LIX Congresso SLI, Bologna, 10-12/9/2026

Johanna Miecznikowski (USI)
Elena Battaglia (Università di Bologna)

Il corpus TIGR: una risorsa multimodale per studiare l'italiano parlato

Studiare il parlato in una prospettiva interazionale

La linguistica interazionale esamina il linguaggio, soprattutto il parlato, tenendo conto del **ruolo costitutivo dell'interazione sociale organizzata sequenzialmente** (Mondada 2001, Couper-Kuhlen & Selting 2017). Seguendo l'insegnamento etnometodologico (Fele 2002), l'analisi cerca di rimanere vicina al punto di vista delle/i partecipanti. Essa è attenta alla temporalità, la prosodia, l'alternanza dei turni e lo sviluppo sequenziale in quanto fenomeni che influiscono sulla formazione e interpretazione di azioni e strutture linguistiche.

L'integrazione della **dimensione incorporata** è uno sviluppo naturale di questo paradigma. Essa presuppone almeno la **videoregistrazione** degli incontri studiati (p.es. Mondada 2013, Deppermann 2014, Pekarek Doehler 2022), con possibile aggiunta di dati rilevati da dispositivi come i sensori indossabili. I dati video permettono di osservare aspetti multimodali come gli sguardi, la mimica, i gesti, l'orientamento e il movimento del corpo e la manipolazione di oggetti. Mettono in prospettiva molti temi rilevanti nello studio del parlato e incoraggiano a precisare la nozione stessa di sequenzialità (Mondada, Tekin, Koda 2025).

Nel contesto di una scienza cumulativa e sostenibile, e in vista di indagini diacroniche e comparate, è auspicabile la raccolta e condivisione di **videocorpora** per la linguistica interazionale, in dialogo con il paradigma esistente della linguistica dei corpora testuali e orali audioregistrati. Questa impresa solleva tuttavia problemi metodologici a più livelli, che qui illustriamo discutendo l'esempio del corpus di italiano parlato **TIGR** (Miecznikowski, Battaglia Geddo 2026, Miecznikowski *et al*. in corso di valutazione).

Il TIGR: dati chiave

- Un corpus raccolto per studiare l'evidenzialità nel parlato (Battaglia 2025, Geddo 2025)
- Parlanti e istituzioni in Ticino e nel Grigioni contattati in 2021-2022 tramite mail e passaparola
- Dichiarazioni di consenso e risposta a un breve questionario sociolinguistico prima della registrazione
- Audio/video-registrazione di 23 eventi per un totale di 23.5 ore e 223'076 parole
- Trascrizione in ELAN secondo GAT 2, tokenizzazione, conversione in vari formati
- Misure di de-identificazione nei vari sotto-insiemi di dati
- Condivisione su LaRS @ SWISSUbase per il riuso a scopo scientifico.

Il TIGR rappresenta una novità nel panorama italiano dove i principali corpora accessibili di parlato sono audioregistrati (p.es. LIP, De Mauro *et al.* 1993; CLIPS, Albano Leoni 2007; KIParla, Ballarè *et al.* 2022).

Impostazione delle riprese audio/video

In vista di analisi sequenziali e multimodali dettagliate, **la registrazione deve assicurare una buona qualità tecnica** e al contempo, nonostante la presenza di ricercatori e di apparecchi, mantenere il quadro complessivo di un'**interazione autentica**, che è gestita dalle/i partecipanti e ha senso per loro.

Nel TIGR la qualità tecnica delle riprese video è elevata (**full HD**), anche se la quantità d'informazione è limitata dall'uso di sole due angolazioni. La qualità del suono è molto alta grazie all'uso di **registratori tascabili con microfoni da bavero** + microfoni ambiente e grazie alla **sincronizzazione** di tutti i dispositivi all'inizio dell'evento mediante generatori di timecode. Il **gruppo di ricerca** era presente ai margini dell'evento per montare e smontare i dispositivi e far compilare i documenti cartacei. Si fermava in aula durante le interazioni didattiche, mentre si allontanava durante gli altri eventi, tranne l'intervistatore/trice nel caso delle interviste (per una vista d'insieme dei generi d'interazione v. sotto).

