

This work is licensed under a Creative Commons Attribution 4.0 International License.

Ovaj rad dostupan je za upotrebu pod međunarodnom licencom Creative Commons Attribution 4.0.




<https://doi.org/10.31820/f.33.1.12>

*Gordana Hržica, Sara Košutar, Kristina Posavec*

## KONEKTORI I DRUGE DISKURSNE OZNAKE U PISANOME I SPONTANOME GOVORENOM JEZIKU

*dr. sc. Gordana Hržica, Sveučilište u Zagrebu, Edukacijsko-rehabilitacijski fakultet*

*gordana.hrzica@erf.unizg.hr*  [orcid.org/0000-0001-6067-9148](https://orcid.org/0000-0001-6067-9148)

*Sara Košutar, mag. educ. philol. croat. i franc., Sveučilište u Zagrebu, Edukacijsko-  
-rehabilitacijski fakultet, sara.kosutar@erf.unizg.hr*  [orcid.org/0000-0002-8882-9600](https://orcid.org/0000-0002-8882-9600)

*dr. sc. Kristina Posavec, Sveučilište u Zagrebu, Edukacijsko-rehabilitacijski fakultet*

*kristina.posavec@erf.unizg.hr*  [orcid.org/0000-0003-0449-6325](https://orcid.org/0000-0003-0449-6325)

*izvorni znanstveni članak*

UDK 811.163.42'42

rukopis primljen: 24. veljače 2021; prihvaćen za tisak: 12. svibnja 2021.

*Govorenje i pisanje dva su modaliteta jezične proizvodnje različitih obilježja koji pružaju različite izražajne mogućnosti te se zasnivaju na različitim jezičnim zakonitostima. Razlike između modaliteta očituju se u uporabi jezičnih sredstava kojima se uspostavljaju odnosi na nadređeničkoj razini, to jest konektora i ostalih diskursnih oznaka. Dosadašnja su istraživanja pokazala da se uporaba konektora i ostalih diskursnih oznaka razlikuje s obzirom na jezični modalitet. Mnogi oblici upotrebljavaju se i u pisanome i govorenome modalitetu, ali s različitim funkcijama, dok su neki drugi ograničeni na jedan jezični modalitet. S obzirom na nedostatak sličnih istraživanja u hrvatskome, u ovome su se istraživanju željele ispitati razlike u uporabi konektora i ostalih diskursnih oznaka s obzirom na jezični modalitet. Sukladno prijašnjim istraživanjima, očekivale su se razlike u čestoti konektora i diskursnih oznaka s obzirom na jezični modalitet. Istraživanje je provedeno na Hrvatskome nacionalnom korpusu (pisani jezik) te na Hrvatskome korpusu govornog jezika (govoreni jezik). Dobiveni su rezultati većim dijelom sukladni dosadašnjim istraživanjima, to jest pronađena je veća*

*čestota diskursnih oznaka koje nisu konektori u govorenome korpusu te su se pojedini konektori i diskursne oznake pojavili u samo jednome od korpusa. Utvrđeno je i da se konektori pojavljuju češće u govorenome korpusu, a taj je rezultat u suprotnosti s prvotnom pretpostavkom o većoj čestoti tih sredstava u pisanome korpusu. Rezultate provedenoga istraživanja potrebno je tumačiti u svjetlu žanrovskih osobitosti, stupnja formalnosti i interakcije.*

**Ključne riječi:** *konektori; diskursne oznake; pisani jezik; govoreni jezik; korpusna analiza*

## 1. Uvod

Govorenje i pisanje dva su modaliteta jezične proizvodnje različitih obilježja. Pružaju različite izražajne mogućnosti i temelje se na različitim zakonitostima. Modaliteti se razlikuju i u uporabi jezičnih sredstava za uspostavljanje odnosa na nadrećeničnoj razini, to jest konektora i ostalih diskursnih oznaka. Mnoga se jezična sredstva upotrebljavaju i u govorenome i pisanome modalitetu, ali s različitim funkcijama, dok su pojedina sredstva ograničena na jedan jezični modalitet. Razlike u uporabi konektora i diskursnih oznaka s obzirom na jezični modalitet u nekim su jezicima opisane, dok u hrvatskome takva istraživanja nedostaju.

### 1.1. Obilježja pisanoga i govorenoga jezika

Temeljna razlika između pisanoga i govorenoga jezika leži u spontanosti jezične proizvodnje (Yabuuchi 1998). Govoreni se jezik odlikuje nepripremljenošću jer nastaje u neposrednome trenutku (Mello 2014). Kako se odvija neposredno tijekom govorenja, prisutna su brojna oklijevanja, prekidi i ponavljanja (Ghasemi i Jahromi 2014). S druge strane, pisanje ne uključuje fizičku prisutnost čitatelja pa pisatelju ostaje više vremena da pažljivo oblikuje informaciju i uspješnije je prenese (Ochs 1979; Brown i Yule 1983; Redeker 1984; Yabuuchi 1998). S obzirom na neposredni kontekst i visok stupanj interakcije, usmena se komunikacija ostvaruje i različitim izvanjezičnim elementima, primjerice intonacijom, mimikom i gestama (Badurina i Kovačević 2001: 154). Ti su izvanjezični elementi u pisanome jeziku nadomješteni interpunkcijskim znakovima i opsežnijim opisima (Halliday 1989; Bartsch 1997).

Razlike između pisanoga i govorenoga jezika očituju se i na formalnojezičnoj razini. Sintaksa govorenoga jezika u pravilu je necjelovita, prevladavaju

jednostavnije strukture, različite vrste elipsa i nezavisnosložene rečenice koje su često parataksički ustrojene (Brown i Yule 1983; Miller i Weinert 1998). Za pisani je jezik karakteristična cjelovitija sintaksa uz prevlast zavisnosloženih rečenica (Paltridge 2006; Zhang 2013). Važnu ulogu u uspostavljanju odnosa na nadrečeničnoj razini u govorenome jeziku imaju deikse i diskursne oznake (Castellà 2004; Silić 2006), dok se u pisanome jeziku odnosi među rečenicama ostvaruju modalnim riječima i konektorima (Badurina i Kovačević 2001). Govoreni jezik od pisanoga razlikuje i čestota pojavljivanja pojedinih leksema. Uporabom mjere leksičke gustoće (engl. *lexical density*; omjer punoznačnih i nepunoznačnih riječi) utvrđeno je da pisani jezik sadrži veći broj punoznačnih riječi poput imenica i glagola, a govoreni jezik nepunoznačnih riječi kao što su veznici, zamjenice i prijedlozi (Biber 1988; Paltridge 2006; Olujić i Matić 2017). Na razlike u leksičkoj gustoći mogu utjecati i tipološke osobitosti jezika unutar kojega se modaliteti uspoređuju. Morfološki složeniji jezici imat će manji udio punoznačnih riječi (Ure 1971; Stromqvist i sur. 2002).

Katkada nepodudarnosti između pisanoga i govorenoga jezika proizlaze iz utjecaja drugih čimbenika te je teško odrediti jesu li određena obilježja svojstvena isključivo jednomu od modaliteta ili se oni djelomice preklapaju. Biber (1988) tvrdi da konačne razlike između modaliteta ne postoje, a sličnosti su zapravo rezultat obilježja žanrova kao višedimenzionalnih konstrukata. U pojedinim se pisanim žanrovima mogu prepoznati obilježja govorenih žanrova i obrnuto. Primjerice, elektronička pošta sadrži obilježja neformalne govorene komunikacije, dok govoreni tekst može biti unaprijed pripremljen i imati cjelovitu strukturu baš poput pisanoga teksta (Badurina i Kovačević 2001). Slično zaključuje i Crystal (2001) na primjeru elektroničkoga čavrljanja. Premda se u čavrljanju može govoriti o stanovnome planiranju, vrijeme koje pošiljatelj ima na raspolaganju nije jednako vremenu kojim on raspolaže u monološkome pisanju. Zbog toga je sintaksa elektroničkoga čavrljanja jednostavnija, a leksik manje raznolik, čime se ono približava spontanoj neplaniranoj konverzaciji. Isti žanrovi mogu čak pokazivati slična obilježja, neovisno o modalitetu kojemu pripadaju. Primjerice, Beaman (1984) je usporedila sintaktičku složenost (mjereno kao postotak zavisnosloženih rečenica) u pisanome i govorenome pripovijedanju te je pronašla sličan broj zavisnosloženih rečenica.

## 1.2. Diskursne oznake i konektori

I za govoreni i pisani jezik karakteristična je uporaba jezičnih sredstva kojima se uspostavljaju odnosi na nadrećeničnoj razini. Iako je interes istraživača za načine povezivanja na nadrećeničnoj razini velik, još uvijek ne postoji jedinstveni naziv niti su ujednačeni kriteriji o opsegu jezičnih sredstava s ulogom nadrećeničnoga povezivanja (Badurina 2008; Nigoević 2010). Nepodudarnosti u nazivlju odraz su različitih teorijskih pristupa u okviru kojih se takva jezična sredstva različito određuju. Kao najrašireniji se izdvaja naziv *diskursna oznaka* (engl. *discourse marker*), a s njime supostoji naziv *konektor* (engl. *connective*) (Pons Bordería 2006; Badurina 2008; Nigoević 2010). U skladu s time, različito se naziva i razina na kojoj se odnosi uspostavljaju – konektorima se uspostavljaju odnosi u *tekstu*, a diskursnim oznakama u *diskursu*. Uopćeno se može govoriti o trima temeljnim teorijskim pristupima konektorima i/li diskursnim oznakama.

