

Tratti fonetici e funzionali delle pause, e gesti concomitanti, nel parlato delle guide turistiche

Violetta Cataldo¹, Loredana Schettino¹, Renata Savy¹, Isabella Poggi², Antonio Origlia³, Alessandro Ansani², Isora Sessa², Alessandra Chiera²

¹ Università degli Studi di Salerno; ² Università di RomaTre; ³ Università di Napoli "Federico II"

Obiettivi e motivazioni del lavoro.

Il contributo presenta una descrizione e analisi dei fenomeni di disfluenza nelle *performance* orali di guide turistiche durante l'accompagnamento delle visite museali.

Il tipo di parlato in oggetto costituisce un campo di indagine particolarmente interessante sotto l'aspetto delle disfluenze: accanto ai tipici fenomeni legati alla programmazione dell'enunciato, caratteristici dell'oralità *tout court*, infatti, lo stile d'eloquio della guida turistica presenta altri tratti legati ad una complessità di fattori, tra cui possiamo annoverare i seguenti:

- 1) funzione 'retorica' delle pause e delle ripetizioni: uso delle pause (parzialmente) consapevole e/o volontario, funzionale a catturare l'attenzione del pubblico, a sottolineare passaggi e termini salienti, a creare aspettativa;
- 2) carattere 'recitativo' del parlato: uso di una minima tecnica teatrale/recitativa in cui i fenomeni disfluenti si accompagnano e coordinano a gesti specifici, cinetica e mimica facciale;
- 3) tipologia 'testuale': testo descrittivo, parzialmente strutturato e ripetitivo, ricco di terminologia tecnica;
- 4) contesto di interazione: presenza di un pubblico eterogeneo, a volte distratto e 'in movimento';
- 5) contesto spaziale/situazionale: necessario ricorso alla deissi spaziale, sia verbale che gestuale, che di frequente interrompe il flusso dell'eloquio.

Per questi motivi, le 'disfluenze' della guida turistica assumono valori strutturali e funzionali, interessanti da indagare e da modellizzare.

Il lavoro si inserisce, infatti, nelle linee di ricerca del progetto CHROME - *Cultural Heritage Resources Orienting Multimodal Experience* (PRIN 2015) che ha come oggetto la definizione e sperimentazione di una metodologia di raccolta, analisi e modellizzazione di dati multimodali per la messa a punto di un "agente virtuale" in grado di presentare beni museali, attraverso un sistema di dialogo uomo-macchina "antropomorfo". La nostra ricerca ha, dunque, come scopo una migliore descrizione e modellizzazione di alcuni aspetti del comportamento verbale che contribuiscono alla naturalezza dell'eloquio e alla sua adeguatezza al contesto situazionale. In quest'ottica, fenomeni di ipospecificazione del significante fonico, di coarticolazione e di disfluenza vengono osservati sul versante 'positivo', alla ricerca di caratteristiche e regolarità implementabili per la sintesi "*text to speech*" (TTS) al fine di migliorarne le prestazioni.

All'analisi verbale si affianca l'osservazione puntuale degli elementi gestuali che cooccorrono con l'enunciazione. I gesti (in particolar modo quelli di tipo iconico e deittico) prodotti in concomitanza con le pause, infatti, oltre ad una funzione orientata alla produzione e all'interazione, possono svolgerne anche una orientata alla comprensione (Stam e Tellier, 2017), evidenziando e sottolineando, ad esempio, le parole/concetti più importanti all'interno del discorso, isolando parole chiave, guidando l'attenzione su elementi contestuali.

Corpus e metodologia

Il materiale a disposizione consiste di registrazioni audio-visive di visite guidate condotte da guide esperte nel campo artistico-culturale. Il *corpus* completo è costituito da 3h30' di parlato per ciascuna delle 3 guide (tutte donne), articolato in 6 porzioni che corrispondono a diversi 'luoghi' della visita (punti di interesse).

Per il presente lavoro, sono stati presi in considerazione 1 punto di interesse di tre guide (35' circa).

Sul materiale selezionato sono stati etichettati tutti i fenomeni di disfluenza, adottando un sistema di codifica che fa riferimento agli studi di modellizzazione di Shriberg (1994), Hiecke (1981) e Lickley (1998) adattato all'italiano parlato. Tale schema di annotazione si muove su quattro livelli (*Disfluency Type, Disfluency Function, Disfluency Model, Disfluency Components*) che agiscono parallelamente su domini separati dell'occorrenza.

Nello studio qui proposto portiamo avanti un'analisi limitata alle pause, considerate secondo la seguente classificazione:

- pausa vuota (*silent pause*), distinta in pausa breve (*short pause*, <sp>) o pausa lunga (*long pause*, <lp>);
- pausa piena (*filled pause*, FP);
- pausa vocalizzata (*vocalized filled pause*, VFP), derivante da allungamento di parola.

