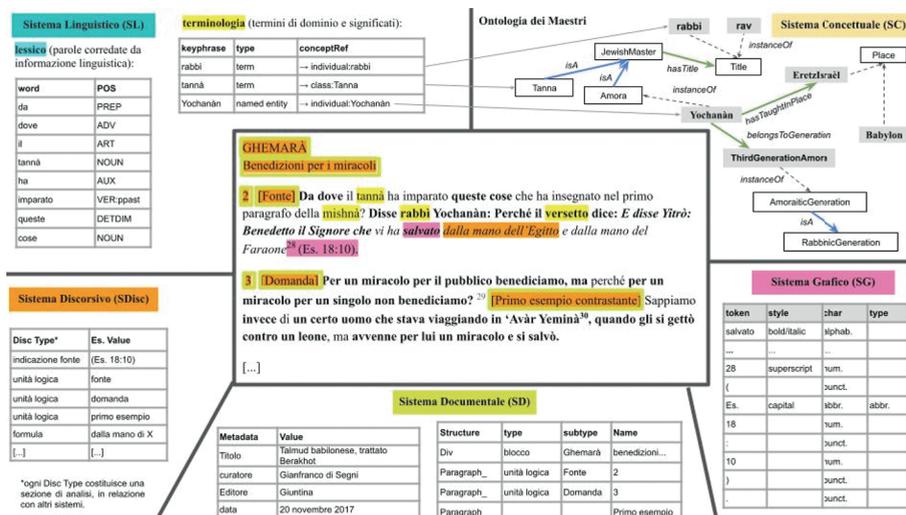
Nel paragrafo successivo mostreremo alcuni casi d'uso di applicazione del modello, utili per fornire un primo inquadramento dei possibili dati con cui è possibile popolare risorse basate su di esso. I casi d'uso illustrati sono stati presi da due progetti che hanno coinvolto il trattamento di risorse testuali e linguistiche e sulla base dei quali sono state inizialmente concepite alcune parti del modello stesso.

3. Casi d'uso

3.1 La traduzione del Talmud Babilonese

Il primo caso d'uso illustrato tratta l'esplicitazione e la strutturazione, nei cinque sistemi del modello, dell'informazione contenuta nel testo della traduzione italiana del Talmud babilonese. Ogni elemento da esplicitare appare annotato nel testo e, allo stesso tempo, strutturato nel sistema e nella dimensione opportuna. L'annotazione del testo, nel nostro paradigma di esplicitazione, non viene effettuata *inline* – ovvero come tipologia di *mark-up* in cui gli elementi sono disposti nella stringa di testo – ma ogni elemento viene annotato mediante riferimenti localizzati esternamente al testo (*stand-off markup*) che identificano, con opportuni indici, l'inizio e la fine dell'annotazione nel testo. Nell'illustrazione di figura 2, ogni elemento esplicitato viene collocato in una tabella posta in una sezione relativa a uno specifico sistema.

Figura 2 - Esempio di esplicitazione e rappresentazione dell'informazione testuale nel caso d'uso della traduzione del Talmud



Nel primo riquadro in alto a sinistra è rappresentato il sistema linguistico, strutturato attraverso una tabella che raccoglie gli elementi della dimensione lessicale (le

parole corredate da informazione linguistica) e della dimensione terminologica (i termini di dominio⁷, le entità nominate e i rispettivi sensi). Ricordiamo che il sistema linguistico può comprendere anche aspetti di morfologia, sintassi, ecc., da formalizzarsi mediante apposite dimensioni.

La tabella del lessico mostra, a sinistra, la colonna *word* dedicata alle forme di superficie delle parole, a destra, la relativa parte del discorso (PoS), mentre la terminologia è organizzata in *keyphrase* (la forma canonica di superficie), *type* (l'appartenenza a una determinata classe, come "term" o "named entity") e *conceptRef* (la relazione tra la parola e la relativa entità concettuale).

La parola *tannà*, per esempio, può essere dunque annotata ed esplicitata, secondo una differente prospettiva: come *word* (elemento del lessico) o *term* (termine di dominio); in questo secondo caso, il termine "tannà" viene associato, tramite *conceptRef*, alla classe ontologica dei Tannà (saggi rabbini del periodo mishnaico), che, nel sistema concettuale, è codificata come sottoclasse di *JewishMaster*. La parte ontologica, con le classi, gli individui e le relazioni che descrivono la natura del mondo riferito dal testo in esame, viene descritta e formalizzata nel sistema concettuale. I termini sono espressi dunque come lessicalizzazioni, collocate nel testo, di individui o classi, le cui singole proprietà concettuali sono descritte nel riquadro in alto a destra.

Nel riquadro in basso, sempre a destra, sono rappresentati gli elementi del sistema grafico; sebbene in un testo a stampa queste informazioni potrebbero sembrare più contenute o meno rilevanti, è comunque possibile catalogare e immagazzinare informazione, per esempio, relativa al font e allo stile del carattere. Difatti, nell'ambito del progetto Talmud, il grassetto (*Bold*, nel campo "Style" nel sistema grafico) ha un valore distintivo, in quanto indica le porzioni di testo tradotte letteralmente dal Talmud. Per poter sfruttare questo tipo di informazioni, a ogni token, come si può notare nella prima colonna disposta in alto a sinistra, vengono associate informazioni relative a carattere, stile e tipologia, in maniera tale da avere una rappresentazione formale anche di segni di interpunzione, cifre e abbreviazioni.

La sezione in basso al centro è dedicata al sistema documentale; comprende i metadati del testo in esame, organizzati nella tabella di sinistra, e le informazioni pertinenti alla struttura del testo, secondo una divisione per unità logiche. Ogni valore è definito attraverso l'associazione con *type*, *subtype* e *name*.

La sezione in basso a sinistra dell'immagine rappresenta il sistema discorsivo; all'interno di questa sezione si annotano i valori che, nel sistema documentale, corrispondono a delle unità di struttura del testo; queste ultime, tuttavia, possono essere associate anche alla classe *DiscType*, in quanto rappresentano una caratteristica del genere. La divisione in unità logiche del testo talmudico difatti non risponde solo a criteri editoriali di struttura materiale della pagina, ma le componenti costituiscono vere e proprie divisioni, strutturate secondo regole precise, legate all'andamento

⁷ Il campione di 4166 termini di dominio utilizzato per i primi test di modello è stato estratto grazie al software T2K (Dell'Orletta et al. 2014); la procedura nel dettaglio è descritta in Giovannetti et al. 2020.

del discorso nell'ambito delle discussioni rabbiniche. Lo schema "Fonte-Domanda-Esempio" viene scomposto nei suoi elementi fondamentali, che costituiscono valori nella tabella del sistema discorsivo. Non riguarda dunque solo la struttura del testo come fisicamente disposto nell'impaginazione, ma risponde a regole precise del genere testuale, in quanto pattern proprio del testo talmudico. Le annotazioni sovrapposte al testo sottolineano la duplice natura informativa del segmento di attestazione scelto.

La selezione dei casi d'uso mira a fornire una panoramica delle possibilità di utilizzo del modello. Di fronte alla ricchezza dell'esempio fornito dal testo talmudico, potrebbe infatti essere lecito chiedersi se l'espressività del modello debba trovare sua giustificazione solo in un utilizzo completo di tutta l'architettura dei sistemi.

L'esempio che segue, pertanto, descrive una situazione che non richiede l'uso di tutti i sistemi, mostrando quanto il modello sia flessibile e adattabile a diversi contesti di studio.

3.2 Un'entrata nel DiTMAO

Nel caso del DiTMAO, il dizionario di antico occitano⁸ dei termini medico-botanici, si considera come esempio il termine *assafetida* e la sua attestazione, contenuta nel manoscritto Vat. Ebr. 550, alla carta 115r. Il manoscritto contiene il testo *Tractatulus de Pestilentia*, che ha una sua edizione di riferimento, ma l'occorrenza di *assafetida* si registra nel successivo glossario non edito. Di conseguenza l'attestazione è data solo dallo spoglio manuale del manoscritto, disponibile in formato immagine⁹ non ricercabile. Si è scelto di presentare questo esempio proprio per la natura del supporto e per mostrare come il modello si adatti a un approccio multilingue, utile nell'ottica di risorse lessicografiche.