Nakon Hallidayjeve i Hasanine (1976) uvodne studije o koheziji stvoren je okvir za proučavanje konektora u tekstu. Prema tom okviru konektori čine dio konjunktivne kohezije, a obuhvaćaju gramatička sredstva koja na rećeničnoj razini imaju ulogu veznika, primjerice *i*, *ali*, *osim toga*, *onda*, *zato* i dr. Kohezija je semantički koncept koji nastaje kada tumačenje jednoga elementa u tekstu zavisi o tumačenju nekoga drugoga elementa. S druge strane, Schiffrin (1987) smatra da kohezija nije sama po sebi dostatna za potpunu interpretaciju teksta pa u analizu uključuje koherenciju (Tanskanen 2006). Koherencija se odnosi na cjelokupni smisao diskursa. Njezin se pristup temelji na poimanju diskursa kao jezične jedinice i procesa društvene interakcije (Schiffrin 2001: 56). Odnosi u diskursu uspostavljaju se diskursnim oznakama koje Schiffrin (2001: 31) određuje kao „sekvencijski uvjetovane elemente koji omeđuju jedinice govora”. One ne pripadaju jedinstvenoj gramatičkoj kategoriji, već čine heterogenu skupinu jezičnih jedinica kao što su konektori *i*, *ali*, *ili*, uzvik *oh*, priloz *sada* i *onda*, čestica *pa*, glagolski izrazi *znaš* i *mislim* te neverbalni elementi poput intonacije. Nadalje, vodeći se gramatičko-pragmatičkim kriterijima, Fraser (1990, 1999) i konektore i diskursne oznake (izuzev uzvika i izvanjezičnih elemenata) određuje kao pragmatičke oznake. Pragmatičkim se oznakama iskazuje pragmatičko značenje koje se odnosi na govornikovu komunikacijsku namjeru.

U hrvatskome je jeziku konektore prvi proučavao Silić (1984) iz sintaktičko-semantičke perspektive. Konektori djeluju kao „signali kontekstualne uključenosti rećenice”, a obuhvaćaju rećenične veznike (*ali*, *no*), modalne rije-

či (*štoviše, uostalom*), propozicionalne izraze (*s obzirom na to da*), ponavljanja, stilske figure, red riječi itd. (Silić 1984: 109). Nadalje, Velčić (1987) je razmatrala konektore kao jedinice povezivanja strukture teksta iz sintaktičke, semantičke i pragmatičke perspektive. Autorica konektore dijeli na relativne (*što, koji*), vezničke (*i, ili*), priložne (*zato, zapravo*), frazeologizirane (*u prvom redu*) i propozicionalne konektore (*to znači da*). Konektori za Velčić (1987) čine otvoreni skup jedinica, pri čemu granice među kategorijama nisu strogo određene. Ipak, prema Velčić (1987) konektori mogu povezivati dijelove na razini složene rečenice, dok je Silić (1984: 110) takvo poimanje konektora odbacio. Glovacki-Bernardi (2004) se također usredotočila na opis konektora, ali poimajući ih kao sredstva uspostavljanja anaforičkih i kataforičkih odnosa u tekstu. Njezin se opis konektora, kao i Velčićin (1987), temelji prvenstveno na pisanoj građi. O konektorima kao višefunkcionalnim sredstvima iz pragmatičke su perspektive promišljale Badurina i Palašić (2012), dok su Badurina i Matešić (2013) govorile o poštapalicama kao posebnoj vrsti diskursnih oznaka. Prvo veće istraživanje diskursnih oznaka u hrvatskome provela je Nigoević (2010) u okviru teorije prototipa. Taj je teorijski okvir omogućio da se diskursne oznake odrede kao šira kategorija s različitim funkcijama koja obuhvaća i konektore. Nigoević (2010) konektore ubraja u skupinu diskursnih oznaka s tekstualnom funkcijom (*dakle, zato, ali* i sl.) koja se odnosi na organizaciju i strukturu teksta. Diskursne oznake s interakcijskom funkcijom (*mislim, sad, recimo*) upućuju na odnos među sugovornicima te izražavanje njihova stava spram sadržaja iskaza. Takva podjela omogućava da se diskursne oznake postave na kontinuum s obzirom na njihove funkcije i kategorije kojima pripadaju (vidi Pons Bordería 2006). Poput ostalih hrvatskih istraživača Nigoević (2010) svoju analizu diskursnih oznaka temelji na ulomcima iz književnih djela i novinskih članaka.

S obzirom na neujednačenost nazivlja i teorijska određenja, u ovome će se radu rabiti nazivi konektor i diskursna oznaka imajući u vidu da se njihove funkcije mogu razlikovati. U skladu s podjelom iz Nigoević (2010), konektori će se poimati kao podskupina diskursnih oznaka s tekstualnom funkcijom, dok će ostale diskursne oznake imati interakcijsku funkciju. Nadalje, neki autori pojmove *tekst* i *diskurs* jasno razgraničuju, dok drugi ističu njihova preklapanja (pregled: Badurina 2008). Slično Nigoević (2010: 13) u ovome se radu *tekst* i *diskurs* smatraju komplementarnima – tekst je objekt formalne i strukturne analize, a diskurs objekt interpretacije. Pri tomu i tekst i diskurs podrazumijevaju svaku jezičnost djelatnost koja pretpostavlja smislenu komunikacijsku cjelinu, bez obzira na modalitet.

### 1.3. Uporaba diskursnih oznaka i konektora s obzirom na jezični modalitet

Kao jedno od obilježja diskursnih oznaka ističe se njihova pripadnost govorenomu jeziku (Brinton 1996). Stoga se diskursnim oznakama često pristupalo s razine spontanoga govorenog jezika. Poticaje za takva istraživanja pružili su podatci korpusnih proučavanja spontane konverzacije. Pronađena je visoka čestota diskursnih oznaka u govorenome jeziku, primjerice u francuskome (Chanet 2003), engleskome (Carter i McCarthy 2006; Fung i Carter 2007; O'Keeffe, McCarthy i Carter 2007) i slovenskome (Verdonik i Rojc 2006; Verdonik, Rojc i Stabej 2007). Međutim, Schourup (1999) ističe kako nije opravdano tvrditi da su diskursne oznake obilježje isključivo govorenoga jezika jer ih često nalazimo i u pisanome jeziku. Ako se diskursne oznake mogu pojaviti u oba modaliteta, postavlja se pitanje je li njihova čestota jednaka te imaju li jednake funkcije u govorenju i pisanju (Östman 1982). S obzirom na formalna obilježja, kraći oblici češće se ostvaruju u govorenome jeziku. Za govoreni je jezik karakteristično i supojavljivanje nekoliko različitih diskursnih oznaka u jednome iskazu (usp. Crible i Cuenca 2017). Čestota pojedine diskursne oznake u govorenome ili pisanome jeziku ovisit će i o stupnju formalnosti/neformalnosti. Neke diskursne oznake zahtijevaju visok stupanj planiranja i načelno se upotrebljavaju u pisanome jeziku. Ostale diskursne oznake karakterizira visok stupanj familijarnosti i nedostatak vremena za planiranje iskaza te se češće pojavljuju u spontanome govorenju (Östman 1982; Schourup 1999). Crible i Cuenca (2017) su u govorenim žanrovima s visokim stupnjem planiranja (politički govor i vijesti) pronašle više različitih diskursnih oznaka nego u spontanoj govorenoj konverzaciji s niskim stupnjem planiranja.

Uporaba diskursnih oznaka zavisit će o funkcijama koje one ostvaruju unutar pojedinoga modaliteta (vidi Taboada i Gómez-González 2012). Naime, Crible i Cuenca (2017) utvrdile su da je za pisani jezik karakteristična veća raznolikost diskursnih oznaka (uključujući i konektore) jer se jednim oblikom načelno može izraziti više funkcija. U govorenome je jeziku zastupljeno manje diskursnih oznaka, ali se one ostvaruju u mnogo širem rasponu funkcija. S druge strane, pojedine se diskursne oznake mogu pojaviti unutar istoga žanra i njihove se funkcije ne moraju razlikovati s obzirom na modalitet. To je potvrdila Fox Tree (2015) koja je usporedila čestotu i funkcije diskursnih oznaka *oh*, *pa*, *zapravo* i *stvarno* u

spontanome govorenom i pisanom dijalogu. Spomenute se diskursne oznake pojavljuju u oba žanra (u pisanome dijalogu u nešto manjem broju), no njihove su funkcije s obzirom na modalitet slične. Naposljetku, Biber (2006) je uočio da su neke diskursne oznake (*ok, pa, sad, znaš, mislim* i sl.) ograničene na spontanu govorenu komunikaciju te se rijetko pojavljuju u pisanome jeziku (usp. Redeker 1984). Brinton (1996) ističe kako je uporaba diskursnih oznaka karakterističnih za govoreni jezik stilistički obilježena i negativno interpretirana upotrijebe li se one u pisanim formalnim žanrovima.

