

Korpusi kao sredstvo za postizanje autonomije u učenju stranog jezika

Maja Miličević i Tanja Samardžić

(Univerzitet u Beogradu)

In: J. Vučo and B. Milatović (Eds), *Autonomija učenika i nastavnika u nastavi jezika i književnosti*. Nikšić: Faculty of Philosophy. 387-399.

This is a pre-print version. The paper is under copyright; for permission to re-use or reprint the material in any form please contact the editors.

Maja Miličević
Tanja Samardžić
Filološki fakultet
Beograd

Korpusi kao sredstvo za postizanje autonomije u učenju stranog jezika

U kontekstu sve raširenije upotrebe računara i Interneta u nastavi stranih jezika, jedno od raspoloživih sredstava u nastavnom procesu jesu i jezički korpusi. Kao skupovi autentičnih tekstova, korpusi predstavljaju dragocen izvor podataka za učenike i nastavnike, a pored toga učenicima mogu pomoći i u postizanju veće autonomije u učenju. Budući da broj korpusa dostupnih preko Interneta stalno raste, učenici bi uz osnovnu obuku i period vođenog korišćenja brzo mogli da se osposobe za samostalan rad sa njima. U ovom radu ilustrujemo upotrebu korpusa u italijanskom i srpskom kao stranim jezicima, na primerima vežbanja u kojima se koriste različite vrste pretrage i dobijaju različite vrste podataka.

Ključne reči: autonomija učenika, jezički korpus, konkordanca, regularni izraz, kolokacija

1. Uvod

Predmet ovog rada je upotreba jezičkih korpusa u nastavi stranog jezika. Radi se o temi kojoj su u poslednje dve decenije posvećeni brojni radovi, ali i pored zalaganja stručnjaka iz oblasti korpusne lingvistike, pedagoške gramatike i metodike nastave, korpusi se još uvek retko koriste u učionici, što ukazuje na potrebu većeg približavanja teorije praksi. Jedan od načina da se to ostvari jeste kroz dodatne konkretne primere vežbanja koja bi se uz početnu obuku nastavnika i učenika mogla raditi na času ili zadavati za samostalni rad kod kuće. Naš cilj u ovom radu jeste da se približimo tom cilju kroz navođenje primera vežbanja koja bi se mogla primeniti u nastavi srpskog i italijanskog jezika. Korpusi iz kojih su primeri uzeti jesu *Korpus savremenog srpskog jezika* (Univerzitet u Beogradu) i *Corpus di riferimento dell'italiano scritto* (Univerzitet u Bolonji). Pre navođenja praktičnih primera, tema se smešta u širi pedagoški kontekst, kao i u okvire razmatranja korpusne lingvistike. Iako se u zaključnom delu ukazuje i na neke praktične probleme, rad treba shvatiti kao još jedan poziv na svrstavanje korpusa među sredstva koja redovno upotrebljavaju i nastavnici i učenici.

2. Jezički korpusi kao nastavno sredstvo

Brojni radovi posvećeni korišćenju elektronskih korpusa u nastavi jezika svedoče o velikom potencijalu ove metode. O upotrebi jezičkih korpusa u pedagoške svrhe najčešće se govori u okviru zalaganja za tzv. „učenje zasnovano na podacima“ (eng. *Data-Driven Learning*, skraćeno DDL, v. Johns 1990, 2002). Osnovni postulat DDL pristupa jeste da učenici znanje stranog jezika treba da stiču pre svega induktivnim putem, kroz korišćenje autentičnih podataka i samostalno otkrivanje pravila u njima, a jezički korpusi svakako predstavljaju jedno od najznačajnijih sredstava pomoću kojih se to može postići. Budući da induktivno učenje zahteva visok nivo samostalnosti, jasno je da upotreba korpusa ujedno može značajno doprineti i razvoju autonomije učenika u procesu učenja jezika.

Mogući načini upotrebe jezičkih korpusa u nastavi vrlo su raznovrsni i mogu se podeliti na više tipova. Pre svega je bitno napraviti razliku između korpusa maternjeg jezika (eng. *native corpora*), koji sadrže tekstove pisanim ili usmenim putem proizvedene od strane izvornih govornika određenog jezika, i korpusa koje čini pisana ili usmena produkcija učenika koji neki jezik usvajaju kao strani (*learner corpora*). Iako obe vrste korpusa mogu naći primenu u nastavnom procesu, do sada je mnogo više pažnje posvećeno korpusima maternjeg jezika, kojima se bavimo i u ovom radu.

U upotrebi korpusa maternjeg jezika mogu se razlikovati *indirektni* i *direktni* pristup. Indirektni pristup podrazumeva da se nastavnik oslanja na podatke iz korpusa prilikom kreiranja nastavnog plana i programa, dok su u slučaju direktnog pristupa učenici u neposrednom kontaktu sa podacima, bilo u nekontrolisanom obliku (putem direktnog pretraživanja korpusa) ili u donekle kontrolisanom uslovima, kroz vežbanja koja nastavnik pripremi na osnovu materijala preuzetih iz korpusa (v. Bernardini 2004, Römer 2006). Meunier (2002) posebno izdvaja i upotrebu materijala iz korpusa kao referentne građe u kojoj se mogu proveriti određeni jezički obrasci i pravila - korpus u tom slučaju predstavlja dopunu rečnicima i gramatikama.

