

# KORIŠTENJE PARALELNOG KORPUSA IZ CESTOVNOG PROMETA U NASTAVI JEZIKA STRUKE U VISOKOM OBRAZOVANJU

STRUČNI RAD / PROFESSIONAL PAPER

**Andrej Smodlaka**

*Sveučilište u Zagrebu, Fakultet prometnih znanosti, Zagreb, Hrvatska  
asmoslaka@fpz.unizg.hr*

## Sažetak

Korpusi se relativno odnedavno koriste u poučavanju stranih jezika, što je izraženo i u nastavi jezika struke (JS). Posebno su zanimljivi specijalizirani paralelni korpusi iz određenog područja koji predstavljaju vrlo vrijedne resurse za obrazovne svrhe. Nastavnici koji poučavaju engleski jezik struke često se suočavaju s izazovom vezanim uz poučavanje vokabulara, odnosno s nedostatkom autentičnih nastavnih materijala. Ovaj se rad bavi poučavanjem engleskog jezika za transport i logistiku na prijediplomskom studiju na Fakultetu prometnih znanosti Sveučilišta u Zagrebu. Zbog specifičnih poteškoća u prikupljanju specijaliziranog paralelnog korpusa — u ovom slučaju englesko-hrvatskog paralelnog korpusa iz područja cestovnog prometa — prikupljen je relativno mali korpus, ali je istovremeno uravnotežen, pažljivo sastavljen te prilagođen potrebama studenata. Takav korpus može pružiti bogatstvo informacija relevantnih za specijaliziranu varijantu jezika. Cilj je ovog rada opisati proces izrade paralelnog korpusa temeljenog na podacima iz cestovnog prometa te istražiti mogućnosti njegove uporabe u obrazovnom kontekstu. U ovom se slučaju koristio jezični alat CorAl (CORpus ALigner), razvijen na Fakultetu elektrotehnike i računarstva u Zagrebu, a pogodan upravo za rad s paralelnim korpusima, koji su se pokazali vrijednim alatom u poučavanju stranog jezika, uz tradicionalne metode poučavanja.

**Ključne riječi:** cestovni promet, jezik struke, paralelni korpusi, sravnjivanje, vokabular

## 1. UVOD

Korpus je digitalna zbirka autentičnih tekstova sastavljena prema specifičnim kriterijima za određenu svrhu kojoj se može pristupiti putem specijaliziranog softvera za lingvističku analizu. Hunston (2002: 2) definira korpus kao „zbirku prirodno pojavljujućih primjera jezika koji su prikupljeni za lingvističko proučavanje”. Danas su te zbirke tekstova strojno čitljive, što znači da se pohranjuju i koriste u elektroničkom obliku. Korpusi se koriste u različite svrhe, poput komparativnih jezičnih studija,

studija prevođenja, crpljenja i analize terminologije, poučavanja i učenja stranih jezika i mnogih drugih.

Posljednjih godina korpusi postaju sve priznatiji i vrijedniji alat u istraživanju i poučavanju stranih jezika. Kontinuirani razvoj istraživanja korpusa doveo je do stvaranja velikog broja specijaliziranih korpusa i širokog spektra primjena ove tehnologije. Jedan od novijih razvojnih pravaca u ovom području jest uporaba paralelnih korpusa. Iako se paralelni korpusi već više od dva desetljeća koriste u studijama prevođenja i komparativnim jezičnim istraživanjima, njihova integracija u poučavanje stranih jezika noviji je trend (McEnery i Xiao 2008). Glavni razlog je u tome što su, zahvaljujući tehnološkom napretku i korisnički pristupačnijim sučeljima korpusa, nedovoljno iskorišteni potencijali paralelnih korpusa tek nedavno počeli dolaziti u prvi plan.

Ovaj rad pruža prijedloge za primjenu specijaliziranih paralelnih korpusa u osmišljavanju zadataka i vježbi za potrebe engleskog jezika struke (EJS). Naglašava mogućnost izravne interakcije studenata s korpusom u okviru učenja temeljenog na podacima, s ciljem poticanja njihove veće samostalnosti.

## **2. TEORIJSKA OSNOVA**

### **2.1. Specijalizirani paralelni korpus**

Postoje različite vrste jezičnih korpusa, a odabir određene vrste ovisi o lingvističkoj analizi koju želimo provesti. Opći jezični korpus zbirka je pisanog i/ili govornog jezika koja ima za cilj prikazati jezik u njegovoj cjelovitosti. Nasuprot tome, specijalizirani korpus, poznat i kao korpus za specifične svrhe ili domenski korpus, elektronički je dostupna zbirka tekstova usmjerena na određeno područje komunikacije, koja odražava specifičnu domenu jezične uporabe. Takvi se specijalizirani korpusi obično sastoje od tekstova specifičnih za određeni žanr ili profesionalno područje, predstavljajući samo jednu znanstvenu ili stručnu disciplinu (Anđelković i dr. 2018: 8).

