

II. PRIKUPLJANJE PODATAKA

U OVOM ĆETE DIJELU KNJIGE NAUĀITI:



- razlikovati eksperimentalna istraživanja od opservacijskih
- razlikovati eksperimentalno i kvaziekperimentalno istraživanje
- razlikovati metodu nasumiĀnog i prigodnog uzorkovanja
- procijeniti eksperimentalno istraživanje
- isplanirati eksperimentalno istraživanje
- razlikovati nesudioniĀko od sudioniĀkog promatranja
- procijeniti prednosti i nedostatke eksperimenata i opservacije
- uoĀiti potencijalna etiĀka pitanja u vlastitom naĀinu prikupljanja podataka
- odrediti prednosti i nedostatke prikupljanja podataka pomoću upitnika, intervjua i fokusne skupine
- odabrati naĀin prikupljanja podataka najprikladniji za vlastito istraživanje
- razlikovati pitanja otvorenog i zatvorenog tipa
- razlikovati nestrukturirani, strukturirani i polustrukturirani intervju
- procijeniti pitanja sastavljena u svrhu prikupljanja podataka
- razlikovati gotove elektroniĀke korpuse od zbirke tekstova
- razlikovati vrste korpusa
- procijeniti koja vrsta korpusa ili zbirka tekstova odgovara nekom istraživanju
- prilagoditi operacionalizaciju podacima dostupnim u korpusu
- procijeniti postoji li potreba za triangulacijom pri prikupljanju podataka.

7. Prikupljanje podataka eksperimentima i opservacijom

U OVOM ĆETE POGLAVLJU NAUĀITI:



- razlikovati eksperimentalna istraživanja od opservacijskih
- razlikovati eksperimentalno i kvaziekperimentalno istraživanje
- razlikovati metodu nasumiĀnog i prigodnog uzorkovanja
- procijeniti eksperimentalno istraživanje
- isplanirati eksperimentalno istraživanje
- razlikovati nesudioniĀko od sudioniĀkog promatranja
- procijeniti prednosti i nedostatke eksperimenata i opservacije
- uoĀiti potencijalna etička pitanja u vlastitom naĀinu prikupljanja podataka.

7.1. Eksperimenti

7.1.1. ZnaĀajke eksperimenata

U eksperimentalnim istraživanjima provodi se pokus ili niz pokusa u kojima se izoliraju varijable (v. 6.) koje se Źeli prouĀiti, dok se istodobno nastoji ukloniti ili neutralizirati sve druge aspekte pojave koji bi mogli utjecati na rezultat, a koji nisu u fokusu istraživanja. Za eksperimentalno se istraživanje stoga kaŹe da se provodi u „kontroliranim uvjetima“.

Na primjer, zamislimo da nas zanima utjecaj pristupa izvornom tekstu na uoĀavanje pogreŹaka pri redakturi strojnih prijevoda (SP). Kao Źto smo

vidjeli u poglavlju o varijablama, zavisna varijabla bile bi uočene pogreške, a nezavisna varijabla pristup izvornom tekstu. Nezavisnu bismo varijablu mogli konceptualizirati kao binarnu (jednostavnu) s dvije vrijednosti: pristup izvorniku postoji ili ne postoji. Zavisna bi varijabla mogla biti operacionalizirana tako da se pogreška smatra uočenom ako ju je sudionik pokušao ispraviti, a podaci o pokušajima ispravljanja pogreške mogli bi se prikupiti pomoću opcije „evidentiraj promjene“ u programu MS Word (kao u Antunović i Pavlović 2019).

Kako bismo provjerili odnos između tih dviju varijabli, mogli bismo provesti pokus u obliku zadatka s dvjema grupama sudionika, pri čemu bi jedna grupa redigirala prijevod uz pristup izvorniku, a druga bez pristupa izvorniku. Budući da pristup izvorniku u ovom slučaju smatramo „intervencijom“, prva se skupina naziva eksperimentalnom skupinom, a druga kontrolnom. Kontroliranje uvjeta značilo bi da su dva pokusa po svemu jednaka osim po intervenciji (nezavisnoj varijabli), dakle, trebalo bi se pobrinuti da obje grupe za obavljanje zadatka imaju na raspolaganju jednako vremena, uz iste upute, i da redigiraju isti tekst preveden istim sustavom za SP u istom računalnom programu, dakle, da se njihovi pokusi razlikuju samo kad je riječ o intervenciji – pristupu izvorniku.

Također, sami sudionici u dvjema skupinama ne bi se smjeli previše razlikovati kad je riječ o znanju i vještinama. U idealnim okolnostima (tzv. „istinska“ eksperimentalna istraživanja) sudionici bi pripadali istoj populaciji, a u skupine bi bili raspoređeni metodom nasumičnog (probabilističkog) uzorkovanja (engl. *random/probability sampling*). To bi značilo da svaki pripadnik neke populacije (npr. populacija profesionalnih prevoditelja u Hrvatskoj; populacija studenata prevoditeljstva) ima jednake izgleda za uključanje u jednu od skupina u istraživanju. Svakom bi pripadniku populacije bio dodijeljen broj te bi se pomoću računalnog programa dobio niz nasumičnih brojeva pomoću kojeg bi se odabrali sudionici istraživanja (v. Saldanha i O'Brien 2013: 33). U biomedicinskim se znanostima takva prava eksperimentalna istraživanja nazivaju randomizirana kontrolirana ispitivanja (engl. *randomized controlled trials*) i smatraju se „zlatnim standardom“.

U društvenim i humanističkim znanostima nasumično uzorkovanje često nije izvedivo pa se sudionici odabiru na različite druge načine, kao što je metoda prigodnog uzorka (recimo, u istraživanju sudjeluju studenti koji su već otprije raspoređeni u grupe za vježbe ili prevoditelji koji se dobrovoljno jave za sudjelovanje u istraživanju). Istraživanja s takvim skupinama sudionika,

a u kojima se inače primjenjuju sva druga obilježja eksperimentalne metode, nazivaju se kvaziekperimentalna istraživanja.⁷

Osim što se sudionici (eksperimentalna i kontrolna skupina) mogu uspoređivati međusobno, isti se sudionici mogu ispitati u različitim kontroliranim uvjetima. U gornjem primjeru, ista bi skupina sudionika dobila zadatak da redigira strojni prijevod jednom uz pristup izvorniku, a drugi put bez pristupa izvorniku. Time bi se riješio problem usporedivosti sudionika, ali je jasno da se u tom slučaju ne bi mogao u objema eksperimentalnim situacijama koristiti isti tekst (osim ako bi između dvaju pokusa proteklo dovoljno vremena). Istraživač bi se tada morao pobrinuti za usporedivost tekstova, odnosno, kontrolirati varijablu težine zadatka odabirući istu vrstu, duljinu i težinu teksta, isto područje, približno jednak broj pogrešaka na broj riječi u jednom i drugom strojnom prijevodu itd.

Kao što smo vidjeli, eksperimentalna istraživanja imaju sljedeće značajke (usp. Mildner 2019: 728):

- zavisnu varijablu i bar jednu nezavisnu varijablu
- kontroliranje ili neutraliziranje ostalih, ometajućih varijabli
- eksperimentalnu i kontrolnu skupinu
- odgovarajući odabir sudionika i njihovo raspoređivanje u skupine
- replikabilnost (ponovljivost).

Potonja točka znači da bi istraživanje provedeno na takav način i u takvim uvjetima, s usporedivim sudionicima, mogli ponoviti i drugi (ili isti) istraživači te provjeriti rezultate. U tu je svrhu osobito važno pri izvještavanju o istraživanju detaljno opisati metodologiju.

⁷ Na koncu, moguće je provesti eksperimentalno istraživanje i samo s jednom skupinom sudionika (bez kontrolne skupine), što se obično čini u izviđajnim istraživanjima.

OKVIR C. NEKE EKSPERIMENTALNE METODE U ISTRAŽIVANJIMA KOGNITIVNIH PROCESA

Verbalizacija

Verbalizacija je način prikupljanja podataka pri kojem sudionici usmeno izražavaju svoje misaone procese kako bi istraživači dobili uvid u kognitivno procesiranje koje je uključeno u obavljanje nekog zadatka. Najpoznatija takva metoda, preuzeta iz psihologije (Ericsson i Simon 1984/1993), jest metoda razmišljanja naglas (engl. *think-aloud method*), a od sudionika se traži da verbaliziraju svoje misli istodobno s obavljanjem zadatka. To iziskuje pripremu sudionika u obliku „uvježbavanja“, a glavni je nedostatak upitna ekološka valjanost (v. 7.1.2.) te utjecaj na sam proces koji se istražuje. Druga je metoda naknadna (retrospektivna) verbalizacija (v. okvir D), koja se odvija nakon obavljena zadatka, pri čemu kao pomoć dosjećanju može poslužiti ponovni prikaz zadatka zabilježen jednom od metoda opisanih u nastavku. Ta metoda ne utječe na proces, ali podložna je kako teškoćama s dosjećanjem tako i racionalizaciji (sudionici kažu ono što misle da se događalo, a ne nužno ono što se doista događalo). Obje metode iziskuju puno vremena za obradu podataka, ali podaci su bogati.

