

PROTOCOLLO DI TRASCRIZIONE (su ELAN)

Una prima riga (*tier*) indipendente per ogni parlante. La riga andrà nominata utilizzando la sigla alfanumerica del parlante, come concordato per ogni sub-corpus (es. COS01_TRAS).

Si aggiungerà poi una riga “Other_” che dovrà contenere solo indicazioni relativamente a rumori esterni nell’ambiente circostante. L’etichetta [NOISE] sarà inserita in corrispondenza del rumore.

RIGA 1 NOME_TRAS

La riga 1 NOME_TRAS contiene la trascrizione ortografica e conversazionale. Ogni unità intonativa è rappresentata da una singola annotazione.

Le pause (indicate tra parentesi tonde) verranno poi completate automaticamente dal software ELAN. SI intende con pausa, in generale, una unità superiore a 20 msec.

Si segue sempre l’ortografia della lingua corrispondente, rispettando abbreviazioni e troncamenti presenti nell’audio originale.

NOTA: gli elementi contrassegnati con * nell’elenco dei simboli indica che questi elementi devono essere introdotti nella trascrizione con estrema parsimonia e solo qualora non vi siano dubbi, da parte del trascrittore, sulla loro rilevanza sia intonativa sia conversazionale.

Simboli

C’è Luigi. Indica una intonazione discendente*

C’è Luigi? Indica una intonazione interrogativa, anche retorica*

C’è Luigi! Indica chiaramente una esclamazione*

ma: I : indicano l’allungamento inconsueto di un fono.

Ca- Il – indica il troncamento di una parola. Vanno trascritti tutti i foni chiaramente udibili prima del troncamento

/ Isolato (es. *poi / no ecco*) indica un cambio di pianificazione all’interno della stessa unità intonativa. ATTENZIONE: se a questo cambio di pianificazione corrisponde una pausa, non si inserirà il simbolo / ma si spezzerà la frase in due o più unità intonative inframmezzate da una pausa.

#...# Indica parlato sovrapposto a un altro interlocutore. ATTENZIONE: vanno inclusi tra #...# le esatte sillabe, foni o parole che sono sovrapposte

ALTRI FENOMENI

PAUSE PIENE

mh, uhm, ehm **eh, ah, uh**

Attenzione a utilizzare solo questi tipi di pause piene e a distinguere tra pause che contengono una nasale o meno.

Numeri Sono da trascrivere per esteso, in ortografia.

La stessa cosa i simboli che sono da evitare (€ = euro)

[...] Tra parentesi quadre andranno inseriti altri fenomeni indicati con lettere maiuscole, facenti parte della produzione del soggetto da trascrivere.

[XXX] elementi non udibili

[BREATH] sospiri, prese di fiato o emissioni di fiato significative, sbadigli

[CRY] pianto

[LAUGH] risata

[CLICK] effettuato o con la lingua o con le labbra in maniera udibile e saliente

Nota: trascrivere il dialetto

Gli eventuali inserti dialettali andranno trascritti il più possibile vicino alla pronuncia attualmente utilizzata da parte del locutore. Queste parti andranno identificate con una apposita etichetta nel secondo tier di annotazione.

SECONDO TIER DI ANNOTAZIONE – WORD1

Un secondo tier di annotazione andrà creato in maniera automatica tokenizzato il tier precedente. Questo secondo tier sarà come DIPENDENTE dal primo e avrà nome “WORD1”.

Su questo si baserà l’etichettatura linguistica del terzo tier.

L’allineamento parola-audio andrà a essere effettuato in un secondo momento sul software PRAAT. In questa fase, dunque, l’allineamento trascrizione-audio rimarrà ancorato al primo tier.

TERZO TIER DI ANNOTAZIONE - LANG

Un terzo tier di annotazione andrà creato come DIPENDENTE dal secondo e avrà nome “LANG”.

Un vocabolario controllato sarà creato per l'intero progetto, di modo che su LANG possano comparire solo queste etichette:

- ITA, italiano*
- PIEM, piemontese**
- BIEL, biellese**
- SARD***
- SIC***
- VEN***
- LOMB***, lombardo
- ENG, inglese (non si applicano a prestiti ormai comuni es. computer)
- FREN, francese (non si applicano a prestiti ormai comuni)

*questa etichetta è da utilizzare solamente nelle porzioni

**l'etichetta PIEM o BIEL andrà usata a seconda di quanto dichiarato dallo stesso parlante, ossia se il parlante dichiara apertamente di parlare Biellese; in assenza di specifica, sarà indicato con il più generico PIEM.

*** nel caso il parlante fornisca una specifica sulla varietà peculiare di dialetto parlato questa verrà inserita nella stessa annotazione come _Luogo premettendo uno spazio rispetto alla prima etichetta (es. SIC _Trapani)

Le etichette si dovranno quindi legare alla singola parola, in modo da segnalare casi di code-mixing e code-switching tra diversi idiomi.

GESTI (TIER 4-5-6)

I tier da 4 a 6 riguardano invece l'annotazione dei gesti rispettivamente di mani (4°), viso (5°) e altre parti del corpo (6°).

Dato che il legame temporale con la riga di trascrizione potrebbe essere diverso (es. gesto che inizia prima del parlato, oppure che sostituisce interamente il parlato), i tier 4, 5 e 6 saranno tier indipendenti e faranno riferimento solamente ai gesti compiuti dall'intervistato (anche perché l'intervistatore non si vede nella registrazione).

I tier per l'annotazione dei gesti saranno quindi i seguenti

- **TIER 4 “Hand_Gesture”**: contiene una descrizione sintetica del gesto performato con le mani, utilizzando una combinazione di etichette e parole:
 - Tipo di gesto, secondo la classificazione di McNeill (1992) e dividendo quindi tra
 - ICO = iconici
 - META = metaforici
 - DEI = deittici
 - BEAT = batonici
 - Con quale mano viene performato il gesto (etichette: SX, DX, BOTH)

- Il referente indicato dal gesto (tramite parole, es. “movimento”, “saltello”, ecc.); per gesti deittici qualora non sia possibile identificare un referente nominale/pronominale preciso o vi sia incertezza di assegnazione, si etichetterà solamente come “deittico”.
- La direzione del movimento (nel caso non vi sia una direzione, es. per un beat, questa etichetta va lasciata bianca)
 - SX_DX vs. DX_SX = se il gesto va dalla sinistra alla destra del parlante e viceversa;
 - UP_DOWN vs. DOWN_UP = se il gesto va dall’alto al basso o viceversa
- Il luogo in cui viene performed il gesto (es. petto, ginocchio, dx, spalla destra, tutto il corpo, fianco dx ecc.)

- **Tier 5 “Face_Gesture”**

Adattati semplificando dal protocollo FACS

- SOP_UP = sopracciglia alzate
- SOP_DOWN = sopracciglia abbassate/aggrottate
- EYES_OPEN = occhi che si allargano
- EYES_CLOSED = occhi che si chiudono
- MOUTH = qualsiasi gesto che coinvolga la bocca

Se due gesti co-occorrono andranno annotati entrambi, separandoli con uno spazio vuoto: es. “SOP_UP EYES_OPEN”. Andrà riportato per primo il gesto ritenuto come principale: nell’esempio precedente l’innalzamento delle sopracciglia è ritenuto come principale, ossia con funzione comunicativa, rispetto all’apertura degli occhi che ne costituisce quasi una conseguenza fisica.

- **Tier 6 “Other_gesture”**

Qui andranno segnati gesti altri, prodotti con altre parti del corpo e ritenuti però pertinenti. Si avrà cura di segnalare la parte del corpo coinvolta.