Poput diskursnih oznaka i konektori se pojavljuje u govorenome i pisanome jeziku, iako su u pisanome jeziku u pravilu češći (Biber 2006). Pojedini su konektori (*štoviše, naime, stoga* i sl.) svojstveni prvenstveno pisanomu jeziku te se načelno ostvaruju u pisanim žanrovima s visokim stupnjem formalnosti i informativnosti poput akademskoga pisma, službenih dokumenata i poslovnih pisama. S druge strane, neki su konektori češći u govorenome jeziku te upravo u govorenome jeziku ostvaruju više različitih funkcija, primjerice *i* ili *pa* u spontanoj govorenoj konverzaciji ili pripovijedanju (Schiffrin 1987; Crible i Cuenca 2017).

Potrebno je napomenuti da istraživanja razlika između govorenoga i pisanoga jezika, pa tako i ona usmjerena na konektore i druge diskursne oznake, većinom uspoređuju korpuse koji su različiti. Korpusi pisanoga jezika uključuju raznovrsne tekstove niza žanrova. Korpusi govorenoga jezika često se temelje na konverzaciji. Žanrovskim se sličnostima i razlikama između dvaju modaliteta te razlikama u žanrovskoj raznolikosti unutar dvaju modaliteta ne pridaje uvijek pažnja, već se polazi od pretpostavki da su razlike u načinu proizvodnje one koje dovode do razlika u uporabi jezičnih elementa. Naime, pisani se tekstovi u većoj mjeri temelje na planiranju, pokazuju konvencionalna obilježja te su obično formalniji (primjerice, Redeker 1984).

## 2. Ciljevi istraživanja i hipoteze

Proučavanja diskursnih oznaka u hrvatskome jeziku poglavito se temelje na opisima konektora u pisanim tekstovima (Velčić 1987; Glovacki-Bernardi 2004). Prvo opsežnije istraživanje diskursnih oznaka (uključujući i konektore) provela je Nigoević (2010), ali ih nije razmatrala s obzirom na razlike u jezičnome modalitetu. Daljnje analize konektora i ostalih diskursnih oznaka s obzirom na njihovu čestotu u pisanome i govorenome modalitetu nisu provedene. Središnji je cilj ovoga rada ispitati uporabu konektora i ostalih



diskursnih oznaka s obzirom na jezični modalitet u hrvatskome jeziku. U skladu s ciljem mogu se izdvojiti dva istraživačka pitanja. Prvo je istraživačko pitanje utvrditi postoje li razlike u čestoti konektora i ostalih diskursnih oznaka u pisanome i govorenome modalitetu? Drugo je istraživačko pitanje utvrditi jesu li određeni konektori i diskursne oznake ograničeni na jednu vrstu jezičnoga modaliteta? S obzirom na spoznaje dosadašnjih istraživanja (Biber 2006; Fox Tree 2015; Crible i Cuenca 2017), oblikovane su sljedeće hipoteze:

- H1: Čestota konektora bit će veća u pisanome jeziku, dok će čestota diskursnih oznaka biti veća u govorenome jeziku.
- H2: Određene diskursne oznake prvenstveno će biti vezane uz govoreni jezik, dok će se neke druge, one iz skupine konektora, pojavljivati isključivo u pisanome jeziku.

### 3. Metodologija istraživanja

Jezični uzorci spontanoga govorenog i pisanog jezika preuzeti su iz računalnih korpusa, u prvome slučaju korpusa spontane konverzacije odraslih govornika, a u drugome korpusa pisanoga jezika. Korpusi su pretraživani pomoću popisa konektora i ostalih diskursnih oznaka prikupljenih iz različitih opisa hrvatskoga jezika. S obzirom na to da se ti izvori temelje primarno na pisanome jeziku, transkripti izdvojeni iz korpusa govorenoga jezika dodatno su pročitani kako bi se izdvojili jezični oblici koji se ne nalaze na polaznome popisu, a dio su korpusa. Iako se diskursne oznake mogu pojaviti na različitim mjestima u rečenici/iskazu, više je istraživanja pokazalo da se češće pojavljuju na početku rečenice/iskaza (Nigoević 2010). Stoga su iz korpusa izdvojeni samo oni primjeri u kojima se tražena riječ nalazila na početku rečenice/iskaza.

#### 3.1. Materijal

##### 3.1.1. Govoreni jezik

Diskursne oznake (konektori i ostale diskursne oznake) u govorenome jeziku izdvojene su iz transkripata *Hrvatskoga korpusa govornog jezika odraslih* (HrAL; Kuvač Kraljević i Hržica 2016)<sup>1</sup>. HrAL je objavljen u bazi korpusa

---

<sup>1</sup> HrAL je dostupan za pretraživanje na internetskoj poveznici: <https://ca.talkbank.org/access/Croatian.html>.



govorenih jezika TalkBank (<https://talkbank.org>) (MacWhinney 2007). Korpus se sastoji od 165 jezičnih transkripata spontane konverzacije snimljenih u neformalnim situacijama. Ukupno sadrži oko 280 000 pojavnica i 121 000 različenica. Iz HrAL-a je izdvojeno 30 jezičnih uzoraka odraslih govornika, od toga 11 muških i 19 ženskih, u rasponu od 20 do 86 godina ( $M = 37,9$ ;  $SD = 17,98$ ;  $Min = 20$ ,  $Max = 86$ ). Prilikom odabira jezičnih uzoraka sudionici su ujednačeni s obzirom na prosječan broj iskaza koje su proizveli ( $M = 249,93$ ;  $SD = 41,32$ ;  $Min = 201$ ,  $Max = 340$ ). Govornici su iz korpusa odabrani nasumično, a kao jedini kriterij uzet je broj iskaza (minimalno 200). Konačan uzorak bio je raznovrstan s obzirom na geografske i dijalektalne karakteristike. Uzorak sadrži ukupno 7498 iskaza i 36 043 pojavnice. Sociodemografski podatci o sudionicima priloženi su u Tablici 1.

Tablica 1. Sociodemografski podatci o sudionicima

Podrijetlo		Dob						Spol	
Županija	N	20-29	30-39	40-49	50-59	60-69	70 +	F	M
Splitsko-dalmatinska	5	4			1			2	3
Varaždinska	4	2		2					4
Krapinsko-zagorska	3	1		1	1			3	
Brodsko-posavska	2			2				2	
Karlovačka	3			2			1	2	1
Bjelovarsko-bilogorska	1	1							1
Međimurska	3	2		1				2	1
Osječko-baranjska	3	2					1	2	1
Koprivničko-križevačka	1			1				1	
Sisačko-moslavačka	1			1				1	
Istarska	1	1						1	
Zagrebačka	1	1						1	
Grad Zagreb	2			1		1		2	
<b>UKUPNO</b>	<b>30</b>	<b>14</b>		<b>11</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>19</b>	<b>11</b>

### 3.1.2. Pisani jezik

Izvor podataka za pisani jezik bio je *Hrvatski nacionalni korpus* (HNK; Tadić 1996, 2003, 2009)<sup>2</sup>, beta verzija v30. HNK je izabran za pretraživanje zato što je riječ o općem, referentnom korpusu hrvatskoga jezika. HNK čini zbirka tekstova različitih žanrova i stilova, područja i tematike kao što su informativni (novine, revije, časopisi, znanstvena publikacija), imaginativni (romani, pripovijetke, eseji, dnevnici, (auto)biografije) i miješani testovi (Tadić 2003). Mrežna inačica trenutačno broji 170 milijuna pojavnica. Dok je govoreni korpus analiziran na temelju 30 izdvojenih jezičnih uzoraka, pisani je korpus obuhvaćen u cijelosti.

### 3.2. Postupak

Kako bi se istražila čestota konektora i ostalih diskursnih oznaka u govorenome i pisanome korpusu, načinjena su dva popisa jezičnih jedinica. Jedan popis sadrži konektore, a drugi ostale diskursne oznake. Konektori su preuzeti iz istraživanja o zastupljenosti konektora u spontanome govorenom jeziku (Košutar i Hržica 2019). Autorice su popis načinile na temelju sljedećih jezičnih izvora: *Od rečenice do teksta: (teoretsko-metodološke pretpostavke nadrečeničnog jedinstva)* (Silić 1984), *Uvod u lingvistiku teksta* (Velčić 1987), *Hrvatska skladnja: rasprave iz sintakse hrvatskoga standardnog jezika* (Pranjeković 1991), *Druga hrvatska skladnja: Sintaktičke rasprave* (Pranjeković 2001), *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta* (Silić i Pranjeković 2007) i *Hrvatski pravopis* (Badurina, Marković i Mićanović 2008). Kako je dio konektora naknadno utvrđen prilikom toga istraživanja, ti su konektori pribrojani popisu jezičnih jedinica za potrebe ovoga istraživanja. Diskursne oznake koje nisu konektori prikupljene su iz doktorske disertacije *Diskursne oznake u hrvatskom i talijanskom jeziku (kontrastivna analiza)* Nigoević (2010). No i u ovome je istraživanju nekoliko diskursnih oznaka pronađeno prilikom pretraživanja korpusa. Te su riječi također pribrojene konačnomu zbroju traženih jezičnih oblika u obama korpusima. Popis konektora sadrži ukupno 131 oblik (vidi Prilog 1), a popis ostalih diskursnih oznaka 54 oblika (vidi Prilog 2).