Međutim, i pored postojanja navedenih mogućnosti, kao i velikog broja teorijskih radova koji se njima bave, korpusi se u praksi još uvek retko koriste. Ova tvrdnja važi čak i za nastavu engleskog jezika, koji je predmet velike veličine istraživanja (v. npr. Römer 2006 za prikaz situacije u nastavi engleskog jezika u Nemačkoj). Osim ovog problema, treba istaći i jedan pretežno tehnički detalj. Naime, u postojećoj literaturi se u kontekstu

nastavnih materijala preuzetih iz korpusa daleko najčešće govori o upotrebi konkordanci, dok mnoge druge opcije koje korpusi pružaju ostaju ili potpuno zanemarene, ili se njihova složenost navodi kao razlog za njihovo retko korišćenje.

Cilj našeg rada jeste približavanje jezičkih korpusa nastavnoj praksi davanjem primera jednostavnih vežbanja koja bi se mogla raditi na času ili kao domaći zadatak. Takođe želimo da ukažemo na neke od dodatnih tehničkih mogućnosti koje elektronski korpusi pružaju, a koje bi se, uz osnovnu obuku u korišćenju, mogle lako primeniti i u nastavi. Pažnju posvećujemo dvama jezicima koji su ređe proučavani u ovom svetlu, srpskom i italijanskom. Koliko nam je poznato, ne postoji nijedno istraživanje koje se bavi upotrebom elektronskih korpusa u nastavi srpskog jezika kao stranog. Što se tiče italijanskog, određeni broj radova o njemu je već objavljen (npr. Đorović 2008 govori o upotrebi konkordanci u nastavi italijanskog kao jezika struke), ali mi želimo da ih dopunimo prikazom mogućnosti koje se u njima ne navode.

U narednim odeljcima prvo će biti uvedeni osnovni nivoi i tipovi pretrage koje korpusi nude, zatim će uslediti primeri vežbanja koja se mogu raditi na srpskom i italijanskom korpusu, posle čega će biti dato nekoliko zaključnih napomena u kojima ćemo se posebno osvrnuti na to kako upotreba korpusa u nastavi stranog jezika podstiče autonomiju učenika. U zaključnom delu takođe se navode prednosti elektronskih korpusa nad naizgled sličnim izvorima podataka (npr. Internet pretraživačima), a ističu se i neki praktični problemi i mogućnosti za njihovo prevazilaženje.

3. Nivoi upotrebe jezičkih korpusa

Elektronski korpusi pretražuju se preko upita koji se unose u odgovarajuće polje u korisničkom interfejsu. Upiti mogu biti različite složenosti, i shodno tome omogućavaju veću ili manju fleksibilnost pretrage i daju određeni tip rezultata. U zavisnosti od vrste upita i vrste dobijenih rezultata mogu se izdvojiti tri osnovna nivoa upotrebe korpusa - konkordance, regularni izrazi i statističke mere. Neutralni termin „nivoi“ koristimo zbog toga što se u većini slučajeva zapravo radi o kombinaciji tipa upita i vrste i načina prikaza rezultata. Navedeni nivoi su sažeto opisani u naredna tri odeljka, nakon čega se njihova upotreba ilustruje konkretnim primerima vežbanja.

3.1 Konkordance

Konkordance predstavljaju najčešći i najjednostavniji sistem za pretragu elektronskih korpusa. Dobijaju se kao odgovor na ma koji tip upita, zbog čega su pri korišćenju korpusa gotovo nezaobilazne. Definišu se kao prikaz “ključne reči u kontekstu” (eng. KWIC, *Key Word in Context*), budući da se u dobijenim rezultatima tražena reč (ili traženih više reči) pojavljuje u centralnom delu ekrana, tako da se vidi kakav rečenični i/ili diskursni kontekst je okružuje. Najčešće se može izabrati koliko širok levi i desni kontekst će biti prikazan, a kod većine korpusa je moguće i sortirati rezultate prema reči koja prethodi traženoj, ili dolazi posle nje, što znatno olakšava dalju analizu dobijenih primera. Konkordance kao najpristupačniji nivo upotrebe korpusa ujedno imaju i najširu primenu u nastavi jezika (v. Bernardini 2004, Đorović 2008).

3.2 Regularni izrazi

Za razliku od konkordanci, koje predstavljaju format odgovora na upit, regularni izrazi predstavljaju format upita. Međutim, i u slučaju kada je upit formulisan u vidu regularnog izraza, rezultati imaju oblik konkordanci, što ukazuje na to da se ova dva nivoa upotrebe ne mogu strogo odvojiti. Najveća prednost upita zasnovanih na regularnim izrazima sastoji se u tome što omogućavaju fleksibilno pretraživanje korpusa, što znači da se njihovim korišćenjem putem jedinstvenog upita može pronaći više reči, ili više oblika iste reči. Drugim rečima, jedan upit formulisan u vidu regularnog izraza odgovara većem broju pojedinačnih upita koji ne koriste regularne izraze.