Brojne su prednosti korištenja manjeg specijaliziranog korpusa u odnosu na korpus općeg jezika. Jedna je od njih mogućnost preciznijeg uvida u korištenje jezičnih obrazaca u točno određenom okruženju, jer specijalizirani korpusi imaju manju raznolikost mogućih okruženja. Stoga je pregledavanje konteksta promatrane jezične jedinice jednostavnije. Također, dok veliki opći korpus pri analizi zahtijeva uporabu uzoraka vrlo učestalih pojava, manji korpus omogućuje detaljno proučavanje svih relevantnih primjera, a ne samo nasumično odabranih (Carter i McCarthy 1995: 143). Stoga se može reći da manji specijalizirani korpusi nedostatak kvantitete nadoknađuju kvalitetom. Uz to, sastavljač korpusa najčešće je i osoba koja provodi analizu, pa je stoga dobro upoznat s kontekstom.

Međutim, pri izboru korpusa u ovom radu otišlo se korak dalje te se prikupio specijalizirani paralelni korpus. Slijedeći terminologiju Bakera (1995: 230) te McEneryja i Wilsona (1996: 58), paralelni korpus odnosi se na zbirku tekstova na jednom ili više jezika, zajedno s njihovim prijevodima na drugi jezik ili jezike, a sve je pohranjeno u strojno čitljivom formatu. To uključuje izvorne tekstove i njihove prijevode, što znači da je isti sadržaj dostupan na različitim jezicima.

## 2.2. Primjena paralelnih korpusa u nastavi jezika struke

Primjena paralelnih korpusa u obrazovnom kontekstu, osobito u nastavi jezika struke, posljednjih je godina sve više u središtu interesa istraživača. Simoes (2004: 7) navodi da se paralelni korpusi mogu koristiti za terminološke studije, automatsko prevođenje, višejezični dohvat podataka i poučavanje drugog jezika. Međutim, McEnery i Xiao (2008) tvrde da, iako su paralelni korpusi intenzivno i učinkovito korišteni u lingvistici i istraživanjima prevođenja, njihove potencijalne primjene u poučavanju i učenju drugog jezika (J2) ostaju uglavnom neistražene. Iako je broj takvih istraživanja malen, neki su se autori bavili primjenom paralelnih korpusa u nastavi. Prema Barlow (2000), pri korištenju paralelnih korpusa mogu se usporediti ekvivalenti između J1 i J2, poput engleske riječi *head* i francuske riječi *chef*, te uočiti da nemaju jednoznačnu ekvivalenciju. Uključivanje J1 konkordancijskih linija iz izvornog jezika zajedno s onima na ciljnom jeziku može pomoći učenicima u razumijevanju sadržaja na ciljnom jeziku (Chujo i dr. 2016). Nadalje, paralelni korpusi mogu poboljšati razumijevanje J2, osobito kada jednojezični ili dvojezični rječnici ne uspijevaju učenicima dati dovoljno relevantnih informacija (Frankenberg-Garcia 2005). Istraživanje koje je provela Satake (2022) pokazalo je da su studenti koji su koristili korpus postigli veći uspjeh u učenju kolokacija od studenata koji su koristili rječnike, dok je Gilquin (2021) otišla korak dalje i pokazala da se korpusi mogu koristiti i za poučavanje većih jezičnih jedinica, kao što su sintaktičke konstrukcije.

Paralelni korpusi mogu poslužiti kao vrijedni izvori autentičnih fraza i rečenica koje želimo da učenici nauče, pritom ilustrirajući kako se može unaprijediti poučavanje stranih jezika. Jedinstvena je vrijednost paralelnih korpusa njihova uloga kao resursa za učenike koji se suočavaju s kompleksnim strukturama, terminima i konceptima, pomažući im da te izazovne elemente razumiju na materinskom jeziku. U nastavi engleskog jezika struke uporaba korpusa može znatno olakšati nastavnicima poučavanje vokabulara. Korištenje paralelnih tekstova omogućuje studentima stjecanje jezičnog znanja, kao i znanja iz specifičnih područja (u ovom slučaju iz područja cestovnog prometa), proširujući njihove profesionalne perspektive, potičući samostalnost u učenju te unaprjeđujući njihove vještine rješavanja problema. Podatci iz korpusa mogu se koristiti za izradu nastavnih materijala, za pripremu aktivnosti vezanih uz vokabular i gramatiku, istovremeno izlažući učenike autentičnom sadržaju koji im pokazuje stvarni jezik i potiče ih da testiraju hipoteze o načinu na koji se jezik

koristi (Borucinsky i Jelčić Čolakovac 2020). Upravo se usko specijalizirano područje brodstrojarstva pokazalo prikladnim za istraživanje Borucinsky i Tominac-Coslovich (2021), u kojem su autorice pokazale kako se korpusi mogu primijeniti u učionici kako bi studentima pružili nove i inovativne tehnike za istraživanje jezika.

Međutim, korištenje paralelnih korpusa u nastavi ima i određene nedostatke. Nekim je učenicima velika količina informacija koje korpus pruža prevelika, posebno onima nižih jezičnih kompetencija ili slabijih sposobnosti uopćavanja i sažimanja. Stoga bi nastavnici trebali postupno poticati učenike na razvijanje navike samostalnog učenja pomoću korpusa, u skladu s njihovim individualnim sposobnostima. Primjena paralelnih korpusa u nastavi trenutačno je ograničena jer ih je relativno malo, a mnogi od postojećih korpusa razvijeni su prvenstveno za lingvistička istraživanja, a ne kao lako dostupni obrazovni alati.

