

Il corpus CAIL2 (*Corpus di Apprendenti di Italiano L2*) è parte della tesi di Dottorato di **Leontyna Bratankova** *Le collocazioni Verbo + Nome in apprendenti di italiano L2*, discussa nell'aprile del 2015 all'Università per Stranieri di Perugia (Tutori: Prof.ssa Stefania Spina, Università per Stranieri di Perugia, e Prof. Pavel Štichauer, Università Carolina di Praga).

Il corpus ha un'ampiezza di 237.000 tokens, ed è costituito da 400 testi scritti, redatti da altrettanti studenti della lingua italiana ad un livello intermedio-avanzato di competenza linguistica e con 33 L1 differenti (arabo, armeno, bengalese, bielorusso, bulgaro, ceco, cinese, coreano, farsi, francese, gallego, georgiano, giapponese, greco, hindi, inglese, mongolo, montenegrino, nederlandese, norvegese, polacco, portoghese, punjabi, romeno, russo, slovacco, sloveno, spagnolo, svedese, tedesco, turco, ucraino, ungherese)

I dati sono stati raccolti presso tre università: l'Università per Stranieri di Perugia, l'Università Carolina di Praga e l'Università della Boemia meridionale di České Budějovice.