Bilježenje aktivnosti tipkovnice

Programima poput Transloga, Inputloga i Scriptloga prikupljaju se podaci o aktivnosti tipkovnice pri zadacima pisanja, prevođenja, redakture i sl. (http://www.writingpro.eu/logging_programs.php). To istraživačima omogućava ponovno pregledavanje cijelog procesa, kao i analiziranje prikupljenih podataka (npr. ispravaka, brzine pisanja, trajanja pauza). Istraživač svakako od sudionika (i poslodavca, ako se istraživanje provodi na radnome mjestu) mora dobiti dopuštenje za instalaciju i upotrebu takvog softvera. Prednost je ove metode što je manje intruzivna od drugih – ne utječe u većoj mjeri na procese koji se istražuju – i što daje brojčane podatke. Nedostatak je taj što se dobiveni podaci mogu tumačiti na različite načine, npr. dulje pauze između aktivnosti tipkovnice smatraju se pokazateljem kognitivnog napora povezanog sa segmentom teksta koji dolazi nakon pauze, ali moguće je da je sudionik izgubio koncentraciju i razmišljao o nečem posve drugom.

Snimanje aktivnosti na zaslonu

Ovom se metodom podaci s računalnog zaslona pohranjuju u obliku videozapisa kako bi se proces mogao naknadno analizirati. Za razliku od programa za bilježenje aktivnosti tipkovnice, ova vrsta softvera (engl. *screen recording software*) omogućava samo pregledavanje procesa, no ne pruža i brojčane podatke, što otežava analizu. Datoteke mogu biti vrlo velike pa istraživač treba osigurati dovoljno prostora za pohranu.

Praćenje pokreta očiju

U novije vrijeme koristi se i metoda praćenja pokreta očiju (engl. *eye tracking*). Ta se metoda temelji na postavci da su pokreti očiju, ukupan broj fiksacija, trajanje fiksacija i širina zjenica indikatori kognitivnih procesa i kognitivnog napora pri obavljanju zadatka. Uređaj pomoću kamere i infracrvenog svjetla bilježi pokrete oka, a softver omogućava naknadno pregledavanje i vizualizaciju procesa te pruža brojne statističke podatke o relevantnim indikatorima. Noviji uređaji nisu intruzivni, što omogućava relativno prirodne uvjete tijekom pokusa, ali zbog cijene obično nisu dostupni pojedincima, nego samo većim istraživačkim timovima. Metoda se, osim u psihologiji, psiholingvistici i znanosti o prevođenju, koristi i u marketingu, ali i kao asistivna (pomoćna) tehnologija za osobe s invaliditetom.

Ispitivanje vremena reakcije

U ispitivanjima vremena reakcije prati se vrijeme koje je sudioniku potrebno da reagira na neki podražaj. Na primjer, na ekranu se pojavi neka slika ili riječ, a sudionik treba pritisnuti tipku čim se ona pojavi ili čim prepozna pripada li navedena slika ili riječ nekoj kategoriji. Ti se eksperimenti temelje na pretpostavci da brzina reakcije odražava mentalno procesiranje, pa se tako sporije reakcije obično povezuju sa složenijim/drugačijim mentalnim procesima, bilo zbog karakteristika podražaja ili sudionika. Istraživanja vremena reakcije omogućuju istraživanje jednog mentalnog procesa (npr. razumijevanje značenja neke riječi, kategorizaciju neke riječi i slično), a glavna je kritika nedostatak ekološke valjanosti (v. 7.1.2.).

Budući da svaka od navedenih metoda ima prednosti i nedostatke, često se koriste u kombinaciji, npr. bilježenje aktivnosti tipkovnice, praćenje pokreta očiju i verbalizacija mogu se koristiti u istom pokusu. Prikupljanje podataka pomoću više metoda naziva se triangulacija podataka (v. 9.5.).

7.1.2. Prednosti i nedostaci eksperimentalnih istraživanja

Osnovna prednost eksperimentalnih istraživanja sastoji se u tome što istraživač ima veći stupanj kontrole nad varijablama, kako nezavisnima tako i ometajućima. Uslijed toga je i dokazivanje kauzalnosti (v. 6.2.) lakše nego u opservacijskim istraživanjima. Eksperimentalna se istraživanja mogu ponoviti (replicirati), čime se dodatno provjeravaju rezultati.

S druge strane, eksperimentalni nacrt istraživanja po definiciji se ne odvija u „prirodnim“ uvjetima. Ako se uvjeti eksperimenta u većoj mjeri razlikuju od uvjeta u „stvarnom životu“, to dovodi u pitanje *ekološku valjanost* istraživanja – primjenjivost rezultata dobivenih istraživanjem na stvarne situacije (Mildner 2019: 31). Ekološka valjanost narušava se samom činjenicom da istraživači nastoje kontrolirati varijable, čega u stvarnom životu nema, ali i činjenicom da sudionici znaju da sudjeluju u pokusu, što utječe na njihovo ponašanje („paradoks promatrača“).⁸ Prijetnja ekološkoj valjanosti može se donekle ublažiti stvaranjem što realističnijih uvjeta tijekom pokusa, kao i stvaranjem uvjeta u kojima sudionici donekle „zaborave“ da sudjeluju u eksperimentu. Važno je naglasiti da sudionici *uvijek moraju biti obaviješteni* o tome da sudjeluju u istraživanju,⁹ iako im se ne moraju unaprijed otkriti baš sve pojedinosti ako bi one mogle utjecati na rezultat. Glavni izazov eksperimentalnih istraživanja tako se može opisati u smislu dvaju međusobno suprotstavljenih zahtjeva: stvoriti što realističnije uvjete, a istodobno kontrolirati varijable.

⁸ Metoda zaslijepljivanja (engl. *blinding*), koja se često koristi u medicini, teško je izvediva ili neizvediva u pokusima koji se tiču jezika i prevođenja. U medicinskim istraživanjima može se koristiti placebo (lažni lijek) koji se daje kontrolnoj skupini kad se želi provjeriti djelovanje pravog lijeka. Na taj način sudionici zapravo ne znaju kojoj skupini pripadaju pa to znanje ne može utjecati na promatrani ishod.

⁹ Danas je uobičajeno dobiti informirani pristanak sudionika u pisanom obliku (v. 17.1.1.). Nikad nije prihvatljivo snimati sudionike bez njihova znanja, kao ni vršiti pritisak na ljude da sudjeluju u istraživanju.

7.2. Opservacija

7.2.1. Značajke opservacije

U opservacijskim (opažajnim) istraživanjima sudionici se promatraju u njihovom prirodnom okruženju, što je ujedno i najveća prednost takvih istraživanja. Istraživač ne uvodi intervenciju (v. 7.1.1.) nego samo promatra što se zbiva sa sudionicima pod utjecajem faktora iz njihove okoline. Za razliku od eksperimentalnih istraživanja, varijable se ne kontroliraju (iako se prate, opisuju i analiziraju), ali je ekološka valjanost (v. 7.1.2.) takvih istraživanja u načelu veća nego u slučaju eksperimenata. Opservacijska istraživanja mogu biti dio kvalitativne tradicije (v. 11.1.), gdje se obično promatra ponašanje pojedinaca ili skupine ljudi u njihovom okruženju, a u obzir se uzima i istraživačeva subjektivnost. Nasuprot tomu, u nekim se opservacijskim istraživanjima sudionicima pristupa iz perspektive koja pokušava objektivizirati proces promatranja i prikupljanje podataka. Kao što ćemo vidjeti u nastavku, opservacijska se istraživanja mogu podijeliti prema uključenosti istraživača (sudioničko i nesudioničko) i vrsti promatranja (strukturirano i nestrukturirano), pri čemu je sudioničko, nestrukturirano promatranje bliže kvalitativnom polu, a strukturirano kvantitativnom („objektivnom“).

7.2.2. Vrste opservacijskih istraživanja

S obzirom na uključenost istraživača, razlikuju se dvije glavne podvrste opservacijskih istraživanja: *nesudioničko* i *sudioničko* promatranje.¹⁰ Prva vrsta „počiva na premisi da je istraživač distanciran od situacije promatranja, iako ne i prikriven, te se suzdržava od interakcije“ (Potkonjak 2014: 71), pri čemu distanca ne mora nužno biti fizička. Pri sudioničkom promatranju istraživač postaje (ili otprije jest) dio skupine koju proučava te sudjeluje u njezinim aktivnostima, obično tijekom duljeg vremenskog razdoblja, istodobno zadržavajući ulogu

¹⁰ Sudioničko promatranje naziva se i promatranjem sa sudjelovanjem (engl. *participant observation*), a tipično je za etnografiju. Nesudioničko se promatranje naziva i naturalističkim promatranjem. Između tih dviju krajnosti mogući su i različiti stupnjevi uključenosti istraživača.

istraživača.¹¹ Ako bismo promatrali organizaciju rada, aktivnosti i uloge različitih aktera u prevoditeljskoj agenciji nastojeći se držati što više po strani i nastojeći da nas sudionici što više „zaborave“, to bi bio primjer nesudioničkog promatranja. Ako bismo pak provodili istraživanje u agenciji u kojoj i sami radimo, to bi bio primjer sudioničkog promatranja. Izuzetno je važno da istraživač vrlo detaljno izvjesti o svojoj ulozi, stupnju sudjelovanja i svim subjektivnim aspektima istraživanja.

