

ITALSERB govorni korpus srbofonih studenata italijanskog kao stranog jezika

Korpusi drugog jezika ili učenički korpusi (engl. *L2 corpora, learner corpora*) predstavljaju poslednjih par decenija dragocen teorijsko-metodološki instrument za proučavanje procesa usvajanja drugog ili stranog jezika kod nematernjih govornika. Jedan takav govorni korpus italijanskog kao stranog jezika pod nazivom ITALSERB (ital. *ITALiano dei SERBofoni* – italijanski srbofonih govornika; Ceković-Rakonjac 2012, 2013, Ceković 2016) kreiran je 2010. godine na Katedri za italijanski jezik i književnost Filološkog fakulteta Univerziteta u Beogradu, u saradnji s Univerzitetom za strance u Sijeni.

Građa za korpus sakupljena je od 2010. do 2014. godine i čini je približno 25 sati audio snimaka asimetrične, poluspontane, monološke i dijaloške produkcije. Uzorak čini 170 studenata koji su u periodu prikupljanja materijala uz pomoć certifikacionih testova CILS pohađali I, II, III ili IV godinu osnovnih studija na Katedri za italijanistiku. Korpus se odlikuje transverzalnim i longitudinalnim karakterom te omogućava kako poprečni presek (uvid u datom trenutku u veštinu govorne produkcije studenata sve četiri godine studija: od ciljnog nivoa jezičke kompetencije, prema Zajedničkom evropskom okviru, A2 za I, B1 za II, B2 za III i C1 za IV godinu studija), tako i longitudinalni (praćenje razvoja govorne veštine kod jedne generacije studenata od početka do završetka studija). Sistem normi za anotiranu ortografsku transkripciju primenjen na ITALSERB usklađen je sa standardima korpusa maternjih govornika italijanskog CLIPS, oformljenog na Univerzitetu u Napulju, i smernicama grupe EAGLES (Savy 2005, Ceković-Rakonjac 2012).

ITALSERB se trenutno nalazi u fazi revizije transkripcionog procesa i njegov očekivani obim iznosi 200.000 reči, što ga potencijalno čini zbirkom velikih dimenzija u kategoriji učeničkih korpusa. Tri i po sata transkribovanog materijala ili 30.000 reči dostupno je od 2016. godine u doktorskoj disertaciji deponovanoj u repozitorijumu NaRDuS (Ceković 2016: 289-371). Paralelno s revizijom, 2022. godine započet je i rad na anotaciji materijala, tj. tagiranju prema gramatičkim kategorijama (*Part-of-Speech tagging*). Namera autora korpusa jeste da se u bliskoj budućnosti kompletan ITALSERB digitalizuje i učini javno dostupnim naučnoj i stručnoj publici putem odgovarajuće onlajn platforme.

U izlaganju koje predlažemo na ovu temu nameravamo da prezentujemo metodološke pretpostavke za nastanak ITALSERB korpusa i modalitete prikupljanja podataka (odabir tehnika za elicitaciju jezičkih podataka i procedura za ortografsku transkripciju tekstova, beleženje sociolingvističkih podataka), uz poseban osvrt na aplikativne aspekte rada na konstituisanju korpusa. U tom smislu, namera nam je i da ukratko predočimo rezultate nekolicine naučnih istraživanja sprovedenih na korpusnim podacima (*corpus-based*; Ceković 2014, 2016, 2020, Ceković & Radojević 2015, Ceković & Vučo 2020), kako bismo skrenuli pažnju na pedagoške implikacije i korisnost učeničkih korpusa za potrebe različitih aktera naučno-nastavnog procesa, odnosno ukazali na značaj ovakvih tipova korpusa u oblasti ne samo korpusne već i primenjene lingvistike, a naročito teorije usvajanja drugog jezika i glotodidaktike.

- Ceković, N. (2014). I segnali discorsivi nell'interlingua degli studenti universitari di italiano L2. *Italica Belgradensia*, 2, 93-110. [<http://italicabelgradensia.fil.bg.ac.rs/Italica%20Belgradensia%202-2014%20ceo%20broj%20integralno%20stamp.pdf>]
- Ceković, N. (2016). *Diskursni markeri u govornoj produkciji na italijanskom kao drugom jeziku* (Neobjavljena doktorska disertacija). Beograd: Filološki fakultet. [<https://nardus.mpn.gov.rs/handle/123456789/6470>]
- Ceković, N. (2020). Metatekstualna kompetencija učenika na srednjem nivou znanja italijanskog kao drugog jezika. U V. Polovina & B. Kovačević (prir.), *Primenjena lingvistika danas - jezik, književnost i interdisciplinarnost: zbornik radova sa Šestog međunarodnog kongresa „Primenjena lingvistika danas - jezik, književnost i interdisciplinarnost“*, Beograd, 12-13. oktobar 2018. (str. 197-209). Beograd: Društvo za primenjenu lingvistiku Srbije/Filološki fakultet Univerziteta u Beogradu/Filozofski fakultet Univerziteta u Novom Sadu. [http://doi.fil.bg.ac.rs/volume.php?pt=eb_ser&issue=dpls_pld-2020-6&i=12]
- Ceković-Rakonjac, N. (2012). Ortografska transkripcija govornog korpusa ESNAKIT. U A. Vraneš, LJ. Marković & G. Alexander (prir.), *Digitalizacija kulturne i naučne baštine, univerzitetski repozitorijumi i učenje na daljinu, knj. 3* (str. 163-182). Beograd: Filološki fakultet. [<https://www.academia.edu/88345656>]
- Ceković-Rakonjac, N. (2013). ITALBEG corpus parlato di italiano L2. *Italica Belgradensia*, 1, 336-348. [https://www.academia.edu/38701496/Cekovi%C4%87_Rakonjac_N_2013_ITALBEG_corpus_parlato_di_italiano_L2]
- Ceković, N. & Radojević, D. (2015). Didattica con i corpora orali di italiano L2. In C. Ramsey-Portolano (ed.), *The Future of Italian Teaching: Media, New Technologies and Multi-Disciplinary Perspectives* (pp. 96-106). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Ceković, N. & Vučo, J. (2020). Uno studio longitudinale come base per la verifica dei criteri di valutazione della competenza discorsiva in italiano L2. *Nasleđe*, 46, 97-107. [http://www.filum.kg.ac.rs/index.php?option=com_content&view=article&id=129&Itemid=223&lang=sr]
- Savy, R. (2005). Specifiche per la trascrizione ortografica annotata dei testi. In F. Albano Leoni & R. Giordano (a cura di), *Italiano Parlato, Analisi di un dialogo*. Napoli: Liguori. [<http://www.clips.unina.it/it/documenti.jsp>]