Le/i partecipanti prestano attenzione a dispositivi, ricercatori e futuri utenti del corpus, come mostrano gesti e commenti che si riferiscono alla registrazione p.es. in termini di contributo alla scienza, presentazione di tradizioni locali a un pubblico interessato o *reality show.* Tali categorie influiscono sulle scelte linguistiche e stilistiche. Non dominano però il *framing* complessivo degli eventi, anche perché l'attenzione delle persone dalla registrazione è diminuita dal loro coinvolgimento nell'interazione principale e dall'assuefazione.

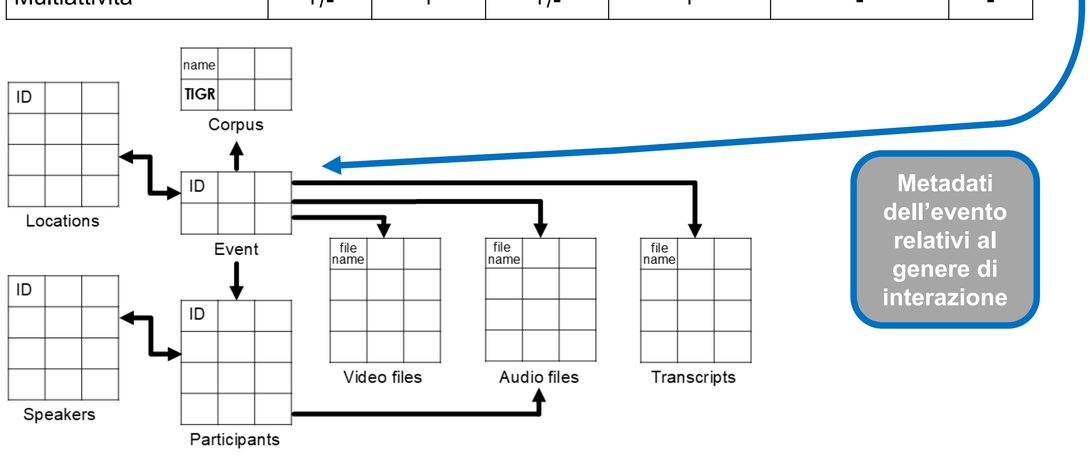
Organizzazione del corpus e metadati

I videocorpora di interazioni autentiche condividono con corpora audioregistrati simili l'**evento sociale** come unità centrale. Possono essere organizzati secondo categorie relative ai parlanti o basate su tratti situazionali – categorizzazioni inevitabilmente semplificanti, ma utili per un primo approccio ai dati.

Il ***corpus design*** del TIGR privilegia i tratti situazionali. Gli eventi sono classificati come istanze di **generi di interazione**, che differiscono quanto a parametri sociali e parametri relativi alle attività pratiche (p.es. la presenza di multiattività). Questi ultimi, poco considerati nella linguistica dei corpora, diventano osservabili nei videocorpora e si rivelano pertinenti a vari livelli, anche perché influiscono sulle dinamiche prosodiche e sequenziali (cfr. Miecznikowski *et al.* in corso di valutazione).

I generi di interazione si ritrovano nella descrizione degli eventi tramite i **metadati**. Questa descrive inoltre le parti costitutive di ogni evento, cioè i suoi **partecipanti** e i **documenti** associati, mentre il **luogo di registrazione** è determinato tramite riferimento a una lista di luoghi. Le caratteristiche sociobiografiche dei partecipanti sono descritte tramite rinvio alla tabella dei **parlanti**, che registra i dati rilevati tramite questionario; la relazione tra parlanti e partecipanti può essere uno-a-uno o uno-a-molti.