A differenza di quanto è stato fatto nell'esempio del testo talmudico, l'annotazione del testo non può essere effettuata specificando un carattere di inizio e di fine all'interno di una sequenza testuale, ma deve implementarsi attraverso la definizione di una porzione di immagine relativa al manoscritto. In questo caso specifico, la carta 115r si presenta divisa in tre colonne: la prima colonna identifica il lemma; la seconda comprende la formula – struttura linguistica introdotta in 2.1 – che collega il lemma alla glossa, presente nella terza colonna. Questa struttura del folio "a colonne" si riflette nel sistema documentale, descritto nelle tabelle che seguono. Le tre colonne vengono descritte nei campi *structure*, *type* e *subtype* (tabella 1), i cui valori comprendono la struttura materiale del testo (in questo caso *column*, le colonne in cui è divisa la pagina) la tipologia testuale in cui sono inserite (*List*) e infine il *subtype*, che indica l'elemento presente ("word", "formula", "gloss"). La *word* descrive la forma di superficie dell'entrata descritta nel dizionario, mentre "formula" registra che l'occorrenza compare sempre preceduta da "esso è" (nel testo "איה", tabella 2). La formula compare nella struttura del testo, in quanto colonna, e al tempo stesso è compresa

⁸ Per la composizione del corpus: <https://www.uni-goettingen.de/en/the+texts/487591.html>.

⁹ https://digi.vatlib.it/view/MSS_Vat.ebr.550/0233.

nei valori di *DiscType* in quanto costituisce una specifica caratteristica della tipologia testuale del glossario. Nel sistema grafico vengono aggiunti i dati relativi alla *scripta*, alla trascrizione in caratteri ebraici e alla traslitterazione del token.

Tabella 1 - *Intabellamento dei valori nel sistema documentale.*

<i>Structure</i>	<i>type</i>	<i>subtype</i>
Column_1	_list	word
Column_2	_list	formula
Column_3	_list	gloss

Tabella 2 - *Intabellamento dei valori per il DiscType nel sistema discorsivo*

<i>DiscType</i>	<i>value</i>
word_source	list
word_gloss	list
formula_gloss	“esso è”

La relazione tra termine e classe si esprime attraverso la *conceptRef*, in quanto istanza di una classe dell’ontologia “Substance”; si sceglie di rimanere nella dimensione terminologica, senza ulteriori collegamenti al sistema concettuale. In questo modo il modello consente la modellazione dell’entrata lessicale secondo il vocabolario *lemon*, ma consente anche di “spacchettare” l’informazione relativa a una delle attestazioni, dividerla in singole componenti che possono essere messe in relazione con l’entrata stessa e utilizzate in diversi processi. Esempi tipici sono l’*optical character recognition* (OCR) o il *pattern-disambiguation*, che possono in questo modo avvalersi di informazione relativa sia al contesto della parola e alla sua posizione in uno schema fondato – secondo la tipologia testuale – sia al significato disambiguato della parola precedente.

4. *A che punto siamo? Prospettive future*

In questo intervento abbiamo voluto rappresentare i primi passi volti alla realizzazione di un modello olistico per la rappresentazione computazionale del testo e la strutturazione di informazione testuale eterogenea.

Abbiamo passato in rassegna i principali vocabolari e standard attualmente esistenti per la rappresentazione del testo in formato digitale per differenti scopi; successivamente abbiamo introdotto la nozione di processo di esplicitazione computazionale dell’informazione testuale, da ordinare secondo un modello organizzato in sistemi e dimensioni. Infine abbiamo illustrato due casi d’uso provenienti da contesti molto differenti tra loro (la traduzione di un testo complesso e multilingua come il Talmud e la costruzione di una terminologia per l’occitano medievale) per dimostrare le differenti possibilità di utilizzo del modello.

Trattandosi di una ricerca tuttora in corso, alcuni aspetti progettuali e architeturali relativi alla definizione di specifiche componenti del modello e delle possibili implementazioni delle risorse su di esso basate devono ancora essere chiariti. Da un punto di vista più teorico, e coerentemente alla nostra volontà di i) riutilizzare, laddove possibile, modelli locali (lessicali, ontologici, ecc.) già disponibili e, allo stesso tempo, ii) aderire alle buone pratiche del paradigma dei Linked Data, intendiamo innanzitutto consolidare la definizione dei sistemi linguistico e concettuale, il primo da basarsi sul già citato lemon, e il secondo sul linguaggio OWL (Web Ontology Language), con interventi di adattamento qualora necessari. Inoltre, riteniamo prioritario definire formalmente le modalità di collegamento di ognuno degli elementi presenti nella risorsa alle relative porzioni di testo in cui essi sono stati esplicitati; ciò potrebbe essere possibile attraverso l'adozione del modulo di lemon dedicato alle attestazioni (non appena sarà reso disponibile) oppure, più in generale, collegando gli elementi esplicitati, tramite i rispettivi URI, alle annotazioni definite sul testo, mediante lo stand-off markup.

Le prime risorse che saranno realizzate sulla base del modello presentato (e pubblicate come LLOD) afferiranno al dominio religioso, e verranno allestite attingendo ai testi e ai dati trattati nell'ambito del citato Progetto Traduzione del Talmud Babilonese e del progetto PRIN 2017 "Representing religious diversity in Europe: past and present features". Per quanto riguarda le applicazioni pratiche del modello, si stanno conducendo primi esperimenti nell'ambito di task di full text search, condotto sul caso studio costituito dal testo del Talmud. Un modello fortemente orientato al testo e volto all'organizzazione di informazioni di varia natura, si rivela particolarmente vantaggioso su testi complessi, che necessitano di un sistema di interrogazione sofisticato e che combini diverse caratteristiche (semantiche, morfologiche, ontologiche, ecc.ecc.) al fine di recuperare informazioni a grana molto fine.

Bibliografia

- Ackoff, Russel. 1971. Towards a system of systems concepts. *Management Sciences* 17(11). 661-671.
- Bellandi, Andrea. 2021. LexO: An Open-source System for Managing OntoLex-Lemon Resources. *Language Resources & Evaluation*, 55, 1093-1126. <https://doi.org/10.1007/s10579-021-09546-4>.
- Bellandi, Andrea & Giovannetti, Emiliano & Weingart, Anja. 2018. Multilingual and Multiword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon. *Information* 9(3). doi:10.3390/info9030052.
- Bellotto, Anna. 2020. Medieval manuscript descriptions and the Semantic Web: analysing the impact of CIDOC CRM on Italian codicological-paleographical data. *DHQ: Digital Humanities Quarterly* 14(1). digitalhumanities.org/dhq/vol/14/1/000449/000449.html.
- Chiarcos, Christian & Ionov, Maxim & de Does, Jesse & Depuydt, Katrien & Khan, Anas Fahad & Stolk, Sander & Declerck, Thierry & McCrae, John P. 2020. Modelling Frequency and Attestations for OntoLex-Lemon. In *Proceedings of the Globalex Workshop on Linked*

Lexicography, Language Resources and Evaluation Conference (LREC 2020), Marseille, 11-16 May 2020. 1-9.

Cimiano, Philip & Chiarcos, Christian & McCrae, John P. & Gracia, Jorge. 2020. Linguistic linked open data cloud. In Cimiano, Philip (ed.) *Linguistic Linked Data*, 29-41. Cham: Springer.

Ciotti, Fabio. 2014. Tematologia e metodi digitali: dal markup alle ontologie. In Alfonzetti, Beatrice & Baldassarri, Guido & Tomasi, Franco (a cura di) *I cantieri dell'italianistica. Ricerca, didattica e organizzazione agli inizi del XXI secolo. Atti del XVII congresso dell'ADI – Associazione degli Italianisti (Roma Sapienza, 18-21 settembre 2013)*, 1-10. Roma: Adi editore.

Ciotti, Fabio & Tomasi, Francesca. 2016. Formal ontologies, Linked Data and TEI Semantics. *Journal of the Text Encoding Initiative* 9. 1-23. <https://doi.org/10.4000/jtei.1480>.

Dell'Orletta Felice & Venturi Giulia & Cimino Andrea & Montemagni Simonetta. 2014. T2K²: a System for Automatically Extracting and Organizing Knowledge from Texts. *Proceedings of the 9th Edition of International Conference on Language Resources and Evaluation (LREC 2014)*, 26-31 May, Reykjavik, Iceland, 2062-2070.

Doerr, Martin. 2003. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24(3). 75-92. doi:10.1609/aimag.v24i3.1720.

Giovannetti, Emiliano & Albanesi, Davide & Bellandi, Andrea & Benotto, Giulia. 2016. Traduco: A collaborative web-based CAT environment for the interpretation and translation of texts. *Digital Scholarship in the Humanities*, I47-i62. doi:10.1093/llc/fqw054.