Promatranje može biti nestrukturirano i strukturirano. Zamislimo da promatramo rad simultanih prevoditelja na nekoj konferenciji. U nestrukturiranom nas promatranju zanimaju svi aspekti situacije te bilježimo sve što uočimo, uključujući i vlastite osjećaje i ulogu. Bilješke iz nestrukturiranog promatranja mogu biti temelj kasnije kvalitativne analize (v. 11.2.), a mogu poslužiti i kao način strukturiranja kasnijeg promatranja.

Strukturirano promatranje odnosi se na sustavno promatranje i bilježenje nekih odabranih aspekata situacije. Jedna je od mogućnosti da na temelju analize bilježaka iz nestrukturiranog promatranja izdvojimo temu koja će nas posebno zanimati u daljnjem istraživanju, npr. način na koji se prevoditelji služe pomagalima tijekom prevođenja. U idućem, strukturiranom promatranju, usredotočit ćemo se upravo na taj aspekt situacije. Pri strukturiranom promatranju stoga je poželjno imati razrađen okvirni skup pitanja kao pomoć pri bilježenju. Strukturirano nesudioničko promatranje može se usredotočiti i na praćenje učestalosti određene pojave, na primjer praćenje korištenja diskursnih oznaka među učenicima u razredu. Ponekad se u takvim istraživanjima teži većoj objektivizaciji, tako da se točno definira vrsta pojave koja se promatra te da u promatranju sudjeluje više istraživača. U opservacijskim nesudioničkim istraživanjima mogu se prikupljati i podaci koji nisu povezani s ponašanjem i interakcijom sudionika u stvarnom vremenu, nego su „izvanjski“, kao što je uspjeh neke skupine učenika u provjerama znanja, uspjeh studenata na ispitima u studiju, zarada prevoditelja koji rade u agencijama i kao samostalni prevoditelji i sl.

Opservacijska nesudionička istraživanja mogu se, s obzirom na nacrt, podijeliti na presječna istraživanja, kohortna istraživanja, istraživanja parova

¹¹ O tom načinu prikupljanja podataka postoji veći broj izvora iz područja etnologije, između ostaloga i na hrvatskom jeziku (npr. Potkonjak 2014).

(engl. *case-control study*) i studije slučaja.¹² Presječnima se istražuju podaci u nekom konkretnom trenutku i bez vremenske dimenzije. U istraživanju prevođenja mogli bismo, na primjer, istražiti koliko u nekom trenutku ima registriranih prevoditeljskih poduzeća i obrta u Republici Hrvatskoj. Tim se tipom istraživanja ne može utvrditi uzročno-posljedična veza, nego se samo „snima“ situacija (zbog toga se katkad na engl. naziva *snapshot study*). Kohortnim se istraživanjima prati skupina sudionika (kohorta) s obzirom na neki ishod. U vremenskom smislu mogu biti retrospektivna ili prospektivna, pri čemu je druga situacija češća. Primjerice, mogli bismo vremenski pratiti kohortu studenata kako bismo vidjeli postoji li korelacija između njihovog uspjeha na prijemnom ispitu i kasnijeg uspjeha na studiju, ili uspjeha na studiju i kasnijeg uspjeha u karijeri, ili svih triju varijabli. (Pojam uspjeha trebalo bi operacionalizirati! (v. 5.2.)). Istraživanje parova uvijek je retrospektivno, to jest, temelji se na prethodno prikupljenim podacima, a sudionici se svrstavaju u dvije skupine s obzirom na ishod koji nas zanima. Npr. glavna skupina „slučajeva“ mogla bi se sastojati od studenata koji su po završenom studiju postali uspješni poduzetnici. Kontrolna skupina sastojala bi se od usporedivih sudionika koji nisu postali poduzetnici. Istraživanjem bi se pokušalo utvrditi koji se faktori mogu dovesti u korelaciju s jednim ili drugim ishodom (mogući kandidati: uspjeh na studiju, demografski podaci ako su isti institucijski prikupljeni i sl.). Valja napomenuti da istraživanja parova nisu česta u društveno-humanističkom području. Daleko su češće studije slučaja. U studiji slučaja promatra se pojedinac, manja skupina ili organizacija, koji se dubinski analiziraju kvalitativnom (v. 11.2.) metodom.

Valja naglasiti da se i u opservacijskim istraživanjima može pojaviti paradoks promatrača (v. 7.1.2.), osobito ako je istraživač vidljivo prisutan u okruženju u kojem se nalaze sudionici. No, zbog rada u uobičajenim, neeksperimentalnim uvjetima, sudionici, iako promatrani, mogu lakše „zaboraviti“ da sudjeluju u istraživanju. Moguće je da su sudionici pristali da se podaci o njima prikupljaju u istraživačke svrhe prije toliko vremena da informiranost o sudjelovanju u istraživanju više i ne utječe na rezultate istraživanja. U svakom je slučaju nužno ishoditi pristanak sudionika te, prema potrebi, dopuštenje institucije ili poduzeća. S druge strane, moguće je da će se istraživač poslužiti podacima koje

¹² Navedene su vrste istraživanja u mnogočemu tipične za medicinu, osobito epidemiologiju (Song i Chung 2010), ali primjenjive su i na društveno-humanističko područje.

je prikupila neka ustanova (koja se pobrinula za pristanak) ili podacima koji su u javnoj domeni i za koje nije potreban poseban pristanak.

ZADACI: 

1. Koji bi način prikupljanja podataka – eksperimentalni ili opservacijski – bio prikladniji u svjetlu vašeg cilja istraživanja i hipoteza? Bi li bilo moguće prikupiti relevantne podatke i na onaj drugi način i, ako bi, bi li to zahtijevalo preformuliranje cilja i hipoteza?
2. Osmislite i provedite manji, probni eksperiment (može i samo s dvoje sudionika), po mogućnosti povezan s vašom temom istraživanja. Na što sve trebate paziti? Što ste naučili iz tog pokusa?
3. Provedite opservaciju manjeg opsega povezanu sa svojom temom istraživanja. Je li opservacija bila sudionička ili nesudionička? Strukturirana ili nestrukturirana? Što ste naučili iz tog promatranja?

PROČITAJTE VIŠE: 

U vezi s opservacijskim istraživanjima preporučujemo knjigu na hrvatskome jeziku Sanje Potkonjak (2014) te poglavlje o opservaciji u Matthews i Ross (2010: 254-264). O eksperimentalnim istraživanjima prijevodnih procesa na pristupačan način govore Saldanha i O'Brien (2013: 109-148), koje se posebno bave i studijama slučaja (2013: 205-232). Više o studijama slučaja potražite u Meyer (2016: 177-184). Pojediniosti o razmišljanju naglas u kontekstu istraživanja prijevodnih procesa v. u Jakobsen (2003).

8. Prikupljanje podataka pomoću upitnika, intervjua i fokusne skupine

U OVOM ĆETE POGLAVLJU NAUČITI:



- odrediti prednosti i nedostatke prikupljanja podataka pomoću upitnika, intervjua i fokusne skupine
- odabrati način prikupljanja podataka najprikladniji za vlastito istraživanje
- razlikovati pitanja otvorenog i zatvorenog tipa
- razlikovati nestrukturirani, strukturirani i polustrukturirani intervjui
- procijeniti pitanja sastavljena u svrhu prikupljanja podataka.

U ovom će poglavlju biti riječi o trima načinima prikupljanja podataka: upitniku, intervjui i fokusnoj skupini. Zajedničko im je obilježje to da se podaci prikupljaju postavljanjem pitanja odabranom uzorku osoba. Sva se tri načina stoga oslanjaju na samoizvještavanje (engl. *self-reporting*) pa su najprikladniji za prikupljanje podataka o mišljenju i stavovima sudionika, zbog čega su osobito popularni u društvenim znanostima. Te se metode koriste i u istraživanjima tržišta, ispitivanjima javnoga mnijenja i dr.