Regularni izrazi zasnovani su na posebnim oznakama za različite operacije. Česta je, na primer, upotreba asteriska (*) kao zamene za ma koju reč ili slovo, kao i upotreba vertikalne crte (|) kao simbola za logičku operaciju „ili“. Regularni izrazi mogu funkcionisati na nivou pojedinačnih reči, kao i na nivou spojeva reči. Tačnije, izrazi na nivou pojedinačnih reči omogućavaju, na primer, dobijanje svih padeških oblika neke imenice, kao i dobijanje čitave porodice reči izvedenih iz istog korena, dok izrazi na nivou spojeva reči omogućavaju automatsko pronalaženje dve ili više reči koje nisu jedna do druge, ili se ne javljaju uvek istim redosledom. Važno je istaći da različiti elektronski korpusi koriste različite regularne izraze, zbog čega se pre upotrebe određenog korpusa mora upoznati sa pravilima i simbolima koje upotrebljava.

3.3 Statističke mere

Treći nivo upotrebe korpusa predstavlja automatsko izdvajanje kolokacija na osnovu različitih statističkih mera, od kojih je najčešća tzv. „uzajamna informacija“ (eng. *mutual information*), čija vrednost se računa putem formule koja poredi verovatnoću slučajnog zajedničkog pojavljivanja dve reči sa verovatnoćom da se te reči pojavljuju zajedno zbog toga što zaista predstavljaju fiksni izraz (v. McEnery & Wilson 1996, Hunston 2002). Rezultat ovakve statističke obrade prikazuje se u vidu tabele ili liste u kojoj se kolokati određene reči sortiraju prema statistički određenoj snazi veze sa rečju unetom kao upit.

Ovaj nivo pretrage nije omogućen u svim korpusima, što može biti i delimični uzrok njegovog nenavođenja u okviru pedagoških razmatranja. Međutim, ukoliko ga korpus dozvoljava, on učenicima može biti vrlo dragocen, budući da je pokazano da DDL pristup ima najveći učinak u uočavanju i učenju kolokacija, odnosno u graničnom području između leksike i sintakse (Johns 2002: 109). Statističkim kolokatima donekle se može zameriti tabelarni prikaz lišen konteksta, ali njihova značajna prednost sastoji se u brzini dostupnosti, jer automatski prikazuju ono što bi učenici inače morali sami da traže. Pored toga, automatski dobijene podatke vrlo je lako kontekstualizovati u konkordancama, koje se dobijaju uporedo sa tabelarno uređenim kolokatima.

4. Primeri upotrebe korpusa u nastavi jezika

Naredni deo rada posvećen je konkretnim primerima za različite nivoe upotrebe korpusa. Primeri koje navodimo treba da posluže kao vrlo uopštena ilustracija mogućnosti koje elektronski korpusi pružaju, i mada se neki mogu koristiti već u radu sa učenicima na početnom nivou znanja, dok su neki primereniji višim nivoima, nisu eksplicitno označeni kao predlozi vežbanja za određeni nivo znanja stranog jezika. Prvo se govori o korpusu srpskog jezika, zatim o korpusu italijanskog jezika.

4.1 Korpus savremenog srpskog jezika

Korpus iz koga su uzeti primeri za nastavu srpskog jezika kao stranog jeste *Korpus savremenog srpskog jezika*, sastavljen na Matematičkom fakultetu u Beogradu. Radi se o elektronskoj zbirci pisanih tekstova koja sadrži četiri potkorpusa: *Neetiketirani korpus*

srpskog jezika, *Izborna kriza 2000. (neetiketirani tekstovi Politike)*, *Henning-ov etiketirani korpus srpskog jezika* i *Neetiketirani korpus Vukovih poslovice*. Potkorpusi se moraju pretraživati pojedinačno, a najobimniji među njima jeste *Neetiketirani korpus srpskog jezika*, koji sadrži 22.226.437 reči, od čega 20.870.654 reči pripada narativnim književnim tekstovima, dok je preostalih 1.355.783 reči iz štampe. Naši primeri su preuzeti iz ovog potkorpusa.

Korpus savremenog srpskog jezika može se, nakon neophodne registracije, pretraživati putem Interneta, preko odgovarajućeg korisničkog interfejsa, prikazanog na slici 1. Željeni broj rezultata može se slobodno podešavati, kao i željeni broj karaktera u levom i desnom kontekstu tražene reči. Korpus ima veoma razvijen sistem regularnih izraza, koji omogućavaju dobijanje različitih morfoloških oblika iste reči, morfoloških porodica reči i morfoloških dubleta, kao i zajedničkih pojavljivanja reči bez obzira na njihovu međusobnu udaljenost. Nasuprot tome, ne pruža mogućnost automatskog dobijanja kolokacija na osnovu statističkih mera.

