

ANALIZIRANJE ANGLICIZAMA SA REČJU VEB POMOĆU ELEKTRONSKOG KORPUSA CLASSLA-WEB.SR U SVRHU USVAJANJA INŽENJERSKE TERMINOLOGIJE

Aleksandra Kardoš Stojanović¹, Jelisaveta Šafrančević², Jelena Zivlak³

^{1,2,3}Univerzitet u Novom Sadu, Fakultet tehničkih nauka, Novi Sad, Srbija

¹sandra.kardos@uns.ac.rs, ²savetas@uns.ac.rs, ³zivlakjelena@uns.ac.rs

Kratak sadržaj: U skladu sa digitalizacijom svih oblasti života, i jezička istraživanja se sve više baziraju na elektronskim korpusima, koji se definišu kao skupovi tekstova u pisanom ili usmenom obliku dostupni na internetu ili pak računaru u vidu softvera. Kako su lako dostupni i jednostavni za upotrebu, korisni su ne samo lingvistima za analiziranje svih nivoa jezika, već i ostalim korisnicima jezika, pa tako i studentima inženjerstva i inženjerima, koji ne moraju biti jezički stručnjaci. Među elektronskim korpusima nalazi se i korpus CLASSLA-web.sr, kojeg je razvila slovenačka digitalna infrastruktura CLARIN.SI (Zajednička infrastruktura jezičkih resursa i tehnologija), čiji je glavni cilj da obezbedi lakši pristup jezičkim resursima i alatima za istraživanja u domenu društvenih i humanističkih nauka, najviše u južnoslovenskim jezicima. Cilj ovog rada je da prikaže kako se korpus CLASSLA-web.sr može koristiti za analiziranje anglicizama, odnosno reči preuzetih iz engleskog jezika i prilagođenih sistemu srpskog jezika, u domenu informacionih tehnologija, i to na primeru porodice reči sa osnovnom rečju veb koja pokazuje veliku produktivnost u građenju anglicizama. Analizom reprezentativnog uzorka iz korpusa dolazimo do saznanja da reč veb gradi mahom imenice koje predstavljaju preoblikovane ili mešovite anglicizme, od kojih neki nisu u skladu sa jasno utvrđenim pravilima pisanja anglicizama u srpskom jeziku. Može se zaključiti da elektronski korpus CLASSLA-web.sr predstavlja koristan alat za usvajanje inženjerske terminologije, ali da se za proveru pravilnog pisanja anglicizama u srpskom ipak mora koristiti stručna lingvistička literatura.

Ključne reči: anglicizmi, informacione tehnologije, elektronski korpus, CLASSLA-web.sr, inženjerska terminologija

ANALYZING VEB-RELATED ANGLICISMS BY MEANS OF THE CLASSLA-WEB.SR ELECTRONIC CORPUS FOR THE ACQUISITION OF ENGINEERING TERMINOLOGY

Abstract: In line with the digitalization of various fields of life, linguistic research has increasingly relied on electronic corpora, which are defined as collections of written or spoken texts accessible on the internet or via software. These corpora, being easily accessible and user-friendly, are valuable not only to linguists for analyzing all levels of language, but also to general language users, including engineering students and engineers, who do not need to be linguistic experts. One such corpus is the CLASSLA-web.sr corpus, developed by the Slovene CLARIN.SI infrastructure (Common Language Resources and Technology Infrastructure). This infrastructure aims to facilitate access to language resources and tools for research in the Humanities and Social Sciences, primarily focusing on the South Slavic languages. This paper explores how the CLASSLA-web.sr corpus can be used to analyze Anglicisms – words borrowed from English and adapted to the Serbian language system – in the field of Information Technology, specifically examining the word family derived from the base word veb which shows significant productivity in the formation of Anglicisms. An analysis of a representative sample from the corpus reveals that the word veb predominantly forms nouns, which can be classified as either transshaped or mixed Anglicisms. Some of these Anglicisms do not adhere to the clearly defined rules for writing Anglicisms in the Serbian language. Therefore, the findings suggest that the CLASSLA-web.sr electronic corpus represents a valuable tool for acquiring engineering terminology, but that relevant linguistic literature should be consulted to ensure the correct spelling of Anglicisms in Serbian.