Giovannetti, Emiliano & Bellandi, Andrea & Dattilo, David & Del Grosso, Angelo Maria & Marchi, Simone & Pecchioli, Alessandra & Piccini, Silvia. 2020. The Terminology of the Babylonian Talmud: Extraction, Representation and Use in the Context of Computational Linguistics. *Materia Giudaica* XXV. 61-74.

McCrae, John P. & Bosque-Gil, Julia & Gracia, Jorge & Buitelaar, Paul & Cimiano, Philipp. 2017. *The OntoLex-Lemon Model: development and applications*, <http://john.mccrae.ac/papers/mccrae2017ontolex.pdf>.

Monella, Paolo & Rosselli Del Turco, Roberto. 2020. Extending the DSE: LOD support and TEI/IIIF integration in EVT. In Marras, Cristina & Passarotti, Marco & Franzini, Greta & Litta, Eleonora (a cura di) *Atti del IX Convegno annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*. 148-155.

Orlandi, Tito. 2010. *Informatica testuale. teoria e prassi*. Roma: Laterza.

Segre, Cesare. 2001. *Ritorno alla critica*. Torino: Einaudi.

Skyttner, Lars. 2001. *General systems theory*, Singapore: World Scientific Publishing Co.

Wray, Alison. 2013. Formulaic Language. *Language Teaching* 46(3). 316-334. doi: 10.1017/S0261444813000013.

LORENZO SPREAFICO¹

Fare linguistica applicata con le *digital humanities*: interfaccia di metodo o di comodo uso?

Abstract

In this paper the approach to digital humanities taken by the authors of the papers in the volume is discussed. It emerges that two possible modalities of doing applied linguistics with digital humanities coexist. The first one focuses on the use of IT tools in dealing with linguistic problems. The second one deals instead with the digital approach to the construction of data, although also in this second case the main motivation seems to be that of verifying the usability of certain tools to make it easier to solve practical problems such as those of transcribing spoken language automatically. In light of this, it is discussed whether the adoption of digital humanities in applied linguistics really changes its paradigm, contributing to characterize it as distinct from that of other forms of linguistic inquiry that make use of computational tools.

1. *Introduzione*

La prefazione all'undicesimo volume della collana Studi AItLA si chiudeva con un rimando a quel che secondo le curatrici la "linguistica applicata è chiamata a fare: una operazione di scambio e di sintesi tra speculazione teorica, ricerca e conseguente proposta di azioni e interventi concreti" (Marra & Dal Negro 2020: 10). La postfazione a questo quattordicesimo volume della collana si apre con una domanda, che quel rimando riprende: si dà uno scambio tra linguistica applicata e *digital humanities*? Ovvero, per parafrasare il titolo del volume in cui questo contributo è inserito, cosa significa fare linguistica applicata con le *digital humanities* in Italia oggi? Si cercherà di rispondere alla domanda sia riflettendo brevemente sulla relazione tra le due discipline (§ 2), sia analizzando le scelte fatte dagli autori dei contributi raccolti in questo volume (§ 3).

2. *Le due discipline*

La relazione tra linguistica applicata e *digital humanities* può essere indagata riprendendo brevemente un tema ricorrente in letteratura, quello relativo a che cosa sia e

¹ Libera Università di Bolzano.

di che cosa si occupi ciascuna di queste due discipline². Il compito è notoriamente difficile, viste le diverse e talvolta inconciliabili posizioni assunte dagli esponenti dei due campi di sapere, e per ragioni di brevità e convenienza lo si svolgerà facendo riferimento alle sole dimensioni del dominio e della metodologia delle due materie.

2.1 Linguistica applicata

Dominio. La discussione sugli oggetti di interesse della linguistica applicata data alle origini stesse della disciplina e nel corso degli anni le opinioni sono molto cambiate. Infatti, se inizialmente l'interesse dei praticanti la linguistica applicata era rivolto soprattutto all'insegnamento delle lingue straniere (Corder 1973), col tempo le tematiche di ricerca si sono modificate. Per quanto riguarda il contesto italiano, queste trasformazioni, così come le difficoltà di cogliere e circoscrivere lo statuto della linguistica applicata, sono state documentate da Carli & Favilla (2005) che hanno trattato della "amplissima varietà di discipline (o sottodiscipline) e della molteplicità dei metodi investigativi impiegati" (Carli & Favilla 2005: 416) dai praticanti la materia³. Più precisamente, i due autori hanno documentato come parecchi tra i partecipanti al IV congresso dell'Associazione Italiana di Linguistica Applicata (AItLA) riconoscessero come rilevante per la linguistica applicata ciascuno dei 37 possibili settori di indagine allora ipotizzati⁴. Tuttavia, secondo Carli & Favilla

² Prima di utilizzare questo termine, occorrerebbe inverosimilmente discutere se linguistica applicata e *digital humanities* possano essere definite discipline (accademiche) oppure no, visto che da anni si dà un dibattito al riguardo [cfr. Grabe (2010) per la linguistica applicata e l'accenno in Ciotti (2017) o la critica in Schnapp (2011) per le *digital humanities*]. Con riferimento alla sola situazione italiana, e alla luce di una definizione enciclopedica di disciplina come quella di Cappello et al. (1932) – secondo cui "il termine "disciplina" passò a significare a un tempo l'oggetto stesso dell'apprendere e dell'insegnare, quella che fu detta anche la materia dell'insegnamento, onde si ebbero le varie discipline" – la risposta non può che essere affermativa [seppur non necessariamente positiva, cfr. in tal senso la parte terza di Foucault (1976)] vista la presenza tanto di insegnamenti denominati "linguistica applicata", quanto di corsi di laurea, di sezioni o dipartimenti universitari in diversi atenei, così come di associazioni scientifiche quali l'*Associazione Italiana di Linguistica Applicata* (AItLA) o di riviste quali la *Rassegna Italiana di Linguistica Applicata* e *Studi Italiani di Linguistica Applicata*. *Mutatis mutandis* lo stesso vale per le *digital humanities*, registrate come insegnamento e corso di laurea in alcune università italiane, ma non come associazione o rivista scientifica, forse anche stante la preferenza per una delle possibili versioni italiane del termine inglese come nel caso della *Associazione per l'informatica umanistica e la cultura digitale* e della rivista *Umanistica Digitale*, che ne è uno degli organi di informazione.

³ Per un approccio emico alla definizione di linguistica applicata, cfr. l'esperienza riportata alla pagina web https://www.cambridge.org/elt/resources/appliedlinguistics/reading/Applied_ART_WhatIsAL.pdf.

⁴ I settori proposti nel questionario commentato da Carli & Favilla (2005: 416) sono: "Acquisizione/apprendimento linguistico L1- L2; Didattica delle lingue/Glottodidattica; Politica linguistica; Pianificazione linguistica; Educazione bilingue; linguistica informatica/computazionale; Standardizzazione ortografica; Multi/bilinguismo; Educazione linguistica culturale; Ricerca sulla leggibilità dei testi; Lingue di specialità; Terminologia professionale; Analisi del discorso/Analisi conversazionale; Traduzione e Interpretazione; Sociolinguistica; Pragmatica; Lessicografia; Comunicazione pubblicitaria; Comunicazione massmediale; Comunicazione interculturale; Linguistica contrastiva; Comunicazione politica; Comunicazione istituzionale; Comunicazione audio-visiva; Linguistica femminista (per es.: usi non sessisti della lingua); Linguistica clinica, Logopedia; Linguistica dei segni;

(2005) l'estrema diversificazione del dominio non costituirebbe un problema, ma semplicemente un tratto caratterizzante tutte le discipline nomotetiche, cosa che però la linguistica applicata non è. Al contrario, secondo altri commentatori tra cui Davies (2007) l'esplosione degli oggetti di legittima indagine da parte della linguistica applicata costituirebbe invece un problema tanto teorico, quanto pratico.

Il problema sarebbe anzitutto teorico – e più precisamente ontologico – perché al crescere dell'estensione degli oggetti di studio, l'intensione della collocazione <linguistica applicata> si ridurrebbe al punto da non aver più un riferimento e determinarne quindi l'inservibilità, così come denunciato da Cook (2015) che sostiene la tesi della “estinzione” della linguistica applicata, criticando la legittimità dell'uso di una espressione ormai svuotata di significato:

The name is not particularly important. The name of any discipline is more likely to enshrine a past reality than reflect the present one, if the two have diverged. So by all means let all the different descendants call themselves applied linguists if that is convenient, and even share mail lists and associations; my point is more that the conceptual unity is not there. What many contemporary 'applied linguists' are doing in reality is selecting the conferences, conference papers, email posts, articles, and reviews which match their own independent and coherent area, while ignoring what is going on in others.

