

LOREDANA SCHETTINO¹, EMANUELA CAMPISI², ANTONIO ORIGLIA¹

Disfluenze bimodali.

Forme, funzioni e pattern di fenomeni di disfluenza e gesti per la gestione del parlato di guide turistiche.³

Abstract

The literature provides increasing evidence that co-verbal gestures partake in utterances. This position is supported by the observation, in several languages, that a suspension of speech due to planning issues often corresponds to a suspension of gestural activity. However, studies on the correlation between speech disfluency phenomena and co-occurring gestures differ considerably over methodological approaches and mostly concern dyadic interactions produced under highly controlled conditions. A comprehensive and dynamic view of the simultaneous development of speech and gesture production at places in communication where speakers intervene to manage production issues is still lacking. This study is based on productions from a real-life communicative context, namely, visits of museum heritage led by expert guides. The representative efficacy of a method for the analysis of disfluency phenomena in a bimodal perspective, at both the phono-articulatory and gestural levels, is proposed and verified.

1. Introduzione

La comunicazione umana nella modalità parlata può essere descritta come un'attività complessa in cui i parlanti sono coinvolti contemporaneamente in diversi processi continui: la pianificazione del messaggio che si vuole trasmettere, la sua codifica e produzione linguistica, la ricezione e percezione della produzione propria e dell'interlocutore (Levelt 1989). Si tratta di operazioni continue ma non per forza lineari in quanto i parlanti co-costruiscono l'interazione in corso d'opera e possono pertanto avere bisogno di sospendere la produzione per guadagnare tempo aggiuntivo per l'elaborazione di quanto vogliono dire o correggere quanto già detto se riconosciuto come problematico per la comunicazione (Lickley 2015). In questo senso fenomeni come pause, riempitivi, prolungamenti di segmenti fonici di parola, ripetizioni, auto-correzioni, cambi di progetto, spesso descritti come "fenomeni di disfluenza", rappresentano elementi normali del parlato che lo caratterizzano in

¹ Università degli Studi di Napoli Federico II.

² Università di Catania.

³ Questo articolo è il risultato della collaborazione tra gli autori. Tuttavia, a fini esclusivamente accademici Loredana Schettino è responsabile delle §§ 1, 2.1, 2.3, 3.1, 3.2.1, 3.3 e 4, Emanuela Campisi delle §§ 2.2, 2.3 e 3.2.2, Antonio Origlia della § 3.4. Tutti gli autori sono responsabili della § 5.

quanto strumenti linguistici a disposizione dei parlanti per gestire in maniera efficace i processi complessi sottesi alla comunicazione parlata (Allwood *et al.* 1990, Voghera 2017).

Quest'ultima, in contesti naturali, non si basa solo sull'uso del canale fonico-uditivo e sulla produzione di segni vocali, ma tende a sfruttare in maniera cooperativa e integrata diverse modalità a disposizione date le specifiche condizioni di produzione e ricezione. Per modalità comunicativa si può intendere un meccanismo semiotico costituito dal complesso insieme di relazioni fra diverse componenti, ovvero le condizioni enunciative e pragmatiche e i canali a disposizione, le quali contribuiscono a veicolare significati attraverso manifestazioni espressive loro correlate (Voghera 2017). In quest'ottica è possibile descrivere l'attività comunicativa parlata come intrinsecamente multimodale. In particolare, nel parlato le informazioni acustiche e visive possono essere integrate con la produzione di gesti delle mani (i cosiddetti "gesti co-verbali" o "*co-speech gestures*") che accompagnano il discorso per definire, completare o rafforzare il significato degli enunciati (cfr. Campisi 2018). Sebbene sia stato dibattuto il ruolo dei gesti rispetto alla produzione linguistica parlata, moltissimi studi hanno fornito prove empiriche a supporto della cooperazione e integrazione fra gesti e parlato nella trasmissione di significati piuttosto che dell'uso dei gesti come mero supporto in caso di difficoltà. È stato quindi ampiamente dimostrato che i gesti co-verbali partecipano allo sforzo comunicativo del parlante interagendo con il parlato sul piano prosodico, sintattico, semantico e pragmatico (McNeill 1992, Kendon 2004).

A tal proposito, sono rilevanti gli studi sulla relazione fra parlato e gesti in corrispondenza di fenomeni di disfluenza. Nello specifico, è stato possibile osservare che in tutte le lingue studiate le disfluenze nel parlato sono raramente accompagnate da un gesto e spesso corrispondono all'interruzione anche del gesto in corso e del mantenimento della posizione al momento dell'interruzione del movimento (Seyfeddinipur 2006, Graziano, Gullberg 2018). Ciò ha messo in evidenza come indagini sulle regolarità d'occorrenza di eventi disfluenti e sulla loro co-occorrenza con gesti co-verbali possano fornire indicazioni sui processi alla base della significazione e della gestione della propria produzione comunicativa (Graziano, Gullberg 2018).

Tuttavia, sebbene sia presente in letteratura una sempre maggiore spinta a cogliere la dinamica della relazione fra gesti e parlato attraverso lo studio della loro interazione in casi di disfluenza, questi studi differiscono notevolmente per la varietà degli approcci metodologici adottati, ad esempio rispetto alla selezione dei fenomeni identificati come "disfluenze", all'identificazione dei gesti e delle fasi gestuali considerate per l'allineamento con le unità disfluenti nel parlato. Inoltre, essi riguardano per lo più interazioni diadiche prodotte in laboratorio, oppure in condizioni altamente controllate. Quindi manca ancora una descrizione coerente delle disfluenze in prospettiva multimodale. In un lavoro recente, Kosmala (2022) propone un approccio multimodale all'analisi delle disfluenze supportato da osservazioni qualitative secondo l'impianto derivante dalla linguistica interazionale, ma

risulta ancora difficile quantificare le osservazioni qualitative in modo da cogliere regolarità nella dinamica dell'interazione fra fenomeni di disfluenza e gesti nel processo continuo della comunicazione parlata. Questo studio mira ad osservare il ruolo svolto dalle disfluenze nella gestione della produzione parlata in italiano e la loro interazione con gesti co-verbali sia sul piano formale che funzionale considerando produzioni in un contesto comunicativo reale specifico, vale a dire il parlato divulgativo prodotto da guide turistiche per la descrizione di beni museali, e proponendo un metodo per l'osservazione di regolarità nella co-produzione di fenomeni vocali e gestuali caratterizzanti eventi disfluenti nella continuità del parlato.