Key Words: Anglicisms, information technology, electronic corpus, CLASSLA-web.sr, engineering terminology

1. UVOD

Među mnogobrojnim jezičkim disciplinama se u kontekstu digitalizacije 21. veka sve više razvija korpusna lingvistika, koja se definiše kao oblast lingvistike koja počiva na empirijskim izučavanjima jezika baziranim na velikom broju kontekstualizovanih primera upotrebe jezika, čijom se analizom dolazi do opštih obrazaca u jezičkom sistemu, kao i do odstupanja od njih [5]. U samoj srži korpusne lingvistike nalazi se korpus, koji predstavlja skup prikupljenih tekstualnih podataka koji služe kao reprezentativni uzorak jezika i mašinski su čitljivi [1]. Iako je u

prošlosti korpus obuhvatao bilo kakvu kolekciju tekstova koja je služila kao izvor za jezička istraživanja, danas se pod korpusom sve više podrazumevaju samo njegove elektronske verzije [2]. Na osnovu dosadašnjih istraživanja, poznato je da su elektronski korpusi našli upotrebu i kao korisna alatka za učenje jezika, gde se učenjem zasnovanim na pretraživanju podataka povećava motivacija i postiže bolji efekat kod učenika i studenata [8]. Cilj ovog rada je da prikaže upotrebu elektronskih korpusa u svrhu istraživanja anglicizama sa osnovnom rečju *veb* u domenu informacionih tehnologija, i to na primeru korpusa CLASSLA-web.sr, koji, zahvaljujući svojoj lakoj dostupnosti i prilagođenosti korisniku, može biti od koristi ne samo lingvistima koji mogu sprovesti analize različitih nivoa jezika, već i studentima inženjerstva i inženjerima koji žele da prošire vokabular u relevantnoj oblasti jezika struke.

2. ELEKTRONSKI KORPUS CLASSLA-WEB.SR

U sklopu serije korpusa južnoslovenskih jezika CLASSLA-web, korpus srpskog jezika CLASSLA-web.sr sadrži 2.342.626.265 reči, čime predstavlja najobimniji korpus savremenog srpskog jezika. Sve korpuse CLASSLA-web razvila je slovenačka digitalna infrastruktura CLARIN.SI (Zajednička infrastruktura jezičkih resursa i tehnologija), koja je u sklopu projekta MaCoCu prikupila jednojezičke i paralelne korpuse za više od deset evropskih jezika sa slabijim jezičkim resursima i učinila ih javno dostupnim [6]. Korpusi su prikupljeni putem pretraživanja nacionalnih internet domena najvišeg nivoa, ali i generičkih domena povezanih sa veb-sajtovima na nacionalnim domenima. Važno je napomenuti da tekstovi pokazuju ne samo standardnu, već i nestandardnu upotrebu jezika, čime daju uvid i u deo vokabulara koji se ređe sreće u rečnicima, kao i da su prikupljeni tokom 2021. i 2022. godine, te uključuju i primere novijeg datuma [6], što je od posebnog značaja za IT inženjere koji rade u oblasti koja se stalno menja i moraju biti upoznati sa najnovijim terminima kako bi pratili novine u struci. Većina korpusa, osim lingvističke anotacije, sadrži i metapodatke, poput veb domena i URL adrese teksta, kao i žanra kojem tekst pripada [7].

Korisnici mogu pristupiti korpusu CLASSLA-web.sr na veb-sajtu CLARIN.SI (<https://www.clarin.si>), gde u sklopu padajućeg menija *Concordancers* mogu da odaberu jedan od ponuđenih tri konkordansera, odnosno korpusna pretraživača, koji predstavljaju kompjuterske programe za pretraživanje i statističku obradu podataka u velikim skupovima tekstova [6]. Korpusni pretraživač korišćen za potrebe ovog rada jeste noSketchEngine, koji je jednostavan za upotrebu i pregledan za pretraživanje i analiziranje rezultata. Nakon izabranog korpusnog pretraživača, sledeći korak jeste odabir samog korpusa, gde se među mnoštvom dostupnih korpusa nudi i CLASSLA-web.sr. Ulaskom u sam korpus, pojavljuju se dostupni korpusni alati – *Concordance*, *Wordlist*, *Keywords* i *Text type analysis*. Prva ponuđena alatka *Concordance* služi za pretraživanje reči i fraza u kontekstima, te prikazivanje svih dobijenih rezultata koji se potom mogu dalje filtrirati i sortirati. Što se tiče alata *Wordlist*, on generiše spisak frekventnosti traženih reči po različitim kriterijumima (npr. vrsti reči ili slovima koje sadrži), dok *Keywords* služi za ekstrakovanje ključnih reči u okviru korpusa, dokumenta ili teksta, odnosno izdvajanje terminološki bitnih reči. Četvrta alatka *Text type analysis* pruža uvid u statistiku tipova teksta uključenih u korpus. Dakle, u zavisnosti od potreba korisnika korpusa, na raspolaganju su kako jednostavnije, tako i naprednije opcije pretrage i obrade podataka.