Korpus savremenog srpskog jezika

Regularni izraz koji se traži
korpus koji se pretražuje

sortiranje
maksimalan broj rezultata
(0 je neograničeno)
kontekst rezultata
(levi i desni; 0 podrazumevana vrednost - default)
grupisanje rezultata

[konfiguracija](#) | [administracija](#) | [pregled statistike](#) | [izlogujte se](#)
(potrebno je da brauzer podržava "kukije")

Primeri pretraga:

- istina
- [mM]atemat[i-a-z]*
- [rR]acyunar[a-z]*

Naša slova

ć	Č	č	Ć	đ	Đ	š	Š	ž	Ž
cx	Cx	cy	Cy	dx	Dx	sx	Sx	zx	Zx

Slika 1. Korisnički interfejs za pristup *Korpusu savremenog srpskog jezika*.

U daljem tekstu navešćemo tri primera vežbanja, od kojih prva dva ilustruju kako se vrlo jednostavni upiti mogu upotrebiti u različite svrhe, dok treći ukazuje na prednosti

upotrebe regularnih izraza. Za svaki primer navodimo kraći uvod, sam upit (onako kako se unosi u polje za pretragu), deo dobijenih rezultata i predlog zadataka za učenike.

Primer 1 - konkordance: višeznačnost (različite vrste reči)

Učenje višeznačnih reči u stranom jeziku smatra se jednim od težih zadataka, naročito ukoliko se radi o oblicima koji pripadaju različitim vrstama reči, na primer imenicama i glagolima (up. *predstave*, imenica u genitivu jednine ili nominativu/akuzativu množine, ili glagol u trećem licu množine prezenta). Unošenjem odgovarajućeg upita, učenici sami mogu doći do velikog broja primera u kojima se određena višeznačna reč javlja u kontekstu (format konkordance), što omogućava prepoznavanje vrste i tačnog oblika reči u svakoj rečenici, a pruža i uvid u to koja je vrsta reči učestalija za dati oblik.

Upit: predstave

Dobijeni rezultati:

473353: anas se sećam jedne <predstave> Gunoovog " Fausta "
470310: na scenu . Posle ove <predstave> i dobre kritike , s
488577: m , mogli izbeći da <predstave> i istupanja Milana
468855: Posetio sam ga posle <predstave> i on me je zadržao
593588: tojanje da se praoci <predstave> i u svojoj neposred
422220: koreografije iz neke <predstave> ili televizijskog s
243790: išta . I pozorišne <predstave> imale su po pravilu
476413: u , pevala sam razne <predstave> , između ostalih t
470482: je još pre početka <predstave> jasno da ćemo u tr
466096: tika " , povodom ove <predstave> je napisao sledeću
581121: ka Srbije za režiju <predstave> " Kabare " 1994 ; "
469661: . Pred sam početak <predstave> , kada sam ga poset
371257: xeleli su time da se <predstave> kao prognanici komu
472709: ružili i solisti za <predstave> " Knez Igor " i " D
422483: pe iz brodvejske hit <predstave> " Kosa " . Istina j

Zadaci: Koji oblici su glagolski, a koji imenički? Na osnovu čega se to može zaključiti?

Primer 2 - konkordance: višeznačnost (ista vrsta reči), kolokacije

Učenje kolokacija u stranom jeziku takođe je poznato kao vrlo problematično. Iako *Korpus savremenog srpskog jezika* ne pruža mogućnost automatskog dobijanja najčešćih kolokata, putem jednostavnih upita i konkordanci koje se dobiju kao odgovor, učenici mogu sami da utvrde da li se, i koliko često, određene reči javljaju zajedno.

Upit: korpus

Dobijeni rezultati:

5879159: ujedno predstavljaju <korpus> dokumentovane građ
5912615: e države uključi u <korpus> Evropske unije . Vi
4924183: sku (Prvi krajiški <korpus> iz Banjaluke) u ra
4532767: i osnivaju Zaštitni <korpus> koji je dostigao 12
4893380: tet , i koji ulaze u <korpus> nacionalnog obrazov
7131974: čeno 429 učenika . <Korpus> ovog rada čini 320
2315746: a stekao je i srpski <korpus> pod zapovedništvom
5904753: stić) . Najzad , u <korpus> polnih stereotipa s
4165377: ao je da je najveći <korpus> preislamske poezije
132433: lno . Jedan nemački <korpus> razbijen i uhvaćen
4943867: lne patologije . Sav <korpus> različitih znanja
3669457: lavonsko - baranjski <korpus> Srpske Vojske Kraji
4713595: postoji elektronski <korpus> srpskog jezika od j
4698680: sutan prepoznatljivi <korpus> stereotipa o Srbiji
2756235: povelja , kao i ceo <korpus> zakona koji treba d

Zadaci: Pronađite najčešće kolokate uz reč *korpus*. Da li se ova reč češće javlja u značenju „skup tekstova“ ili „vojna jedinica“?