2. *Quadro teorico*

2.1 Fenomeni di disfluenza per la gestione del parlato

Il termine “disfluenze” è usato in letteratura per fare riferimento a una classe eterogenea di fenomeni tipici della lingua parlata come cambi di progetto, auto-correzioni, ripetizioni, riempitivi, allungamenti segmentali, pause che sembrerebbero alterare l'ideale flusso del parlato. Trattandosi di fenomeni oggetto di interesse per studiosi di diversi ambiti, è possibile trovare notevoli divergenze nella loro definizione terminologica, le quali riflettono differenze nella concettualizzazione di “fluenza” e nella definizione tassonomica degli elementi inclusi fra i cosiddetti “fenomeni di disfluenza” a seconda di scopi di studio specifici e relativi approcci di analisi. Per citare qualche esempio⁴, si parte dal termine *dysfluency*, usato per far riferimento a fenomeni di disfluenza *atipici* legati a condizioni patologiche rispetto a quelli *tipici* della lingua parlata (Wingate 1984), per passare da “*own speech management phenomena*”, per far riferimento all'insieme di fenomeni impiegabili come strumento di gestione della propria produzione linguistica (Allwood *et al.* 1990), a *fluenzemi*, per identificare questi fenomeni come indicatori quantificabili della percezione di fluenza nel parlato di apprendenti rispetto a quello di nativi di una lingua (Götz 2013), fino all'impiego di parentesi tonde per il prefisso *dis-* di (*dis*)*fluencies*, per sottolineare l'ambiguità della valenza funzionale di questi fenomeni a seconda del contesto d'occorrenza (Crible 2018, Kosmala 2021).

Nonostante tale mancanza di accordo, è ormai comunemente riconosciuto che l'occorrenza dei fenomeni di disfluenza rappresenti la norma piuttosto che l'eccezione nel parlato spontaneo, con una frequenza media di circa 6-10 fenomeni ogni 100 parole (cfr. Bortfeld *et al.* 2001 e Lickley 2015), e che l'idea della fluenza nel parlato come corrispondente a un ininterrotto e regolare di un flusso di parole in enunciati ben formati possa essere considerata irrealistica come un movimento privo di attrito (Ginzburg *et al.* 2014).

In effetti, oltre alla pervasività è stato possibile osservare che l'occorrenza di questi fenomeni segue schemi regolari risultanti dal loro impiego nell'attività *self-monitoring* svolta dal parlante (Levelt 1989, Shriberg 1994). Secondo il modello del

⁴ Per una rassegna terminologica si vedano Eklund (2004), Crible (2017).

linguaggio proposto da Levelt (1983, 1989), i parlanti sono in grado di monitorare e controllare l'appropriatezza intenzionale e contestuale e la correttezza linguistica della propria produzione su un doppio circuito, *double perceptual loop*: l'*inner speech* non ancora pronunciato (dunque durante le fasi di pianificazione o di codifica linguistica, prima dell'articolazione) e l'*overt speech* pronunciato (dunque durante o dopo la fase di articolazione). Sulla base di questa attività, i parlanti possono decidere di sospendere temporaneamente l'erogazione del discorso e guadagnare tempo per la soluzione di problemi di produzione pre-articolatoria, ad esempio attraverso silenzi, prolungamenti segmentali, riempimenti, o alterare sequenze già pronunciate, ad esempio attraverso l'inserimento, la cancellazione, la sostituzione di alcuni elementi. A questi ultimi può essere attribuita una funzione retrospettiva e sono stati descritti pertanto come *backward-looking disfluencies*, mentre ai fenomeni realizzati per coprire tempi aggiuntivi necessari per la pianificazione e codifica del messaggio è attribuibile una funzione prospettiva, e sono stati descritti come *forward-looking disfluencies* (Hieke 1981, Ginzburg *et al.* 2014).

In quanto strumenti linguistici di cui i parlanti dispongono per gestire efficacemente i processi online di pianificazione, codifica, articolazione e ricezione del discorso (Allwood *et al.* 1990; Levelt 1989), i fenomeni di disfluenza possono rientrare fra i correlati linguistici funzionali della modalità parlata (Voghera 2017) ed essere considerati come *finestra* sulla natura dei processi sottostanti la produzione e percezione della comunicazione parlata (Chafe 1980, Shriberg 2001).

2.2 Gesti co-verbali e produzione parlata

C'è ormai ampio consenso in letteratura sul fatto che i gesti co-verbali sono da considerare a pieno titolo parte integrante dell'enunciato. Dopo una prima fase in cui i *gesture studies* avevano come obiettivo principale dimostrare come tali movimenti contribuiscano in maniera pervasiva alla costruzione del significato, la crescente presa di consapevolezza della loro integrazione col parlato a livello semantico e temporale ha avuto come conseguenza un aumento di interesse nei confronti dei meccanismi soggiacenti alla produzione di un messaggio multimodale che integri componenti proposizionali con quelle legate alle immagini mentali (McNeill 1992).