Generi d'interazione / Tratti	Conversazioni a tavola	Preparazione di cibo	Interazioni didattiche: Tutoring in architettura	Interazioni didattiche: Lezioni di musica, teatro, restauro	Interazioni didattiche: Lezioni di cultura generale e di didattica	Interviste
Istituzionalità	-	-	+	+	+	+
No. di partecipanti	3-4	4	2-3	6-11	8-19	2
Ruoli asimmetrici	+/-	+/-	+	+	+	+
Persone in movimento	+/-	+	+/-	+	+/-	-
Manip. di oggetti	+	+	+	+	+/-	-
Multiattività	+/-	+	+/-	+	-	-



Accessibilità del corpus e protezione dei dati personali

Per rispettare i principi etici e le leggi vigenti, **i dati personali delle/i partecipanti vanno protetti** nella misura del possibile e compatibilmente con i principali scopi scientifici del corpus. Il **consenso informato** delle persone coinvolte ne è la più importante garanzia. Concorrono inoltre le misure di **deidentificazione** parziale e una **calibrazione prudente delle modalità di condivisione** (Profazi & Miecznikowski 2023).

Nel TIGR i dati personali sono stati protetti come segue:

- Dichiarazioni di consenso che limitano l'uso dei dati a scopi scientifici, con opzioni di deidentificazione.
- Misure di deidentificazione:
  - Nei file video, effetto di sfocatura gaussiana per mascherare alcune facce, loghi, schermi, passanti.
  - Nei file audio, alterazione di alcune voci, silenziamento di conversazioni con passanti esterni.
  - Nei file audio, silenziamento di nomi propri ecc. Procedura: annotazione dei segmenti da silenziare in ELAN, importazione del loro timecode in Praat per silenziare i passi audio corrispondenti.
  - Nelle trascrizioni, sostituzione di nomi di persone, istituzioni, luoghi, date con pseudonimi.
  - Nei dati sociolinguistici, anonimizzazione di nomi e luoghi, sostituzioni di informazioni specifiche con categorie più generiche (p.es. regioni geografiche, fasce d'età).
- Scaricamento dei dati dal repositorio LaRS con licenza che limita l'uso a soli scopi di ricerca scientifica.

Trascrizione

La **trascrizione dettagliata dei fenomeni verbali, prosodici e multimodali** e il loro **ancoraggio preciso nel tempo** sono prerequisiti per l'analisi sequenziale. La trascrizione è però anche un atto di **interpretazione** e varia secondo gli scopi di ricerca. Perciò in un videocorpus destinato al riuso conviene fornire una **trascrizione di base** che può essere modificata da altri ricercatori secondo i bisogni.

Per rispondere a questa esigenza nel TIGR sono state fatte le scelte seguenti:

- I dati sono stati trascritti manualmente in ELAN secondo le **convenzioni GAT 2 per la trascrizione fine** (Selting et al. 2011), fornendo dunque una quantità relativamente alta di informazione prosodica. Ciò era necessario entro il progetto InfinIta. Chi riusa i dati è libero di semplificare o correggere la trascrizione.
- L'informazione temporale in ELAN si inserisce definendo dei **segmenti ancorati a una linea del tempo**. Nel TIGR la segmentazione è stata fatta così da ottimizzare il processo di trascrizione, osservando alcuni principi: (a) garantire una granularità di base dell'informazione temporale evitando segmenti molto lunghi; (b) indicare con precisione i passi sovrapposti tra due o più parlanti non solo tramite riferimenti testuali (parentesi quadre), ma anche definendo confini di segmenti ancorati alla timeline; (c) per non complicare la tokenizzazione, evitare confini di segmenti all'interno di parole. In conversazioni con più partecipanti e molte sovrapposizioni o addirittura scissione in più sottogruppi, sono stati frequenti i conflitti tra (b) e (c).
- Una traccia in ELAN è stata riservata all'annotazione di **fenomeni udibili nell'ambiente**.
- **Non è invece stata fornita nessuna trascrizione multimodale**. Una trascrizione multimodale può essere aggiunta da chi usa il corpus (per un esempio trascritto secondo Mondada 2018 v. l'estratto *infra*).