---

<sup>2</sup> HNK je dostupan za pretraživanje internetskoj poveznici: [http://filip.ffzg.hr/cgi-bin/run.cgi/first\\_form?corpname=HNK\\_v30;align=](http://filip.ffzg.hr/cgi-bin/run.cgi/first_form?corpname=HNK_v30;align=) .

### 3.2.1. Izdvajanje diskursnih oznaka iz korpusa govorenoga jezika

Jezični uzorci iz HrAL-a (Kuvač Kraljević i Hržica 2016) analizirani su tako da su pomoću programa Computerized Language Analysis (CLAN; MacWhinney 2000), to jest njegove naredbe KWAL, izdvojeni svi iskazi koji sadrže riječi s popisa konektora te popisa ostalih diskursnih oznaka. Kako se mnoge riječi i izrazi s popisa ne moraju nužno ostvarivati u ulozi konektora i ostalih diskursnih oznaka, ručno su izdvojeni svi iskazi u kojima se takve funkcije ostvaruju. Ostale su naredbe CLAN-a omogućile preciznije prebrojavanje jezičnih jedinica. Nakon što su konektori i ostale diskursne oznake izdvojeni iz korpusa govorenoga jezika, dodatno su ručno pročitani svi transkripti kako bi se izdvojili primjeri u kojima pronađene riječi ili nemaju ulogu diskursne oznake, ili ne čine dio popisa.

### 3.2.2. Izdvajanje diskursnih oznaka iz korpusa pisanoga jezika

HNK (Tadić 1996, 2003, 2009) je pretraživan aplikacijom *NoSkechEngine*. Pri pretraživanju se koristilo vrstom upita *Corpus Query Language* (CQL)<sup>3</sup>, kako bi se ograničile pojavnosti traženih oblika isključivo na početak rečenica, dok je pretpostavljeni CQL atribut pretraživanja bila riječ. Rezultati pretraživanja dobiveni su u obliku konkordancijskih nizova te su filtrirani i pročišćeni od pojava koje su se odnosile na inicijale osobnih imena, oznake pravnih članaka i nazive ustanova. Budući da diskursne oznake mogu imati jednake oblike poput drugih vrsta riječi, kao i u korpusu govorenoga jezika, pretraga je ograničena na početak rečenice (Nigoević 2010) te, s obzirom na pravopisnu normu, na primjere u kojima se izravno nakon ciljne riječi pojavljuje zarez.

### 3.2.3. Usporedba čestote konektora i ostalih diskursnih oznaka u korpusima

Budući da su kao građa za analizu čestote konektora i ostalih diskursnih oznaka izabrani korpusi različitih veličina, prilikom analize izračunate su relativne frekvencije svakoga konektora i ostale diskursne oznake s popisa, to jest njihovi udjeli u milijun pojava u svakome od korpusa te su uspoređene prosječne vrijednosti tih udjela. Na kraju su izračunati ukupni brojevi različitih konektora i ostalih diskursnih oznaka u jezičnim uzorcima.

---

3 CQL je upitni programski jezik koji se koristi za pretraživanje korpusa s označenim riječima.

#### 4. Rezultati istraživanja

Cilj je ovoga rada bio ispitati razlike u uporabi konektora i drugih diskursnih oznaka s obzirom na jezični modalitet u hrvatskome jeziku. Analiza je provedena na pisanome korpusu hrvatskoga jezika HNK (Tadić 1996, 2003, 2009) te dijelu govorenoga korpusa HrAL (Kuvač Kraljević i Hržica 2016). Prvo je istraživačko pitanje rada bilo postoje li razlike u čestoti konektora i ostalih diskursnih oznaka u dvama jezičnim modalitetima u hrvatskome jeziku? Kako bi podatci o čestoti bili usporedivi, izračunati su udjeli svakoga konektora i druge diskursne oznake u milijun pojavnica. Podatci o čestoti konektora u pisanome i govorenome korpusu nalaze se u Prilogu 1, a podatci o ostalim diskursnim oznakama u Prilogu 2. Tablice sadrže broj pojavnica svakoga konektora i diskursne oznake s popisa, podatak o njihovu udjelu u milijun pojavnica te prosječnoj vrijednosti u obama korpusima.

Usporedbom prosječnih vrijednosti udjela konektora i ostalih diskursnih oznaka u pisanome i govorenome korpusu utvrđeno je da se konektori češće upotrebljavaju u govorenome korpusu, nego u pisanome korpusu. U govorenome je korpusu prosječan udio konektora 70,61 na milijun pojavnica, a u pisanome 4,54 na milijun pojavnica. Što se tiče ostalih diskursnih oznaka, i one se u govorenome korpusu pojavljuju češće nego u pisanome. Prosječan je udio diskursnih oznaka koje nisu konektori u govorenome korpusu 78,45, a u pisanome tek 0,64. Promotri li se podatci o čestoti konektora i ostalih diskursnih oznaka unutar pojedinih modaliteta, u pisanome korpusu pojavljuje se više konektora (4,54 u milijun pojavnica) i mnogo manje ostalih diskursnih oznaka (0,64 u milijun pojavnica). U pisanome se korpusu od konektora među prvih deset prema čestoti pojavljuju *iako* (194,6 u milijun pojavnica), *osim* (154,6 u milijun pojavnica), *još* (52,49 u milijun pojavnica), *nakon toga* (37,37 u milijun pojavnica), *u svakom slučaju* (20,16 u milijun pojavnica), *doduše* (18,29 u milijun pojavnica), *jedino* (17,15 u milijun pojavnica), *neka* (10,88 u milijun pojavnica), *uglavnom* (9,97 u milijun pojavnica) i *ma* (8,52 u milijun pojavnica). S druge strane, u govorenome su korpusu najčešći *i* (2132,1 u milijun pojavnica), *a* (1842,9 u milijun pojavnica), *pa* (1360,7 u milijun pojavnica), *onda* (864,29 u milijun pojavnica), *ali* (832,14 u milijun pojavnica), *znači* (178,57 u milijun pojavnica), *baš* (171,43 u milijun pojavnica), *nego* (167,86 u milijun pojavnica), *jer* (139,29 u milijun pojavnica) i *zato* (125 u milijun pojavnica). Nadalje, u govorenome korpusu diskursne oznake koje nisu konektori (78,45 u milijun

pojavnica) prednjače nad konektorima (70,61 u milijun pojava), iako je razlika mnogo manja nego u pisanome korpusu. Od diskursnih su oznaka u govorenome korpusu najčešće *da* (1517,9 u milijun pojava), *e* (535,7 u milijun pojava), *dobro* (346,4 u milijun pojava), *no* (325 u milijun pojava), *kužiš* (232,1 u milijun pojava), *znaš* (217,9 u milijun pojava), *mislim* (192,9 u milijun pojava), *ono* (128,6 u milijun pojava), *ovoga* (107,1 u milijun pojava) i *joj* (85,7 u milijun pojava). U pisanome se korpusu od diskursnih oznaka najčešće pojavljuju *eto* (6,3 u milijun pojava), *da* (5 u milijun pojava), *recimo* (4 u milijun pojava), *e* (2,6 u milijun pojava), *sada* (3,9 u milijun pojava), *evo* (3,8 u milijun pojava), *ovako* (3,3 u milijun pojava), *točno* (1,6 u milijun pojava), *čini se* (1,5 u milijun pojava) i *sad* (0,7 u milijun pojava).

Drugo je istraživačko pitanje rada bilo utvrditi jesu li pojedini konektori i ostale diskursne oznake ograničeni na određenu vrstu jezičnoga modaliteta? Podatci o broju natuknica konektora i ostalih diskursnih oznaka u govorenome i pisanome korpusu nalaze se u Prilozima 1 i 2. Rezultati o pojavnostima određenih konektora i ostalih diskursnih oznaka u dvama jezičnim modalitetima pokazuju da se u pisanome korpusu pojavljuje 72 konektora koji nisu pronađeni u govorenome korpusu, primjerice *drugim riječima*, *pored toga*, *iz toga proizlazi da*, *s jedne strane*, *stoga*, *treće* itd. Samo je 5 konektora koji se pojavljuju u govorenome korpusu, ali ne i u pisanome (*bez obzira na to*, *isto*, *pak*, *ustvari*, *znači*). U pisanome se korpusu pojavljuje 14 diskursnih oznaka koje nisu pronađene u govorenome korpusu (*eno*, *hajde*, *hoću reći*, *izgleda*, *kao*, *ništa*, *ok*, *okej*, *razumijem*, *reklo bi se*, *sada*, *što ćeš*, *što ja znam*, *tako reći*). U govorenome se korpusu pak pojavljuje 14 diskursnih oznaka koje izostaju iz pisanoga korpusa (*aj*, *ajd*, *ček*, *čuješ*, *glej*, *kužim*, *kužiš*, *no*, *onak*, *ovoga*, *razumiješ*, *reko*, *šta ćeš*, *viš*). Naposljetku, rezultati o raznolikosti konektora i diskursnih oznaka pokazuju da se korpusi pisanoga i govorenoga jezika razlikuju. Veća se razlika pritom ostvaruje u broju natuknica konektora. Tako je u korpusu pisanoga jezika pronađeno 126 različitih konektora, a u korpusu govorenoga jezika 59. No iako su istraživani korpusi različitih veličina, utvrđeno je da se oni ne razlikuju brojem natuknica diskursnih oznaka. U oba je korpusa pronađeno 47 diskursnih oznaka.