Il problema della moltiplicazione degli interessi della linguistica applicata sarebbe però anche pratico perché, come fatto intendere da Hellermann (2015) tra gli altri, in assenza di accordi per quanto generici sugli oggetti di studio della disciplina diventerebbe difficile definire e inquadrare le tematiche di legittima indagine ad esempio per una rivista, una società scientifica o un percorso formativo che alla linguistica applicata⁵ fossero dedicati.

Forse proprio per queste ragioni nel corso degli anni la natura delle definizioni della linguistica applicata è cambiata, passando dall'essere orientata a circoscrivere gli oggetti di indagine, all'essere diretta a piuttosto definirne le finalità, come testimoniano proposte quali quella di Davies & Elder (2004) secondo cui la linguistica applicata è “concerned with solving or at least ameliorating social problems involving language”; oppure quella di Berns & Matsuda (2006: 394) per cui “applied linguistics can be described as a broad interdisciplinary field of study concerned with solutions to problems or the improvement of situations

Linguistica statistica; Linguistica matematica; Fonetica e Fonologia sperimentale; Etnolinguistica; Teoria critica del linguaggio (*Critical Linguistics*); Retorica; Linguistica forense; Psicolinguistica; Neurolinguistica; Stilistica linguistica”.

⁵ In ottica valutativa può ad esempio risultare difficile selezionare gli argomenti di ricerca che legittimano l'inclusione o l'esclusione di un candidato che concorra per una posizione in linguistica applicata. In Italia il problema non si pone visto che la collocazione <linguistica applicata> non compare nella declaratoria dei settori concorsuali o scientifico disciplinari previsti dal Ministero per l'Università e la ricerca (a differenza delle molte altre scienze “applicate” contemplate nei documenti ministeriali sui settori, come nel caso di fisica, geologia, biologia, igiene, meccanica, economia, o sociologia).

involving language and its users and uses”⁶. Per quanto queste definizioni di linguistica applicata – così come molte altre a loro simili – permettano di risolvere, eludendolo, il problema dell’enumerazione degli oggetti di indagine, la loro formulazione non è certo priva di criticità, anzitutto perché rendono la linguistica applicata una disciplina di pratiche più e prima che di ricerca e insegnamento, una disciplina basata sull’impiego di saperi creati e ripresi da altri ambiti tra cui – ma non solo, e magari non principalmente – la linguistica non applicata, proprio come sostenuto da Corder (1973: 10) secondo cui:

the application of linguistic knowledge to some object – or applied linguistics, as its name implies – is an activity. It is not a theoretical study. It makes use of the findings of theoretical studies.

In tal senso la linguistica applicata risulterebbe essere una disciplina che adotta un approccio pragmatico – nel senso di rivolto alla prassi – il che le consentirebbe di rinunciare tanto a una teoria (cfr. a tal riguardo McNamara 2015), quanto a una metodologia della ricerca proprie, potendosi dunque eventualmente limitare a documentare le attività svolte.

Metodologia. In effetti, secondo McKinley (2000) per lungo tempo chi opera nella linguistica applicata avrebbe mancato di riflettere sui metodi impiegati nello svolgimento delle diverse pratiche e ricerche, forse in conseguenza della – più o meno meditata – accettazione della linguistica applicata quale beneficiaria del bagaglio teorico della linguistica generale di cui l’applicata sarebbe, in qualche modo, una branca⁷. Tuttavia, negli ultimi anni l’interesse per la dimensione metodologica della linguistica applicata è aumentato, come testimonia il numero crescente di iniziative editoriali dedicato alla tematica⁸. Ciò potrebbe essere dovuto a due ragioni. Da una parte, il proposito di reagire alla messa in discussione della linguistica applicata come disciplina, dunque la volontà di affermarla come scienza indipendente (anzitutto dalla linguistica *non* applicata) cui necessariamente occorrono un paradigma teorico e un proprio e sistematico apparato metodologico per la raccolta, la preparazione e l’analisi dei dati. Dall’altra – e per converso – la necessità di riflettere su come adattare alle particolari finalità pratiche della linguistica applicata le metodologie delle molte altre discipline che concorrono alla risoluzione dei problemi di cui questa si interessa. Un approccio del primo tipo è quello che si ritrova per

⁶ Per inciso, se si accetta che la linguistica applicata sia una materia interdisciplinare, allora paradossalmente va considerata essere antidisciplinare perché viola la compartimentazione dei saperi corrente.

⁷ Si veda in tal senso Corder (1973: 10) che afferma: “I am enough of a purist to believe that ‘applied linguistics’ presuppose ‘linguistics’; that one cannot apply what one does not possess”; oppure quanto discusso da McNamara (2015). Che la linguistica applicata sia una branca della linguistica generale è però argomento dibattuto.

⁸ Tra le varie iniziative si ricorda la recente (2021) comparsa della rivista *Research Methods in Applied Linguistics* che nella propria pagina web dichiara essere “the first and only journal devoted exclusively to research methods in applied linguistics” e specifica che “the journal does not exclude submissions investigating methods and tools that are also applicable to, or introduced from, other disciplines, but they must be examined from the perspective of applied linguistics and for the purpose of solving problems in this discipline” <https://www.journals.elsevier.com/research-methods-in-applied-linguistics>).

esempio in McKinley & Rose (2020), dove la presentazione dei diversi metodi di elicitazione e analisi dei dati è accompagnata da riflessioni che mostrano – per dirla nel solco di Brown (2004) – come le diverse metodologie possano essere declinate a più livelli anche in funzione di fattori contestuali che influiscono sulla ricerca stessa, per esempio la disponibilità o la fonte di eventuali finanziamenti, oppure sulle sue finalità, che possono variare dal voler osservare al voler misurare oppure al voler controllare una situazione. Invece, un approccio del secondo tipo è quello che affiora per esempio in Block (2017: 32) là dove si sostiene che “political economy should be adopted as a frame for research and discussion in applied linguistics as part of a general social turn which has taken hold in the field over the past three decades”.

Quale che sia la ragione delle recenti riflessioni metodologiche in linguistica applicata, e quale che ne sia l'impostazione, trattarne con riferimento alle *digital humanities* può sicuramente essere utile, perlomeno nella misura in cui si accetti che queste ultime sono caratterizzate, anche e forse anzitutto dalla loro dimensione metodologica.

2.2 Digital humanities

Proporre una definizione di *digital humanities*⁹ è almeno tanto problematico quanto offrirne una per la linguistica applicata, come dimostrano i numerosi tentativi in merito fatti negli ultimi due decenni (Nyhan et al. 2013). Tra le varie proposte avanzate, vi sono quelle di quanti si siano focalizzati sulle pratiche caratteristiche di chi operi nel quadro della disciplina come per esempio la digitalizzazione delle fonti o il ricorso a metodi di analisi quantitativi. Altri invece hanno enfatizzato i vantaggi delle *digital humanities* rispetto agli approcci non digitali, per esempio con riferimento alla capacità di analizzare velocemente grandi quantità di dati oppure di codificare le procedure di analisi così che possano essere testate, replicate e modificate (Nerbonne & Tonelli 2017). Altri ancora, infine, hanno trattato della dimensione epistemologica disquisendo se l'umanistica digitale debba essere intesa come una disciplina a sé stante (Rockwell 1999), come una comunità di pratica (Fragaszy Troyano & Rhody 2013), oppure come una metodologia di lavoro (Levenberg et al. 2018). Pertanto, sebbene la collocazione <*digital humanities*> sia sempre più diffusa e, in qualche maniera, di gran moda (Terras 2011), una definizione largamente condivisa della disciplina pare ancora lontana dall'essere trovata, soprattutto con riferimento alla questione dell'essere le *digital humanities* un approccio teorico oppure un approccio metodologico per lo studio delle tematiche di interesse. Maggiore accordo¹⁰ invece c'è su quello che sia il dominio di indagine.

Dominio: Gli oggetti di interesse delle *digital humanities* sono – almeno nominalmente – facilmente identificati: si tratta infatti di tutte le manifestazioni di cui si interessano le varie discipline umanistiche, ovvero coreutica, filosofia, letteratura,

⁹ Non è questa la sede per discutere della (storia della) denominazione della disciplina, conosciuta in Italia anche con le espressioni <informatica umanistica> e <umanistica digitale>.