Elaborazione dei documenti per vari tipi di analisi

La linguistica interazionale adotta **una pluralità di procedure analitiche** perlopiù qualitative, che includono compiti come l'esplorazione del corpus sulla base dei metadati, la visione e lettura di documenti audio/video e di trascrizioni, l'ulteriore annotazione o trascrizione multimodale di documenti ed estratti, l'interrogazione automatica di trascrizioni e annotazioni, la selezione e il raggruppamento di estratti per formare collezioni di casi simili. Tali compiti si eseguono con l'aiuto di diversi strumenti digitali (cfr. Frick & Schmidt 2025), fra i quali ambienti integrati che allineano documenti audio/video e trascrizioni offline (p.es. ELAN) o online (p.es. CLAPI, v. Baldauf-Quillatre et al. 2016, AGD presso il Leibniz-Institut für Deutsche Sprache, ZuMult), ma anche svariate applicazioni per l'elaborazione e l'analisi di singoli tipi di documenti.

In vista del **riuso dei dati in IL**, i documenti del TIGR sono perciò stati convertiti in **formati interoperabili** (Miecznikowski et al. in corso di valutazione):

- **Dalle trascrizioni EAF originali è stata derivata una versione TXT in stile dialogo teatrale con una riga di timecode all'incirca ogni 10 secondi**, adottando una procedura che alterna l'elaborazione automatica, interventi manuali e la revisione manuale assistita da script (TIGRformat 2025). TXT codifica poca informazione esplicitamente, ma è compatibile con molti programmi, inclusi software di annotazione come p.es. INCEpTION (Klie et al. 2018). L'impaginazione dialogica è più compatta della visualizzazione a spartito in ELAN ed è comunemente usata nell'analisi sequenziale del discorso in interazione. Una quantità ridotta di *timecode* facilita, senza appesantire il testo, la navigazione tra testo e documenti audio/video quando non si dispone di software che assicuri l'allineamento testo-video automatico.
- **Le trascrizioni EAF sono state tokenizzate e convertite in documenti XML conformi allo standard TEI/ISO 20624:2016** (Hedeland & Schmidt 2022). Questi conservono le indicazioni temporali originali e le completano per interpolazione. Contengono inoltre i metadati nel *header*, trasformano in elementi o attributi XML l'informazione codificata dalle convenzioni GAT 2 e raggruppano segmenti consecutivi pronunciati dallo/a stesso/a parlante in contributi, ottenendo una struttura dialogica. Il formato XML scelto è congruo con le convenzioni TEI, ampiamente usate nella linguistica dei corpora, ed è ricco di informazione esplicita leggibile dalle macchine. Si presta all'interrogazione automatica tramite script, è adatto per l'archiviazione e consente la modellazione dei dati in piattaforme digitali per i corpora.
- **Sulla base delle trascrizioni XML tokenizzate, i dati sono stati modellati per una futura rappresentazione nella piattaforma LCP-Videoscope** (Graën et al. 2024). LCP-Videoscope è un ambiente multimediale integrato online che è stato sviluppato recentemente dal Linguistic Research Infrastructure LiRI presso l'Università di Zurigo. Esso nella sua forma attuale permette la visualizzazione a spartito delle trascrizioni allineate ai documenti audio/video e la loro interrogazione tramite un apposito linguaggio, DQD.
- La ripresa video da due angolazioni e l'uso di più microfoni produce da quattro a otto documenti audio/video per evento. Ciò rende possibile l'ascolto focalizzato su singoli parlanti e la considerazione di dettagli visivi, per esempio in vista di analisi e annotazioni prosodiche e multimodali avanzate oppure per ritrascrivere singoli passi così da descrivere con più precisione l'alternanza dei turni. **Oltre ai documenti singoli, per ogni evento è stato creato un video unico con immagine compresso, schermo diviso e suono miscelato**. Tale versione è più leggera dell'insieme dei documenti audio/video singoli, può essere riprodotta da lettori di uso comune ed è adatta quando è sufficiente una visione globale dell'evento.