Primer 3 - regularni izrazi: morfološke porodice reči

Zbog velikog broja prefiksa koji nemaju uvek isto značenje, učenje morfoloških porodica glagola u srpskom može biti teško. Upotreba regularnih izraza omogućava brzo dolaženje do rečenica koje sadrže glagole izvedene od iste osnove, različitim prefiksima.

Upit: [a-z]+pisa(o|la|lo)

Dobijeni rezultati:

201707: godine svom biskupu <napisao> sledeće : Ovaj dat
175190: tu i tamo " kradom " <napisao> više poetskih dela
141038: om prizoru , koji je <opisao> dr Đorđe Nešić
200127: (62 - 113) koji je <opisao> i smrt svog strica
200208: grad Pompeju koji je <opisao> Plinije Stariji . I
224253: , koju je vrlo dobro <opisao> profesor Grubač .
239750: avio umetnik koji se <potpisao> kao : Fidiija , sin
224737: , a kamoli da sam ga <potpisao> . Žalio sam se Ust
203380: ojanović je u Odesi <prepisao> iz Cetinjskog zborn
199752: ne obale Krita Evans <pripisao> varvarskim osvajač
103314: vniku . Kada bi se " <raspisao> " , čemu , inače
98602: sa stola naočari i <upisao> mu - odličnu ocjen
154466: ije crnim slovima je <zapisala> 2 . juni 1969 . god
233337: , Ljubinka Kojić je <zapisala> : " Bili su to zais
201403: SAD gospođa Prentis <zapisala> je : Miris sumpora

Zadaci: Koji prefiksi se mogu kombinovati sa glagolom *pisati*? Koji prefiks se najčešće javlja? Kako prefiksi utiču na značenje?

Ovim primerima nisu naravno obuhvaćene sve mogućnosti koje nudi *Korpus savremenog srpskog jezika*. Međutim, već iz tri jednostavna vežbanja jasno je da korpusi mogu učenicima pružiti jedan novi način saznavanja, koji ne samo da je zasnovan na autentičnim podacima i induktivnom metodu učenja, već je i praktičniji od uobičajenog traženja primera u udžbeniku, rečnicima i gramatikama.

4.2 CORIS - *Corpus di riferimento dell'italiano scritto*

Korpus iz koga su preuzeti primeri vežbanja za nastavu italijanskog jezika sastavljen je na Univerzitetu u Bolonji i nosi naziv CORIS (*CO*rpus di Riferimento dell'*I*taliano Scritto 'Referentni korpus pisanog italijanskog jezika'). Pored osnovne, ovaj korpus ima i verziju koju je moguće pretraživati po žanrovima tekstova, CODIS (*CO*rpus Dinamico dell'*I*taliano Scritto 'Dinamični korpus pisanog italijanskog jezika'). Kao što se vidi već iz samog naziva, i u ovom slučaju je u pitanju zbirka pisanih tekstova. CORIS sadrži približno 100 miliona reči, a tekstovi koji ga čine pripadaju sledećim žanrovima: štampa - 38 miliona reči (38%), književna proza - 25 miliona reči (25%), naučna proza - 12 miliona reči (12%), administrativna proza - 10 miliona reči (10%), razno - 10 miliona reči (10%), efemera (pisma, prospekti, uputstva) - 5 miliona reči (5%).

Obe verzije korpusa dostupne su na Internetu i pristupa im se putem namenskog korisničkog interfejsa (v. slike 2 i 3). CODIS verzija omogućava pretragu ne samo po žanrovima, već i u četiri manja potkorpusa unutar određenog žanra. Kao i kod srpskog korpusa, za pretragu CORIS/CODIS-a neophodna je prethodna registracija. Treba istaći i da ovaj korpus nameće jedno specifično ograničenje u pretraživanju: iako se prilikom svake pretrage dobija podatak o ukupnom broju rezultata za tu pretragu, broj rezultata koji se zaista mogu videti ograničen je na 300. Što se tiče ostalih opcija, moguće je izabrati prema kojoj reči u kontekstu koji okružuje traženu reč će rezultati biti složeni, a može se podesiti i širina konteksta. Korpus dozvoljava upotrebu manjeg broja regularnih izraza, ali oni omogućavaju samo povezivanje više odvojenih reči, a ne i automatsko dobijanje morfoloških paradigmi pojedinačnih reči, ili morfološki sličnih reči. Međutim, jedna od značajnijih pozitivnih osobina CORIS-a jeste to što dozvoljava automatsko dobijanje kolokacija, poređanih prema jednoj od četiri ponuđene statističke mere.

CORIS - Corpus query form

User Authentication Username <input type="text"/> Password <input type="password"/>		Query <input type="text"/> Query Language Help <input checked="" type="checkbox"/> Case insensitive search	
Concordance Options Reduce to max <input type="radio"/> 30 <input type="radio"/> 100 lines. <input checked="" type="radio"/> 1 every n-th <input type="radio"/> Random <input type="radio"/> 300		<input type="text" value="Unsorted"/>	Context of <input checked="" type="radio"/> 80 <input type="radio"/> 120 characters. <input type="radio"/> 160
Collocations Get Collocates? <input checked="" type="radio"/> NO! <input type="radio"/> Yes		Sort using <input checked="" type="radio"/> Log-Likelihood Ratio. <input type="radio"/> Mutual Information. <input type="radio"/> T-score. <input type="radio"/> Raw frequency.	
<input type="button" value="Esegui"/> <input type="button" value="Cancella"/>			

Slika 2. Korisnički interfejs za pristup CORIS-u.