## 5. Rasprava

U ovome istraživanju promatrana je uloga konektora i ostalih diskursnih oznaka u pisanome i govorenome jeziku. Analizom dvaju korpusa

hrvatskoga jezika pokušalo se odgovoriti na dva temeljna istraživačka pitanja. Iako su se mnogi autori bavili razlikama u uporabi ovih jezičnih sredstava s obzirom na modalitet, još je uvijek malen broj korpusnih istraživanja (Crible i Cuenca 2017).

Rezultati ovoga istraživanja pokazali su da razlike u uporabi konektora u pisanome i govorenome jeziku. Udio konektora znatno je veći u govorenome jeziku. Uočene su i razlike u čestoti ostalih diskursnih oznaka u pisanome spram govorenomu jeziku. Te su diskursne oznake u pisanome jeziku zastupljene znatno manje nego u govorenome jeziku. S obzirom na dobivene rezultate, prihvaćena je hipoteza o većem udjelu konektora u pisanome korpusu nasuprot većem udjelu ostalih diskursnih oznaka u govorenome korpusu. Rezultati dobiveni ovim istraživanjem u skladu su s rezultatima ranijih istraživanja koja pokazuju visoku zastupljenost diskursnih oznaka u govorenome jeziku (engleski: Carter i McCarthy 2006; Fung i Carter 2007; O'Keeffe, McCarthy i Carter 2007; francuski: Chanet 2003; slovenski: Verdonik i Rojc 2006, Verdonik, Rojc i Stabej 2007).

Obično se naglašava važna uloga konektora u povezivanju pisanoga teksta te njihova specifičnost upravo za pisani modalitet (primjerice, Biber 2006). Međutim, u ovome je istraživanju znatno veći udio konektora pronađen u govorenome jeziku. Mogući je razlog visoke čestote konektora u govorenome jeziku činjenica da govornik mora na neki način povezati iskaze u govoru kako bi uspostavio koherentnost među njima, a to češće čini konektorima. Na taj način govornici nadomještaju ostala jezična sredstva koja zbog nedostatka vremena za planiranje iskaza i spontanosti situacije u govorenome jeziku ne bi upotrijebili. I ovo istraživanje pokazuje da se govornici češće oslanjaju na kraće oblike (vidi Košutar i Hržica 2019), što je u skladu s temeljnim obilježjima govorenoga jezika koji načelno obiluje kraćim oblicima i necjelovitim strukturama.

Zanimljivo je i da diskursne oznake koje nisu konektori neki autori (primjerice, Biber 1988) smatraju isključivo obilježjem govorenoga jezika. Međutim, u ovome su istraživanju mnoge takve diskursne oznake pronađene u oba modaliteta. Drugim riječima, većina diskursnih oznaka koje nisu konektori i pronađene su u govorenome korpusu pojavljuje se i u pisanome korpusu (vidi Schourup 1999), iako je riječ o korpusu profesionalno pisanih autorskih tekstova. Ukupno se oba korpusa podudaraju u više od 80% takvih diskursnih oznaka, to jest manje od 20% diskursnih oznaka pojavljuje se u samo jednome od dvaju modaliteta. Takva su podudaranja primijećena i u drugim jezicima, primjerice u francuskome (Mosegaard Hansen 2006).

Crible i Zufferey (2015) napominju da se u opisima diskursnih oznaka često polazi iz perspektive jednoga jezičnog modaliteta, premda to nije uvijek opravdano jer se mnoge diskursne oznake upotrebljavaju u oba modaliteta podjednako, ali katkada s različitim funkcijama. Nadalje, u govorenome su jeziku diskursne oznake manje raznolike (Crible i Cuenca 2017). Manja se raznolikost nadoknađuje većim brojem funkcija diskursnih oznaka, što svakako otvara nove mogućnosti istraživanja hrvatskih korpusa, naročito korpusa govorenoga jezika (vidi Košutar i Hržica 2019).

U prijašnjim su istraživanjima pronađeni konektori i ostale diskursne oznake koji se pojavljuju isključivo u govorenome jeziku, dok su neki drugi vezani prvenstveno uz pisani jezik, primjerice u engleskome (Biber 2006). I ovo istraživanje pokazuje da postoje konektori i ostale diskursne oznake koji se pojavljuju u samo jednome od dvaju modaliteta ili je u jednome od modaliteta njihova čestota znatno veća. Ograničena uporaba nekih konektora i ostalih diskursnih oznaka u jednome modalitetu odrazom je različitih funkcija koje se tim jezičnim sredstvima unutar pojedinoga modaliteta ostvaruju. Primjerice, konektori *prvo, nadalje, naposljetku, na tome mjestu, u prvom redu* i sl., koji su u ovome istraživanju pronađeni samo u pisanome korpusu, upućuju na linearnu povezanost sadržaja pisanoga teksta te bi ih govornik rjeđe upotrijebio u spontanoj konverzaciji koja ne pretpostavlja strogu organizaciju sadržaja. Slično funkcioniraju i konektori *na tome mjestu, pored toga, prije toga, iz toga proizlazi, to znači da, s tim ciljem* i sl. koji sadrže anaforičku pokaznu zamjenu, to jest upućuju na sadržaj koji u tekstu prethodi doprinoseći njegovoj povezanosti. S druge strane, ostale diskursne oznake kao što su *ček* ili *viš* upućuju na dobivanje na vremenu pri osmišljavanju iskaza tijekom neplaniranoga spontanoga govorenja. S obzirom na fonološka obilježja tih jezičnih sredstava i činjenicu da tijekom pisanja pisac ima vremena posložiti misli koje želi prenijeti, takve bi se diskursne oznake mogle rjeđe pojaviti u pisanome jeziku. Rezultati ovoga istraživanja dakle upućuju na postojanje kontinuuma na koji se konektori i ostale diskursne oznake mogu postaviti s obzirom na njihovu funkciju u pisanome ili govorenome modalitetu (vidi Nigoević 2010). Na krajnjim se točkama kontinuuma nalaze jezična sredstva koja mogu teže prelaziti iz jedne kategorije u drugu, dok su pojedina sredstva fleksibilnija te se njima mogu izraziti različite funkcije neovisno o obilježjima modaliteta (vidi i Koch i Oesterreicher 1990; Mosegaard Hansen 1998; Mosegaard Hansen 2006). Nadalje, Crible i Cuenca (2017) upozoravaju na govorene tekstove koji su obilježjima slični pisanima, ali i obrnuto, pa se u njima mogu pojavi-



ti ista jezična sredstva s istom funkcijom. Primjerice, konektori poput *pored toga, iz toga proizlazi da, s tim ciljem* itd., koji u ovome istraživanju nisu pronađeni u spontanoj konverzaciji, mogli bi se pojaviti u akademskom diskursu kao što je sveučilišno predavanje, koje je unaprijed pripremljeno i teži visokoj strukturiranosti (usp. Badurina i Matešić 2013).

Rezultati ovoga istraživanja ocrtavaju razlike u uporabi konektora i ostalih diskursnih oznaka u korpusu profesionalnoga pisanog jezika i u konverzacijskome korpusu. Potrebno je napomenuti da su se dosadašnja istraživanja razlika u uporabi konektora i ostalih diskursnih oznaka s obzirom na modalitet provodila usporedbom jezičnih korpusa različitih modaliteta. Pri tome su se kao građa za proučavanje govorenoga jezika uglavnom rabili uzorci spontane konverzacije, dok su za pisani jezik uključeni žanrovski raznolikiji tekstovi. Tomu je tako jer pisani korpusi općenito imaju za cilj biti visoko reprezentativni. Tadić (1998) upozorava da se reprezentativnost postiže uključivanjem u korpus različitih područja uporabe jezika, različitih tipova tekstova, različitih duljina tekstova, različitih žanrova, različitih medija ostvarivanja jezične poruke, različitih osobina autora te različitih vremena nastanka teksta. No zbog metodoloških ograničenja, to jest teže dostupnosti govorene građe i znatno duljega vremena koje je potrebno za njezinu pripremu, korpusi govorenoga jezika redovito su znatno manji opsegom i žanrovskom raznolikošću od korpusa pisanoga jezika. Najčešće su ograničeni na samo jedan govoreni žanr, primjerice na konverzaciju. Slijedom toga, u korpuse su pisanoga jezika uključeni žanrovski raznolikiji tekstovi, dok su u korpusima govorenoga jezika tekstovi manje raznoliki. Uzme li se u obzir razlika u veličini uspoređivanih korpusa i njihova raznolikost (mnogo pisanih žanrova vs. isključivo konverzacija), razlike u uporabi konektora i diskursnih oznaka dobivene u ovome istraživanju moguće je sagledati iz drugačije perspektive, to jest ne nužno kao posljedicu razlika u modalitetima. Dodatno, svakako bi trebalo uzeti u obzir i funkcionalna obilježja diskursnih oznaka jer iste diskursne oznake mogu u različitim modalitetima imati slične ili različite funkcije.

