

## Abstract

The present work deals with learners' Italian lexical competence and it is focused on a specific linguistic phenomenon: Verb + Noun collocations, i.e. restricted word combinations whose semantic and lexical properties cannot be fully predicted from those of its components. In reason of their lexico-semantic status collocations are generally hardly acquired by non natives.

For the purposes of the present work we built the Corpus of Learner Italian CAIL2 (*Corpus di Apprendenti di Italiano L2*) which contains 400 essays for 237 000 tokens written by intermediate and advanced learners of Italian from 33 mother tongue backgrounds. Collocations produced by learners were studied from quantitative and qualitative perspective and were compared to those of natives contained in the written corpus SCUOLA, a section of the PEC corpus (Perugia Corpus), a representative corpus of contemporary Italian.

Empirical collocations, i.e. collocational candidates, were extracted from corpus CAIL2 and corpus SCUOLA by means of lexical association measures (Mutual Information  $\geq 3$ , t-score  $\geq 2$ ) and frequency threshold (fr.  $\geq 10$ ).

First of all, we analysed empirical collocations by a quantitative point of view. We compared their amount in native and non native texts and in those of different groups of learners identified by two variables: time of study Italian and time spent in Italy. The goal of quantitative analysis was to find out if any substantial differences among groups of learners and/ or between the two varieties of natives and non natives exist. For this reason tests of statistical significance were performed.

Secondly, we studied empirical collocations from linguistic perspective: we evaluated which V+N combinations extracted from corpora were Italian collocations and we described their prominent features. Inter alia, this allow us to evaluate to what extent frequency and lexical association thresholds adopted were suitable in extracting collocations from corpora of natives and learners Italian.

Thirdly, learners Italian as second language productions were investigated from an additional outlook: concordance analysis relating to a group of selected Verb + Noun combinations was conducted in order to estimate, among other, the influence that time of teaching and exposure to Italian have on collocational competence.

**Keywords:** Italian language, Verb + Noun collocations, Learner Corpus, lexical association measures, Second Language Acquisition