CODIS - Corpus query form

User Authentication Username <input type="text"/> Password <input type="password"/>		Query <input type="text"/> Query Language Help <input checked="" type="checkbox"/> Case insensitive search																																								
Subcorpora selection <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th rowspan="2">Subcorpus</th> <th colspan="4">Size (in Mw)</th> </tr> <tr> <th>20</th> <th>10</th> <th>5</th> <th>3</th> </tr> </thead> <tbody> <tr> <td>STAMPA</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> <tr> <td>NARRATIVA</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> <tr> <td>PROSA ACCADEMICA</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> <tr> <td>PROSA GIURIDICO-AMM.</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> <tr> <td>MISCELLANEA</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> <tr> <td>EPHEMERA</td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </tbody> </table>				Subcorpus	Size (in Mw)				20	10	5	3	STAMPA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NARRATIVA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	PROSA ACCADEMICA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	PROSA GIURIDICO-AMM.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	MISCELLANEA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	EPHEMERA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Subcorpus	Size (in Mw)																																									
	20	10	5	3																																						
STAMPA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
NARRATIVA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
PROSA ACCADEMICA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
PROSA GIURIDICO-AMM.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
MISCELLANEA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
EPHEMERA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																						
Concordance Options Reduce to max <input type="radio"/> 30 <input type="radio"/> 100 lines. <input checked="" type="radio"/> 1 every n-th <input type="radio"/> Random <input type="radio"/> 300		<input type="text" value="Unsorted"/>	Context of <input checked="" type="radio"/> 80 <input type="radio"/> 120 characters. <input type="radio"/> 160																																							
Collocations Get Collocates? <input checked="" type="radio"/> NO! <input type="radio"/> Yes		Sort using <input checked="" type="radio"/> Log-Likelihood Ratio. <input type="radio"/> Mutual Information. <input type="radio"/> T-score. <input type="radio"/> Raw frequency.																																								

Slika 3. Korisnički interfejs za pristup CODIS-u.

Kao i za srpski, navešćemo tri primera vežbanja koja se mogu upotrebiti u nastavi italijanskog jezika. Prvi primer ilustruje vrlo jednostavnu upotrebu konkordanci, drugi upotrebu regularnih izraza, a treću upotrebu statističkih mera za pronalaženje kolokacija.

Primer 1 - konkordance: vrsta objekta

Za učenike može biti vrlo problematično kada se rekcija nekog glagola u stranom jeziku razlikuje od rekcije odgovarajućeg glagola u njihovom maternjem jeziku.

Pronalaženjem takvih glagola u korpusu i njihovim proučavanjem u kontekstu rečenice vrlo brzo se može utvrditi koju vrstu objekta zahtevaju, kao i da li se uvek radi o istoj vrsti objekta, ili izbor zavisi i od nekih dodatnih faktora. Kao primer navodimo italijanski glagol *perdonare*, koji se, za razliku od njegovog srpskog parnjaka *oprostiti*, koristi sa direktnim objektom, osim u slučaju kada je upotrebljen kao ditranzitivni glagol.

Upit: perdonare

Dobijeni rezultati:

STAMPAQuot: assini chiedono perdono . Li si può perdonare solo dopo una condanna esemplare "

STAMPAQuot: on diciamo sciocchezze ! Non si può perdonare chi uccide così . Troppo facile ,

STAMPAQuot: razia : un obiettivo cui non si può perdonare nulla . So quali sono i miei nemic

STAMPAQuot: ustizia e dubita che la mafia possa perdonare uomini passati all ' altra parte .

STAMPAQuot: perché . A queste condizioni si può perdonare qualsiasi fregola : anche quella c

STAMPAQuot: a vedova Montinaro che non riesce a perdonare i pentiti ? Io penso che quella do

STAMPAQuot: credo per i mercati . Si può dunque perdonare a Cardoso di averla sposata anche

STAMPAQuot: eminato dolore . Ritengo sia giusto perdonare chi sbaglia , ma è veramente ingiu

STAMPAPeri: è inutile . Molte donne pensano che perdonare chi le ha ferite significhi negare

STAMPAPeri: " è liberata dalla sua prigionia può perdonare i nemici . E accorgersi che allora

STAMPAPeri: alle calunnie e alle insinuazioni , perdonare chi ha sbagliato . Sono le indispe

STAMPASupp: ee . Peccati veniali che si possono perdonare . La differenza è che i coreani im

STAMPASupp: n Pastoral , è che oggi nessuno può perdonare a uno scrittore di apprezzare la n

NARRATRoma: ande bestemmia che nessun umano può perdonare ad un cyborg ! Per questo li uccid

NARRATRoma: a stardi rotti in culo ... - Li devi perdonare , signor segretario generale - int

Zadaci: Utvrdite kakvi se objekti javljaju uz glagol *perdonare*. Da li je uvek u pitanju isti tip, i ako nije, od čega zavisi izbor?

