

## 9.

### RAZVOJ SPECIJALIZIRANOGA KORPUSA DIREKTIVNIH GOVORNIH ČINOVA

Za potrebe istraživanja predstavljenoga u Cjelini 8 – posvećenoga utvrđivanju načina i sredstava izražavanja direktivnosti te njihovih pragmatičkih obilježja i funkcija – razvili smo specijalizirani korpus direktivnih govornih činova hrvatskoga jezika. Korpus se sastoji od 800 elicitiranih direktivnih govornih činova prikupljenih putem *online*-upitnika, a proveden je među 100 govornika hrvatskoga jezika (o metodi prikupljanja jezične građe i sudionicima istraživanja v. Cjelinu 8). Veličine je 12 676 pojavnica i 1692 različnica. Korpus je zakodiran prema aktualnim Smjernicama TEI (*Text Encoding Initiative*)<sup>117</sup>, a primijenjena je shema oznaka koju smo razvili za potrebe označavanja direktivnih govornih činova.

[TEI] konzorcij razvija i održava smjernice (*Guidelines*) koje su namijenjene svima koji se bave pripremom i/ili obradom tekstualnih resursa u digitalnom obliku. Smjernice su ustvari preporuke konzorcija kako na odgovarajući način prikazati svojstva onih tekstualnih resursa koji se radi lakše računalne obrade moraju izričito označiti. Preporuke dolaze u obliku detaljnoga opisa oznaka (*tag*) koje se unose u digitalni prikaz teksta, čime se obilježava struktura teksta i moguća druga svojstva koja se radi kasnije obrade žele eksplicitno označiti. Na taj se način implicitna tekstualna svojstva, u procesu koji se naziva kodiranje (*encoding*), čine eksplicitnima. Kodiranje prema unaprijed definiranoj shemi omogućava razmjenu podataka bez gubitaka informacija i bez mogućnosti da će neka informacija biti krivo interpretirana. (Bago 2014b)

Smjernice se mogu primijeniti na tekstove svih prirodnih jezika, iz bilo kojega razdoblja te bez ikakvoga ograničenja na oblik i sadržaj (*ibid.*).

Sadržaj smjernica vrlo je širok pa obuhvaća općenita poglavlja namijenjena svima koji se bave kodiranjem teksta, ali i specifična poglavlja za određene vrste tekstova te razne druge sadržaje potrebne korisnicima. (*ibid.*)

Smjernice se temelje na XML-u (*eXtensible Markup Language*), jeziku za označavanje koji definira skup pravila za kodiranje teksta (prema Bray et al. 2000). Riječ je o vrlo jednostavnome i fleksibilnome jeziku koji je čitljiv računalima i ljudima, s naglaskom na strukturi teksta, a ne na njegovoj prezentaciji. Koristi se za razmjenu podataka i podatkovnih struktura među aplikacijama. XML je otvoren standard koji je neovisan o operativnim sustavima, hardveru i softveru, a održava ga W3C (*World Wide Web Consortium*). Datoteke sadržavaju ekstenziju *.xml*, a s obzirom na

<sup>117</sup> <https://tei-c.org/>

činjenicu da su te datoteke ustvari čisti tekst, mogu se otvoriti sa programima koji otvaraju .txt-datoteke te s preglednicima. XML ima definiranu sintaksu, ali ne i semantiku jezika, odnosno definirana su pravila o tome kako treba biti organiziran XML-dokument te pravila imenovanja temeljnih jedinica (elemenata i atributa). Svaki dokument koji se pridržava propisanih pravila dobro je oblikovan XML-dokument. Budući da korisnici sami definiraju nazive elemenata i atributa, ali i strukturu dokumenta, mnogo je jezika za označavanje u različitim područjima (od zdravstva do bankarstva) koji se temelje na XML-u.<sup>118, 119</sup> Jedan od takvih jezika upravo su Smjernice TEI.

### 9.1. Prikaz sheme za označavanje direktivnih govornih činova

Korpus direktivnih govornih činova ručno je označen na razini govornoga čina, a svaki govorni čin može sadržavati sljedeće slojeve informacija: (1) identifikacijsku oznaku ispitanika; (2) informaciju o tipu situacije u kojoj je ostvaren govorni čin; (3) informaciju o tipu iskaza od kojega se sastoji govorni čin; (4) informaciju o zastupljenosti direktivnoga performativnoga glagola u 1. licu u sklopu govornoga čina; (5) informaciju o eksplicitnosti/implicitnosti ilokucijske snage govornoga čina; (6) informaciju o eksplicitnosti/implicitnosti propozicijskoga sadržaja govornoga čina; (7) informaciju o obraćanju na *ti* ili *Vi*; (8) informaciju o zastupljenosti adhorativa u sklopu govornoga čina. Informacije su prikazane kao atributi elementa govornoga čina. U Tablici 55 na narednim stranicama prikazani su slojevi informacija (osim identifikacije oznake ispitanika) s vrijednostima koje mogu poprimiti (za tekst zadataka iz *online*-upitnika v. Tablicu 36).

---

<sup>118</sup> V. više na poveznici [https://en.wikipedia.org/wiki/List\\_of\\_XML\\_markup\\_languages](https://en.wikipedia.org/wiki/List_of_XML_markup_languages).

<sup>119</sup> Uvodni tutorijal o XML-u dostupan je na poveznici <https://www.w3schools.com/xml/default.asp>.