Römer (2011: 207) razlikuje dvije primjene korpusa u poučavanju jezika: izravnu i neizravnu. Autorica navodi da nastavnici mogu neizravno koristiti jezične korpusne uključivanjem rezultata svojih analiza u oblikovanje nastavnih kurikuluma, referentnih izvora (kao što su rječnici i gramatike) i obrazovnih materijala. U tom slučaju studenti nemaju pristup samom korpusu. Učenici i nastavnici mogu izravno koristiti ove alate u praksi poučavanja temeljenoj na podacima, (engl. *Data-Driven Learning*, DDL), gdje su učenici u izravnom kontaktu s korpusima, što je možda još značajnije od njihova neizravnog korištenja jezičnih korpusa. Poučavanje temeljeno na podacima odnosi se na pristup poučavanju koji uključuje korištenje velike količine autentičnih jezičnih podataka za istraživanje i otkrivanje određenih jezičnih obrazaca te razumijevanje njihove upotrebe i značenja. Pritom učenik otkriva značenje, donosi zaključke i pronalazi pravila i principe koji opisuju stvarne jezične obrasce. Istovremeno, učitelj je tu kako bi olakšao proces učenja, a ne bio krajnji izvor informacija.

Jedna od prednosti DDL-a jest element istraživanja, koji može proces učenja učiniti motivirajućim i zanimljivim. Učenici se potiču na analiziranje podataka iz korpusa, postavljanje pretpostavki i oblikovanje pravila kako bi bolje razumjeli jezik ili provjerili točnost pravila iz svojih udžbenika. Osim toga, korištenjem DDL-a učenici mogu steći ili usavršiti niz ključnih vještina učenja: „predviđanje, opažanje, primjećivanje, razmišljanje, zaključivanje, analiziranje, interpretiranje, promišljanje, istraživanje, donošenje zaključaka (induktivno ili deduktivno), usmjeravanje pažnje, nagađanje, uspoređivanje, razlikovanje, teorijsko promišljanje, postavljanje hipoteza i provjeravanje“ (O’Sullivan 2007: 277). Mnoge od ovih navedenih vještina navedene su u ishodima učenja koje učenici trebaju ostvariti u raznim područjima kompetencija.

S druge strane, česta kritika poučavanja temeljenog na podacima i izravne uporabe korpusa u učionici jest da mnogi, ako ne i većina korpusa dostupnih danas, nisu prikladni za učenike jer je jezik u njima previše složen za razumijevanje ili primjenu. Također, alati za pretraživanje i obradu korpusa obično se smatraju nepristupačnima

za učenike, jer su primarno osmišljeni za istraživače. Osim toga, logistika se često navodi kao jedan od najvećih problema DDL-a. Za uporabu korpusa u učionici potrebna su računala (idealno jedno po učeniku ili barem jedno za svako dvoje ili troje učenika), kao i korpusi te softver za pretraživanje teksta. Navedeni su uvjeti češće dostupni u visokoobrazovnim ustanovama, pa Farr i Hagen Karlsen (2023) primjećuju da se DDL uglavnom primjenjuje upravo u visokom obrazovanju, osobito u područjima akademskog pisanja i obrazovanja nastavnika. Jedno od takvih istraživanja proveli su Kegalj i Borucinsky (2025), u kojem su prikazali kako se korpusi (zbirke tekstova), korpusni alati (softverski paketi) i korpusne metode (tehnike za analizu korpusnih podataka) mogu primijeniti u razvoju studentskih vještina pisanja, a istovremeno doprinijeti unapređenju njihovih digitalnih kompetencija, u skladu s modernim obrazovnim trendovima.

Unatoč navedenim nedostacima, izravna uporaba korpusa ima potencijal značajno poboljšati samostalnost učenika potičući njihovu sposobnost usvajanja terminologije, fraza, kolokacija, gramatičkih obrazaca i drugih struktura specifičnih za specijalizirani diskurs, a ne samo za istraživanje jezika. Svako od navedenih istraživanja pokazuje primjenu korpusa u različitim područjima, a ovaj rad predlaže uvođenje specijaliziranog paralelnog korpusa za engleski jezik u području cestovnog prometa, što do sada nije bilo dovoljno istraženo.

### **2.3. Engleski jezik struke na Fakultetu prometnih znanosti u Zagrebu**

Kako bi se bolje razumjele mogućnosti primjene specijaliziranog paralelnog korpusa, potrebno je ukratko predstaviti status engleskog jezika struke na Fakultetu prometnih znanosti u Zagrebu. Na navedenoj ustanovi, engleski jezik struke integriran je u prijeddiplomske programe. Studenti su obvezni završiti tri jednosemestralna kolegija tijekom prve i druge godine studija: Engleski jezik 1, Engleski jezik 2 i Engleski jezik 3. Predavači engleskog jezika struke na ovom fakultetu suočavaju se s izazovom poučavanja velikih grupa studenata različitih razina predznanja engleskog jezika. Nedostatak inicijalnog testiranja za razvrstavanje po razinama znanja rezultira značajnim razlikama u jezičnim vještinama studenata.

Osim toga, izostanak standardiziranih udžbenika i prikladnih nastavnih materijala za tako raznolike skupine znači da se nastavni materijali moraju izrađivati i prilagođavati specifičnim potrebama studenata. Često su se koristili materijali iz drugih kolegija te su se uključivali autentični tekstovi u izvornom ili prilagođenom obliku. Međutim, ovi su pristupi vremenski zahtjevni, osobito kad je riječ o odabiru i prilagodbi autentičnih materijala, te često ne donose željene rezultate.