8.1. Prikupljanje podataka pomoću upitnika

Upitnik je mjerni instrument koji se sastoji od niza pitanja kojima se mogu prikupiti standardizirani, relativno strukturirani podaci od većeg broja sudionika

(Matthews i Ross 2010: 201). Kad je riječ o strukturiranosti u prikupljanju podataka, pojam se odnosi na stupanj u kojem su pitanja ista za sve sudionike ili se pak razlikuju. Upitnici se obično odlikuju visokim stupnjem strukturiranosti, što znači da se većina pitanja iz upitnika postavlja svim sudionicima, i to u posve istom obliku i istim redoslijedom. Ako je riječ o zatvorenim pitanjima, to jest, pitanjima s već ponuđenim odgovorima među kojima sudionici biraju, svima su ponuđeni isti odgovori. Uz zatvorena pitanja, u upitnicima se mogu postaviti i otvorena pitanja, na koja sudionici mogu odgovoriti kako žele (Matthews i Ross 2010: 201–202), a ograničiti se može samo dopušten broj riječi. Veći stupanj strukturiranosti u prikupljanju podataka olakšava analizu, ali s druge strane može predstavljati ograničenje ako, uz ponuđene odgovore, nije ostavljena mogućnost slobodnog dodavanja odgovora. Zbog toga se u upitnicima često kombiniraju zatvoreni i otvoreni tip pitanja tako da je uz ponuđeni niz odgovora dopušteno i upisivanje vlastitih odgovora. Na taj se način prikupljaju polustrukturirani podaci.

Standardiziranost se odnosi na činjenicu da se upitnik koji je sastavio i testirao jedan istraživač ili istraživački tim može koristiti i u drugim sredinama te se rezultati mogu uspoređivati s velikom dozom „objektivnosti“. To osobito vrijedi za instrumente koji se koriste u psihologiji jer se oni podvrgavaju vrlo rigoroznom i dugotrajnom testiranju prije nego što postanu prihvaćeni.¹³

Uz strukturiranost, prednost je upitnika u tome što omogućava prikupljanje podataka od velikog broja sudionika u relativno kratkom vremenu. To je osobito točno u današnje vrijeme kad se upitnici šalju i ispunjavaju putem interneta te mogu doći do velikog broja ljudi. Pritom je važno pobrinuti se da sudionici od kojih prikupljamo podatke doista pripadaju populaciji (v. 10.1.) koja nas zanima. Na primjer, ako nas zanimaju stavovi profesionalnih prevoditelja o sustavima za strojno prevođenje, trebali bismo uzeti u obzir samo pripadnike populacije profesionalnih prevoditelja. Budući da može biti teško kontrolirati tko će sve ispuniti upitnik, pri analizi se mogu isključiti odgovori onih sudionika koji se ne ubrajaju u željenu populaciju.

¹³ Eysenckov upitnik ličnosti primjer je psihološkog standardiziranog upitnika kojim se mjere osnovne dimenzije ličnosti, a na pitanja se odgovara zaokruživanjem odgovora *da* ili *ne*. Upitnik je u izvorniku sastavljen na engleskom, a prijevod na hrvatski standardiziran je na 5700 sudionika (<https://www.nakladaslap.com/pds/pregled/b5404754443e94434a10ab96510433279>).

To se može doznati iz podataka koji se upitnikom prikupljaju o samim sudionicima – tzv. demografskih podataka.¹⁴ Demografski podaci omogućavaju istraživaču da odgovore na ostala pitanja dovede u korelaciju (v. 6.2.) s jednom ili više varijabli koje se tiču samog sudionika. Recimo, moglo bi nas zanimati utječe li dob sudionika, studij koji je sudionik završio ili vrsta radnog odnosa na njegove stavove o strojnom prevođenju.

Ukoliko je uzorak sudionika obuhvaćen anketnim istraživanjem reprezentativan, utoliko je moguće rezultate generalizirati na čitavu populaciju (v. 10.1.), koju obično nije moguće ispitati cijelu. *Reprezentativnost* znači da se u uzorak uključuju pojedinci koji dobro predstavljaju određenu populaciju. Na primjer, osobe koje prevode samo povremeno i prevođenje im nije glavni izvor prihoda ne bi bili dobri predstavnici populacije „profesionalni prevoditelji“, koliko god možda imali zanimljive stavove. Na reprezentativnost utječe način odabira uzorka (v. 7.1.1. i 10.1.), veličina uzorka i njegov sastav.

Nije uvijek jednostavno dobiti odgovarajući uzorak, što može umanjiti vrijednost rezultata. Tako bi u istraživanju o stavovima o strojnom prevođenju u kojem bi se upitnik distribuirao isključivo putem društvenih mreža postojao velik rizik od pristranosti, to jest, postojala bi velika vjerojatnost da sudionici do kojih će upitnik doći budu mlađe ili srednje generacije, što znači da njihovi odgovori ne bi bili reprezentativni za cijelu populaciju prevoditelja.

Drugi nedostatak neodvojiv je od same prirode prikupljanja podataka pomoću samoizvještavanja, a to je da nam podaci prikupljeni na ovaj način zapravo daju uvid u *percepciju* sudionika, koja se može ali i ne mora podudarati sa stvarnim stanjem. To ne znači da sudionici namjerno lažu; možda jednostavno nemaju pravi odgovor (npr. ako ih se pita o nečemu što ne znaju ili razlozima zbog kojih nešto čine, kojih ne moraju biti u potpunosti svjesni¹⁵) ili da kažu ono što misle da je ispravan odgovor ili što misle da se od njih očekuje (npr. ako se pitanje tiče postupanja u nekoj situaciji). Zbog toga se istraživači u prikazivanju rezultata anketnih istraživanja često ograđuju („Trećina sudionika *navodi da...*“, „20 % sudionika *tvrdi da...*“ i sl.).

¹⁴ Demografski podaci traže se na početku ili na kraju upitnika, pri čemu se znanstvenici ne slažu oko toga koja je od tih dviju mogućnosti bolja. Prikupljanjem demografskih podataka na početku upitnika smanjuje se odustajanje sudionika, dok se smještanjem demografskih pitanja na kraj upitnika umanjuje njihov mogući utjecaj na odgovore.

¹⁵ O tom problemu vidi opširnije u Nisbett i Wilson 1977.

Valja još naglasiti da prikupljanje podataka pomoću upitnika može biti presječno (v. 7.2.2.) ili longitudinalno. U drugom se slučaju isti upitnik šalje istim ili nekim drugim sudionicima nakon protoka vremena, jednom ili više puta.

8.1.1. Sastavljanje upitnika

O sastavljanju upitnika napisane su mnoge upute (npr. Saldanha i O'Brien 2013; Matthews i Ross 2010; Visser, Krosnick i Lavrakas 2000), a svi se autori slažu u jednom: dobar upitnik nije jednostavno sastaviti! Loše sastavljen upitnik dovodi u pitanje cjelokupne rezultate pa tom dijelu istraživanja treba posvetiti najviše pažnje.

Upitnik ne bi smio biti predugačak jer će inače sudionici odustati. Uputno je zapitati se: „Koliko će vremena sudioniku trebati da ispuni upitnik?“ O tome ne bi trebalo nagađati, nego provjeriti (na pokusnim sudionicima) te navesti taj podatak na početku upitnika. U nekim je slučajevima pitanja moguće filtrirati kako bi se upitnik skratio, npr. ako je odgovor na neko pitanje negativan, sudionik može preskočiti idućih nekoliko pitanja. Istraživač bi se trebao zapitati: „Hoće li mi ovo pitanje pomoći da odgovorim na svoje istraživačko pitanje?“ Ako neće, pitanje treba izostaviti, koliko god možda bilo zanimljivo (Saldanha i O'Brien 2013: 153–154).

Jednako je važno imati jasnu predodžbu o tome što se upitnikom želi istražiti pa pitanja formulirati s obzirom na taj cilj: „Kako će mi ovo pitanje pomoći da odgovorim na svoje istraživačko pitanje ili da provjerim svoju hipotezu?“ (Saldanha i O'Brien 2013: 153). To je pak povezano s operacionalizacijom ključnih pojmova. Recimo da provodimo anketno istraživanje stavova profesionalnih prevoditelja o korisnosti sustava za strojno prevođenje u svakodnevним prijevodnim procesima. Jedna bi mogućnost bila da sudionike jednostavno zamolimo da procijene korisnost sustava, npr. na ljestvici (v. 8.1.2.) od 1 do 5 pri čemu je 1 = posve beskoristan, a 5 = izuzetno koristan. I na taj način bismo dobili nekakav podatak, ali on bi bio prilično neprecizan i vjerojatno bi se kretao oko sredine ljestvice. Bilo bi ga teško interpretirati, osim u najopćenitijem smislu (npr. „sudionici smatraju sustave donekle korisnima“ ili „neki sudionici smatraju sustave korisnima, a drugi ne“). Pojam korisnosti bilo bi puno bolje operacionalizirati uz pomoć nekoliko parametara – npr. brzina

rada, ideje (kreativnost), standardizirani izrazi, terminologija – a zatim osmisliti pitanja koja se tiču svakog od tih parametara te na taj način izmjeriti „korisnost“. Podaci bi u tom slučaju vjerojatnije bili iznijansirani i pružili bi potpuniju sliku o onome što se proučava.