3. ISTRAŽIVANJE ANGLICIZAMA POMOĆU KORPUSA CLASSLA-WEB.SR

Kako engleski jezik ne predstavlja samo opšti jezik sporazumevanja na globalnom nivou, već je u poslednjih dvadesetak godina stekao status odomaćenog stranog jezika zbog svoje lake audio-vizuelne dostupnosti, dvojnog usvajanja i funkcije dopunskog jezika [9], dolazi do velikog priliva leksičkog i drugog materijala iz engleskog jezika u sve jezike sa kojima on dolazi u kontakt, uključujući i srpski. Reči koje su iz engleskog kao izvornog jezika preuzete u srpski kao ciljni jezik, i koje su, nakon adaptacije na nivou izgovora, pisanja, oblika i značenja, dostigle bar delimičnu integraciju u sistem srpskog jezika, nazivaju se anglicizmima [10]. Iako su anglicizmi prisutni u svim tematskim oblastima vokabulara u manjoj ili većoj meri, njihov broj je posebno velik u oblasti informacionih tehnologija. To je u skladu sa činjenicom da se razvoj informacionih tehnologija kao novije grane nauke poklapa sa razvojem engleskog kao globalnog jezika, zbog čega se dosta termina preuzima u vidu anglicizama iz engleskog jezika u druge jezike [3]. Među anglicizmima u ovoj oblasti nalazi se i reč *veb*, nastala preuzimanjem i adaptacijom engleske reči *web*. Kako anglicizam *veb* kao osnovna reč učestvuje u građenju mnogih drugih anglicizama, u nastavku ćemo istražiti njenu produktivnost pomoću korpusa CLASSLA-web.sr.

Da bismo došli do spiska svih anglicizama koji spadaju u porodicu reči sa *veb*, ali i do podatka koliko su frekventni, korpusna alatka koja nam je potrebna jeste *Wordlist*, i to opcije osnovne pretrage. Kako bismo izbegli da nam se kao odvojeni rezultati prikazuju različiti padežni oblici iste reči, biramo traženje leme, tj. samo osnovnog oblika reči. Takođe, kako *veb* gradi dosta polusloženica i složenica gde stoji kao prednji element, opredeljujemo se za pretragu reči koje počinju sa *veb*. Pretraga daje 222 pogotka, ali neke od njih moramo odbaciti jer, recimo, ne dolaze od istog korena, ili se pak odnose na nazive nekih veb-sajtova, softvera i slično, čime ne spadaju u opštu leksiku. Pored spiska reči i njihove frekventnosti, rezultati daju i mogućnost prikaza svih reči u različitim kontekstima. U tabeli 1 biće navedeno 30 najfrekventnijih relevantnih rezultata pretrage, zajedno sa podatkom o njihovoj učestalosti upotrebe.

Tabela 1 – Prikaz 30 najfrekventnijih anglicizama sa *veb* po rezultatima pretrage korpusa CLASSLA-web.sr

ANGLICIZAM	FREKVENCIJA UPOTREBE U KORPUSU
veb	47912
vebsajt	6402
veb-sajt	6332
veb-stranica	1654
veb-sajtov	1411
vebmaster	298
veb-alat	264
veb-prezentacija	219
veb-lokacija	201
veb-kamera	188
veb-aplikacija	179
veb-portal	144
veb-adresa	128
veb-servis	120
veb-hosting	109