¹⁰ Utilizzo il termine anche con riferimento alla pratica di co-costruzione delle definizioni diffusa nelle *digital humanities* come testimonia il lavoro di Svensson (2009, 2010).

linguistica, musicologia, storia e storia dell'arte, tra le altre. Vista la numerosità delle discipline umanistiche pertinenti, alto è anche il numero dei fenomeni indagati e variabile la natura degli oggetti osservati, che possono essere tanto materiali quanto culturali. Tuttavia, e per quanto eterogenei, nella prospettiva delle *digital humanities* questi osservabili sono accomunati dal poter essere tutti ugualmente indagati usando strumenti computazionali così da facilitarne e migliorarne la comprensione per le finalità più diverse. Difatti, le *digital humanities* operano “at the intersection of computational methods and humanities materials” (Drucker 2021: 1), al fine di dare risposta a domande di ricerca aperte, oppure di stimolare nuove domande e riflessioni.

Metodologia: Lo stesso Drucker (2021: xii) sostiene che nelle *digital humanities* le pratiche di approntamento del dato siano ormai standardizzate – il che pare però valere solo per la procedura, non anche per gli strumenti – e prevedano una prima fase relativa alla generazione di dati digitali (che possono essere o nativamente tali, o rielaborati a partire da dati analogici tramite un processo di rimediazione digitale); una seconda fase di organizzazione e modellizzazione dei dati digital(izzati), ciò che prende talvolta il nome di datificazione; una terza fase di analisi dei dati modellizzati; una quarta fase di diffusione dei dati e dei risultati dell'analisi, possibilmente in modalità liberamente accessibili.

Queste operazioni fanno sì che l'intero processo sia caratterizzato dalla presenza dell'informatica, da intendersi non tanto nei termini di un ricorso agli elaboratori elettronici da parte dei ricercatori, quanto piuttosto dell'adozione di un approccio computazionale in ciascuna delle fasi del processo, così che fenomeni interessanti per le scienze umane vengano analizzati ricorrendo ad approcci digitali e algoritmici. Ciò è particolarmente vero per la tradizione italiana delle *digital humanities* (cfr. Ciotti 2018) che nel corso degli anni ha promosso una riflessione sulla natura dei dati sin dalle fasi di elicitazione così da rendere i ricercatori consapevoli che le tecnologie informatiche impiegate per la raccolta delle informazioni condizionano il contesto – dunque gli osservabili – e che i dati che a quest'ultimi si riferiscono, da quest'ultimi si distinguono poiché si tratta di artefatti interpretati e processati (Owens 2011), dunque di oggetti manipolati e categorizzati dall'osservatore alla luce della propria ideologia e mediante apparati informatici. In tal senso, le *digital humanities* definiscono anzitutto un approccio alla ricerca che enfatizza la componente metodologica rispetto a quella teorica e applicativa, focalizzandosi sull'approntamento critico della base di dati e sulla sua analisi. Quest'ultima, infine, è spesso operata in chiave quantitativa considerato l'approccio computazionale mirato a identificare schemi nei dati e visto che “a pattern is based on regularity, and regularity is reflected in frequency” (Ebensgaard Jensen 2014: 126). Questa quantificazione – che si vorrebbe fosse sempre meno centrale, a rivendicare il valore delle indagini qualitative (Schnapp 2011) – è resa possibile dalla potenza di calcolo degli elaboratori prodotti negli ultimi decenni anche quando applicata ad ampie basi di dati ed è solitamente operata sfruttando analisi statistiche anche quando applicata ad

ampie basi di dati sviluppate a comprendere strumenti per l'analisi di fenomeni con distribuzione non normale tipici delle manifestazioni indagati dalle scienze umane.

3. *La relazione tra le due discipline*

3.1 Nella letteratura

Se la linguistica applicata si caratterizza più per le sue finalità che per il metodo di studio, e se al contrario le *digital humanities* si contraddistinguono più per la loro metodologia che per i fenomeni di indagine, l'interazione tra le due discipline è potenzialmente fruttuosa. Tuttavia, diversamente da quanto atteso, in letteratura si trovano pochi espliciti resoconti di esperienze o considerazioni sul rapporto tra le due materie¹¹.

Quando la si consideri dal punto di vista della linguistica, la faccenda non sorprende, perché molte delle possibili riflessioni in merito sono assorbite dal dibattito su differenze (Jensen 2014) e rapporti (Sprugnoli et al. 2019) tra informatica umanistica e linguistica computazionale. Quest'ultima, infatti, non solo rappresenta la prima manifestazione dell'umanistica digitale, ma anche quella che per definizione persegue lo studio scientifico del linguaggio in prospettiva computazionale, indipendentemente dal fatto che l'obiettivo sia teorico e generale o pratico e applicato. Al contrario, se considerata dal punto di vista delle *digital humanities*, la scarsità di riflessioni sul rapporto tra le due discipline colpisce, perché il primo impiego dell'informatica umanistica è proprio stato teso ad automatizzare il procedimento della traduzione (Orlandi, 2007), dunque a risolvere un classico problema di linguistica applicata che ha a suo fondamento la capacità di codificare – di qui l'utilità della riflessione computazionale – come onomasiologicamente equivalenti oggetti semasiologicamente differenti.

Peraltro, molte delle considerazioni sul rapporto tra linguistica applicata e *digital humanities* sono relative alla discussione di applicazioni informatiche per la risoluzione di problemi pratici – *in primis* l'insegnamento delle lingue – o per l'analisi di usi linguistici in contesti digitali perché mediati dal computer, piuttosto che al trattamento computazionale di fenomeni linguistici. In tal senso, queste riflessioni riguardano un'accezione di *digital humanities* ormai superata come esplicitato da Unsworth (2002) secondo cui “[digital humanities] is entirely distinct from using the computer when it models the typewriter, or the telephone, or the phonograph, or any of the many other things it can be.” In tal senso si può però notare che, se con riferimento agli esempi riportati nella citazione di Unsworth in cui il computer imita digitalmente il funzionamento e l'uso di strumenti analogici l'affermazione è condivisibile, diversamente con riguardo alle situazioni in cui il computer emula invece l'uso di sistemi linguistici – dunque di capacità semiotiche e cognitive – come nel caso dei chatbot, la situazione si complica: non si può infatti ignorare che la simulazione del comportamento linguistico che la macchina fa è basata su una mo-

¹¹ Si vedano ad esempio le osservazioni riportate alla pagina web <https://eadh.org/linguistics>.

dellizzazione computazionale della lingua dell'interazione e del sistema simbolico-culturale di interagenti umani.

Nella maggior parte dei lavori che – per lo più implicitamente – trattino del rapporto tra linguistica applicata e *digital humanities* emerge quindi una visione dell'informatica quale disciplina ingegneristica mirata a offrire soluzioni, piuttosto che come scienza teorica per la rappresentazione delle informazioni. In tal senso – e nella misura in cui mira a semplificare o velocizzare talune operazioni piuttosto che a trattare (delle conseguenze) della messa in forma digitale degli osservabili o della formalizzazione e algoritmizzazione delle procedure per analizzarli – l'interfaccia tra linguistica applicata e *digital humanities* pare definire una relazione di comodo strumento¹², piuttosto che di metodo.

Non è questa una considerazione valoriale, tanto più nell'ottica di chi aderisse a una visione della linguistica applicata quale disciplina indipendente finalizzata alla sola risoluzione dei problemi, a prescindere dalla loro analisi (come alle volte – legittimamente – si dà nella glottodidattica non sperimentale). È piuttosto un richiamo al rischio, insito in ogni attività di ricerca basata sull'uso di strumenti (anche computazionali), che l'ignorare i fondamenti di funzionamento degli stessi comporta¹³, vale a dire il ritenere che siano componenti neutri del processo di scoperta, quando invece trasformano le osservazioni (e in prima istanza gli osservabili), come ben testimoniato dal fatto che solo usandoli si svelano informazioni inedite.

3.2 Nel volume

Con riferimento alle categorie presentate poco sopra, il volume raccoglie contributi riconducibili a entrambe le declinazioni del fare linguistica applicata con le *digital humanities*. Più precisamente – e premesso che la distribuzione sarebbe meglio caratterizzata come gradiente piuttosto che come categoriale – tre dei nove contributi, ovvero Cutugno & Campisi, De Iacovo & Palena, Revelli et al. Sono tendenzialmente riconducibili alla prima tipologia (quella di comodo), perché presentano l'applicazione di strumenti informatici alla risoluzione di problemi di linguistica applicata e/o di procedure di ordinamento e calcolo numerico che rendono possibile svolgere operazioni tutto sommato semplici su basi di dati numerose e complesse; cinque, ovvero Caruso & Presta, Del Rosso & Brambilla, Draxler, Maffia

¹² Questa relazione strumentale è analoga a quella che promuovono i corsi di laurea magistrale in Metodologie informatiche per le discipline umanistiche (LM-43) che mirano alla formazione di esperti nell'“uso di strumenti informatici in ambito umanistico” (Decreto Ministeriale 16 marzo 2007: 171) e non, come nel caso dei corsi di laurea magistrale in informatica o ingegneria informatica di specialisti nella pianificazione, la progettazione, lo sviluppo, la direzione lavori, la stima, il collaudo e la gestione di impianti e sistemi complessi o innovativi per la generazione, la trasmissione e l'elaborazione delle informazioni, anche quando implicino l'uso di metodologie avanzate, innovative o sperimentali” (Decreto Ministeriale 16 marzo 2007: 78).