Esempio: configurazioni multimodali con *ah* come risorse evidenziali

Il corpus TIGR è stato recentemente utilizzato in uno studio sul segnale discorsivo *ah* come risorsa evidenziale indessicale, cioè che segnala l'acquisizione *in situ* di conoscenza da elementi presenti nell'ambiente o processi mentali interni e il conseguente cambiamento di stato epistemico nel parlante (cfr. Miecznikowski e Battaglia in corso di valutazione). Tale interpretazione è resa possibile dallo scrutinio dei dati in una procedura analitica di ispirazione conversazionale, facilitata non solo dalla mera disponibilità delle videoregistrazioni, ma anche dalle caratteristiche tecniche del corpus (formato delle trascrizioni, allineamento temporale ecc.).

ev17a, 00:27:58.832 (tutoring in architettura, discussione di un progetto e modello di Clara)

```
Assistente  #+(2.63) ±secondo me; ↑o                ±lo lo ingrandisci come fai di sopra, (1.68)
             +gz plan------------------------------------------------------------------------------>
                    ±L points twd upper part of plan ±R draws on plan-------------------------->
fig          #1
            e allora diventa %questo%          l'appartamento, (0.77)
            ---------------------------------------------------------->
            ---------------------------------------------------------->
clara                          %turns twd model%
Assistente  +±<<dim> tanto non;  ± (.) fa #nie%nte-> ±ah +#NO lì     ±non c'è il +±tetto-% (0.40)
            +gz model-----------------------------------+gz plan----------------+gz model-->
             ±R taps w pen twice ±R palm half up-----±R points model ±R drops down±
clara                                          %.............bends twd model,,,,,,,,,,,,,%
fig                                      #2              #3
Clara       <<p> eh no- >
```

fig. 1

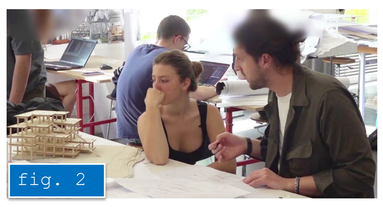


fig. 2

fig. 3

L'estratto sopra illustra l'uso di *ah* concomitante con una verifica percettiva nell'ambiente. L'assistente e la studentessa discutono del progetto di un edificio e, come si vede dalla trascrizione multimodale (aggiunta qui alla trascrizione TXT del TIGR) e dai fermo immagine (Fig. 1-3), si orientano con lo sguardo e i gesti verso il disegno e il plastico. *Ah* introduce un'autoriparazione dell'assistente. Segue l'accesso visivo al plastico, incoraggiato a sua volta da cambiamenti di postura della studentessa che indicano una focalizzazione dell'attenzione verso l'oggetto, ed è accompagnato da un gesto indicale. Ne risulta una configurazione sequenziale e multimodale, peraltro ricorrente nel corpus, di cui si riconosce il valore evidenziale soltanto considerando in maniera integrata i dati provenienti dal video e dalle trascrizioni.

L'esempio mostra come la produzione verbale sia coordinata in maniera fine al comportamento corporeo dei partecipanti e sia ancorata alle attività in corso. A livello teorico, la presa in considerazione dei dati multimodali in questo studio ha contribuito alla rivisitazione della nozione di evidenzialità in relazione con quella di cambiamento di stato (Heritage 1984). Fra l'altro ha illuminato il funzionamento di risorse evidenziali che sono vaghe rispetto al tipo di fonte (percezione, inferenza, sentito dire) e ha resa evidente l'importanza della distinzione tra fonti nel qui e ora e nel passato.

La costruzione e l'uso di videocorpora non rappresenta dunque solo un perno nella metodologia di indagine, ma ha anche ricadute teoriche di ampia portata. Contribuisce infatti al superamento dei limiti di un approccio logocentrico e favorisce la nostra comprensione della lingua come sistema semiotico complesso.

Biblio