

Il Corpus of Chinese Learners of Italian (COLI)

Il progetto

Il *Corpus of Chinese Learners of Italian (COLI)* è un corpus **scritto e parlato** che comprende produzioni in italiano di apprendenti cinesi.

L'autrice è **Carmen Coli**, che ha lavorato alla costituzione del corpus nell'ambito di un progetto POR Umbria (Fondo sociale europeo), intitolato *L'apprendimento dell'italiano L2 da parte di studenti sinofoni: raccolta di dati linguistici per un'indagine sulle strategie di apprendimento* (Maggio 2009-giugno 2010).

Il coordinamento scientifico del progetto, all'interno dell'Università per Stranieri di Perugia, è stato assicurato dalle due tutor, **Lidia Costamagna** e **Stefania Spina**.

La predisposizione dei dati per la versione di rete è stata curata da **Stefania Spina**.

Composizione del corpus

Gli apprendenti

I dati inclusi nel corpus sono stati prodotti da **30 studenti cinesi** dell'Università per Stranieri di Perugia (età media: **22 anni**), e sono stati raccolti dall'autrice nel periodo Luglio-Settembre 2009:

- **10 di livello B1** (Corsi Marco Polo; 5 M e 5 F). Sono in Italia in media da 5 mesi. Studiano la lingua italiana in media da 7 mesi;
- **10 di livello B2** (Corsi Marco Polo; 5 M e 5 F). Sono in Italia in media da 5 mesi. Studiano la lingua italiana in media da 9 mesi;
- **10 di livello C1** (Corso di lingua italiana trimestrale; 10 F). Sono in Italia in media da 2 anni. Studiano la lingua italiana in media da 2 anni.

Il corpus contiene in totale 82.300 tokens, 4093 lemmi e 8572 frasi. È stato etichettato per categoria grammaticale ([questo è il tagset](#), e questa è una [lista di espressioni formate da più di una parola](#), annotate come parole singole) e lemmatizzato.

I task

I dati sono stati raccolti attraverso:

- 2 PROVE SCRITTE
- Un'INTERVISTA ORALE

- **1a prova scritta:**
 - 5 domande a risposta aperta (min. 10 righe max 20 righe)

- Descrizione della giornata di Silvia (immagine)
- Commentare una pubblicità progresso e fornire suggerimenti.
- **2a prova scritta:**
 - 5 domande a risposta aperta (min. 10 righe max 20 righe)
 - Dare suggerimenti e consigli in due differenti situazioni.
- **intervista orale** con la medesima struttura:
 - Commento della scheda sociolinguistica.
 - Inventare una storia partendo da tre immagini.
 - Commentare un quadro.
 - Descrivere due immagini o un'immagine con più personaggi.
 - Trovare le differenze tra due immagini/fumetti.
 - Analizzare delle sequenze fotografiche e ricostruire l'accaduto.
 - Domande di carattere culturale.

Durata totale interviste B1: 255,97 minuti
 Media per ogni parlante B1: 25,597 minuti

Durata totale interviste B2: 272,97 minuti
 Media per ogni parlante B2: 27,297 minuti

Durata totale interviste C1: 359,4 minuti
 Media per ogni parlante C1: 35,94 minuti

DURATA TOTALE INTERVISTE: 888,34 minuti (14,6 ore)

MEDIA TOTALE PER PARLANTI: 29,6113 minuti