S obzirom na istraživanja koja naglašavaju sve veću primjenu specijaliziranih korpusa u nastavi jezika struke, logično je istražiti potencijalne prednosti korištenja specijaliziranog paralelnog korpusa u nastavi engleskog jezika struke na Fakultetu prometnih znanosti.

Takav korpus mogao bi biti ključan za poboljšanje usvajanja vokabulara, unaprijeđenje vještina pisanja te podršku u usporednim i prijevodnim vježbama.

### 3. ISTRAŽIVANJE

#### 3.1. Sastavljanje korpusa

Model korišten za izradu korpusa opisan je u radu de Carvalho i dr. (2021). Autori navode da se za izradu vlastitog korpusa mogu koristiti različiti izvori, poput interneta, računalnih baza podataka sastavljača te tekstovi iz dotične domene. Literatura iz prometa i logistike, a posebno cestovnog prometa, opsežna je i raznolika. S obzirom na raznolikost terminologije u ovom području, u korpus je uključen širok raspon vrsta tekstova među kojima su članci iz znanstvenih časopisa, izvještaji prometnih udruga, bilteni, savjeti za vozače iz različitih prometnih časopisa, znanstveni i stručni članci, prometni zakoni, upute, izvješća sa stručnih konferencija, poglavlja iz knjiga, bilješke s predavanja, diplomski radovi, propisi, odredbe te članci iz popularnih časopisa. Jedan dio prikupljenih tekstova osigurali su nastavnici Fakulteta prometnih znanosti tj. stručnjaci iz predmetnog područja. Drugi dio tekstova ti su nastavnici odobrili kao prikladne, što doprinosi relevantnosti tekstova i pomaže da korpus bude što je moguće reprezentativniji. Ključno je bilo da tekstovi budu javno dostupni u digitalnom formatu i relevantni za cestovni promet kako bi se osigurala terminološka točnost.

Tekstovi su se, također, prikupljali i na internetu, gdje postoji velik izbor paralelnih tekstova u različitim jezičnim parovima. Raspoloživost uvelike ovisi o području proučavanja, ali problem predstavlja činjenica da visokokvalitetni stručni i znanstveni članci najčešće nisu besplatno javno dostupni. Osim toga, vrlo je teško pronaći paralelne tekstove u jezičnim parovima rjeđe korištenih jezika. Razumljivo, najčešći jezik u jezičnom paru je engleski. Čak i kada se pronađu prijevodi internetskih stranica na traženom jeziku, oni su često loše kvalitete i ne mogu se koristiti u paralelnom korpusu.

Engleski potkorpus ima 58 datoteka i ukupno 346 945 pojavnica. Hrvatski potkorpus također ima 58 datoteka, ali 289 933 pojavnica što se može objasniti flektivnom prirodom hrvatskog jezika. Ovaj je obujam podataka dovoljan uzimajući u obzir usku žanrovsku namjenu predmetnog korpusa.

U tablici 1 prikazan je broj tekstova za svaki od uključenih žanrova, razvrstanih prema tematskim cjelinama. Stupac s postotcima prikazuje udio svake tematske cjeline u ukupnom korpusu, kao i udio pojedinog žanra unutar te cjeline.

**Tablica 1.** Broj tekstova i postotak svakog pojedinog žanra u predmetnom korpusu

| Žanr                                                    | Broj tekstova | Postotak      |
|---------------------------------------------------------|---------------|---------------|
| <b>Tematska cjelina: Znanstveni i stručni radovi</b>    | <b>15</b>     | <b>25,8 %</b> |
| Znanstveni radovi                                       | 4             | 7 %           |
| Članci iz znanstvenih časopisa                          | 3             | 5 %           |
| Diplomski radovi                                        | 2             | 3,4 %         |
| Poglavlja iz knjiga                                     | 2             | 3,4 %         |
| Bilješke s predavanja                                   | 4             | 7 %           |
| <b>Tematska cjelina: Propisi i službeni dokumenti</b>   | <b>27</b>     | <b>46,5 %</b> |
| Prometni zakoni                                         | 4             | 7 %           |
| Izveštaji prometnih udruga                              | 5             | 8,6 %         |
| Izvešća s konferencija                                  | 3             | 5 %           |
| Upute                                                   | 4             | 7 %           |
| Odredbe                                                 | 5             | 8,6 %         |
| Propisi                                                 | 6             | 10,3 %        |
| <b>Tematska cjelina: Popularni i praktični tekstovi</b> | <b>16</b>     | <b>27,5 %</b> |
| Bilteni                                                 | 6             | 10,3 %        |
| Savjeti za vozače                                       | 7             | 12 %          |
| Članci iz znanstvenih časopisa                          | 3             | 5 %           |
| <b>UKUPNO</b>                                           | <b>58</b>     |               |

Iz tablice je vidljivo da tekstovi potječu iz 14 različitih žanrova unutar područja cestovnog prometa, što korpusu osigurava visoku razinu raznolikosti. Najveći broj tekstova pripada tematskoj cjelini propisa i službenih dokumenata, što je očekivano s obzirom na njihovu dostupnost. Kao što je prethodno istaknuto, relevantnost korpusa potvrđena je činjenicom da su nastavnici, stručnjaci za predmetno područje, odobrili odabrane tekstove kao prikladne.