A differenza di alcune teorie che ricercano l'architettura cognitiva responsabile della produzione dei gesti nei meccanismi soggiacenti alla produzione di azioni indipendenti dal parlato (Hostetter & Alibali 2008), alcune proposte suggeriscono che, quando è coinvolto nella produzione di gesti, il sistema motorio è influenzato anche da processi di pianificazione del parlato (Kita & Özyürek 2003); altre ancora, in modo più radicale, partono dal presupposto che parole e gesti, in quanto componenti dell'atto comunicativo, sono entrambi gestiti da un processo centrale che controlla la produzione linguistica (che nel modello di Levelt corrisponde al *conceptualizer*) e, come tali, pianificati insieme (de Ruiter 2000; 2007).

Una visione di questo tipo consente di superare un approccio riduttivo alla gestualità, in cui ad esempio ci si chiede se il gesto abbia esclusivamente una finalità compensativa, oppure in cui ci si interroga solo sul contributo che il singolo gesto

apporta al corrispondente immediato nel parlato, per soffermarsi invece sul messaggio nel suo insieme: un flusso più o meno lungo in cui le due modalità si intrecciano in modo dinamico e la cui organizzazione dipende dall'intreccio di intenzioni comunicative, condizioni del discorso, adattamento al destinatario e problemi di produzione.

2.3 Fenomeni di disfluenza e gesti co-occorrenti

Lo studio della co-occorrenza fra gesti e disfluenze nel parlato è stato riconosciuto come particolarmente utile per fornire indicazioni sul tipo di relazione che sussiste fra lingua parlata e gestualità. In letteratura è possibile trovare osservazioni discordanti sulla relazione temporale fra la produzione di gesti e marcatori di disfluenza spesso legate a questioni di natura metodologica, considerato che questi studi possono essersi basati sull'osservazione di diversi tipi e parti di gesto (Graziano, Gullberg 2018). Tuttavia, si riscontra concordanza nell'osservare che i gesti vengono prodotti per lo più in corrispondenza di parlato "fluente" e molto meno frequentemente in sovrapposizione con fenomeni di disfluenza (Kosmala 2021). Nondimeno è interessante studiare questa relazione in quanto può fornire indicazioni sulle dinamiche alla base dell'attività comunicativa fonico-gestuale intesa come processo continuo e integrato.

I primi studi di questo genere si basano sull'ipotesi che i gesti rappresentino uno strumento di compensazione in caso di difficoltà nel linguaggio (Krauss, Hadar 1999) o che, considerata la sincronizzazione fra sospensioni nel parlato e nei movimenti gestuali, questi possano segnalare l'arrivo di una pausa di pianificazione linguistica (Butterworth e Beattie 1987). In particolare, Graziano e Gullberg (2018) sottolineano che la sospensione dell'attività gestuale in corrispondenza di disfluenze nel parlato supporta l'idea di un sistema vocale-gestuale integrato e che l'osservazione di casi di co-occorrenza sia con gesti referenziali che pragmatici (Kendon 2004) indica che, in questi casi, i parlanti possono non solo cercare di compensare le loro difficoltà espressive attraverso gesti referenziali, ma anche usare gesti per marcare e fornire informazioni sul discorso. Ciò sembrerebbe in linea con l'interpretazione proposta da de Ruiter (2000: 293), basata sul modello di Levelt, per cui il modulo di controllo della produzione linguistica (*conceptualizer*) può individuare (attraverso i *feedback* interni ed esterni derivanti dal *double-perceptual loop*) problemi temporanei nel parlato e compensare trasmettendo parte dell'intenzione comunicativa attraverso la modalità gestuale.

Questi studi tendono a concentrarsi su alcuni tipi di fenomeni generalmente rientranti nella categoria "disfluenze", cioè su elementi di sospensione o "interruzione" del parlato (fenomeni di esitazione come pause, sia piene che silenti, prolungamenti, interruzioni di parola) mentre minore attenzione è riservata alla loro distinzione funzionale, eccezion fatta per lo studio in cui Seyfenddinipur (2006) osserva che fenomeni di disfluenza con funzione diversa possono correlare con diversi tipi di gesti: una sospensione del parlato può correlare con una sospensione anche nei

gesti mentre un'autocorrezione può correlare con una ripetizione del gesto prodotto durante l'articolazione vocale della parte "corretta".

Manca ancora una visione globale e dinamica del contemporaneo sviluppo della produzione vocale e gestuale nei punti della comunicazione in cui i parlanti intervengono sotto la spinta della propria attività di controllo (*monitoring*, Levelt 1989). Nel tentativo di raggiungere tale scopo Kosmala (2021, 2022) propone di descrivere fenomeni di disfluenza identificandone i *marker* nel contesto d'occorrenza rispetto a più livelli di analisi (vocale, sintattica, gestuale) e supportandone l'analisi quantitativa con una più dettagliata analisi qualitativa.

Per rispondere all'esigenza di individuare sistematicità nella relazione fra fenomeni gestuali e vocali e tenerne in considerazione il carattere continuo, questo studio avanza una proposta metodologica basata su un'osservazione di sequenze regolari di co-occorrenza che permetta di ricostruire *pattern* di sviluppo della produzione vocale e gestuale in prossimità di disfluenze.