Pri formuliranju pitanja valja izbjegavati nejasnoće i dvosmislenosti, uključujući dvostruke negacije, kao i pretjeranu upotrebu stručnog žargona. Sve pokrate trebalo bi objasniti, a sve ključne pojmove definirati kako bismo bili sigurni da sudionicima znače ono što je mišljeno (Saldanha i O'Brien 2013: 155). Ni u kojem slučaju nije dobro istim pitanjem tražiti odgovor na dvije stvari (Saldanha i O'Brien 2013: 156). Sam format upitnika trebao bi pridonositi jasnoći, primjerice, podjelom upitnika u dijelove, preglednošću pitanja i odgovora i sl. Prilikom testiranja upitnika od pokusnih se sudionika traži povratna informacija o tome koliko su pitanja jasna pa ne bi trebalo preskočiti taj korak u istraživanju.

Pitanja bi nadalje trebalo sastaviti na takav način da ne navode sudionike na odgovor. Formuliranje neutralnih pitanja, kao i odgovora kojima se sudionici neće navoditi, zna biti zahtjevno. Potrebno je provjeriti jesu li u pitanju ili u ponuđenim odgovorima upotrijebljene riječi koje imaju negativne ili pozitivne konotacije, je li redosljed odgovora takav da daje prednost određenom tipu odgovora, iščitava li se iz pitanja istraživačev stav i sl.

Valja imati na umu da pitanja kojima se od sudionika traži da se svrstaju u određenu kategoriju mogu utjecati na odgovore na pitanja u nastavku upitnika (Saldanha i O'Brien 2013: 154). Takva bi pitanja, ako postoje, trebalo staviti u zadnji dio upitnika te po mogućnosti onemogućiti povratak na prethodna pitanja. Recimo, Babić (2019) i Miličević (2019) uspoređuju stavove sudionika o terminologiji na hrvatskom jeziku s njihovim preferencijama kad je riječ o konkretnim terminima. Zadaci kojima se ispituju preferencije termina nalaze se u prvom dijelu upitnika, a pitanja o stavovima sudionika ispituju se u drugom dijelu upitnika, nakon što su zadaci već ispunjeni i na njih se ne može vratiti.

Pitanjima o hipotetskim situacijama obično se dobiva ono što sudionici misle da bi trebali učiniti ili što misle da istraživač želi čuti, a ne ono što bi zaista učinili. Takva bi pitanja trebalo izbjeći ako se želi doznati što ljudi zaista čine u određenim situacijama, ali se ona mogu koristiti ako nas zanima njihov stav ili percepcija vlastitog ponašanja. Upravo se na taj način u sociolingvistici istražuje imaju li neki nestandardni jezični oblici prikriveni prestiž. Promatraju

se razlike između toga koliko sudionici doista koriste određeni nestandardni oblik (u snimljenom govoru) u odnosu na to koliko smatraju da ga koriste (što se mjeri upitnikom). Ako sudionici smatraju da nestandardni oblik koriste manje nego što ga doista koriste, taj oblik ima prikriveni prestiž (Trudgill 2003: 30): sudionicima je važan u komunikaciji, ali oni tu važnost ne izražavaju u upitniku.

8.1.2. Vrste pitanja

Već smo spomenuli dvije osnovne vrste pitanja – zatvorena pitanja, u kojima se nude odgovori, i otvorena, u kojima sudionik može odgovoriti po želji. Zatvorena je pitanja lakše obraditi kvantitativno, ali odgovorima je potrebno predvidjeti sve mogućnosti i, u slučaju kad je dopušten samo jedan odgovor, izbjeći preklapanja. Otvorena pitanja sudionicima nude priliku da napišu i nešto što nismo predvidjeli te tako omogućavaju bolji uvid u proučavanu pojavu. S druge strane, takva pitanja iziskuju kvalitativnu analizu (v. 11.2.), koja je obično zahtjevnija. Uobičajene su kombinacije otvorenih i zatvorenih pitanja ili postavljanje najmanje jednog otvorenog pitanja na kraju upitnika.

Među zatvorenim pitanjima razlikujemo nekoliko čestih tipova (Matthews i Ross 2010: 207–210):

- dihotomni odgovor: da/ne/(ne znam)
- odabir među ponuđenim kategorijama: npr. M/Ž, dobne skupine
- višestruki odabir: odabir jednog ili više odgovora među ponuđenima
- stupanj slaganja s tvrdnjom (Likertova ljestvica): potpuno slaganje, slaganje, neodlučnost/neutralnost, neslaganje, potpuno neslaganje
- druge vrste procjena na ljestvici (ljestvica „Likertovog tipa“): npr. nikad, rijetko, povremeno, često, uvijek
- rangiranje: sudionici moraju poredati odgovore, npr. prema važnosti.

Kad je riječ o ljestvicama, moguće je odabrati paran ili neparan broj odgovora (3, 4, 5 ili više stupnjeva). Obje mogućnosti imaju prednosti i nedostatke (Sun 2016: 270). U slučaju neparnog broja, srednji je stupanj neutralan odgovor. Neki istraživači preferiraju paran broj odgovora kako bi se sudionike „natjeralo“ da se izjasne, no time ih se možda navodi na odabir koji ne odgovara stvarnome stanju.

8.1.3. Etička pitanja u prikupljanju podataka upitnikom

Jedno od najvažnijih etičkih pitanja (v. 17.) kad je riječ o upitnicima tiče se anonimnosti sudionika. U upitnicima se podaci najčešće prikupljaju na anonimnoj bazi, što može značiti ili da sudionik nigdje neće navesti svoje ime, prezime, adresu e-pošte i sl. ili, ako će ti podaci i biti poznati istraživaču, da se oni neće spomenuti u izvještaju o istraživanju i, što je još važnije, da se neće moći dovesti u vezu s pojedinim odgovorima. Anonimnost može biti ugrožena ako istraživač ne vodi dovoljno računa o sigurnosti podataka ili pak ako je populacija toliko malena i specifična da se već prema nekim odgovorima može zaključiti koji je identitet osobe (Saldanha i O'Brien 2013: 161–162). Očuvanje anonimnosti osobito je važno ako se radi o osjetljivim pitanjima, recimo, o iskustvima prevoditelja u ratnoj ili drugoj kriznoj situaciji, ili o povjerljivim podacima, kao što su prihodi. Osjetljiva pitanja moraju se formulirati tako da sudioniku ne izazovu nelagodu (Matthews i Ross 2010: 206), a preporučljivo je i unaprijed upozoriti sudionike da bi upitnik mogao sadržavati takva pitanja (Saldanha i O'Brien 2013: 155).

8.1.4. Savjeti o korištenju upitnika

Prije slanja ciljnoj populaciji, upitnik bi trebalo testirati na odgovarajućem uzorku kako bi se:

- izmjerilo vrijeme potrebno za ispunjavanje
- preformulirala nejasna pitanja i odgovori
- izbjeglo navođenje ispitanika
- ponudili još neki mogući odgovori
- izmijenio redoslijed pitanja
- izmjerila pouzdanost i valjanost upitnika.

Jednom kad je upitnik poslan, nema povratka!

Kako biste se pobrinuli za to da što više sudionika ispuni upitnik:

- nastojte da bude što kraći i što jasniji
- u uvodu i/ili popratnom pismu napišite istinitu informaciju o tome koliko traje ispunjavanje

- motivirajte sudionike da ispune upitnik, npr. tako što ćete im objasniti relevantnost istraživanja
- zajamčite anonimnost
- nakon nekog vremena podsjetite sudionike da ispune upitnik ako još nisu.

8.1.5. Softver za izradu upitnika

Danas se upitnici uglavnom šalju sudionicima putem interneta. Za izradu upitnika koriste se softverska rješenja koja, osim što omogućavaju sastavljanje pitanja, prikupljaju odgovore u bazu podataka, odakle se mogu „izvesti“ u formatu koji podržavaju programi za statističku obradu (v. 10.5.). Među poznatim sustavima za anketna istraživanja možemo izdvojiti LimeSurvey¹⁶, Google Forms i SurveyMonkey, a stalno se razvijaju novi sustavi.

8.2. Značajke intervjua kao načina prikupljanja podataka

Uz upitnike, intervjui su vrlo popularan način prikupljanja podataka, osobito u društvenim znanostima, a sežu još u 19. stoljeće (Matthews i Ross 2010: 219). Kao i upitnicima, intervjui se prvenstveno prikupljaju podaci o mišljenjima, stavovima i iskustvima sudionika, iako se mogu prikupljati i činjenični podaci. Intervjui mogu biti prikladan način prikupljanja podataka i za eksplanatorna istraživanja, u kojima istraživač nastoji bolje shvatiti ili objasniti ponašanje ili stavove sudionika iz njihove vlastite perspektive (Matthews i Ross 2010: 223). U intervjui su, za razliku od upitnika, pitanja obično otvorenog tipa pa se mogu prikupiti sadržajno bogatiji podaci. Još je jedna prednost intervjua što se njima u većoj mjeri nego upitnicima, upravo zbog otvorenosti pitanja, mogu prikupiti i neočekivani podaci, što istraživaču omogućava dobivanje uvida koji bi se možda inače previdjeli. Zahvaljujući dvosmjernoj interakciji sa sugovornikom, istraživač u intervjuu može postavljati i potpitanja, a tijekom

¹⁶ LimeSurvey dostupan je članovima akademske zajednice u Hrvatskoj preko Srca, na ovoj adresi: <https://limesurvey.srce.hr/>.

intervjua prilagoditi odgovorima koje dobiva, čime se povećava mogućnost dubljeg poniranja u temu.