### 3.2. Obrada korpusa

Nakon sastavljanja korpusa pristupa se njegovoj obradi, koja obuhvaća glavne korake poput pretvorbe datoteka, uparivanja tekstova te sravnjivanja na razini paragrafa i rečenice.

Seljan i dr. (2009) istraživali su automatsko usklađivanje rečenica u hrvatsko-engleskim paralelnim korpusima pomoću različitih alata za sravnjivanje rečenica, kao što su *SDL Trados WinAlign* (alat za sravnjivanje razvila britanska tvrtka), *Vanilla Aligner* (osnovni algoritam za sravnjivanje), *Hunalign* (besplatno sravnjivanje rečenica) i *CorAl* (domaće rješenje razvijeno na FER-u). Cilj je bio odrediti koji je od tih alata najprikladniji za sravnjivanje korpusa hrvatskog i drugih jezika u budućim projektima. Odabran je *CorAl* (CORpus ALigner) jer je algoritam Gale-Church, na kojem se temelji *CorAl*,

dokazano vrlo učinkovit (preciznost je 97–98 %) u sravnjivanju segmenata u odnosu 1:1. Pritom je uzeto u obzir da su tekstovi relativno kratki — s prosječno 6000 pojava na engleskom jeziku i 5000 pojava na hrvatskom jeziku — i jednostavni u pogledu složenosti ručnog sravnjivanja. *CorAl* je jezični alat razvijen na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Služi za sravnjivanje dvaju dokumenata na razini rečenice, tj. stvaranje prijevodne memorije. Za njegovo pokretanje potreban je samo Java Runtime Environment, što znači da se može pokrenuti na bilo kojem operativnom sustavu. Alat je jednostavan za korištenje i ne zahtijeva posebno računalno predznanje.

Nakon odabira tekstova, sljedeći je korak bio njihovo pretvaranje u oblik običnog teksta kako bi bili kompatibilni sa softverom za obradu korpusa. budući su datoteke bile u raznim formatima (npr. PDF, DOC, HTML), konvertirane su u UTF-8 format s ekstenzijom .txt, jer je odabrani softver zahtijevao taj format. Datoteke su također preimenovane radi lakšeg sortiranja i upravljanja. Završni korak, koji je opcionalan, ali je u ovom slučaju i proveden, uključuje ručno uklanjanje svih netekstualnih sadržaja iz svake datoteke. Premda je postupak vremenski zahtjevan, s obzirom na ograničeni opseg predmetnog korpusa, proveden je radi povećanja učinkovitosti sravnjivanja.

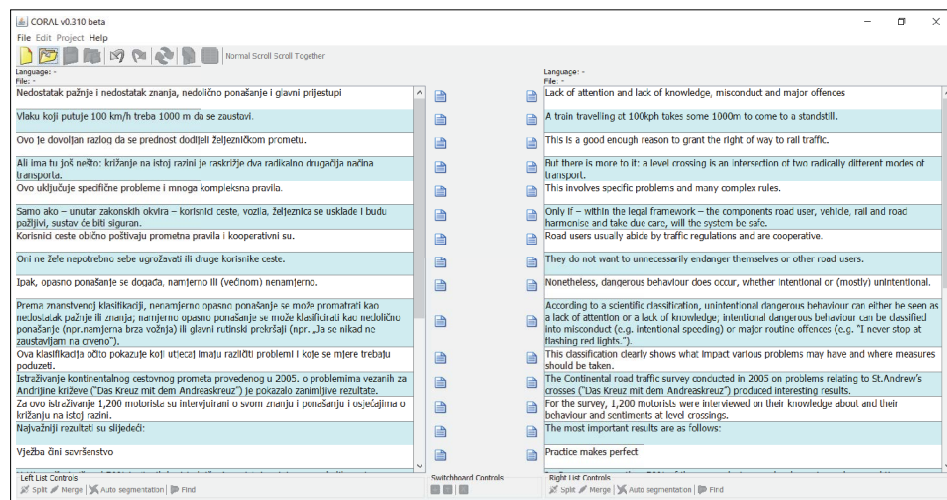
Tijekom ovih procesa povremeno su se pojavljivali sljedeći problemi: prilikom konverzije u .txt dokumente, ponekad su nedostajali određeni znakovi i simboli iz izvornih PDF datoteka, osobito dijakritički znakovi u hrvatskim tekstovima (svi su pisani latiničnim pismom). Osim toga, prilikom konvertiranja PDF datoteka u obični tekst, dva stupca su se ponekad spajala u jedan. Kako bi se smanjile ove i slične pogreške, neke su datoteke prvo konvertirane u Microsoft Word format (.doc), zatim ručno ispravljene i tek potom spremljene kao običan tekst. Kako je već i navedeno, nakon konverzije teksta, svi elementi koji nisu relevantni za jezičnu analizu (npr. tablice, grafikoni, dijagrami, formule, referencije, sadržaji, zaglavlja itd.) uklonjeni su iz korpusa.