¹³ Schmidt (2016) discute del problema con riferimento alla eventuale necessità dei ricercatori in *digital humanities* di comprendere gli algoritmi informatici cui ricorrono concludendo che, vista l'alta specializzazione degli stessi, questi non debbano farlo se non vogliono, ma non possono rinunciare a capire in che modo quelli che usano modifichino i dati su cui operano.

et al., Sciolette & Giovannetti alla seconda tipologia (quella di metodo) perché trattano di procedure computazionali di costruzione e categorizzazione dei dati; e uno, Ciccolone & Grosso, né all'una né all'altra categoria perché modificando la prospettiva trasformano il fare linguistica applicata con le *digital humanities* in analizzare le *digital humanities* con la linguistica applicata. Questa macroscopica distinzione è in larga parte conseguenza del fatto che – nel quadro dell'organizzazione della ricerca nelle *digital humanities* proposta da Drucker (2021) – i contributi del primo gruppo trattano della fase di analisi dei dati, quelli del secondo riguardano le fasi di ri(mediazione), organizzazione e modellizzazione degli stessi.

In tal senso, il lavoro di Cutugno & Campisi dimostra come l'integrazione – tutt'altro che banale sia sul piano teorico che su quello pratico – di singoli strumenti concettuali e computazionali approntati per l'analisi di sistemi semiologici tanto diversi quanto le lingue, i gesti e l'architettura consenta una descrizione multidimensionale dell'interazione asimmetrica tra guida e guidato in contesto museale che seppur primariamente “orientata ad applicazioni informatiche arricchite da sistemi di rappresentazione della conoscenza”, in realtà arricchisce essa stessa le forme di rappresentazione e caratterizzazione digitale della conoscenza e della comunicazione linguistica e non linguistica.

Il contributo di De Iacovo & Palena tratta invece di chatbot e, in particolare, del loro uso in un progetto di didattica della intonazione. In questo caso l'interfaccia tra linguistica applicata e *digital humanities* è data dallo sviluppo di una procedura computazionale nel processo di insegnamento della pronuncia di una L2. Questa procedura – finalizzata alla generazione di un feedback valutativo asincrono – comporta da un lato la sistematizzazione delle fasi del processo glottodidattico (elicitazione delle produzioni dell'apprendente, analisi delle produzioni parlate, comunicazione degli esiti dell'analisi) che ricalca quella tipica delle *digital humanities*; dall'altro l'impiego di programmi elettronici per l'interpretazione e l'annotazione del segnale acustico, l'analisi intonativa algoritmizzata, il confronto su base statistica dei dati elicitati con quelli attesi, nonché in tutte le fasi per l'interazione con l'apprendente che quindi risulta sempre digitale perché mediata dal computer.

Il lavoro di Revelli et al. Si configura invece come una parziale messa alla prova di strategie digitali per la ricerca preliminare di correlati della sottodeterminazione linguistica da impiegare per l'analisi di testi giuridici. Gli strumenti computazionali selezionati per il compito costituiscono senza dubbio un prodotto spendibile per ricerche nel quadro delle *digital humanities*, e pur nella loro applicazione a una base di dati assai ridotta testimoniano del potenziale dell'informatica umanistica, tanto più che la ricerca rimanda almeno idealmente alla linguistica dei corpora che ha ormai adottato il digitale tanto per la costruzione delle basi di dati, quanto per la loro organizzazione e investigazione.

Ciò che caratterizza la ricerca di Revelli et al., così come quelle di De Iacovo & Palena e di Cutugno & Campisi, è che nell'indagine gli strumenti informatici vengono impiegati su unità linguistiche (di base) già identificate, per lo più in maniera non computazionale e alla luce di categorie linguistiche tradizionali.

Diversamente, il lavoro di Caruso & Presta, che si pone alla confluenza tra i contributi che trattano delle fasi di analisi dei dati da una parte e di ri(mediazione), organizzazione e modellizzazione dei dati dall'altra, testimonia l'adozione di un approccio non di comodo alle *digital humanities*, dove queste ultime diventano parte in causa nella costruzione delle categorie di analisi. Le due autrici infatti trattano del fare linguistica applicata con le *digital humanities* descrivendo e riflettendo sulla complessità – gestibile anche grazie agli elaboratori elettronici – di un progetto in cui confluiscono conoscenze e competenze di lessicologia, lessicografia computazionale, linguistica dei corpora e studi sull'interazione tra uomo e macchina che consentono da un lato di elaborare una base di dati lessicografici caratterizzata dalla compresenza di definizioni e descrizioni di espressioni idiomatiche italiane, e dall'altro di valutare come le diverse modalità di presentazione delle stesse in un dizionario per *smartphone*, pur mantenendo macro- e microstruttura, comportamenti tipologici d'interazione diverse da parte degli stessi utenti apprendenti di italiano L2. In tal senso lo studio – che si basa anche sulla ripresa di strumenti computazionali elaborati al di fuori del progetto – permette di apprezzare come la diversa organizzazione e presentazione delle informazioni ne condizioni l'analisi, non solo con riferimento alla dimensione della ricerca, ma anche di quella dell'apprendimento.

Se il lavoro di Caruso & Presta tratta sia di preparazione che di elaborazione del dato, i contributi di Del Rosso & Brambilla, Draxler e Maffia et al. trattano piuttosto della sua discretizzazione così da permetterne la computazione e offrono riflessioni sulle fasi di (ri)mediazione e datificazione. Più precisamente, ciascuna delle tre ricerche si focalizza sull'analisi di produzioni orali e sulle possibili soluzioni offerte dalle *digital humanities* a un problema ricorrente nella linguistica tanto generale quanto applicata, vale a dire quello della trascrizione e annotazione automatica del parlato, anche nel caso di produzioni in ambienti non controllati (Del Rosso & Brambilla) o atipiche (Del Rosso & Brambilla, Maffia et al.).

Nel dettaglio, il contributo di Del Rosso & Brambilla mette a confronto gli esiti della trascrizione di identiche registrazioni quando elaborate percettivamente e manualmente da trascrittori umani oppure strumentalmente e automaticamente da un software commerciale per il riconoscimento automatico del parlato. A tal fine la ricerca tratta della formalizzazione delle procedure per la caratterizzazione quantitativa e qualitativa delle differenze tra i due esiti trascrittori. Così facendo la ricerca evidenzia il potenziale delle *digital humanities* per la linguistica applicata mostrando come i risultati delle ricerche all'intersezione tra queste due discipline consentono di oggettivizzare opinioni diffuse ma non sempre suffragate da dati come quelle su vantaggi e svantaggi della trascrizione automatica, peraltro fornendo ai potenziali interessati chiavi di lettura che permettano di valutare quale approccio tenere in base alla natura (nei termini di forma e contenuto) delle registrazioni disponibili.

Anche il lavoro di Draxler affronta la tematica della delega a strumenti informatici (di parte del processo di creazione) del dato per l'analisi di fenomeni linguistici riprendendo la questione della trascrizione automatica. Più precisamente, nel contributo l'autore amplia il campo di riflessione sul confronto tra procedure algorit-

miche e non algoritmiche di trasposizione del parlato già affrontato da Del Rosso & Brambilla così da includere, oltre che la dimensione ortografica, anche quella fonetica e fonologica. Ciò consente all'autore sia di presentare alcune risorse e procedure informatiche da lui sviluppate e liberamente accessibili – come da spirito delle *digital humanities* – sia di discutere di alcuni limiti etici e tecnici della trascrizione automatica in ambito scientifico, quelli che non consentono ancora di dichiararne la superiorità rispetto alla trascrizione impressionistica.