Primer 2 - regularni izrazi: upotreba člana

Upotreba članova u italijanskom jedan je od većih problema za učenike kojima je maternji jezik srpski, a njihovo (ne)korišćenje uz kombinacije prisvojnih prideva i

imenica koje označavaju rodbinske odnose problem je koji se javlja u učenju italijanskog jezika uopšte. Naime, uz ove kombinacije se određeni član koristi samo u posebnim okolnostima (npr. ukoliko je imenica u deminutivu, u množini, ili određena nekim atributom). Učenici do pravila mogu doći i samostalno, upotrebom regularnog izraza pomoću koga dobijaju rečenice u kojima su prisvojni pridev i imenica jedno do drugog, ali i one u kojima su odvojeni atributom („2:2“ određuje moguću veličinu razmaka).

Upit: sua 2:2 moglie

Dobijeni rezultati:

STAMPAQuot: temibile dello stesso Cancelliere : la sua terza moglie
Hiltrud , che tutti già
STAMPAQuot: e , fece salire l ' aristocratico e la sua bella moglie
sulla diligenza . Invita
STAMPAQuot: ritratti dell ' arciduca Alberto e di sua moglie
Isabella (1615) . Ma ancor p
STAMPAQuot: . A poche settimane dal divorzio dalla sua ultima
moglie Hillu , il 53 enne Schr
STAMPAQuot: na pensare alla famiglia di Loredano . Sua moglie è
molto malata . Lui è sempre
STAMPAQuot: dibilmente tragico . Agli assassini di sua moglie
vorrebbe dire qualcosa ? Spero
STAMPAQuot: interminabili giornate vuote . Guarda sua moglie
Antonietta che piange silenzio
STAMPAQuot: . Valerio Fioravanti torna a guardare sua moglie .
Squillano i telefoni . Il la
STAMPAQuot: tta " Be real " , la linea che disegna sua moglie . Il
cannibale si è messo a di
STAMPAPERi: e stanno davvero le cose . Stefano ama sua moglie e
dopo un breve periodo di cri
STAMPAPERi: un atteggiamento infantile : sono o no sua moglie ?
Perché allora mi tratta come
NARRATRoma: famoso per la sua cretineria e per la sua
insopportabile moglie : una miliardar
NARRATRoma: anocchiava delle ottime sfoglie che la sua bella moglie
preparava anche per noi
NARRATRoma: che dei sopravvissuti . L ' unione con sua moglie -
quanto di più si potesse app
NARRATRoma: Pà " che si perse lungo il corridoio ; sua moglie
invece era indaffarata in came

Zadatak: U kojim slučajevima se uz imenicu *moglie* (‘žena’) javlja određeni član?

Primer 3 - statističke mere: kolokacije

Jedna od vrlo pozitivnih odlika italijanskog korpusa jeste mogućnost automatskog dobijanja kolokata, sortiranih prema statističkom značaju. Ovo predstavlja vrlo efikasan

način za dobijanje određene vrste informacija. Na primer, u italijanskom se značenje srpskog prideva „kratak“ uz neke reči (kako imenice, tako i glagole) iskazuje pridevom *breve*, a uz neke pridevom *corto*. Umesto da pregledaju nizove rečenica u kojima se ova dva prideva javljaju, učenici mogu konsultovati tabele sa kolokatima i vrlo brzo utvrditi koje pojedinačne reči, i koji tipovi reči, se javljaju uz koji pridev. Pri tom mogu sasvim zanemariti statističke podatke u tabeli, i obratiti pažnju samo na dobijene liste reči.

Upiti: breve, corto, corta (pretraga u potkorpusu književne proze)

Dobijeni rezultati:

Collocate	Mutual Information	f(node,colloc)	f(colloc)
lasso	10.7471	13	346
farla	10.5733	85	2552
pausa	10.5526	90	2741
tragitto	10.4033	27	912
esitazione	9.58976	11	653

Tabela 1. Najznačajniji kolokati uz prideve *breve*.

Collocate	Mutual Information	f(node,colloc)	f(colloc)
tagliò	15.5845	109	440
tagliar	15.1826	12	64
taglia	11.5847	26	1679
fiato	11.5277	45	3023
circuito	11.4244	30	2165

Tabela 2. Najznačajniji kolokati uz prideve *corto* (m.r.).

Collocate	Mutual Information	f(node,colloc)	f(colloc)
gonna	13.7405	25	884
barba	11.4295	13	2281
gamba	10.7781	10	2756
memoria	8.6532	10	12021
troppo	7.87788	17	34977

Tabela 3. Najznačajniji kolokati uz pridev *corta* (ž.r.).

Zadaci: Uz koje imenice se tipično koristi pridev *breve*, a uz koje *corto/corta*? U kojim glagolskim izrazima se pojavljuju ovi pridevi?