Iako se paralelnom korpusu mogu dodati morfološka analiza, lematizacija, sintaktička analiza, itd., ta svojstva nisu specifična za paralelne korpus. Prvi je korak u obogaćivanju paralelnog korpusa povećanje paralelizma između jezičnih jedinica u oba teksta, što se naziva sravnjivanjem (engl. *alignment*). Podudarnost između pojedinih tekstualnih jedinica (poput rečenica ili riječi), kao i proces uspostavljanja tih odnosa podudarnosti, naziva se usklađivanjem. Sravnjivanje se može napraviti na razini odlomka, rečenice, segmenata, riječi i znakova, pri čemu je uobičajeno da se radi na razini rečenice i riječi. Sravnjivanje na razini rečenice omogućuje točno utvrđivanje koja rečenica u izvornom tekstu odgovara kojoj u ciljnom tekstu i kakav je njihov odnos. Međutim, prije toga, potrebno je napraviti sparivanje tekstova, što znači pronalazak parova tekstova koji prijevodom odgovaraju jedan drugome. U prikupljenom se paralelnom korpusu može dogoditi da su poneki tekstovi pogrešno spareni. Pogrešno će spareni tekstovi dati vrlo loš rezultat pri sravnjivanju teksta na razini rečenice i takvi se uklanjanju iz korpusa. Nakon sastavljanja paralelnog korpusa i pronalaženja tekstova parnjaka, slijedi sravnjivanje na razini odlomka, a potom i rečenice.

Pri sravnjivanju rečenica, najčešće jedna rečenica u izvornom tekstu odgovara jednoj rečenici u ciljnom tekstu, međutim, to ne mora uvijek biti slučaj. Ponekad se jedna rečenica iz izvornog teksta prevodi s dvije ili tri rečenice u ciljnom tekstu. Nije rijetkost da su neke rečenice čak i izostavljene iz prijevoda (Deléger i dr. 2009: 694). Sravnjivanje na razini rečenice nužan je preduvjet poravnanju teksta na razini riječi, iako taj postupak ovdje nije proveden. Učinkovitost usklađivanja rečenica uvelike ovisi o broju jednoznačnih podudaranja, pri čemu je jedna rečenica na jednom jeziku izravno prevedena jednom rečenicom na drugom jeziku. Kada postoji veliki broj odnosa jedan-prema-više ili rečenica koje nemaju odgovarajuće podudaranje na drugom jeziku učinak sravnjivanja može biti znatno smanjen.

CorAI-ov modul za segmentaciju rečenica može se koristiti za automatsku segmentaciju rečenica. Za uspješnu segmentaciju rečenica potreban je popis kratica koje služe kao iznimke u odgovarajućem jeziku. Također je podržano ručno segmentiranje i spajanje rečenica na ekranu. Ovaj način rada u potpunosti iskorištava napredno grafičko sučelje posebno razvijeno za ergonomsku ručnu izradu poravnanja rečenica. Sučelje je jasno, jednostavno za korištenje, intuitivno te ne zahtijeva posebno predznanje u radu s ovakvim tipom softvera, što se vidi na slici 1.

Na slici 1 prikazana je snimka zaslona programa CorAI. Kao primjer odabran je jedan par tekstova — na lijevoj je strani hrvatski tekst, a na desnoj engleska verzija u tom paru. Navedeni par tekstova sravnjen je na razini rečenice. Rečenice su međusobno razdvojene i svaka se nalazi u svom retku, što znači da je moguće pratiti svaku rečenicu pojedinačno i njen prijevod na drugi jezik u jezičnom paru.



Slika 1. Snimka zaslona programa CorAI

### 3.3. Izravna i neizravna primjena korpusa u nastavi engleskog jezika na Fakultetu prometnih znanosti

Sve nastavne aktivnosti povezane s paralelnim korpusom temelje se na očekivanom poravnanju između pojedinih jedinica korpusa u ciljnom jeziku i materinskom jeziku učenika ili u drugom stranom jeziku kojim studenti dovoljno dobro vladaju. Raspon mogućih aktivnosti širok je i uvelike ovisi o kreativnosti nastavnika. Može se reći da većina aktivnosti poučavanja jezika uz pomoć korpusa uključuje različite vrste konkordancija. Konkordancijske linije mogu biti skraćene (tzv. ključna riječ u kontekstu) ili imati oblik cijele rečenice te se mogu prikazivati na ekranu ili ispisivati na radnim listovima. Studentima se može prikazati konkordancijska lista sortirana abecednim redom i potaknuti ih da prepoznaju ponavljajuće leksičke obrasce ili ih grupiraju na smislen način. Također, mogu dobiti niz konkordancijskih iskaza u kojima je riječ izostavljena, a njihova je zadaća pronaći riječ koja nedostaje. S druge strane, konkordancije mogu dolaziti iz dvojezičnog korpusa, a mogu se izraditi od paralelnog korpusa prikazanog u ovom radu, što je osobno korisno studentima koji dijele isti materinski jezik jer im omogućuje uspostavljanje ekvivalencija između materinskog i ciljanog jezika (Gilquin i Granger 2010). Budući da studenti na FPZ-u povremeno imaju problema s pronalaženjem odgovarajućih prijevoda za višerječne izraze, osobito višeimenične nazive (npr. *road markings retroreflectivity*), upravo se dvojezični paralelni korpus pokazuje kao neprocjenjiv izvor prijevodnih ekvivalenata. Dvojezična struktura daje mu prednost pred jednojezičnim korpusom, kod kojeg je potrebna dodatna uporaba rječnika. U ovom slučaju, studenti mogu dobiti radne listiće s dijelovima korpusa ili raditi s korpusom na računalima, naravno ako za to postoje tehnički uvjeti. Također, kao dio neizravne uporabe korpusa, nastavnik može izraditi ciljano oblikovane nastavne materijale koji mogu uključivati zadatke popunjavanja praznina kako bi studenti istražili stručni vokabular, podudaranje riječi, pronalaženje pogrešaka, ključne fraze i specijalizirane jezične obrasce (uzorke), tekstove za čitanje, zadatke pisanja i prevođenja koji se usredotočuju na ciljani osnovni vokabular, izradu testova itd. Nedavno istraživanje provela je Lyu (2024), koja je istaknula potencijal korištenja vježbi uparivanja u poučavanju tehničkog rječnika, pokazavši da one mogu biti učinkovite dopunske aktivnosti u nastavi engleskog jezika struke, omogućujući studentima da poboljšaju svoje razumijevanje stručnog jezika.