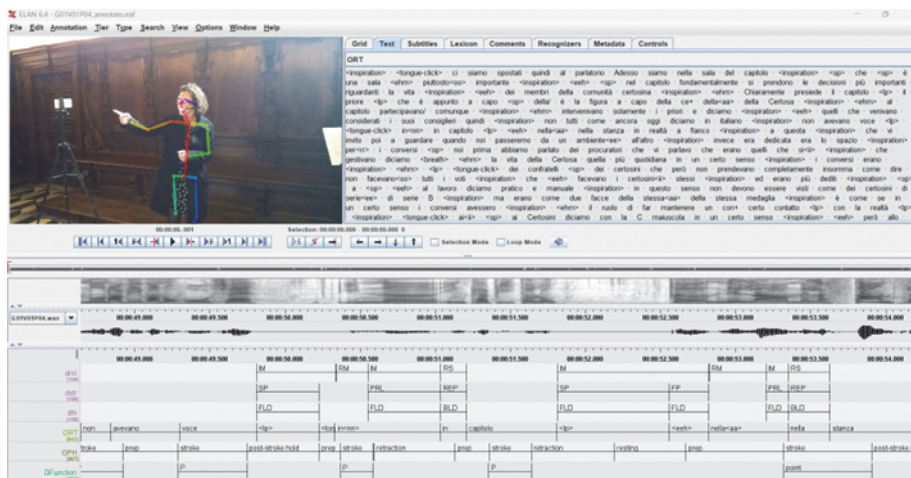
3. Metodo

3.1 Dati

Lo studio che presentiamo si basa sull'osservazione delle regolarità di co-occorrenza di gesti e fenomeni di disfluenza presenti in dati estratti dal corpus CHROME (Origlia *et al.* 2018) il quale è stato raccolto nell'ambito del Progetto PRIN CHROME (*Cultural Heritage Resources Orienting Multimodal Experiences*) per le analisi di dati multimodali da cui derivare modelli di comportamento comunicativo applicabili alla progettazione di agenti virtuali per la fruizione di beni storico-artistici. L'analisi riguarda quindi uno specifico tipo di parlato prodotto in un contesto comunicativo reale, quello delle guide turistiche. Si tratta di parlato per lo più monologico caratterizzato da un elevato grado di pianificazione del discorso e dalla relazione asimmetrica tra la guida e uditorio.

Il corpus consiste di registrazioni audio-video di visite guidate tenute da tre guide esperte presso la Certosa di San Martino a Napoli. Ciascuna guida (G) tiene 4 visite (V) della durata di circa un'ora suddivise in sei punti di interesse (P: *il pronao; il chiostro grande; il parlatorio; la sala del capitolo; il coro; la cappella del tesoro*). Per questo studio esplorativo è stata considerata la registrazione audio-video della descrizione della sala del capitolo tenuta dalla prima guida durante la sua prima visita (G01V01P04). Quest'ultima è composta da 673 parole pronunciate in 5 minuti e 46 secondi (circa 117 parole al minuto). Su questi dati sono state condotte l'annotazione dei fenomeni di disfluenza e quella dei gesti co-verbali in maniera indipendente secondo i criteri descritti nei seguenti paragrafi (cfr. §§ 3.1.1 e 3.1.2). Come strumento di lavoro è stato adottato il software ELAN (2020, Sloetjes, Wittenburg 2008) il quale permette di operare agevolmente annotazioni multilivello di file audio-video (Figura 1).

Figura 1 – Esempio di annotazione multilivello condotta in ELAN



3.2 Annotazione

3.2.1 Annotazione dei fenomeni di disfluenza

L'annotazione dei fenomeni di disfluenza è stata svolta dalla prima autrice considerando solo il segnale acustico e seguendo uno schema articolato su più livelli per integrare l'identificazione delle strutture formali dei fenomeni (Shriberg 1994, Eklund 2004) e delle loro funzioni nel contesto di occorrenza (Ginzburg et al. 2014). Lo schema di annotazione comprende i seguenti livelli (Schettino 2022):

1. *Disfluency Model (dml)*, sul primo livello vengono individuate le regioni in cui può articolarsi un evento disfluente, ovvero la regione della stringa pronunciata che viene corretta (Reparandum, RM), quella dove si realizza la correzione (Reparans, RS) e quella in cui si realizza la sospensione dell'articolazione del messaggio (Interregnum, IM, cfr. Shriberg, 1994);

2. *Disfluency Structure (dstr)*, sul secondo livello vengono individuati i tipi di struttura che realizzano le regioni disfluente (Eklund, 2004), ovvero inserimenti (*Insertion, INS*), cancellazioni (*Deletion, DEL*), sostituzioni (*Substitution, SUB*), ripetizioni (*Repetition, REP*), pause silenti (*Silent Pause, SP*), allungamenti segmentali (*Prolongation, PRL*), pause piene (*Filled Pause, FP*), pausa piene lessicalizzate (*Lexicalized Filled Pause, LFP*);

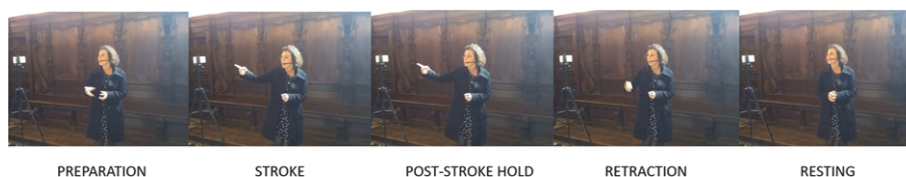
3. *Disfluency Function (dfn)*, sul terzo livello, a ogni elemento disfluente viene assegnata la sua funzione principale, retrospettiva (*Backward-Looking Disfluency, BLD*) o prospettiva (*Forward-Looking Disfluency, FLD*, Ginzburg et al. 2014).

L'identificazione di questi fenomeni non si basa su misure assolute, ma sul contesto d'occorrenza seguendo un approccio pragmatico (Lickley 2015), il che implica un giudizio percettivo soggettivo. Pertanto, l'affidabilità del sistema è stata testata misurando l'accordo fra due annotatori esperti attraverso il calcolo del Kappa di Cohen ($\kappa = 0,87$, *good/almost perfect agreement*, Landis, Koch 1977).