S druge strane, broj sudionika koje je moguće obuhvatiti intervjuom tipično je manji od broja koji je moguće obuhvatiti slanjem upitnika. Samo prikupljanje podataka intervjuima, kao i analiza takve vrste podataka, obično oduzimaju više vremena. Kao i u eksperimentalnim (v. 7.1.) i opservacijskim (v. 7.2.) istraživanjima, u intervjuima prisutnost istraživača može utjecati na sudionikove odgovore. Nadalje, intervjuima je teže, ako ne i nemoguće, sudionike odabrati metodom nasumičnog odabira (v. 7.1.1.) pa se umjesto toga odabire prigodni uzorak (v. 7.1.1.) ili, češće, namjerni uzorak. Potonji termin odnosi se na uzorke sudionika koji su unaprijed definirani prema određenim kriterijima koji se smatraju relevantnima za temu istraživanja. Na primjer, u dijakronijskom istraživanju razvoja prijevodnih normi u Hrvatskoj od 1960-ih do danas Veselica Majhut (2020) provodi intervjuje s relevantnim akterima povezanim s produkcijom prijevoda kriminalističkih romana (koje analizira u drugom dijelu istraživanja) u danom razdoblju, a to su prevoditelji, urednici i izdavači tih romana koji su u trenutku provođenja istraživanja bili živi i dostupni za razgovor.

8.2.1. Vrste intervjua

U poglavlju o upitnicima (v. 8.1.) i o opservaciji (v. 7.2.) spomenut je stupanj strukturiranosti kao jedan od vidova prikupljanja podataka. Pojam strukturiranosti može se u sličnome smislu primijeniti i na prikupljanje podataka intervjuima pa tako razlikujemo strukturirane, polustrukturirane i nestrukturirane intervjuje. U strukturiranim (standardiziranim) intervjuima sva su pitanja unaprijed pripremljena te se ista pitanja istim redoslijedom postavljaju svim sudionicima, katkad i uz ponuđene odgovore između kojih će birati. Takav je intervju poput upitnika koji se provodi usmenim putem. Činjenica da se svim sudionicima postavljaju ista pitanja omogućava usporedivost podataka i olakšava njihovu kvantifikaciju, dok su nedostatak fleksibilnosti, mogućnosti improvizacije i dvosmjerne komunikacije glavne mane te vrste intervjua. Strukturirani intervjuji prvenstveno se koriste u različitim vrstama anketnih istraživanja, primjerice u sociologiji, gdje je cilj standardizirati situaciju i smanjiti utjecaj istraživača na intervju.

Nasuprot tomu, nestrukturirani intervju omogućava upravo veći stupanj fleksibilnosti i spontanosti. Za takav intervju istraživač priprema temu, ali ne i konkretna pitanja te sudioniku prepušta inicijativu i pruža mu priliku da o temi govori iz svoje perspektive. Budući da sudionici imaju veliku slobodu u takvome razgovoru, prikupljene podatke teže je analizirati, a osobito ih je teško uspoređivati među sudionicima. Podaci koji se prikupe na taj način znaju biti neočekivani i dati posve nove uvide u temu pa je takva vrsta intervju pogodna za izviđajna istraživanja ili ranu fazu većeg istraživanja.

Polustrukturiranim se intervjuima nastoje iskoristiti prednosti gornjih dviju vrsta intervju te ublažiti njihovi nedostaci. U takvim su intervjuima pitanja unaprijed pripremljena, kao okosnica intervju, ali dopuštene su varijacije u pitanjima i njihovu redoslijedu ovisno o konkretnim okolnostima, sudioniku i tijeku razgovora. Mogu se postavljati potpitanja i usmjeravati tijek intervju, ali se sudionicima daje mogućnost da o temi govore na svoj način i iz svoje perspektive. Zahvaljujući tim obilježjima polustrukturirani su intervjui vrlo popularna vrsta intervju koja se u lingvistici i znanosti o prevođenju koristi kako bi se istražila iskustva, stavovi i ponašanje sudionika pa će se upute za pripremu i provođenje intervju u nastavku odnositi upravo na tu vrstu intervju.

Osim po stupnju strukturiranosti, intervjui se mogu razlikovati i po stupnju uključenosti istraživača, od aktivnije do pasivnije uloge. U strukturiranim su intervjuima istraživači obično aktivniji, a u nestrukturiranim pasivniji, premda su moguće i druge kombinacije, kao što je kolaborativni intervju (Matthews i Ross 2010: 220–221) u kojem istraživač, zajedno sa sudionikom, sudjeluje u kreiranju narativa (v. [okvir E](#)).

8.2.2. Priprema polustrukturiranog intervju

Istraživač se za intervju, osobito polustrukturirani, mora temeljito pripremiti. Predlaže se sastavljanje „vodiča kroz intervju“ (Matthews i Ross 2010: 227–230), to jest, „hodograma“ (Potkonjak 2014: 72) u kojem se navode teme i pitanja, ali i druge korisne točke, na primjer, kako započeti i završiti intervju, kako sudionika potaknuti na to da odgovori iskreno i otvoreno, moguća potpitanja i dr.

Za pitanja koja se postavljaju u intervjuu vrijedi ono što je rečeno za upitnike: pitanja bi trebala biti kratka i jasna, bez dvostrukih negacija, razumljiva sudioniku (prema potrebi definirati ključne pojmove na početku intervju!), nijedno pitanje

ne bi trebalo sadržavati više od jedne teme i ne bi trebalo postavljati pitanja, koliko god zanimljiva, koja izlaze iz opsega istraživanja. Kao što naglašava Potkonjak (2014: 72), „[p]itanja moraju moći usmjeriti intervju na istraživačko pitanje“. Pitanjima ne bi trebalo sudionika navoditi na odgovore, nego mu dati priliku da odgovori na svoj način pa bi ona trebala biti formulirana neutralno. Matthews i Ross (2010: 231) predlažu neke oblike potpitanja koji su neutralni i potiču daljnji razgovor, npr. „Možete li mi reći nešto više o tome?“, „Na koji način je to bilo dobro/loše?“, „Prije ste rekli da... Možemo li porazgovarati još malo o tome?“.

Tijek intervjuja i pitanja trebalo bi prije samoga intervjuja iskušati na nekoj drugoj osobi te uzeti u obzir njenu povratnu informaciju i samo iskustvo intervjuja, kao i duljinu intervjuja. Sudioniku bi unaprijed trebalo reći koliko će približno intervju trajati. Ako se istraživač odluči na snimanje intervjuja umjesto vođenja bilježaka, treba unaprijed isprobati uređaj i provjeriti je li dostupno dovoljno memorije za spremanje snimke, a vodič kroz intervju može sadržavati podsjetnik da se uređaj uključi jer je to lako zaboraviti.

8.2.3. Provedba polustrukturiranog intervjuja

Kao što je već naglašeno, smisao je intervjuja dati priliku sudioniku da izrazi svoje mišljenje, stav ili osjećaje u vezi s temom istraživanja (Matthews i Ross 2010: 231). U tu bi svrhu istraživač trebao stvoriti ugodnu atmosferu u kojoj će sudionik osjećati da ga se sluša i cijeni. U tome smislu mogu vrlo dobro poslužiti savjeti koji se obično daju u vezi s aktivnim slušanjem (v. npr. University of Adelaide 2014).

Uvodni dio intervjuja vrlo je važan jer će dati ton ostatku razgovora. Na početku intervjuja istraživač bi trebao ukratko predstaviti sebe i istraživanje te dobiti pristanak za intervju (ako to nije već prije učinjeno). Ako namjerava snimati intervju, to mora najaviti. U uvodnom dijelu istraživač može naglasiti da nema točnih i netočnih odgovora, nego da ga isključivo zanima mišljenje i iskustvo sudionika. Ako je potrebno, može se pojasniti što se misli kad je riječ o ključnim terminima kako u razgovoru ne bi došlo do nesporazuma. Preporučuje se temu uvesti jednostavnijim, činjeničnim i opisnim pitanjima, prije nego se prijeđe na složenija ili intimnija pitanja (Potkonjak 2014: 73).