## 4. ZAKLJUČAK

Iako postoji sve veći broj radova posvećen uključivanju korpusa u poučavanje jezika, još uvijek je malo istraživanja o tome kako učinkovito integrirati u svakodnevnu učioničku praksu. Nastavnici često imaju zadani nastavni plan i propisan udžbenik, što im ostavlja malo prostora za dodatne materijale. Stoga, čak i ako su se tijekom

svog obrazovanja ili profesionalnog razvoja upoznali s korpusima, to ne znači da će ih nastavnici primijeniti u svojoj nastavi. Gotovi nastavni materijali za jezik — od kojih su mnogi već temeljeni na korpusima — posebno u području engleskog kao stranog jezika, često se doživljavaju kao praktičnije rješenje. Bilo bi neučinkovito da nastavnici troše vrijeme na izradu dodatnih materijala koji se temelje na korpusima ili na obuku učenika za samostalno korištenje korpusa. Međutim, nastavnici mogu koristiti korpuse kako bi izašli izvan okvira umjetnog jezika prisutnog u udžbenicima, izložili učenike autentičnom jeziku i pomogli im pronaći odgovore na jezična pitanja koja možda nisu dovoljno obrađena u gramatikama i rječnicima. Prvi korak u ovom procesu jest osigurati da nastavnici ne budu obeshrabreni prije nego što počnu koristiti korpuse. Ključno je da shvate kako uključivanje korpusa u nastavu ne zahtijeva kupnju ili instalaciju novog softvera, izradu vlastitog korpusa ili pretjerano ulaganje vremena — ne više nego bilo koja druga dodatna aktivnost u učionici. Također, nije nužno imati računalni laboratorij s individualnim radnim mjestima za svakog učenika. Ispisi iz korpusa mogu učinkovito zamijeniti računala. Ključna poruka koju treba prenijeti jest da uporaba korpusa ne mora biti složena niti odvojena od redovne nastave, već se može integrirati u postojeće nastavne materijale i uobičajene aktivnosti u učionici.

Korištenje paralelnih korpusa kao pedagoškog alata predstavlja novo istraživačko područje koje je još uvijek u svojim počecima, s više pitanja nego odgovora. Postoji značajna potreba ne samo za daljnjim istraživanjima koja koriste paralelne korpuse, već i za stvaranjem novih paralelnih korpusa posebno prilagođenih obrazovnim ciljevima. Štoviše, u ovom je području primjetan nedostatak detaljnih studija slučaja u okviru pedagoških istraživanja. Iako su paralelni korpusi pokazali potencijal kao vrijedni alati za učenje jezika, najučinkovitiji načini za korištenje ove tehnologije još su uvijek uglavnom neistraženi. Jedan od mogućih pravaca budućih istraživanja je korištenje nekog drugog softvera prigodnog za paralelne korpuse, kao što je *AntPConc*. Dodatna se mogućnost otvara uključivanjem studenata u proces izrade korpusa. Umjesto da nastavnici sastavljaju korpus, studenti bi sami mogli prikupljati, razvrstavati i analizirati relevantne tekstove, poput prometnih izvješća, službenih dokumenata o prometu ili radijskih transkripata. Time bi se proces učenja učinio interaktivnijim i zanimljivijim, približavajući ga konceptu projektno utemeljenog učenja.

Zaključno, predmetni hrvatsko-engleski paralelni korpus iz cestovnog prometa može biti koristan resurs za izravnu upotrebu u nastavi. Osim toga, posredno nudi mogućnost razvoja nastavnih materijala i drugih didaktičkih resursa temeljenih na analizama učestalosti riječi, fraza, kolokacija i gramatičkih obrazaca prepoznatih unutar korpusa. Međutim, unatoč prednostima izravne primjene paralelnih korpusa, poučavanje jezika struke ne bi se smjelo temeljiti isključivo na korištenju korpusa, već ga treba uključiti kao dopunu tradicionalnom pristupu.