Uključivanje različitih žanrova govorenoga jezika u analizu moglo bi dovesti do drugačijih rezultata u istraživanju razlika između pisanoga i govorenoga jezika. Ni pisani ni govoreni modalitet nisu jedinstveni te se žanrovi unutar njih znatno razlikuju s obzirom na formalnost i stupanj interakcije. Razlike između pisanoga i govorenoga jezika posebno su velike u usporedbi nekih žanrova. Primjerice, za spontanu je konverzaciju karakterističan vrlo nizak stupanj planiranja i visok stupanj interakcije. Za

pretpostaviti je da će pisani žanrovi različitiiji od konverzacije po stupnju planiranja i interakcije (primjerice, znanstveni radovi) pokazivati drugačija obilježja, a pisani žanrovi sličniji konverzaciji (primjerice, internetsko čavrljanje) slična obilježja. Suprotno tomu, znanstveno izlaganje podrazumijeva visok stupanj planiranja, ali niži stupanj interakcije te se po tome udaljava od govorenih žanrova kao što je spontana konverzacija. Dakle, vrsta jezičnoga modaliteta nije jedini odlučujući kriterij pojavljivanja konektora i ostalih diskursnih oznaka, već je potrebno uzeti u obzir i čimbenike kao što su stupanj formalnosti te razlike u registru i žanru. Buduća bi istraživanja uporabe konektora i ostalih diskursnih oznaka u različitim modalitetima uključivanjem većega broja govorenih žanrova ili kontroliranom usporedbom specifičnih žanrova mogla nadopuniti rezultate ovoga istraživanja.

Dobivene je rezultate važno promatrati i iz perspektive metodoloških ograničenja ovoga rada, a ona se odnose na nejednak broj riječi u dvama uspoređivanim korpusima i žanrovske razlike. Naime, s jedne je strane istražen višemilijunski korpus pisanoga jezika koji obuhvaća različite žanrove, dok je korpus govorenoga jezika znatno manjega opsega te obuhvaća isključivo neformalnu spontanu konverzaciju. Zbog tih su se razloga u istraživanju promatrale prvenstveno relativne čestote uporabe konektora i ostalih diskursnih oznaka, a ne apsolutni brojevi ili broj natuknica. I ranija istraživanja redovito u opisu razlika između dvaju modaliteta ne uspoređuju korpuse jednakih veličina (vidi Biber i sur. 1999). Zbog teškoća u prikupljanju i obradi jezičnih uzoraka govoreni su korpusi redovito manji od pisanih. To nameće pitanja koliko su pouzdane tvrdnje u dosadašnjim istraživanjima o većem broju različitih diskursnih oznaka (posebice konektora) u pisanome jeziku (vidi Biber 1988). Usporedba govorenih i pisanih korpusa koji se razlikuju u veličini i opsegu žanrova koje obuhvaćaju (govoreni se često zadržava samo na zapisima konverzacije), poslužila je za opise razlika između pisanoga i govorenoga jezika u brojnim istraživanjima, bilo zbog metodoloških ograničenja (transkripcija), bilo zbog toga što se smatralo da i takav omjer odražava razlike između dvaju modaliteta, to jest da je pisani jezik nužno žanrovski raznolikiji. Uzevši u obzir da su podatci prijašnjih istraživanja također rezultat metodoloških odluka sličnima ovomu istraživanju (iako s manjim razlikama u veličini korpusa), moguće je zaključiti da je ono što je pronađeno u skladu s tim istraživanjima, ali i da podliježe jednakim metodološkim ograničenjima vezanima uz veličinu te zastupljenost žanrova u korpusu govorenoga jezika.

Dobivene je rezultate potrebno promotriti i u svjetlu drugoga metodološkog ograničenja ovoga rada koje proizlazi iz nemogućnosti izravne usporedbe dvaju modaliteta. Naime, važno je napomenuti da je temeljna sintaktička jedinica govorenoga jezika u ovome radu surečenica. Sukladno ranijim istraživanjima, pravila razdvajanja govorenoga niza na rečenice ili iskaze prema izgovornim kriterijima kao što su pauze ili intonacija smatraju se nedovoljno pouzdanima jer su podložne subjektivnoj procjeni transkriptora (primjerice, Stockman 2010). Stoga je teško razlučiti označuju li pauze u govorenim jezičnim uzorcima doista granice među rečenicama ili čine dio slobodnije rečenične strukture (Crystal 1987). Neki istraživači (primjerice, Halliday 1989; Miller i Weinert 1998) taj problem rješavaju tako da se priklanjaju razdvajanju govorenoga niza na surečenice. No takav pristup ima posljedice na izdvajanje konektora iz govorenoga jezika. Dok se u pisanome jeziku, gdje su granice među sintaktičkim jedinicama jasno određene, konektorima nazivaju sredstva koja povezuju tekst na nadrečeničnoj razini, u govorenome jeziku ona povezuju i surečenice. Takav različit način izdvajanja konektora i ostalih diskursnih oznaka u dvama modalitetima može dovesti i do razlike u broju i udjelu tih jezičnih sredstava.

## 6. Zaključak

Uporaba konektora i ostalih diskursnih oznaka u različitim modalitetima istraživana je tema jer odgovara na općenita pitanja o razlikovanju govorenoga i pisanoga jezika. Dok je u nekim jezicima dobro istražena, u drugima, kao što je hrvatski, sličnih je istraživanja manje. U ovome radu opisano je kako se konektori i druge diskursne oznake rabe u dvama korpusima hrvatskoga jezika. Utvrđeno je da je uporaba i konektora i ostalih diskursnih oznaka češća u govorenome jeziku, ali je pronađena i velika podudarnost oblika u dvama modalitetima. Rezultati su primjenjivi u kontekstu budućih istraživanja. Svoju svrhu mogu pronaći i pri izradi opisa hrvatskoga jezika, primjerice gramatika ili rječnika u kojima se može primijetiti tendencija izostavljanja opisa diskursnih oznaka.

**Zahvale:** Ovaj rad sufinancirala je Hrvatska zaklada za znanost projektom Višerazinski pristup govornom diskursu u jezičnom razvoju (UIP-2017-05-6603). Rad doktorandice Sare Košutar financiran/sufinanciran je iz „Projekta razvoja karijera mladih istraživača – izobrazba novih doktora znanosti” Hrvatske zaklade za znanost koji je financirala Europska

unija iz Europskoga socijalnog fonda. Mišljenja, nalazi i zaključci ili preporuke navedene u ovom materijalu isključiva su odgovornost autora i ne odražavaju nužno stajališta Hrvatske zaklade za znanost i Europske komisije.

## Literatura

- Badurina, Lada (2008) *Između redaka: studije o tekstu i diskursu*, Hrvatska sveučilišna naklada, Zagreb.
- Badurina, Lada, Ivan Marković, Krešimir Mićanović (2008) *Hrvatski pravopis*, Matica hrvatska, Zagreb.
- Badurina, Lada, Marina Kovačević (2001) *Raslojavanje jezične stvarnosti*, Filozofski fakultet u Rijeci, Rijeka.
- Badurina, Lada, Mihaela Matešić (2013) „Poštapalice pod povećalom: o jednoj posebnoj vrsti diskursnih oznaka”, *Tuzla. Grad na zrnu soli*, ur. Amira Turbić-Hadgadžić, Tuzla, 9–24.
- Badurina, Lada, Nikolina Palašić (2012) „Pragmatika veznih sredstava”, *Sarajevski filološki susreti I, Zbornik radova (knjiga 1)*, ur. Ismail Palić, Sarajevo, 252–265.
- Bartsch, Carla (1997) „Oral style, written style, and Bible translation”, *Notes on Translation*, 11, 3, 41–48.
- Beaman, Karen (1984) „Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse”, *Coherence in spoken and written discourse*, ur. Deborah Tannen, Norwood, 45–80.
- Biber, Douglas (1988) *Variation across Speech and Writing*, Cambridge University Press, Cambridge.
- Biber, Douglas (2006), *University Language: A Corpus-Based Study of Spoken and Written Registers*, John Benjamins, Philadelphia.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999) *Longman Grammar of Spoken and Written English*, Longman, London.
- Brinton, Laurel J. (1996) *Pragmatic markers in English*, Mouton de Gruyter, Berlin – New York.
- Brown, Gillian, George Yule (1983) *Discourse analysis*, Cambridge University Press, Cambridge.