3.2.2 Annotazione dei gesti

L'annotazione dei gesti è stata effettuata dalla seconda autrice seguendo Kendon (1980; 2004). Il flusso gestuale è stato segmentato in un primo livello, senza accesso al parlato, in *gesture phrases* (GPH, sulle ragioni per cui abbiamo scelto di non tradurre il termine in italiano si veda Cutugno, Campisi 2022), composte da una fase di *preparazione* (*preparation*), in cui le mani (o la mano) iniziano a muoversi dalla posizione di riposo e ad assumere una certa configurazione, e da uno *stroke*, definito come “la fase dell’escursione in cui le dinamiche del movimento di “sforzo” e di “configurazione” si manifestano con la massima chiarezza” (Kendon 2004, p. 112; trad. nostra). Alle fasi di preparazione e di *stroke* sono state aggiunte una fase di *retroazione* (*retraction* o *recovery*) e una di *riposo* (*resting*), di lunghezza variabile. La seguente sequenza esemplifica le fasi descritte (Figura 2).

Figura 2 – Esempio di segmentazione del flusso gestuale



Gli *stroke* individuati sono stati ulteriormente analizzati in un livello successivo, a seconda della funzione prevalente che assolvono, in gesti con funzione *referenziale*, gesti con funzione *pragmatica*, gesti *deittici* e gesti *batonici* (*beats*). I gesti con funzione referenziale sono quei gesti che “forniscono una rappresentazione di un aspetto del contenuto di un enunciato” (Kendon 2004, p. 160, traduzione nostra): essi possono, ad esempio, modellare la forma di un referente oppure riprodurre un’azione. I gesti con funzione pragmatica, invece, non forniscono una rappresentazione, ma agiscono sull’enunciato, ad esempio fornendo indicazioni su come esso vada interpretato oppure marcando la struttura del discorso (Kendon 2004, p. 225). I gesti deittici proiettano un “vettore, cioè un segmento orientato e dotato di una direzione e di un verso, che usiamo principalmente allo scopo di attirare l’attenzione di qualcuno su qualcosa” (Campisi 2018, p. 38). Infine, i *beats* (o gesti batonici) sono movimenti bifasici, solitamente senza alcuna funzione prevalente se non quella di marcare il ritmo del discorso a cui si accompagnano (McNeill 1992).

3.3 Frequenze d’occorrenza

In primo luogo, è stata svolta un’analisi quantitativa delle occorrenze:

- dei fenomeni di disfluenza, considerandone la frequenza degli elementi disfluenti (annotati sul piano strutturale, ovvero inserimenti, cancellazioni, sostituzioni, pause, ...) e della loro funzione principale (prospettiva e retrospettiva);
- dei gesti co-verbali, considerandone le fasi di articolazione (*gesture phrases*) e funzioni prevalenti associate a ciascuna unità;

- dei casi di sovrapposizione temporale fra unità disfluenti e gestuali (culminanti nella fase di *stroke*).

3.4 Analisi dei pattern

Successivamente, sulla base delle annotazioni, sono state analizzate le regolarità nelle sequenze gestuali considerando la loro co-occorrenza con fenomeni di disfluenza. A questo scopo, il livello di annotazione delle fasi gestuali (GPH) è stato unito al livello di annotazione delle regioni disfluenti (dml) in un unico livello in cui nei casi di sovrapposizione viene aggiunta, come suffisso, l’etichetta della regione disfluente all’etichetta gestuale. Per cui alle etichette relative alle fasi gestuali sovrapposte a disfluenze è aggiunto il suffisso “IM”, in caso di sovrapposizione a un intervallo *Interregnum*, “RM”, in caso di sovrapposizione a un intervallo *Reparandum* o “RS”, per le sovrapposizioni con un intervallo *Reparans*. Le sequenze di etichette presenti su questo nuovo livello integrato sono state analizzate considerando i loro bigrammi, ossia le coppie di etichette gestuali che occorrono consecutivamente. A partire dalla frequenza di ciascun bigramma osservato, è stata generata una matrice di transizione per rappresentare la probabilità di osservare una transizione da un’etichetta a tutte le altre. Dalla matrice di transizione è possibile stimare le sequenze di etichette più probabili che si possono osservare nel materiale annotato, il che permette di rappresentare la sovrapposizione discorso-gesto nel flusso della produzione in corso.

4. Risultati

4.1 Disfluenza bimodale: frequenze d’occorrenza

Nel dataset di riferimento sono stati individuati 141 fenomeni di disfluenza. Come riportato in Tabella 1 (e illustrato dalle colonne blu in Figura 4), si tratta per lo più di pause piene (34%), seguite da allungamenti segmentali (20%), pause silenti (18%), ripetizioni (14%), mentre decisamente meno frequenti risultano casi di sostituzione (4%), inserzione (2%), cancellazione (1%).

Sul piano funzionale, il numero di fenomeni cui è attribuita funzione prospettiva (113, 80%) è nettamente superiore a quello dei fenomeni con funzione retrospettiva (28, 20%).

Tabella 1 – Frequenza d’occorrenza – numero (percentuale sul totale) – dei diversi tipi di fenomeni di disfluenza individuati

	DEL	SUB	INS	REP	SP	FP	PRL	LFP
N(%)	1(1%)	5(4%)	3(2%)	20(14%)	25(18%)	47(34%)	28(20%)	9(7%)

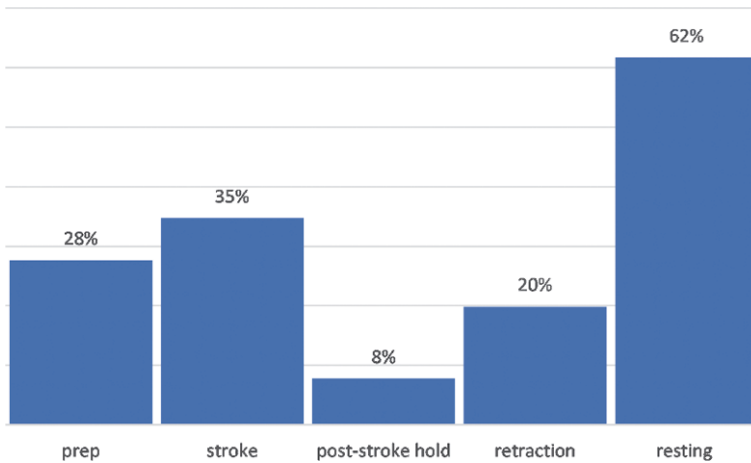
Per quanto riguarda i gesti, sono stati identificati 123 *stroke*, preceduti da altrettante fasi di preparazione e seguiti nel 43% dei casi da fasi di *retraction* (53) e nel 24% dei

casi da fasi di *post-stroke hold* (29). Inoltre, sono stati annotati 41 intervalli corrispondenti a fasi di riposo. A gran parte degli *stroke* è associata funzione pragmatica (74%), un quinto di essi è costituito da gesti deittici (20%) mentre sono state individuate pochissime occorrenze di gesti referenziali (3%) e *beat* (3%).

