

Longitudinal Corpus of Chinese Learners of Italian (LOCCLI)

Stefania Spina (University for Foreigners of Perugia, Italy)

Anna Siyanova-Chanturia (Victoria University of Wellington, New Zealand)

How to cite the LOCCLI:

Spina, S. & Siyanova-Chanturia, A. (in preparation). *The Longitudinal Corpus of Chinese Learners of Italian*.

The *Longitudinal Corpus of Chinese Learners of Italian* (LOCCLI) is made up of **350 essays** written by **175 chinese learners of Italian**, who attended Italian language courses in Perugia for 6-8 months, and it was collected in 2016.

Collection

The LOCCLI is a **longitudinal corpus**: data was collected in **two different points in time**, from the same 175 learners.

Each of the learners contributed **two essays** to the corpus: one essay was written at the beginning of a six-month course of Italian, and the other at the end of the course.

In order to collect corpus data, they were asked to write on three similar essay topics: 1) My first impression of Italy and Italians, 2) My hobbies: what do I usually do in my free time, 3) My last holidays.). They were instructed not to write on the same topic more than once.

Learners

Using a placement test, learners were assigned to one of three proficiency levels: A1 (n=39), A2 (n=86), and B1 (n=50).

All the learners came from China and were between 17 and 33 years of age. On average, they spent 1.7 months in Italy before writing the first essay.

Size and annotation

The corpus was pos-tagged ([this is the tagset](#), and this is a [list of multi-words units](#), annotated as single words), and xml-annotated, and its total size is 97,000 tokens:

- data collection 1: A1=7,126, A2=22,851, B1=15,903;
- data collection 2: A1=9,487, A2=24,117, B1=17,386.