Anche il contributo di Maffia et al. rappresenta un caso esemplare di studio della datificazione con le *digital humanities* per la linguistica applicata. Infatti, gli autori mostrano come al fine di risolvere un problema pratico – quello di analizzare automaticamente il parlato letto per ricavare degli indici spendibili per la diagnosi precoce della Malattia di Parkinson – sia possibile adottare una procedura sperimentale strutturata che prevede anche la digitalizzazione del parlato cui segue una annotazione che può essere tanto manuale, quanto automatica. La disponibilità delle due annotazioni permette poi di operare un raffronto qualitativo da cui emergono quelle che sono le criticità della procedura automatizzata, certo accattivante perché rapida ed economica, che parrebbero poter essere risolte – oltre che con il ricorso ad una diversa digitalizzazione del dato, come suggerito dagli autori – con una revisione dell'algoritmo impiegato.

Nel complesso, i contributi di Draxler, Del Rosso & Brembilla e Maffia et al. testimoniano di come l'informatica umanistica possa essere intesa quale scienza della rappresentazione delle informazioni che comporta, prima ancora che l'annotazione dei dati, la formalizzazione delle procedure di categorizzazione degli stessi. Inoltre, questi lavori adottano procedure computazionali che rendono possibile svolgere operazioni di ordinamento e calcolo numerico anche assai complesse una volta che i dati siano stati opportunamente preparati. È questa una valida testimonianza di che cosa significhi fare linguistica applicata con le *digital humanities*, perché mostra come l'impiego degli strumenti e delle metodologie informatiche per la codifica e la rappresentazione dei fenomeni linguistici permetta di operare con metodo nel quadro di un approccio scientifico e replicabile all'analisi dei fenomeni umanistici, ma richiede che si sia consapevoli dell'influenza dello strumento sull'elaborazione delle informazioni.

I contributi di Del Rosso & Brambilla, Draxler e Maffia et al. trattano di produzioni orali. Al contrario, quello di Sciolette & Giovannetti si occupa – come da prima tradizione delle *digital humanities* – di lingua scritta, focalizzandosi sulle fasi di rimediazione e di datificazione. Più precisamente, la ricerca mira ad indagare due dimensioni: da un lato le possibili strategie di formalizzazione e integrazione in una unica rappresentazione computabile tanto delle informazioni documentali (quasi mai considerate dai linguisti applicati, che tendenzialmente considerano solo i valori assunti dalla categoria *DIAMESIA*) quanto delle informazioni testuali. Dall'altro lato la ricerca mira a valutare una delle possibili modalità di esplicitazione di quella parte di informazione testuale che, pur inclusa nel documento, non è codificata scritturalmente – cioè linguisticamente – ma risulta dalle conoscenze e dalle inferenze del fruitore del testo, così da renderla spendibile

per la traduzione (anche automatica in ultima istanza). In tal senso il contributo testimonia delle ricadute della riflessione teorica svolta in seno alle *digital humanities* nella linguistica applicata tanto nelle fasi di conversione digitale dell'informazione analogica materialmente accessibile, quanto in quelle di rappresentazione materiale e computabile di un dato analogico quale il significato, non lessicalizzato.

Anche il contributo di Ciccolone & Grosso tratta di lingua scritta, ma diversamente dai precedenti non affronta né la tematica delle procedure computazionali di categorizzazione dei dati, né quella di procedure di ordinamento e calcolo numerico, collocandosi così al di fuori delle due categorie di interazione tra linguistica applicata e *digital humanities* ipotizzate sopra. Infatti, la ricerca dei due autori tratta di un caso di analisi di frammenti dell'interlingua simulata di un chatbot. In tal senso, la ricerca è interessante perché si caratterizza per il ribaltamento della relazione tra le due discipline in causa e mostra l'utilità del fare *digital humanities* – e linguistica computazionale più precisamente – con la linguistica applicata, eventualmente anche in prospettiva glottodidattica, adottando una metodologia di indagine impressionistica. La forza di questo approccio sta tutta nel riconoscere implicitamente che i prodotti (almeno in parte) delle *digital humanities* sono a loro volta manifestazioni di interesse delle scienze umane e nel caso specifico che possono essere analizzati con gli strumenti della linguistica applicata.

4. Conclusioni

Nonostante l'invito mosso dai curatori a farlo, dai contributi emerge una scarsa propensione a discutere esplicitamente le motivazioni e gli obiettivi del ricorso alle *digital humanities* comunque intese nel fare linguistica applicata, sia con riferimento ai possibili cambiamenti del paradigma teorico, sia con riferimento all'impatto e alla rilevanza che l'informatica umanistica può avere per il successo nel trattamento dei problemi pratici correlati al linguaggio, le lingue, il loro uso e i loro utilizzatori. È questo un limite del volume nel suo complesso, perché se è vero che ogni attività di ricerca è mediata e condizionata dagli strumenti materiali e concettuali di cui beneficia, del comprendere come quelli informatici influiscano sulle pratiche della linguistica applicata avrebbe beneficiato l'intera comunità scientifica.

Ciononostante, da quanto riferito nei contributi emerge che convivono due possibili declinazioni del fare linguistica applicata con le *digital humanities*, quella di comodo e quella di metodo. La prima – difforme rispetto a quella che le *digital humanities* stesse ormai promuovono – mette a fuoco il ricorso a strumenti informatici nel trattare problemi linguistici. La seconda – in linea con l'approccio dell'informatica umanistica – tratta invece dell'approccio digitale alla costruzione del dato, problematizzandolo. Peraltro, sebbene anche in questo secondo caso la motivazione principale pare essere quella di voler verificare la spendibilità di certi strumenti per rendere più agevole e rapido risolvere alcuni problemi pratici come trascrivere o annotare automaticamente la lingua parlata, il risultato ultimo potrebbe essere quello di modificare – sistematizzandolo – il paradigma della linguistica applicata contribuendo a caratterizzarlo come distinto da quello di altre forme di indagine linguistica.

Ringraziamenti

Ringrazio Giuliano Bernini e Jacopo Saturno per i loro commenti a una versione preliminare di questo contributo. La responsabilità dei contenuti resta esclusivamente mia.

Bibliografia

- Berns, Margie & Matsuda, Paul Key. 2006. Applied Linguistics: Overview and History. In: Brown, Keith & Anderson, Anne (eds.), *The encyclopedia of language & linguistics*, 394-405. Amsterdam: Elsevier.
- Block, David. 2016. Political economy in applied linguistics research. *Language Teaching* 50(1). 32-64.
- Cappello, Felice & Boccaccia, Epimede & Vidari, Giovanni & Vitta, Cino. 1932. Disciplina. *Enciclopedia Italiana*. https://www.treccani.it/enciclopedia/disciplina_%28Enciclopedia-Italiana%29/
- Carli, Augusto & Favilla, Maria Elena, 2005. Che cos'è oggi la linguistica applicata. In: Banti, Giorgio & Marra, Antonietta & Vineis, Edoardo (a cura di), *Atti del IV Congresso di studi dell'Associazione Italiana di Linguistica Applicata*, Modena, 19-20 febbraio 2004, 415-448. Perugia: Guerra Edizioni.
- Ciotti, Fabio. 2017. Una nuova avventura. *Umanistica Digitale* 1 <https://umanisticadigitale.unibo.it/article/view/7248/7000>
- Ciotti, Fabio. 2018. *Dall'Informatica umanistica alle Digital Humanities. Per una storia concettuale delle DH in Italia*. <https://dh2018.adho.org/en/dallinformatica-umanistica-alle-digital-humanities-per-una-storia-concettuale-delle-dh-in-italia/>.
- Cook, Guy. 2015. Birds out of dinosaurs: The death and life of applied linguistics. *Applied Linguistics*, 36(4). 425-433.
- Corder, Stephen. 1973. *Introducing applied linguistics*. Harmondsworth: Penguin.
- Davies, Alan & Elder, Catherine. 2004. Applied Linguistics: Subject to Discipline? In: Davies, Alan & Elder, Catherine (eds.) *The Handbook of Applied Linguistics*, 1-15. Blackwell Publishing Ltd.
- Davies, Alan. 2007. *An introduction to applied linguistics*. Edinburgh: Edinburgh University Press.
- Dean Brown, James. 2004. Research methods for applied linguistics: Scope, characteristics, and standards. In: Davies, Alan & Elderthe, Catherine (eds.) *The handbook of applied linguistics*, 476-500. Blackwell Publishing Ltd.
- Drucker, Johanna. 2021. *The digital humanities coursebook*. New York: Routledge.
- Fragaszy Troyano & Rhody Lisa. 2013. Expanding communities of practice. *Journal of Digital Humanities*, 2(2). <http://journalofdigitalhumanities.org/2-2/expanding-communities-of-practice/>.
- Geoffrey, Rockwell. 1999. Is humanities computing an academic discipline? Institute for Advanced Technology in the Humanities, University of Virginia. <http://jefferson.village.virginia.edu/hcs/rockwell.html>.