Kao i gore za srpski, istaknuto je svega nekoliko mogućnosti upotrebe korpusa u nastavi italijanskog jezika. Međutim, ovi jednostavni primeri mogu da posluže kao model za veliki broj drugih nastavnih aktivnosti.

5. Zaključak

Iz svega navedenog jasno je da elektronski korpusi imaju veliki potencijal, koji se može iskoristiti i u nastavi jezika. Pre svega, upotreba korpusa može omogućiti učenicima da od samog početka sami otkrivaju i uvežbavaju pravila jezika koji uče. U kasnijim fazama učenici takođe mogu koristiti korpus u svrhu samostalnog dolaženja do informacija koje bi teško pronašli u drugim izvorima (rečnicima, gramatikama, udžbenicima), i za koje bi morali da se obrate nastavniku. Dakle, postoje indikacije da bi odgovarajuća upotreba korpusa mogla značajno pomoći učenicima u postepenom dosezanju sve viših nivoa autonomije u učenju.

Pored toga što mogu doprineti razvoju autonomije kod učenika, korpusi su značajni i kao nastavno sredstvo koje ima brojne kvalitativne prednosti u odnosu na druge, donekle slične izvore podataka (npr. Internet pretraživače). Kao zbirke namenski sakupljenih tekstova čija se reprezentativnost pažljivo ispituje i čiji su autori sasvim sigurno izvorni govornici datog jezika (što često nije slučaj sa tekstovima na Internetu), korpusi su izvor jezičkih podataka koji je ujedno autentičan i pouzdan. Zatim, korpusi dozvoljavaju mogućnosti pretrage i prikaza rezultata (konkordance, regularni izrazi, statistički podaci o kolokacijama) koje su daleko ispred onoga što nudi ma koje drugo sredstvo.

Za kraj treba navesti i nekoliko praktičnih problema vezanih za upotrebu korpusa u nastavi. Oba korpusa o kojima se govori u ovom radu, kao i većina drugih, sadrže samo pisane tekstove, dok je za učenike vrlo bitno i susretanje sa govornim jezikom. Pored toga, dostupnost korpusa je često ograničena - potrebna je licenca koja se plaća, ili u najmanju ruku registracija, što može biti problem čak i za studente, a čini pristup korpusima gotovo nemogućim za mlađe učenike. Međutim, smatramo da ovi problemi nikako ne treba da obeshrabre nastavnike. Naime, sudeći po brzini razvoja korpusne lingvistike, može se očekivati brz porast u broju korpusa govornog jezika, a trenutni tokovi ukazuju i na mogućnost povećanja broja javno dostupnih korpusa. Nadamo se da će do ovih promena zaista doći i da će one korpusima najzad otvoriti vrata učionica.

Literatura

- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In: *How to Use Copora in Language Teaching* (J. McH. Sinclair, ed.), Amsterdam/Philadelphia: John Benjamins, 15-36.
- Đorović, D. (2008). Korišćenje konkordanci u nastavi stranog jezika struke. U: *Multidisciplinarnost u nastavi jezika i književnosti. Zbornik radova sa naučnog skupa održanog 2006. godine* (J. Vučo, prir.), Nikšić: Univerzitet Crne Gore, Filozofski fakultet, 251-266.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1990). From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning. *CALL Austria* 10: 14-34.
- Johns, T. (2002). Data-driven Learning: The Perpetual Challenge. In *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000*. (B. Kettemann and G. Marko, eds), Amsterdam: Rodopi, 107-117.
- McEnery, A. M. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In: *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching* (S. Granger, J. Hung and S. Petch-Tyson, eds), Amsterdam/Philadelphia: John Benjamins, 119-141.
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik* 54: 121-134.

Korpusi

- CORIS - Corpus dell'italiano scritto*, Facoltà di Lettere e Filosofia, Università di Bologna, http://corpora.dslo.unibo.it/coris_ita.html. Poslednji pristup 10.12.2009.
- Korpus savremenog srpskog jezika*, Matematički fakultet, Univerzitet u Beogradu, <http://www.korpus.matf.bg.ac.yu/prezentacija/korpus.htm>. Poslednji pristup 10.12.2009.

Maja Miličević and Tanja Samardžić
SUMMARY

Linguistic corpora as means for developing autonomy in second language learning

In recent years, with the expanding use of computers and the Internet in second language teaching, electronic linguistic corpora have become important means at the teachers' disposal. However, linguistic corpora do not only represent a valuable source of authentic language data, but they can also help students become more autonomous in language learning. That is, as the number of freely available online corpora grows, it is becoming increasingly more feasible for students to use them on their own for practising various lexical and grammatical features, as well as for inducing new grammatical rules. In this paper, we specifically illustrate how linguistic corpora could be used to teach Italian and Serbian as second languages. In particular, we give examples of exercises based on different types of corpus searches and aimed at obtaining different types of data, all of which facilitate a more autonomous learning process.

Keywords: learner autonomy, linguistic corpora, concordance, regular expression, collocation