- Carter, Ronald, Michael McCarthy (2006) *Cambridge grammar of English*, Cambridge University Press, Cambridge.
- Castellà, Josep Maria (2004) *Oralitat i Escriptura. Dues Cares de la Complexitat del Llenguatge*, Publicacions de l'Abadia de Montserrat, Barcelona.
- Chanet, Catherine (2003) „Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie”, *Recherches sur le français parlé*, 18, 83, 1–25.
- Crible, Ludivine, Maria Josep Cuenca (2017) „Discourse Markers in Speech: Characteristics and Challenges for Corpus Annotation”, *Dialogue and Discourse*, 8, 2, 149–166.
- Crible, Ludivine, Sandrine Zufferey (2015) „Using a unified taxonomy to annotate discourse markers in speech and writing”, *Proceedings of the 11th Joint ACL – ISO Workshop on Interoperable Semantic Annotation (ISA – 11)*, ur. Harry Bunt, London, 14–22.
- Crystal, David (1987) *The Cambridge Encyclopaedia of Language*, Cambridge University Press, Cambridge.
- Crystal, David (2001) *Language and the Internet*, Cambridge University Press, Cambridge.
- Fischer, Kerstin 2006. „Towards an understanding of the spectrum of approaches to discourse particles: introduction to the volume”, *Approaches to Discourse Particles*, ur. Kerstin Fischer, Amsterdam, 1–20.
- Fox Tree, Jean E. (2015) „Discourse markers in writing. *Discourse Studies*”, 17, 1, 64–82.
- Fraser, Bruce (1990) „An approach to discourse markers”, *Journal of Pragmatics*, 14, 3, 383–395.
- Fraser, Bruce (1999) „What are discourse markers?” *Journal of Pragmatics*, 31, 7, 931–952.
- Fung, Loretta, Ronald Carter (2007) „Discourse Markers and Spoken English: Native and Learner Use in Pedagogic Settings”, *Applied Linguistics*, 28, 3, 410–439.
- Ghasemi, Hamta, Maryam Khoshbouie Jahromi (2014) „The differences between spoken and written discourses in English”, *International Journal of Language Learning and Applied Linguistics World*, 6, 4, 147–155.
- Glovacki-Bernardi, Zrinjka (2004) *O tekstu*, Školska knjiga, Zagreb.

- Halliday, Michael (1989) *Spoken and written language*, Oxford University Press, Oxford.
- Halliday, Michael, Raquel Hasan (1976) *Cohesion in English*, Longman, London.
- Hansen Mosegaard, Maj-Britt (1998) *The Functions of Discourse Particles. A Study with Special Reference to Spoken French*, Benjamins, Amsterdam.
- Hansen Mosegaard, Maj-Britt (2006) „A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*)”, *Approaches to Discourse Particles*, ur. Kerstin Fischer, Amsterdam, 21–41.
- Koch, Peter, Wulf Oesterreicher (1990) *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Niemeyer, Tübingen.
- Košutar, Sara, Gordana Hržica (2019) „Konektori u spontanome govorenom jeziku”, *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 45, 1, 157–183.
- Kuvač Kraljević, Jelena, Gordana Hržica (2016) „Hrvatski korpus govornog jezika odraslih (HrAL)”, *Fluminensia*, 28, 2, <http://talkbank.org/access/CABank/Croatian.html>, posjet 14. siječnja 2020.
- MacWhinney, Brian (2000) *The CHILDES Project: Tools for analyzing talk. Third Edition*, Lawrence Erlbaum Associates, Mahwah, NJ.
- MacWhinney, Brian (2007) „The TalkBank Project”, *Creating and Digitizing Language Corpora: Synchronic Databases, Volume 1: Synchronic Databases*, ur. Joan C. Beal, Karen C. Corrigan i Hermann Mois, Hampshire, 163–180.
- Mello, Heliana (2014) „Methodological issues for spontaneous speech corpora compilation. The case of C-ORAL-BRASIL”, *Spoken Corpora and Linguistic Studies*, ur. Tommaso Raso, Helen Mello, Amsterdam/Philadelphia, 27–68.
- Miller, James Edward, Regina Weinert (1998) *Spontaneous Spoken Language: Syntax and Discourse*, Clarendon, Oxford.
- Nigoević, Magdalena (2010) *Diskursne oznake u hrvatskom i talijanskom jeziku (kontrastivna analiza)*, neobjavljena doktorska disertacija, Sveučilište u Zadru.
- Ochs, Elinor (1979) „Planned and unplanned discourse”, *Syntax and semantics: Vol. 12. Discourse and syntax*, ur. T. Givón, Academic Press, New York, 51–80.

- O’Keeffe, Anne, Michael McCarthy, Ronald Carter (2007) *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge University Press, Cambridge.
- Olujić, Marina, Ana Matić (2017) „Govorni i pisani jezik odraslih: koliko se razlikuju?”, *Govor* 34, 1, 33–52.
- Östman, Jan-Ola (1982) „The Symbiotic Relationship between Pragmatic Particles and Improptu Speech”, *Impromptu Speech : A Symposium. Papers Contributed to a Symposium on Problems in the Linguistic Study of Impromptu Speech*, ur. Nils Erik Enkvist, Abo, Finska, 147–177.
- Paltridge, Brian (2006) *Discourse analysis: an Introduction*, Continuum Discourse Series, London.
- Pons Bordería, Salvador (2006) „A functional approach to the study of discourse markers”, *Approaches to discourse particles*, ur. Kirsten Fisher, Amsterdam, 189–204.
- Pranjković, Ivo (1991) *Hrvatska skladnja: rasprave iz sintakse hrvatskoga standardnog jezika*, Hrvatska sveučilišna naklada, Zagreb.
- Pranjković, Ivo (2001) *Druga hrvatska skladnja: Sintaktičke rasprave*, Hrvatska sveučilišna naklada, Zagreb.
- Redeker, Gisela (1984) „On differences between spoken and written language”, *Discourse Processes* 7, 1, 43–55.
- Schiffrin, Deborah (2001) „Discourse Markers: Language, Meaning, and Context”, *The Handbook of Discourse Analysis*, ur. Deborah Schiffrin, Heidi E. Hamilton, Oxford, 54–75.
- Schiffrin, Deborah (1987) *Discourse markers*, Cambridge University, Cambridge.
- Schourup, Laurel (1999) „Discourse markers”, *Lingua*, 107, 227–265.
- Silić, Josip (1984) *Od rečenice do teksta : (teoretsko-metodološke pretpostavke nadrečeničnog jedinstva)*, Sveučilišna naklada Liber, Zagreb.
- Silić, Josip (2006) *Funkcionalni stilovi hrvatskoga jezika*, Disput, Zagreb.
- Silić, Josip, Ivo Pranjković (2007) *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*, Školska knjiga, Zagreb.
- Stockman, Ida J. (2010) „Listener reliability in assigning utterance boundaries in children’s spontaneous speech”, *Applied Psycholinguistics*, 31, 363–395.



- Strömqvist, Sven, Victoria Johansson, Sarah Kriz, Hrafnhildur Ragnarsdóttir, Aisenman Ravid, Dorit Ravid (2002) „Toward a crosslinguistic comparison of lexical quanta in speech and writing”, *Written language and Literacy*, 5, 1, 45–67.
- Taboada, Maite, María de los Ángeles Gómez-González (2012) „Discourse markers and coherence relations: Comparison across markers, languages and modalities”, *Linguistics and the Human Sciences*, 6, 1-3, 17–41.
- Tadić, Marko (2009) „New version of the Croatian National Corpus”, *After Half a Century of Slavonic Natural Language Processing*, ur. Dana Hlaváčková, Aleš Horák, Klara Osolobě i Pavel Rychlý, Masaryk University, Brno, 199–205. [http://filip.ffzg.hr/cgi-bin/run.cgi/first\\_form?corpname=HNK\\_v30;align=](http://filip.ffzg.hr/cgi-bin/run.cgi/first_form?corpname=HNK_v30;align=), posjet 15. siječnja 2020.
- Tadić, Marko (2003) *Jezične tehnologije i hrvatski jezik*, Ex Libris, Zagreb, [http://filip.ffzg.hr/cgi-bin/run.cgi/first\\_form?corpname=HNK\\_v30;align=](http://filip.ffzg.hr/cgi-bin/run.cgi/first_form?corpname=HNK_v30;align=), posjet 15. siječnja 2020.
- Tadić, Marko (1996) „Računalna obradba hrvatskoga i nacionalni korpus”, *Suvremena lingvistika* 1-2, 41-42, 603–611. [http://filip.ffzg.hr/cgi-bin/run.cgi/first\\_form?corpname=HNK\\_v30;align=](http://filip.ffzg.hr/cgi-bin/run.cgi/first_form?corpname=HNK_v30;align=), posjet 15. siječnja 2020.
- Tanskanen, Sanna-Kaisa (2006) *Collaborating towards Coherence*, John Benjamins, Amsterdam/Philadelphia.
- Tadić, Marko (1998) „Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika”, *Filologija*, 30–31, 337–347.
- Ure, Jean (1971) „Lexical density and register differentiation”, *Applications of Linguistics*, ur. G. Perren i J. L. M. Trim, Cambridge University Press, Cambridge, 443–452.
- Velčić, Mirna (1987) *Uvod u lingvistiku teksta*, Školska knjiga, Zagreb.
- Verdonik, Darinka, Matej Rojc (2006) „Are you ready for a call? – Spontaneous conversations in tourism for speech-to-speech translation systems”, European Language Resources Association (ELRA), [http://www.lreconf.org/proceedings/lrec2006/pdf/279\\_pdf.pdf](http://www.lreconf.org/proceedings/lrec2006/pdf/279_pdf.pdf), posjet 2. veljače 2019.
- Verdonik, Darinka, Matej Rojc, Marko Stabej (2007) „Annotating discourse markers in spontaneous speech corpora on an example for the Slovenian language”, *Language Resources and Evaluation Journal*, 41, 2, 147–180.

- Yabuuchi, Akio (1998) „Spoken and written discourse: What's the true difference?“, *Semiotica*, 120, 1-2, 1–37.
- Zhang, Bei (2013) „An analysis of spoken language and written language and how they affect English language learning and teaching“, *Journal of Language Teaching and Research*, 4, 4, 834–838.

