



Università
della
Svizzera
italiana



The categorization
of information
sources in
face-to-face
interaction.
A study based on
the TIGR corpus
of spoken Italian

Il corpus TIGR

Questo è un documento informativo sul corpus TIGR. Descrive il progetto di ricerca collegato, la composizione e le finalità del corpus stesso, le modalità di trattamento dei dati raccolti presso gli informatori.

Il progetto di ricerca

Il progetto "InfiNlta - La categorizzazione delle fonti di informazione nell'interazione faccia a faccia. Una indagine basata sul corpus di italiano parlato TIGR" (sussidio FNS n° 100012_192771, 2020-2024) viene condotto presso l'Università della Svizzera italiana da Johanna Miecznikowski (responsabile), Elena Battaglia e Christian Geddo (dottorandi).

Nel progetto, si studiano le modalità di espressione delle fonti di informazione nell'italiano parlato. Ci interessa capire quali mezzi verbali e non verbali i parlanti impiegano per comunicare come sono venuti a conoscenza di un'informazione, per esempio per sentito dire, per esperienza diretta (visiva, uditiva, olfattiva, ecc.) o tramite un ragionamento logico, e quanto sia rilevante, nella comunicazione faccia a faccia, esplicitare tali fonti. L'indagine rientra in due campi più ampi di ricerca: la ricerca linguistica sull'evidenzialità, che paragona l'espressione grammaticale e lessicale delle fonti di informazione in varie lingue, e l'analisi della conversazione, che esamina la lingua parlata in situazioni di interazione diretta.

Composizione del corpus

In questo quadro, uno degli obiettivi è la creazione di un nuovo corpus di italiano parlato, il corpus TIGR (Ticino GRigioni). Si tratta di una risorsa innovativa, che si distinguerà da altre pubblicamente disponibili per il contesto geografico della raccolta (la Svizzera italiana), il focus sull'interazione diretta e il suo carattere multimodale. Sarà messo a disposizione della comunità scientifica a scopi di ricerca linguistica. Il corpus comprenderà 28 ore di videoregistrazione in cinque situazioni comunicative (interviste, pause pranzo, lavori in gruppo, preparazione di un pasto, lezioni), le relative trascrizioni e alcuni dati sociolinguistici raccolti tramite un questionario.

Tutela della privacy e trattamento dei dati



Consenso informato. Chi deciderà di aderire riceverà una scheda informativa dettagliata, specifica per la situazione di registrazione scelta, e firmerà un modulo nel quale dichiarerà il proprio consenso a partecipare allo studio. La partecipazione è volontaria ed è sempre possibile ritirarsi dallo studio senza conseguenze.

Protezione dei dati. Non sarà possibile risalire all'identità di un informatore a partire dai dati raccolti, che saranno anonimizzati. Si adotta un protocollo che comprende diverse misure a tutela della *privacy*:



- ✓ compilazione anonima del questionario sociolinguistico;
- ✓ nelle trascrizioni, sostituzione dei nomi di persona con degli pseudonimi, dei nomi di luogo e altre informazioni con una sequenza di caratteri;
- ✓ nella traccia audio, sostituzione degli stessi elementi con dei segnali acustici.

Se la persona lo desidera, verranno applicate delle misure ulteriori:

- ✓ nella traccia audio, alterazione della voce;
- ✓ nella traccia video, applicazione di filtri che rendono irriconoscibili i volti e eventuali elementi sensibili dell'ambiente.



Conservazione dei dati. I documenti audio e video originali saranno conservati in una cartella protetta su un server USI, a cui avrà accesso soltanto l'*équipe* del progetto. Le informazioni personali saranno trattate in maniera confidenziale e in seguito cancellate.

Pubblicazione. Il corpus TIGR comprenderà esclusivamente dati anonimizzati secondo le procedure illustrate al punto *Protezione dei dati*. I dati originali non saranno pubblicati. Il corpus sarà depositato su un repository web sicuro. L'accesso sarà concesso ad altri studiosi solo su richiesta motivata. L'uso a fini commerciali è vietato.



Uso dei dati. Nel quadro del progetto Infilta, il corpus TIGR sarà usato per condurre analisi linguistiche sull'espressione delle fonti di informazione. Dopo la conclusione del progetto, i materiali pubblicati saranno accessibili anche ad altri studiosi della lingua. Fotogrammi ed estratti anonimizzati delle trascrizioni e dei video saranno integrati in pubblicazioni scientifiche quali articoli, comunicazioni a convegni, tesi di dottorato.