<b>Sloj informacije</b>	<b>Vrijednosti atributa</b>	<b>Opis</b>
<b>Tip informacije</b>	<i>NEFAM1</i>	Govorni činovi ostvareni kao odgovor na prvu situaciju sa sugovornicima u nefamilijarnome odnosu.
	<i>NEFAM2</i>	Govorni činovi ostvareni kao odgovor na drugu situaciju sa sugovornicima u nefamilijarnome odnosu.
	<i>NEFAM3</i>	Govorni činovi ostvareni kao odgovor na treću situaciju sa sugovornicima u nefamilijarnome odnosu.
	<i>NEFAM4</i>	Govorni činovi ostvareni kao odgovor na četvrtu situaciju sa sugovornicima u nefamilijarnome odnosu.
	<i>FAM1</i>	Govorni činovi ostvareni kao odgovor na prvu situaciju sa sugovornicima u familijarnome odnosu.
	<i>FAM2</i>	Govorni činovi ostvareni kao odgovor na drugu situaciju sa sugovornicima u familijarnome odnosu.
	<i>FAM3</i>	Govorni činovi ostvareni kao odgovor na treću situaciju sa sugovornicima u familijarnome odnosu.
	<i>FAM4</i>	Govorni činovi ostvareni kao odgovor na četvrtu situaciju sa sugovornicima u familijarnome odnosu.
<b>Tip iskaza</b>	<i>imperativni</i>	Govorni čin sastoji se od imperativnoga iskaza.
	<i>tvrdnja</i>	Govorni čin sastoji se od iskaza kojim se iznosi tvrdnja.
	<i>upitni</i>	Govorni čin sastoji se od upitnoga iskaza.
	<i>elipticni</i>	Govorni čin sastoji se od iskaza eliptične strukture.
	<i>neverbalni</i>	Govorni čin sastoji se samo od neverbalnoga signala.
	<i>izbjegavanje</i>	Govornik je izbjegao izvršavanje govornoga čina.

<b>Direktivni performativni glagol u 1. licu</b>	<i>dpg1Y</i>	Govorni čin sadržava direktivni performativni glagol u 1. licu. (Atribut se odnosi samo na tvrdnje i upitne iskaze kao tipove iskaza.)
	<i>dpg1N</i>	Govorni čin ne sadržava direktivni performativni glagol u 1. licu. (Atribut se odnosi samo na tvrdnje i upitne iskaze kao tipove iskaza.)
<b>Eksplicitnost/implicitnost ilokucijske snage</b>	<i>isE</i>	Ilokucijska snaga govornoga čina eksplicitno je iskazana. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
	<i>isl</i>	Ilokucijska snaga govornoga čina implicitno je iskazana. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
<b>Eksplicitnost/implicitnost propozicijskoga sadržaja</b>	<i>psE</i>	Propozicijski sadržaj govornoga čina eksplicitno je iskazan. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
	<i>psl</i>	Propozicijski sadržaj govornoga čina implicitno je iskazan. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
<b>Obraćanje na <i>ti</i>/<i>Vi</i></b>	<i>ti</i>	Govorni čin sadržava obraćanje na <i>ti</i> . (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
	<i>vi</i>	Govorni čin sadržava obraćanje na <i>Vi</i> . (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)

	<i>persNeodredivo</i>	Nije moguće odrediti sadržava li govorni čin obraćanje na <i>ti</i> ili <i>Vi</i> . (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
<b>Ahorativ</b>	<i>adhorativY</i>	Govorni čin sadržava adhorativ. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)
	<i>adhorativN</i>	Govorni čin ne sadržava adhorativ. (Atribut se odnosi na tipove iskaza koji uključuju verbalna sredstva [ <i>imperativni, tvrdnja, upitni, eliptični</i> ].)

**Tablica 55.** Shema za označavanje direktivnih govornih činova.

#### Primjer jednoga označenoga govornoga čina

```
<u who="#l055" ana="#formalna1 #tvrdnja #dpg1Y #isE #psE #vi #adhorativN">Ispricavam se, ali krivo ste mi vratili novac, molim vas da jos jednom provjerite i vratite mi ostatak novca. </u>
```

Uz označene govorne činove korpus sadržava i zaglavlje u kojemu se nalaze metapodaci vezani uz jezični resurs: naziv korpusa, autorice korpusa, objašnjenje sheme korištene za označavanje direktivnih govornih činova, podaci o ispitanicima te ostale informacije koje su preporučene prema Smjernicama TEI.

Korpus v1.0 javno je dostupan pod licencijom *Imenovanje-Dijeli pod istim uvjetima 4.0 međunarodna (CC BY-SA 4.0)*<sup>120</sup>, a objavljen je na poveznici: <https://github.com/pbago/DirKorp>.

U planu je daljnji rad na korpusu, koji uključuje evaluaciju primjerenosti razvijene sheme za označavanje direktivnih govornih činova, označavanje korpusa na nižim razinama od govornoga čina te unošnje dodatnih oznaka koje smo prikupili *online*-upitnikom: informacija o glagolskome načinu glagola upotrijebljenoga u sklopu govornoga čina, informaciju o zastupljenosti modalnoga glagola u 2. licu u sklopu govornoga čina te informacije o zastupljenosti leksičkih markera molbe, isprike i zahvale te honorifika u sklopu govornoga čina.

<sup>120</sup> <https://creativecommons.org/licenses/by-sa/4.0/deed.hr>