## SUMMARY

Gordana Hržica, Sara Košutar, Kristina Posavec

### CONNECTIVES AND OTHER DISCOURSE MARKERS IN WRITTEN LANGUAGE AND SPONTANEOUS SPEECH

Speaking and writing are two language production modalities that have different characteristics, expressive possibilities, and language principles. They differ in language devices that establish the relations on the level higher than the level of sentence, e.g., connectives and other discourse markers. Previous studies showed that the usage of connectives and other discourse markers differs regarding the language modality. Many forms are used in both spoken and written modality, though with different functions, while others are limited to only one language modality. Given the lack of similar research in Croatian, the aim of this study was to examine the differences in the usage of connectives and other discourse markers with respect to the language modality. Following previous research, differences in the frequency of connectives and other discourse markers in written and spoken modality were expected. It was also expected that some connectives and discourse markers would be limited to only one language modality. Corpus based research was conducted using a written (*Croatian National Corpus*) and a spoken corpus (*Croatian Adult Spoken Language Corpus*). The obtained results are largely in line with previous studies. A higher frequency of discourse markers was found in the spoken corpus, and some connectives and discourse markers were limited to only one or the other of the corpora. Contradictory to the initial assumption that connectives would be more frequent in the written corpus, the frequency of connectives was higher in the spoken corpus. The results of this research should be interpreted in the light of genre characteristics, the degree of formality and interaction.

**Key words:** *connectives; discourse markers; written language; spoken language; corpus analysis*

## Prilozi

### Prilog 1. Konektori u pisanome i govorenome korpusu

Konektor	Broj pojavnica		Udio u milijun pojavnica		Konektor	Broj pojavnica		Udio u milijun pojavnica	
	HNK	HrAL	HNK	HrAL		HNK	HrAL	HNK	HrAL
a	1160	516	6,82	1842,9	pa	68	381	0,4	1360,7
ali	306	233	1,8	832,14	pak		15		53,57
bar	197	1	1,16	3,57	pored toga	3		0,02	
baš	19	48	0,11	171,43	poslije	54		0,32	
bez obzira na to		3		10,71	poslije toga	5		0,03	
čak	75		0,44		pošto	495	1	2,91	3,57
čim	1148	1	6,75	3,57	potom	64		0,38	
dakle	111	20	0,65	71,43	prema tome	22		0,13	
doduše	3109	1	18,29	3,57	prije svega	16		0,09	
drugim riječima	34		0,2		prije toga	7		0,04	
drugo	41		0,24		primjerice	63		0,37	
drukčije rečeno	1		0,01		prvo	5	2	0,03	7,14
i	653	597	3,84	2132,1	s druge strane	73		0,43	
iako	32373	2	190,4	7,14	s jedne strane	9		0,05	
ili	47	32	0,28	114,29	s obzirom na to	12		0,07	
inače	202	11	1,19	39,29	s tim ciljem	2		0,01	
ionako	277	2	1,63	7,14	s tim da	130	3	0,76	10,71
ipak	147	1	0,86	3,57	samo	137	28	0,81	100
isto		4		14,29	stoga	380		2,24	
isto tako	39		0,23		sve u svemu	9		0,05	
iz toga proizlazi da	3		0,02		svejedno	3		0,02	
jedino	2916	1	17,15	3,57	što se toga tiče	1		0,01	
jednom riječju	3		0,02		štoviše	52		0,31	
jednostavnije rečeno	1		0,01		tada	74	1	0,44	3,57
jer	200	39	1,18	139,29	tako	449	31	2,64	110,71

Konektor	Broj pojava		Udio u milijun pojava		Konektor	Broj pojava		Udio u milijun pojava	
jerbo*	4	1	0,02	3,57	tako da	713	20	4,19	71,43
još	8924	14	52,49	50	također	170		1	
k tome	17		0,1		tamo	34	3	0,2	10,71
kao primjer	6		0,04		tek	57	1	0,34	3,57
kao prvo	5		0,03		time	134		0,79	
konačno	13	1	0,08	3,57	to jest	3		0,02	
ma	1448	84	8,52	300	to znači da	45		0,26	
makar	352	5	2,07	17,86	točnije rečeno	1		0,01	
međutim	367	1	2,16	3,57	treće	8		0,05	
na kraju	73	4	0,43	14,29	tu	205	11	1,21	39,29
na osnovi toga	4		0,02		u isto vrijeme	7		0,04	
na primjer	10	3	0,06	10,71	u međuvremenu	36		0,21	
na taj način	33	1	0,19	3,57	u protivnom (slučaju)	9		0,05	
na temelju toga	3		0,02		u prvom redu	8		0,05	
na tome mjestu	1		0,01		u prvom slučaju	1		0,01	
nadalje	42		0,25		u skladu s tim	2		0,01	
naime	480		2,82		u suprotnom	8		0,05	
najprije	21	1	0,12	3,57	u svakom slučaju	3427	3	20,16	10,71
nakon svega toga	25	1	0,15	3,57	u tom slučaju	11		0,06	
nakon toga	6353	2	37,37	7,14	u tu svrhu	3		0,02	
naposljetku	9		0,05		u vezi s tim	8		0,05	
naprotiv	20		0,12		uglavnom	1695	13	9,97	46,43
nasuprot tome	7		0,04		ukratko	18		0,11	
nego	3	47	0,02	167,86	umjesto da	1259	1	7,41	3,57
neka	1850	13	10,88	46,43	unatoč svemu	4		0,02	
ni	60	14	0,35	50	unatoč tomu	23		0,14	
ni	11	5	0,06	17,86	upravo	116		0,68	
no	904		5,32		uslijed toga	2		0,01	
odatle	2		0,01		usprkos tomu	4		0,02	

Konektor	Broj pojava		Udio u milijun pojava		Konektor	Broj pojava		Udio u milijun pojava	
odavde	4		0,02		usto	4		0,02	
odnosno	5	2	0,03	7,14	ustvari		4		14,29
onako*	2	12	0,01	42,86	uz to	57		0,34	
onamo	1		0,01		uza sve to	3		0,02	
onda	6	242	0,04	864,29	uzgred	3		0,02	
ondje	10		0,06		za razliku od	59		0,35	
opet	8	4	0,05	14,29	zapravo	20	12	0,12	42,86
osim	26274	1	154,6	3,57	zatim	24		0,14	
osim toga	158		0,93		zato	165	35	0,97	125
otad	3		0,02		zbog toga	82		0,48	
ovako	13	4	0,08	14,29	znači		50		178,57
ovdje	34	1	0,2	3,57					
					<b>Ukupno</b>	101188	2590		
					<b>Prosječno</b>			4,54	70,61

Zvezdicom (\*) su označeni primjeri konektora koji su u korpusima pronađeni naknadno tijekom pretraživanja (vidi više u Košutar i Hržica 2019).

*Prilog 2. Diskursne oznake u pisanome i govorenome korpusu*

Diskursna oznaka	Broj pojavnica		Udio u milijun pojavnica		Diskursna oznaka	Broj pojavnica		Udio u milijun pojavnica	
	HNK	HrAL	HNK	HrAL		HNK	HrAL	HNK	HrAL
ah	27	1	0,2	3,6	ništa	28		0,2	
aj		6		21,4	no		91		325
ajd		12		42,9	oh	15	1	0,1	3,6
ajde	9	2	0,1	7,1	ok	9		0,1	
ajme	7	7	>0,1	25	okej	4		>0,1	
ček*		10		35,7	onak*		5		17,9
čekaj	3	18	>0,1	64,3	onako*	16	14	0,1	50
čini se	253	1	1,5	3,6	ovoga*		30		107,1
čuj	5	4	>0,1	14,3	ono	49	36	0,3	128,6
čuješ		2		7,1	ovaj	27	10	0,2	35,7
da	858	425	5	1517,9	ovako	556	4	3,3	14,3
daj	4	15	>0,1	53,6	pazi	6	4	>0,1	14,3
dobro	244	97	1,4	346,4	razumijem	1		>0,1	
e	440	150	2,6	535,7	razumiješ		5		17,9
eh	89	3	0,5	10,7	recimo	684	13	4	46,4
eno	5		>0,1		reklo bi se	25		0,1	
eto	1076	14	6,3	50	reko*		4		14,3
evo	641	18	3,8	64,3	sad	123	13	0,7	46,4
gle	10	14	0,1	50	sada	671		3,9	
gledaj	3	3	>0,1	10,7	slušaj	5	3	>0,1	10,7
glej		3		10,7	super	22	8	0,1	28,6
hajde	19		0,1		šta ćeš*		2		7,1
hoću reći	36		0,2		što ćeš	7		>0,1	
izgleda	8		>0,1		što ja znam	1		>0,1	
je	35	49	0,2	175	tako reći	1		>0,1	
joj	13	24	0,1	85,7	točno	275	3	1,6	10,7
kao	81		0,5		u redu	57	1	0,3	3,6
kužim*		3		10,7	vidiš	7	8	>0,1	28,6
kužiš*		65		232,1	viš		5		17,9
mislim	101	54	0,6	192,9	znaš	18	61	0,1	217,9
ne znam	105	19	0,6	67,9					
					<b>Ukupno</b>	6 679	1 340		
					<b>Prosječno</b>			0,64	78,45

Zvezdicom (\*) su označeni primjeri diskursnih oznaka koje su u korpusima pronađene naknadno tijekom pretraživanja.