Passando alla considerazione dei casi di co-occorrenza fra fasi gestuali e fenomeni di disfluenza, la maggior parte delle fasi di *stroke* non si sovrappone ad alcun fenomeno di disfluenza (64%).

Viceversa, si riscontra che nella maggior parte dei casi i fenomeni di disfluenza occorrono in corrispondenza di fasi di riposo (62%), mentre risultano in minoranza casi di accavallamento con fasi di preparazione, raggiungimento dell'apice e ritrazione del gesto (Figura 3). Si noti che le percentuali riportate in Figura 3 sono calcolate sul numero di fenomeni di disfluenza e superano il 100% in quanto ciascun fenomeno può sovrapporsi a più di una fase gestuale.

Figura 3 – Frequenza di co-occorrenza (%) di tutti i fenomeni di disfluenza con le diverse fasi gestuali



Le figure 4 e 5 illustrano la distribuzione delle sovrapposizioni fra fasi gestuali e disfluenze suddivise per tipo (Figura 4) e funzione principale (Figura 5). Per quel che riguarda il tipo, concentrandosi su quelli relativamente più frequenti, ovvero ripetizioni, pause silenziose, pause piene e prolungamenti, è interessante notare che le ripetizioni si distinguono in quanto i casi di sovrapposizione riguardano quasi in egual misura fasi di riposo e *stroke*, mentre non si riscontra co-occorrenza con fasi di preparazione. Pause silenziose, piene e prolungamenti rispettano le tendenze generali (cfr. Figura 3), ma nel caso delle pause piene le occorrenze non sovrapposte a fasi gestuali di riposo, co-occorrono per lo più con fasi di ritrazione o preparazione seguite da *stroke*. Considerando poi i fenomeni di disfluenza raggruppati per funzione principale ad essi associata, si nota che mentre i fenomeni con funzione prospettiva seguono le tendenze generali, o piuttosto contribuiscono significativamente a definire queste tendenze considerata la loro frequenza d'occorrenza nettamente superiore a

quella dei fenomeni con funzione retrospettiva, per questi ultimi non si riscontra la preferenza di sovrapposizione con fasi di riposo, bensì i casi di sovrapposizione riguardano in pari misura fasi di preparazione, *stroke* e riposo.

Figura 4 – Frequenza di occorrenza (%) dei fenomeni di disfluenza e di co-occorrenza con le diverse fasi gestuali per tipo

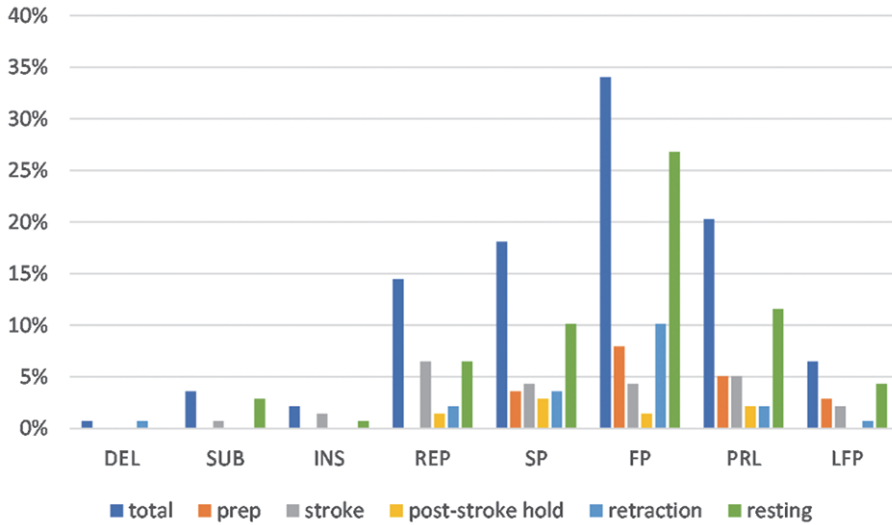
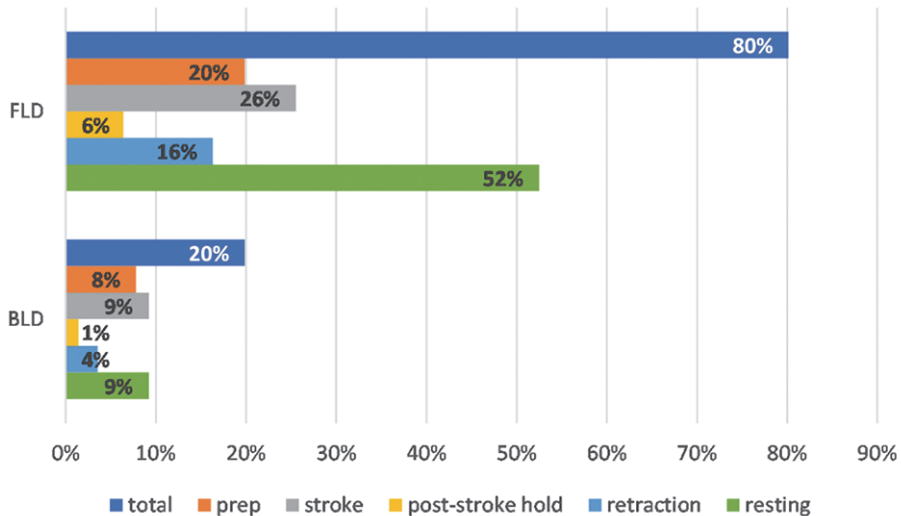