I fenomeni disfluenti corrispondenti a queste categorie sono analizzati:

A. sul piano fonetico acustico, su cui sono state misurate:

- durata
- contenuto segmentale
- timbro vocalico
- profilo tonale (eventuale)
- livello di range (relativo rispetto alla sequenza adiacente)

B. sul piano funzionale, sul quale sono state individuate le seguenti funzioni primarie:

- *demarcativa*: distinguendo le pause fisiologiche da quelle di significato, legate alla scansione grammaticale/linguistica dell'enunciazione;
- *programmativa*: identificando le pause di pianificazione e/o recupero degli elementi lessicali operate dal parlante nel corso dell'enunciazione;
- *esitativa*: identificando le pause che manifestano un certo grado di incertezza nell'esposizione del parlante;
- *strategica*; identificando le pause utilizzate come strumento di controllo dell'interazione al fine di agevolare la produzione e allo stesso tempo la comprensione, permettendo ai parlanti di guadagnare tempo e di gestire l'attenzione dell'interlocutore.

C. *sul piano gestuale*, (specificamente i gesti delle mani) sul quale sono stati individuati:

- forma del gesto
- fasi di movimento (eventuale)
- ampiezza e direzione del gesto
- funzione del gesto
- sincronizzazione

Per l'analisi dei dati sul versante fonetico-acustico è stato utilizzato il software Praat (Boersma, Weenink, 2018), mentre l'etichettatura sia dei fenomeni di disfluenza, sia degli eventi gestuali è stata effettuata tramite ELAN (2018), che permette di visualizzare in sincrono le componenti audio e video del materiale raccolto per un'annotazione multilivello.

Risultati (attesi)

L'elaborazione dei dati, al momento della stesura di questo riassunto, è ancora in corso. Possiamo tuttavia riassumere i risultati preliminari più interessanti come segue:

1. la stragrande maggioranza (81%) delle pause di questo eloquio è costituito da pause piene; quasi la metà (45%) FP, il 36% VFP, da allungamento finale di parola; la loro incidenza è di circa il 10% sul totale delle parole (1 pausa ogni 20 parole circa);
2. la maggior parte delle pause (sia piene che vuote, 66%) è di tipo volontario, e realizza una funzione *demarcativa* o *strategica*, mentre solo il 34% si può addebitare a fenomeni 'esitativi' veri e propri o fisiologici (più difficile è individuare le pause di programmazione e identificarne la volontarietà);
3. circa le loro posizioni, è sorprendente che ben il 73% occorra in posizione interna di unità intonativa e/o sintattica (con valori > 90% per le VFP), solo il 27% invece ad un confine (più spesso iniziale) di unità;
4. sul piano acustico, le pause piene (FP) durano in media il doppio delle altre pause, incluso le piene da allungamento di parola (48ms in media contro i 25ms delle VFP e i 23ms delle <sp>) e sono costituite in prevalenza da vocali nasalizzate;

Per quel che riguarda la relazione tra pause e gesti cooccorrenti (*co-speech gestures*), i dati si limitano al momento a osservazioni da approfondire:

5. in corrispondenza di FP sembrerebbe verificarsi una completa e momentanea interruzione del gesto in corso;
6. mentre, in corrispondenza delle VFP, si realizzano diverse forme di interazione tra il parlante (la guida) e il pubblico presente che sembrano coinvolgere, però, l'espressione facciale e i movimenti della testa (o dell'intero corpo) piuttosto che la gestualità delle mani.

References

- Boersma, P., Weenink, D., 2018, *Praat: doing phonetics by computer* [Computer program]. Version 6.0.40
- Carlmeyer, B., Schlangen, D., Wrede, B., 2016, Exploring self-interruptions as a strategy for regaining the attention of distracted users, in *Proc. of the 1st Workshop on Embodied Interaction with Smart Environments*, p. 4, ACM.
- Chafe, W., 1980, Some reasons for hesitating, *Temporal variables in speech*, pp. 169-180.
- ELAN (Version 5.2), 2018, [Computer software].
- Goldman-Eisler, F., 1958, The predictability of words in context and the length of pauses in Speech, in *Language and Speech*, 1(3), pp. 226-231.
- Goldman-Eisler, F., 1972, Pauses, clauses, sentences, in *Language and speech*, 15(2), pp. 103-113.
- Hieke, A. E., 1981, A content-processing view of hesitation phenomena, *Language and Speech*.
- Lickley, R. J., 1998, *HCRC disfluency coding manual*, HCRC, University of Edinburgh.
- O'Connell, D. C., Kowal, S., 1983, *Pausology*, in Sedelow, W. A., Sedelow, S. Y., *Computers in Language Research 2*, Mouton Publishers, Berlin, New York, Amsterdam, pp. 221-301.
- Reich, S. S., 1980, Significance of pauses for speech perception, in *J. of Psycholinguistic Research*, 9(4), pp. 379-389.
- Shriberg, E. E., 1994, *Preliminaries to a theory of speech disfluencies*, Doctoral dissertation, University of California, Berkeley.
- Stam, G., Tellier, M., 2017, The sound of silence: the functions of gestures in pauses in native and nonnative interaction, in *Why Gesture? How the hands function in speaking, thinking and communicating*, John Benjamins, pp.353-377.