- Hellermann, John. 2015. Three contexts for my work as co-editor: Introduction to the special issue. *Applied linguistics*. 36(4). 419-424.
- Jensen, Kim Ebsengaard. 2014. Linguistics in the digital humanities: (computational) corpus linguistics. *MedieKultur: Journal of media and communication research*. <http://dx.doi.org/10.7146/mediekultur.v30i57.15968>.
- Kirschenbaum Matthew. 2010. *What is Digital Humanities and what's it doing in English departments?* <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbfd1e/section/f5640d43-b8eb-4d49-bc4b-eb31a16f3d06>.
- Levenberg, Lewis & Neilson, Tai & Rheams, David. 2018. *Research methods for the digital humanities*. Cham: Palgrave Macmillan.
- McKinley, Jim. 2020. Introduction. Theorizing research methods in the 'golden age' of applied linguistics research. In: McKinley, Jim & Rose, Heath (eds.) *The Routledge handbook of research methods in applied linguistics*, 1-12. New York: Routledge.
- Mcnamara, Tim. 2015. Applied linguistics: The challenge of theory. *Applied Linguistics* 36(4). 466-477.
- Nerbonne, John. & Tonelli, Sara. 2017. Introduction to the special issue on digital humanities and computational linguistics. *Italian Journal of Computational Linguistics* 2(2). 7-10.
- Nyhan, Julianne & Terras, Melissa & Vanhoutte, Edward. 2013. Introduction. In: Terras, Melissa & Nyhan, Julianne & Vanhoutte, Edward (eds.) *Defining digital humanities: A reader*, 1-10. Farnham: Routledge.
- Orlandi, Tito. 2007. *Un ultimo bilancio dell'informatica umanistica*. <http://www.cmcl.it/~orlandi/pubbli/montevarchi.pdf>.
- Owens, Trevor. 2011. Defining data for humanists: Text, artifact, information or evidence? *Journal of Digital Humanities* 1(1). 6-8.
- Pallotti, Gabriele. 2021. Measuring complexity, accuracy, fluency (CAF). In Winke, Paula & Brunfaut, Tineke (eds.) *The Routledge handbook of second language acquisition and language testing*, 201-210. New York: Routledge.
- Schnapp, Jeffrey. 2011 *The Digital Humanities Manifesto 2.0*. https://jeffreyschnapp.com/wp-content/uploads/2011/10/Manifesto_V2.pdf.
- Sprugnoli, Rachele & Pardelli, Gabriella & Boschetti, Federico & Del Gratta, Riccardo. 2019. Un'analisi multidimensionale della ricerca italiana nel campo delle digital humanities e della linguistica computazionale. *Umanistica Digitale* 5. doi:10.6092/ISSN.2532-8816/8581.
- Svensson, Patrik. 2009. Humanities Computing as Digital Humanities. *Digital Humanities Quarterly* 3.3. <http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html>
- Svensson, Patrik. 2010. The landscape of digital humanities. *Digital Humanities Quarterly* 4.1. <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>.
- Terras, Melissa. 2011. *Stats and the digital humanities*. <http://melissaterras.blogspot.com/2011/11/stats-and-digital-humanities.html>.

Indice autori

Silvia Brambilla
silvia.brambilla@uniroma3.it

Emanuela Campisi
emanuela.campisi@unict.it

Valeria Caruso
vcaruso@unior.it

Simone Ciccolone
simone.ciccolone@unica.it

Franco Cutugno
cutugno@unina.it

Valentina De Iacovo
valentina.deiacovo@unito.it

Giovina Angela del Rosso
giovinangela.delrosso@uniroma3.it

Anna De Meo
ademeo@unior.it

Rosa De Micco
rosita.demicco@gmail.com

Christoph Draxler
draxler@phonetik.uni-muenchen.de

Emiliano Giovannetti
emiliano.giovannetti@ilc.cnr.it

Giulia I. Grosso
giulia.grosso@unica.it

Marta Maffia
mmaffia@unior.it

Antonio Mastropaolo
a.mastropaolo@univda.it

Marco Palena
marco.palena@polito.it

Massimo Pettorino
mpettorino@gmail.com

Roberta Presta
roberta.presta@centrosclenzanuova.it

Daniele Paolo Radicioni
daniele.radicioni@unito.it

Luisa Revelli
l.revelli@univda.it

Flavia Sciolette
flavia.sciolette@ilc.cnr.it

Lorenzo Spreafico
lorenzo.spreafico@unibz.it

Gioacchino Tedeschi
gioacchino.tedeschi@unicampania.it

Alessandro Tessitore
alessandro.tessitore@unicampania.it

La collana “studi AItLA” accorpa le precedenti due collane dell’Associazione, quella degli «Atti» del convegno annuale e gli «Strumenti per la ricerca». La collana, costituita da volumi collettivi e monografie dedicate a temi e problemi della linguistica applicata, è ad accesso libero per tutti gli interessati.

Volumi pubblicati

1. *Varietà dei contesti di apprendimento linguistico*, a cura di Anna De Meo, Mari D’Agostino, Gabriele Iannaccaro e Lorenzo Spreafico, 2014.
2. *Grammatica applicata: apprendimento, patologie, insegnamento*, a cura di Maria Elena Favilla e Elena Nuzzo, 2015.
3. *L’insegnamento della grammatica nella classe multilingue. Un esperimento di focus on form nella scuola primaria*, Anna Whittle e Elena Nuzzo, 2015.
4. *Linguaggio e apprendimento linguistico. Metodi e strumenti tecnologici*, a cura di Francesca Bianchi e Paola Leone, 2016.
5. *Le dinamiche dell’interazione. Prospettive di analisi e contesti applicativi*, a cura di Cecilia Andorno e Roberta Grassi, 2016.
6. *Attorno al soggetto. Percorsi di riflessione tra prassi didattiche, libri di testo e teoria*, a cura di Emilia Calaresu e Silvia Dal Negro, 2018 | Premio Aitla 2017.
7. *Usare le lingue seconde. Comunicazione, tecnologia, disabilità, insegnamento*, a cura di Anna De Meo e Margaret Rasulo, 2018.
8. *Translation And Interpreting for Language Learners (TAIL)*, Lessons in honour of Guy Aston, Anna Ciliberti, Daniela Zorzi, 2018.
9. *Lingua in contesto. La prospettiva pragmatica*, a cura di Elena Nuzzo e Ineke Vedder, 2019.
10. *Motivazione, gioco, lingua. Elementi ludici tra glottodidattica e psicolinguistica*, di Ivan Lombardi, 2019.
11. *Lingue minoritarie tra localismi e globalizzazione*, a cura di Antonietta Marra e Silvia Dal Negro, 2020.
12. *Dalla glottodidattica alla formazione dei docenti. Dall’università alla scuola. In memoria di Cristina Bosisio*, a cura di Maria Cristina Gatti e Silvia Gilardoni, 2021.
13. *Lingua in contatto e linguistica applicata: individui e società*, a cura di Maria Elena Favilla e Sabrina Machetti, 2021.

finito di stampare
nel mese di aprile 2022
presso la LITOGRAFIA SOLARI
Peschiera Borromeo (MI)

Quali sono le motivazioni e gli obiettivi che spingono i linguisti applicati a ricorrere alle *digital humanities*? I contributi di questo volume offrono alcune delle possibili risposte alla domanda, discutendo di strumenti e applicazioni informatiche per la risoluzione dei problemi di utenti e analisti di lingue.

Jacopo Saturno è ricercatore presso l'Università degli Studi di Bergamo, dove si occupa di acquisizione di lingue seconde (con particolare riguardo alla morfosintassi delle lingue slave) e delle applicazioni didattiche, traduttive e interazionali dell'intercomprensione.

Lorenzo Spreafico è professore associato di Didattica delle lingue straniere moderne presso la Libera Università di Bolzano. I suoi attuali interessi di ricerca riguardano l'acquisizione e l'insegnamento della pronuncia in soggetti multilingui.

studi AItLA vol. 14

FARE LINGUISTICA APPLICATA CON LE *DIGITAL HUMANITIES*

a cura di
Jacopo Saturno - Lorenzo Spreafico

AItLA - Associazione Italiana di Linguistica Applicata
via Cartoleria, 5 - 40100 Bologna - Italy
email: info@aitla.it | sito: www.aitla.it

Edizione realizzata da
Officinaventuno
via F.lli Bazzaro, 18 - 20128 Milano - Italy
email: info@officinaventuno.com | sito: www.officinaventuno.com

ISBN: 978-88-97657-50-7



9 788897 657507