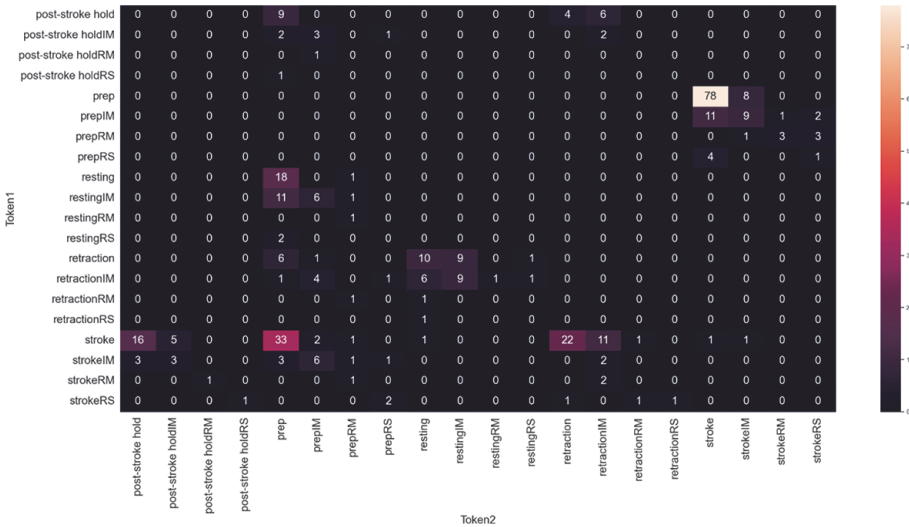
Figura 5 – Frequenza di occorrenza (%) dei fenomeni di disfluenza e di co-occorrenza con le diverse fasi gestuali per funzione



4.2 Analisi dei Pattern di disfluenza in prospettiva bimodale

Per rappresentare la corrispondenza fra fenomeni di disfluenza e fasi gestuali nel flusso della produzione in corso è stata utilizzata una *heatmap* (Figura 6) in modo da mettere in evidenza il grado di frequenza di osservazione delle sequenze da un elemento (*Token 1*) all'altro (*Token 2*). Non sorprende che le sequenze più frequenti riguardano le fasi componenti *gesture phrases*, a partire dalla sequenza *preparation -> stroke* (78), seguita da *stroke -> preparation* (33), *stroke -> retraction* (22), *retraction -> resting* (10), *resting -> preparation* (18), *stroke -> post-stroke hold* (16). Come atteso, le sequenze con casi di sovrapposizione fra fenomeni di disfluenza e fasi gestuali sono meno frequenti e riguardano in particolare fasi di ritrazione e fasi di riposo sovrapposte a sospensioni nell'articolazione del discorso. I primi casi (*retractionIM*) sono relativamente più spesso preceduti da *stroke* (*stroke -> retractionIM*, 11) e seguiti da *restingIM* (*retractionIM -> restingIM*, 9), questi ultimi casi sono relativamente più frequentemente seguiti da fasi di preparazione fuori la fase di sospensione del messaggio (*restingIM -> prep*, 11) o al suo interno (*restingIM -> prepIM*, 6), questi ultimi casi poi sono più spesso seguiti da *stroke* dopo la sospensione (*prepIM -> stroke*, 11), o all'interno della stessa sospensione del messaggio (*prepIM -> strokeIM*, 9).

Figura 6 – Heatmap rappresentante la matrice di transizione. I valori indicano il numero di osservazioni per le sequenze da un elemento all'altro (Token1 -> Token2).



5. Conclusioni

L'analisi della frequenza dei fenomeni di disfluenza nel dataset di riferimento, in linea con quanto osservato in studi precedenti sul parlato di guide turistiche (Schettino 2022), evidenzia una piuttosto bassa esigenza di autocorrezione (can-

cellazione, sostituzione o inserzione di parti del messaggio) a fronte di un uso frequente di elementi come pause piene, prolungamenti segmentali o pause silenti per sospendere l'articolazione del messaggio e guadagnare tempo per la sua pianificazione. La frequenza di questi fenomeni nel parlato, in una certa misura pre-strutturato, di esperti del mestiere ne suggerisce il valore positivo di supporto al raggiungimento dello scopo comunicativo.

Per quanto riguarda i gesti, alla maggior parte degli *stroke* individuati è associato un valore pragmatico, si tratta quindi di gesti impiegati per fornire all'interlocutore indicazioni sulla struttura del discorso e sul modo in cui esso vada interpretato.

Come atteso, solo una minoranza delle fasi di *stroke* si sovrappone a fenomeni di disfluenza, i quali co-occorrono per lo più con fasi di riposo. Tuttavia, è interessante notare come i casi di sovrapposizione tendano a differire per tipo e funzione principale dei fenomeni di disfluenza interessati. Il che può evidenziare differenze nel ruolo dei fenomeni di disfluenza nel processo di produzione e gestione del messaggio a seconda della loro forma (tipo di elementi) e funzione.

Inoltre, la matrice di transizione permette di evidenziare regolarità nella sequenza di fasi gestuali e fenomeni di disfluenza co-occorrenti. In questo caso specifico, emerge una certa regolarità dei casi in cui la guida compie un gesto, poi il movimento è ritratto e si ferma durante una pausa di pianificazione per poi riprendere successivamente (casi di *stroke*, seguiti da posizioni di ritrazione e poi di riposo durante una sospensione nell'articolazione del messaggio, seguiti da una fase di preparazione). Si osserva poi un minor numero di casi in cui alla sospensione nell'articolazione si sovrappongono una fase di preparazione e lo *stroke* successivo.