ANGLICIZAM	FREKVENCIJA UPOTREBE U KORPUSU
veb-mesto	99
veb-dizajn	96
veb-prodavnica	95
veb-platforma	92
veb-portal	88
veb-pregledač	81
veb-dizajner	66
veb-pretraživač	63
vebstranica	53
veb-server	50
vebmastering	48
vebšop	47
veb-sadržaj	44
veb-novinarstvo	39
veb-šop	38

Ono što možemo videti na osnovu reprezentativnog uzorka rezultata pretrage prikazanog u tabeli jeste da osnovna reč *veb* većinski gradi imenice, sa izuzetkom (nepravilno napisanog) primera *veb-sajtov*, koji predstavlja pridev. Broj pogodaka u korpusu pokazuje da se ubedljivo najčešće koristi prvih pet anglicizama, dok je kod ostalih frekvencija upotrebe značajno niža. Još jedno zapažanje tiče se načina adaptacije anglicizama, gde ih po nastanku možemo razvrstati na preoblikovane anglicizme, koji su nastali preuzimanjem novog oblika i njemu pridružene sadržine (npr. *veb-hosting*, *veb-server*, *veb-dizajn*, *veb-šop*), i mešovite anglicizme, kod kojih je deo nastao preoblikovanjem, a deo prevodjenjem (npr. *veb-stranica*, *veb-prodavnica*, *veb-pretraživač*, *veb-sadržaj*). Iako je većina njih pravilno adaptirana sistemu srpskog jezika, pored već spomenutog nepravilno napisanog anglicizma *veb-sajtov*, primećuje se i da se neki anglicizmi javljaju dva puta u tabeli, sa različitim načinima pisanja. U pitanju su anglicizmi *vebsajt* / *veb-sajt*, *veb-stranica* / *vebstranica*, *vebšop* / *veb-šop*. Neujednačenost pisanja anglicizama pokazuje opšte nepoznavanje pravila o njihovom pisanju, što je jedan od dokaza savremenog trenda proizvoljne upotrebe jezika, te nepoštovanja jezičke norme [9]. Samim tim, neophodno je da se na nastavi engleskog i srpskog jezika, uključujući i nastavu engleskog jezika za inženjere, pažnja posveti podučavanju pravila za upotrebu, a naročito pisanje anglicizama [4]. Dakle, slično zaključku do kojeg smo došli izučavajući koliko ChatGPT poznaje pravila o pisanju anglicizama [4], jasno je i da pretraživanjem korpusa CLASSLA-web.sr studenti inženjerstva i inženjeri mogu da prošire vokabular u oblasti informacionih tehnologija, ali da se za proveru pravilnih oblika pisanja anglicizama ipak moraju okrenuti relevantnoj lingvističkoj literaturi. Na osnovu pravila standardizacije anglicizama iz *Srpskog rečnika novijih anglicizama* (2021), pravilni oblici navedenih anglicizama jesu *vebsajt* i *vebsajtov*, koji se pišu kao jedna neprekinuta reč jer predstavljaju jednu sadržinsku celinu, odnosno *veb-stranica* i *veb-šop*, koji se pišu sa crticom jer, iako predstavljaju jednu sadržinsku celinu, u srpskom jeziku se može uočiti značenjska veza između elemenata, te se data veza predstavlja u vidu crtice između tih elemenata.

4. ZAKLJUČAK

Želeći da prikazemo kako se elektronski korpusi mogu koristiti za razne jezičke analize, ali i učenje jezika, te usvajanje inženjerske terminologije, u radu smo na primeru korpusa CLASSLA-web.sr istraživali anglicizme sa osnovnom rečju *web* u domenu informacionih tehnologija. Nakon konciznog pojašnjenja o načinu korišćenja korpusa, tabelarno je prikazano 30 najfrekventnijih anglicizama sa rečju *web*, a potom je usledila njihova sažeta analiza, koja pokazuje da reč *web* mahom učestvuje u građenju imenica koje su u vidu preoblikovanih i mešovityh anglicizama. Dakle, korpus služi kao dobar izvor za mnoštvo primera upotrebe anglicizama u različitim kontekstima, naročito zbog njegove aktuelnosti u poređenju s materijalom u rečnicima, što ga čini posebno korisnim za usvajanje vokabulara koji obuhvata najnovije termine iz jezika za inženjere. Ipak, kako sadrži i neke nepravilno napisane anglicizme, za proveru pravilnog pisanja anglicizama bismo se ipak morali konsultovati sa stručnom lingvističkom literaturom. Možemo zaključiti da bi, uz obuku kako ga treba upotrebljavati i u koje tačno svrhe, korpus CLASSLA-web.sr bio koristan alat u rukama studenata inženjerstva, inženjera, kao i ostalih korisnika jezika.