## Literatura

- Andelković, Jelena; Seničić, Danica; Stanković, Ranka. 2018. Aligned Parallel Corpus for the Domain of Management: Preparation and Potential Applications. *Infotheca* 18. 7–28.
- Baker, Mona. 1995. Corpora in translation studies. An overview and some suggestions for future research. *Target* 7/2. 223–243.
- Barlow, Michael. 2000. Parallel texts in language teaching. *Multilingual Corpora in Teaching and Research*. Amsterdam, Rodopi. 106–115.
- Borucinsky, Mirjana; Jelčić-Čolakovac, Jasmina. 2020. Promoting Authenticity in the ESP Classroom: The Impact of ICT and Use of Authentic Materials on Reading Comprehension. *5th International e-Conference on Studies in Humanities and Social Sciences: Conference Proceedings*. Belgrade. Center for Open Access in Science.
- Borucinsky, Mirjana; Tominac-Coslovich, Sandra. 2021. Introducing Data-Driven Learning into the Marine Engineering English Classroom. *Humanities Science Current Issues* 3. 19–27.
- Carter, Ronald; McCarthy, Michael. 1995. Grammar and the Spoken Language. *Applied Linguistics* 16/2. 141–158.
- Chujo, Kiyomi; Kobayashi, Yuichiro; Mizumoto, Atsushi; Oghigian, Kathryn. 2016. Exploring the Effectiveness of Combined Web-based Corpus Tools for Beginner EFL DDL. *Linguistics and Literature Studies* 4/4. 262–274.
- de Carvalho, Carolina; Laranja, Luana; Pinto, Paula. 2021. Do-it-Yourself Corpora: what are they and who are they for. *Tradterm* 37. 64–87.
- Deléger, Louise; Merkel, Magnus; Zweigenbaum, Pierre. 2009. Translating medical terminologies through word alignment in parallel text corpora. *Journal of Biomedical Informatics* 42. 692–701.
- Farr, Fiona; Hagen Karlsen, Petter. 2023. DDL Pedagogy, Participants, and Perspectives. *The Routledge Handbook of Corpora and English Language Teaching and Learning*. Ur. Jablonkai, Reka Ratkaine; Csomay, Eniko. London and New York: Taylor & Francis. 329–343.
- Frankenberg-Garcia, Ana. 2005. Pedagogical uses of monolingual and parallel concordances. *ELT Journal* 59/3. 189–198.
- Gilquin, Gaëtanelle; Granger, Sylviane. 2010. How can data-driven learning be used in language teaching? *The Routledge Handbook of Corpus Linguistics*. Ur. O'Keeffe, Anne; McCarthy, Michael. Routledge. 359–370.
- Gilquin, Gaëtanelle. 2021. Using Corpora to Foster L2 Construction Learning: A Data-Driven Learning Experiment. *International Journal of Applied Linguistics* 31/2. 229–247.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge University Press.
- Kegalj, Jana; Borucinsky, Mirjana. 2025. Data-Driven Learning for Writing Skills Development. *International Conference on Corpora in Language Learning, Translation and Research*. Zadar. 33–48.

- Lyu, Evgueniya. 2024. Teaching technical vocabulary: Are matching exercises the thing of the past? *The Journal of Teaching English for Specific and Academic Purposes* 12/1. 133–150.
- McEnery, Tony; Wilson, Andrew. 1996. *Corpus Linguistics*. Edinburgh University Press. Edinburgh.
- McEnery, Tony; Xiao, Richard. 2008. Parallel and comparable corpora: What is happening. *Incorporating corpora: The linguist and the translator*. Ur. Anderman, Gunilla; Rogers, Margaret. Clevedon: Multilingual Matters. 18–31.
- O'Sullivan, Íde. 2007. Enhancing a Process-Oriented Approach to Literacy and Language Learning: The Role of Corpus Consultation Literacy. *ReCALL* 19/3. 269–286.
- Römer, Ute. 2011. Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics* 31. 205–225.
- Satako, Yoshiho. 2022. The Effects of Corpus Use on L2 Collocation Learning. *The JALT CALL Journal* 18. 34–53.
- Seljan, Sanja; Agić, Željko; Tadić, Marko. 2009. Evaluating Sentence Alignment on Croatian-English Parallel Corpora. *FASSBL 2008*. Ur: Tadić, Marko; Dimitrova-Vulchanova, Mila; Svetla, Koeva. Hrvatsko društvo za jezične tehnologije. 101–108.
- Simoes, Alberto. 2004. *Parallel Corpora Word Alignment and Applications*. Magistarski rad. Universidade do Minho. Braga. 128 str.

## USING A ROAD TRAFFIC PARALLEL CORPUS IN TEACHING LANGUAGES FOR SPECIFIC PURPOSES IN HIGHER EDUCATION

Corpora have recently been used in foreign language teaching, particularly in Languages for Specific Purposes (LSPs). Specialized parallel corpora within a specific field are especially valuable resources for educational purposes. Teachers of English for Specific Purposes often face challenges related to vocabulary instruction, particularly the lack of authentic teaching materials. This paper focuses on teaching English for Traffic and Transport to undergraduate students at the Faculty of Transport and Traffic Sciences, University of Zagreb. Due to specific difficulties in compiling a specialized parallel corpus - in this case, an English-Croatian parallel corpus in the field of road transport - a relatively small corpus was collected. However, it was carefully balanced, compiled, and adapted to meet students' needs. As such, this corpus can provide a wealth of information relevant to a specialized language variant. The aim of this paper is to describe the process of creating a parallel corpus based on road transport data and to explore its potential use in an educational context. In this study, the language tool CorAl (CORpus ALigner), developed at the Faculty of Electrical Engineering and Computing in Zagreb, was used. CorAl is particularly suitable for working with parallel corpora, which have proven to be valuable tools for foreign language teaching, albeit as a supplement to traditional teaching methods.

**Keywords:** alignment, language for specific purposes, parallel corpora, road transport, vocabulary