Osim što je pomno pripremljen, istraživač bi trebao biti i fleksibilan te se prilagoditi svakom sudioniku i katkad neočekivanom tijeku intervjua. Istodobno, vodič kroz intervju služi nam da ne odlutamo od teme i da ne zaboravimo postaviti nijedno od važnih pitanja koja nas zanimaju. Na samome kraju intervjua sudioniku bismo trebali pružiti mogućnost da nam kaže jesmo li ga propustili upitati nešto što smatra važnim u vezi s temom (Potkonjak 2014: 73) ili da nam postavi pitanje u vezi s istraživanjem. Svakako bi sudioniku trebalo zahvaliti na sudjelovanju i uloženom trudu i vremenu. Može mu se ponuditi prilika da vidi transkript intervjua nakon što isti bude izrađen, a prije analize rezultata.

Osim uživo, intervju se može provesti telefonom ili internetskim videopozivom. Druga su dva načina katkad praktičnija, ali u takvim situacijama uspostavljanje dobrog odnosa sa sudionikom može predstavljati izazov.

Nakon intervjua poželjno je zabilježiti neka opažanja u vezi s intervjuom koje se neće moći iščitati iz zvučnog zapisa, npr. o dojmju koji je sudionik ostavio ili o neverbalnim aspektima komunikacije. Istraživač također može obaviti i refleksiju o vlastitim osjećajima i mislima tijekom intervjua te eventualno unijeti neke izmjene u vodič kroz intervju za buduće razgovore (Matthews i Ross 2010: 232).

OKVIR D. RETROSPEKTIVNI INTERVJU

U širem smislu retrospektivni su svi intervjui u kojima se sa sudionicima razgovara o prošlim iskustvima. U užem smislu u znanosti o prevođenju i lingvistici termin se odnosi na intervjue koji se provode odmah nakon što su sudionici sudjelovali u obavljanju nekog zadatka, obično u sklopu eksperimentalnog istraživanja (v. 7.1.). Budući da verbalizacija tijekom zadatka (razmišljanje naglas; v. okvir C) može utjecati na samo obavljanje zadatka, pa tako i na rezultate pokusa (v. npr. Jakobsen 2003), umjesto te metode katkad se koristi retrospektivni intervju koji ne utječe na sam zadatak. S druge strane, nedostatak je retrospektivnog intervjua u tome što se sudionici možda neće sjetiti svega o čemu su razmišljali tijekom zadatka, a možda i neće znati obrazložiti razloge svojih postupaka (v. Nisbett i Wilson 1977). Kako bi ih se podsjetilo na zadatak, kao „znakovi“ za dosjećanje koriste se video- ili audiosnimke zadatka, snimke zaslona ili zapisi dobiveni bilježenjem aktivnosti tipkovnice tijekom zadatka

i dr. (v. okvir C). Sudionikova se verbalizacija snima (ili katkad sluša iz druge prostorije i bilježi) te zatim kvalitativno analizira.

Sobzirom na uključenost istraživača razlikuju se dvije vrste takve retrospekcije (Hansen 2006). U jednom slučaju sudionik dobiva upute te, uz pomoć znakova za dosjećanje, naglas razmišlja o onome što se događalo tijekom zadatka, bez prisutnosti istraživača. U drugome slučaju riječ je o intervjuu u uobičajenijem obliku, u kojem istraživač sa sudionikom razgovara o zadatku dok se sudionik dosjeća svojih misaonih procesa promatrajući snimku ili zapis zadatka. U prvoj vrsti introspekcije umanjen je utjecaj istraživača (iako taj utjecaj uvijek postoji jer sudionik zna da sudjeluje u istraživanju i da se njegova verbalizacija snima ili sluša), a u drugoj vrsti istraživač ima priliku postavljati potpitanja i od sudionika dobiti informacije o onim aspektima zadatka koji ga najviše zanimaju.

Retrospektivni intervju može se provesti i u kombinaciji s opservacijom (v. npr. Risku 2014).

8.2.4. Etička pitanja u prikupljanju podataka intervjuom

Etička pitanja spomenuta u kontekstu upitnika (v. 8.1.3. i 17.1.) pojavljuju se i u kontekstu intervju. To znači da se snimke i transkripti intervju moraju čuvati na način kojim će se zaštititi povjerljivost. Objavljeni rezultati ne bi smjeli sadržavati informacije po kojima se može prepoznati identitet intervjuirane osobe, osim u slučajevima kad je sa sudionikom dogovoreno da intervju neće biti anoniman (obično je riječ o intervjuima u kojima se od osoba na određenim položajima prikupljaju činjenice o nekoj pojavi). U slučaju osjetljivih tema istraživač se mora pobrinuti da intervjuom sudioniku ne izaziva dodatan stres. Matthews i Ross (2010: 226) savjetuju istraživačima da sudionika unaprijed upoznaju s temom intervju i mogućim pitanjima koja bi se mogla postaviti, da sudioniku dopuste da zaustavi snimanje intervju ako za tim osjeti potrebu, da završe intervju u pozitivnom tonu te da sudionika upute na izvore gdje u slučaju potrebe može pronaći podršku nakon intervju. Sudionici u svakom slučaju prije intervju potpisuju obrazac za informirani pristanak (v. 17.1.1.), u kojem bi trebali biti jasno navedeni mogući rizici. Kao i u drugim vrstama istraživanja, nipošto ne bi trebalo vršiti pritisak na ljude da sudjeluju.

8.3. Značajke fokusne skupine kao načina prikupljanja podataka

Ako se podaci prikupljaju istodobno od više sudionika koji raspravljaju o određenoj temi, govorimo o fokusnoj skupini. Točan broj sudionika nije zadan, ali obično se kreće od pet do 13 osoba (Matthews i Ross 2010: 235). Fokusne se skupine koriste u istraživanjima tržišta od 60-ih godina prošloga stoljeća, a kao metoda prikupljanja podataka u društvenim znanostima postaju popularne 80-ih godina (Matthews i Ross 2010: 235). Do podataka se dolazi raspravom sudionika unutar skupine, pri čemu istraživač igra ulogu moderatora rasprave (a ne ispitivača) i ne bi trebao zauzimati strane u raspravi. Rasprava se vodi oko zadane teme, a obično traje oko jedan do dva sata. Može se odigrati uživo ili u *online* okruženju. Na početku rasprave obično se utvrđuju pravila kojih će se skupina pridržavati (npr. da sudionici ne prekidaju jedni druge, da poštuju tuđe mišljenje i sl.).

Za fokusnu skupinu vrijedi većina onoga što je rečeno za (osobito nestrukturirani) intervju, uz neke dodatne elemente, među kojima je najvažnije pitanje zajedničkog konstruiranja značenja te pitanje dinamike grupe. Zbog tih obilježja fokusna se skupina koristi za prikupljanje podataka, kako o sadržaju rasprave, tako i o načinu na koji se rasprava razvija, to jest, o grupnoj interakciji (Matthews i Ross 2010: 236). Fokusna skupina vrlo je pogodna metoda za izviđajna istraživanja i za formuliranje hipoteza u inicijalnim fazama istraživanja, a katkad se koristi i za uključivanje manje zastupljenih osoba u donošenje odluka, to jest, kao oblik društvenog osnaživanja. Fokusnu skupinu moguće je uključiti i u kasnijim fazama istraživanja, nakon što su dobiveni rezultati, kao pomoć pri njihovoj interpretaciji.

Podaci prikupljeni fokusnim skupinama obrađuju se kvalitativnom analizom (v. 11.2.). Za fokusne skupine vrijede slična etička načela kao za prikupljanje podataka upitnikom i intervjuom (v. 8.1.3. i 8.2.4.), pri čemu bi se i sudionici skupine, a ne samo istraživač, trebali obvezati na čuvanje povjerljivih informacija (v. 17.1.2.) izrečenih u zajedničkoj raspravi.

Dosad je bilo riječi o prikupljanju podataka od sudionika istraživanja, a u idućem ćemo se poglavljju osvrnuti na prikupljanje podataka iz tekstova i korpusa.

OKVIR E. NARATIVI - PODACI NA RAZMEĐI SUDIONIKA I TEKSTOVA

Narativ je način prikupljanja podataka koji je na razmeđi između sudionika (v. 7.1., 7.2. i 8.) i tekstova (v. 9.). Primjerice, u nekim intervjuima sudionik sam ili zajedno s istraživačem pripovijeda o nekom svom iskustvu, na taj način stvarajući narativ. Narative, osim u usmenom kontaktu s jednim ili više sudionika (npr. fokusnom skupinom; v. 8.3.), možemo skupljati i u pisanom obliku. Na primjer, sudionicima se može zadati da vode dnevnik svojih svakodnevnih iskustava. Također, narative možemo skupljati i opservacijskim istraživanjima (v. 7.2.), npr. iz blogova ili foruma. Premda prikupljamo „tekst“, ono što nas zanima u analizi jest kako sudionici doživljavaju iskustva, kako ih strukturiraju u cjelovitu priču, koje riječi koriste, koje stavove izražavaju i slično. Narativi se obično analiziraju kvalitativno (v. 11.2.).