ZAHVALNICA

Autorke se zahvaljuju na podršci u okviru projekta Departmana za opšte discipline u tehnici Fakulteta tehničkih nauka u Novom Sadu koji nosi naziv »Socio-tehnološki aspekti unapređenja nastavnog procesa na engleskom jeziku u opštim disciplinama« (Socio-technological aspects of improving the teaching process in the English language in fundamental disciplines).

5. LITERATURA

- [1] Heidi Agerbo, *Explaining meaning in lexicographical information tools*, In: P. A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, London, New York: Routledge, pp. 59–77, 2018.
- [2] Gertrud Faaß, *Lexicography and corpus linguistics*, In: P. A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, London, New York: Routledge, pp. 123–137, 2018.
- [3] Aleksandra Kardoš Stojanović, Jelisaveta Šafran, Jelena Zivlak, *Upotreba elektronskog izdanja Srpskog rečnika novijih anglicizama u nastavi engleskog jezika za inženjere*, U: B. Dumnić (ur.), *Trendovi razvoja: Univerzitet pred novim izazovima. Zbornik radova*, Novi Sad: Fakultet tehničkih nauka, str. 107–110, 2023. Dostupno na: http://www.trend.uns.ac.rs/stskup/trend_2023/TREND2023-ZBORNIK-RADOVA.pdf. [Pristupljeno: 23.12.2024.]
- [4] Aleksandra Kardoš Stojanović, Jelisaveta Šafran, Jelena Zivlak, *Koliko ChatGPT poznaje pravila o upotrebi anglicizama u srpskom jeziku: mogućnosti i ograničenja njegove primene u nastavi engleskog jezika za inženjere*, U: B. Dumnić (ur.), *Trendovi razvoja: Nastavnici i saradnici kao centar promena u visokom obrazovanju. Zbornik radova*, Novi Sad: Fakultet tehničkih nauka, str. 141–144, 2024. Dostupno na: http://www.trend.uns.ac.rs/stskup/trend_2024/TREND2024-ZBORNIK-RADOVA.pdf. [Pristupljeno: 23.12.2024.]
- [5] Aleksandar Kavgić, *Nastava korpusne lingvistike na studijama anglistike*, Novi Sad: Filozofski fakultet, 2020.
- [6] Taja Kuzman, *CLASSLA-web: Bigger and Better Web Corpora for Croatian, Serbian and Slovenian on CLARIN.SI Concordancers*, Medium, June 23, 2023. Available at: <https://medium.com/@taja.kuzman/classla-web-bigger-and-better-web-corpora-for-croatian-serbian-and-slovenian-on-clarin-si-548fbf0794b2>. [Accessed: 23.12.2024.]
- [7] Nikola Ljubešić, Taja Kuzman, *CLASSLA-web: Comparable Web Corpora of South Slavic Languages Enriched with Linguistic and Genre Annotation*, LREC-COLING, pp. 3271–3282, 2024.
- [8] Tatjana Marjanović, *Some Benefits of Corpora as a Language Learning Tool*, *Journal on English Language Teaching*, vol. 2, no. 4, pp. 7–14, 2012.
- [9] Tvrtko Prčić, *Engleski u srpskom*, Treće, elektronsko, izdanje, Novi Sad: Filozofski fakultet, 2019. Dostupno na: <https://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-512-9>. [Pristupljeno: 23.12.2024.]
- [10] Tvrtko Prčić, Jasmina Dražić, Mira Milić, Milan Ajdžanović, Sonja Filipović Kovačević, Olga Panić Kavgić, Strahinja Stepanov, *Srpski rečnik novijih anglicizama*, Prvo, elektronsko izdanje, Novi Sad: Filozofski fakultet, 2021. Dostupno na: <https://digitalna.ff.uns.ac.rs/sadrzaj/2021/978-86-6065-636-2>. [Pristupljeno: 23.12.2024.]