Pur consapevoli della quantità limitata dei dati analizzati e dell'esigenza di considerare un dataset di riferimento più ampio e vario per poter valutare il grado di generalizzabilità delle osservazioni, questo studio ha permesso di testare il metodo di analisi proposto e l'efficacia della sua capacità rappresentativa per l'analisi di fenomeni di disfluenza in prospettiva bimodale, sul piano fono-articolatorio e gestuale, e del loro sviluppo nel flusso continuo della comunicazione parlata.

Ringraziamenti

Lavoro svolto con il supporto del progetto PRIN "Cultural Heritage Resources Orienting Multimodal Experience (CHROME)" #B52F15000450001.

Bibliografia

- Allwood, Jens & Nivre, Joakim & Ahlsén, Elisabeth. 1990. Speech management – on the non-written life of speech. *Nordic Journal of Linguistics* 13, 3-48. <https://doi.org/10.1017/S0332586500002092>
- Bortfeld, Heather & Leon, Silvia D. & Bloom, Jonathan E. & Schober, Michael. F. & Brennan, Susan E. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech* 44 (2), 123-147.

- Butterworth, Brian & Beattie, Geoffrey. 1978. Gesture and Silence as Indicators of Planning in Speech. In Campbell, Robin N. & Smith, Philip T. (a cura di), *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, 347-360. Boston: Springer.
- Campisi, Emanuela. 2018. *Che cos'è la gestualità*. Roma: Carocci.
- Chafe, Wallace L. 1980. Some reasons for hesitating. In Dechert, Hans W. & Raupach, Mangred (a cura di), *Temporal variables in speech*, 169-180. Berlin, New York: De Gruyter Mouton.
- Crible, Ludivine. 2018. *Discourse markers and (dis)fluency: Forms and functions across languages and registers*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Cutugno, Francesco & Campisi, Emanuela. 2022. Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME. In Saturno, Jacopo & Spreafico, Lorenzo (a cura di), *Fare linguistica applicata con le Digital Humanities*, 9-25. Milano: Officinaventuno.
- de Ruijter, Jan Peter. 2000. The production of gesture and speech. In McNeill, David (a cura di), *Language and gesture*, 284-311. Cambridge, UK: Cambridge University Press.
- de Ruijter, Jan Peter. 2007. Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture* 7(1), 21-38.
- Eklund, Robert. 2004. *Disfluency in Swedish human-human and human-machine travel booking dialogues* (Doctoral dissertation). Linköping University Electronic Press.
- Ginzburg, Jonathan & Fernández, Raquel & Schlagen, David. 2014. Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics* 7 (9), 1-64. <https://doi.org/10.3765/sp.7.9>
- Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Graziano, Maria & Gullberg, Marianne. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology* 9, 879.
- Hieke, Adolf E. 1981. A content-processing view of hesitation phenomena. *Language and Speech* 24 (2), 147-160.
- Hostetter, Autumn B. & Alibali, Martha W. 2008. Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review* 15, 495-514.
- Kendon, Adam. 1980. Gesticulation and speech. Two aspects of the process of utterance. In Key, Mary R. (a cura di), *The relation of verbal and non-verbal communication. Contributions to the Sociology of Language* 25, 207-227. The Hague: De Gruyter Mouton.
- Kendon, Adam. 2004. *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kita, Sotaro, & Özyürek, Asli. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48, 16-32.
- Kosmala, Loulou. 2021. *A multimodal contrastive study of (dis)fluency across languages and settings: Towards a multidimensional scale of inter-(dis)fluency* (Tesi dottorale). Sorbonne Nouvelle.

- Kosmala, Loulou. 2022. Rethinking (dis)fluency within the scope of interactional linguistics and gesture studies. *Studia ubb. philosophia* 67(2), 49-66. <https://doi.org/10.24193/subbphil.2022.2.03>
- Krauss, Robert M. & Hadar, Uri. 1999. The role of speech-related arm/hand gestures in word retrieval. In Messing, Lynn S. & Campbell, Ruth (a cura di), *Gesture, speech and sign*, 93-116. Oxford: Oxford University Press.
- Landis, Richard J., & Koch, Gary. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
- Levelt, Willelm J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14 (1), 41-104.
- Levelt, Willelm J. M. 1989. *Speaking: From intention to articulation*. Cambridge: MIT Press. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.
- Origlia, Antonio & Savy, Renata & Poggi, Isabella & Cutugno, Francesco & Alfano, Iolanda & D'Errico, Francesca & Vincze, Laura & Cataldo, Violetta. 2018. An audiovisual corpus of guided tours in cultural sites: Data collection protocols in the chrome project. 2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage, 2091, 1-4.
- Schettino, Loredana. 2022. *The Role of Disfluencies in Italian Discourse. Modelling and Speech Synthesis Applications*. (Tesi dottorale). Università degli Studi di Salerno, Salerno.
- Seyfeddinipur, Mandana. 2006. *Disfluency: Interrupting speech and gesture* (Tesi dottorale). Radboud University Nijmegen, Nijmegen. doi:10.17617/2.59337.
- Shriberg, Elizabeth E. 1994. *Preliminaries to a theory of speech disfluencies* (Tesi dottorale). University of California.
- Shriberg, Elizabeth E. 2001. To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the international phonetic association* 31 (1), 153-169.
- Slotjes, Han & Wittenburg, Peter. 2008. Annotation by category-ELAN and ISO DCR. In Proceedings of the 6th international Conference on Language Resources and Evaluation (LREC 2008), 816-820.
- Voghera, Miriam. 2017. *Dal parlato alla grammatica*. Roma: Carocci.
- Wingate, Marcel E. 1987. Fluency and disfluency; illusion and identification. *Journal of Fluency Disorders* 12 (2), 79-101.