ZADACI:

1. Na koji biste način u svom istraživanju mogli upotrijebiti upitnik, intervju ili fokusnu skupinu za prikupljanje podataka? Koji bi od tih načina bio najprikladniji za vaše istraživačko pitanje? Koje bi bile prednosti, a koji nedostaci svake od tih metoda u odnosu na vaše istraživačko pitanje?
2. Ako u istraživanju namjeravate koristiti jedan od načina prikupljanja podataka obrađenih u ovom poglavlju, pokušajte ga provesti s nekoliko probnih sudionika. Provjerite jesu li im sva pitanja razumljiva. Što ste naučili iz probnog ispitivanja?
3. Bi li u vašem istraživanju bilo moguće/korisno kombinirati dva od tih načina i što bi se time dobilo? Bi li u vašem istraživanju bilo moguće/korisno kombinirati upitnik, intervju ili fokusnu skupinu s nekim drugim načinom prikupljanja podataka (eksperiment, opservacija, korpusi...) i što bi se time dobilo?

PROČITAJTE VIŠE: 

Savjete o sastavljanju upitnika nude Saldanha i O'Brien (2013: 151–166) te Matthews i Ross (2010: 200–218). Obje knjige također sadrže vrijedne upute o prikupljanju podataka pomoću intervjua, uključujući i temu fokusnih skupina te etičkih pitanja u provođenju intervjua (Saldanha i O'Brien 2013: 168–204; Matthews i Ross 2010: 218–254). U vezi s retrospektivnim intervjuiima v. Hansen (2006). Pojedivosti o narativima v. u Barkhuizen (2015) i u Matthews i Ross (2010: 264–276).

9. Prikupljanje podataka iz tekstova i korpusa

U OVOM ČETE POGLAVLJU NAUČITI:



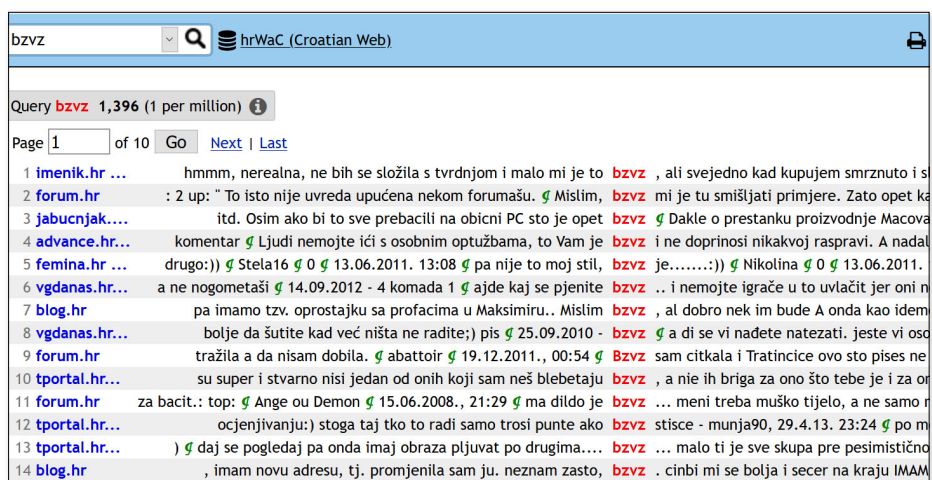
- razlikovati gotove elektroničke korpusne od zbirke tekstova
- razlikovati vrste korpusa
- procijeniti koja vrsta korpusa ili zbirka tekstova odgovara nekom istraživanju
- prilagoditi operacionalizaciju podacima dostupnim u korpusu
- procijeniti postoji li potreba triangulacije pri prikupljanju podataka.

9.1. Računalni korpusi i zbirke tekstova

Kad su nam za istraživanje potrebni jezični podaci (npr. način na koji netko piše/govori ili prevodi), kao izvor podataka možemo koristiti već postojeće računalne korpusne ili jezične podatke koje smo sami prikupili. Pri tome je najvažnije da podaci odgovaraju uzorku (v. 10.1.) koji nam je potreban kako bismo odgovorili na istraživačka pitanja. Na primjer, ako želimo istražiti razvoj prijevodnih normi po razdobljima u pojedinom žanru, potreban nam je uzorak izvornika i prijevoda u okvirima odgovarajućeg žanra i to tako da obuhvaća sva razdoblja koja nas zanimaju. Budući da nijedan postojeći korpus ne zadovoljava te zahtjeve, jasno je da ćemo morati sami prikupiti tekstove za istraživanje. To ujedno znači da će ta zbirka tekstova biti prikladna za donošenje zaključaka u okviru našeg istraživanja. Ipak, na temelju takvog uzorka možemo donositi zaključke samo za navedenu populaciju (v. 10.1.), ali ne i za druge populacije. Ako, s druge strane, želimo istražiti na koji se način u suvremenom pisanom jeziku koristi riječ *бзрвз*, u tom nam je slučaju dostupan računalni korpus hrvatskoga hrWaC (Ljubešić i Klubička 2014), koji će nam za navedeno pitanje

dati odgovarajuće podatke jer je prikupljen tako da sadrži suvremeni jezik dostupan na internetskoj domeni *.hr*.

Između vlastite zbirke tekstova (to jest, okazionalnog korpusa) i postojećeg računalnog korpusa postoje razlike u pretraživanju i obradi teksta. Računalne korpusne možemo preuzeti i na vlastito računalo, ali nam ih je vjerojatno lakše pretraživati pomoću specijaliziranih tražilica, obično dostupnih na internetu. Javno dostupan korpus hrWaC (<http://nlp.ffzg.hr/resources/corpora/hrwac/>) moguće je pretraživati pomoću otvorene tražilice NoSketch Engine (https://www.clarin.si/noske/all.cgi/first_form?corpname=hrwac;align=) i komercijalne tražilice Sketch Engine (<https://www.sketchengine.eu/>). Dostupne tražilice omogućuju različite načine pretrage. Na primjer, želimo li pretraživati riječ *bzvz*, u tražilicu korpusa na odgovarajuće mjesto dovoljno je upisati tu riječ kako bismo dobili sve rečenice u kojima se koristila u tekstovima dostupnim u korpusu. Rezultati se obično prikazuju u obliku KWIC (engl. *key word in context*), tako da je ključna riječ koju smo tražili u sredini (v. [sliku 2](#)). Dakle, prednost je računalnih korpusa što imamo pristup velikom broju tekstova bez mnogo pripreme, ali su nam pretrage ujedno ograničene mogućnostima tražilica.



Slika 2. Rezultati pretrage riječi *bzvz* u korpusu hrWaC pomoću tražilice NoSketch Engine

Za razliku od toga, iz zbirke tekstova koju smo sami izradili podatke možemo prikupljati na koji god način želimo – ručno, pomoću računala ili pomoću korpusnih tražilica. Primjerice, jedno istraživanje u okviru diplomskog rada (Fogec 2014) bavilo se grafičkom i diskursnom analizom najistaknutijih vijesti na naslovnica tiskanih novina *24sata*, za što je trebalo prvo prikupiti naslovnice novina, a zatim iz njih ekstrahirati najistaknutije vijesti. Prikupljanje ciljanih podataka iz naslovnica može ići ručno, bilježenjem odgovarajućih naslova. Ipak, u današnje je vrijeme materijal lako pretvoriti u digitalni format, bilo onaj koji se ne može uređivati (npr. *jpeg*) ili, još bolje, u format koji je čitljiv i može ga se uređivati (npr. *docx*). U oba slučaja, prikupljanje ciljanih podataka dolazi tek nakon digitalizacije, bilo „pješke“ (pregledavanjem svih naslovnica i označavanjem odgovarajućih naslova u programu za uređivanje teksta i slike ili specijaliziranom programu za analizu), bilo pomoću specijalizirane korpusne tražilice. Pri unošenju u specijalizirane tražilice može biti potrebna i dodatna obrada, o čemu će više riječi biti kasnije.

Kad se u današnje vrijeme govori o korpusima, obično se misli na velike računalne korpuse ili na tekstove koje smo sami prikupili te smo ih zatim obradili na takav način da se mogu pretraživati poput računalnih korpusa. Štoviše, korpusna lingvistika, kao i korpusna znanost o prevođenju, obično se povezuju s nekim prevladavajućim (i razmjerno specijaliziranim) načinima obrade podataka koji proizlaze iz frekvencija riječi ili konstrukcija te različitih mjera supojavlivanja izvedenih iz frekvencija (npr. mjere logDice)¹⁷. Prednost je velikih računalnih korpusa u tome da podatak koji nam je zanimljiv (npr. riječ *бэвэ* ili istaknuti naslov na naslovnici) možemo usporediti i s drugim podacima koji se nalaze u korpusu te na taj način doći do novih spoznaja. U takvim se studijama obično koristi mnogo podataka. Premda korpusnu lingvistiku i korpusnu znanost o prevođenju ne možemo nazvati ni teorijom niti metodom, obje se povezuju s nekim istraživačkim tradicijama i konvencijama (v. [Pročitajte više](#)).

¹⁷ LogDice (Rychlý 2008) je jedna od mjera kojima se izražava karakterističnost kolokacija. Druge slične mjere su T-score i MI-score.

9.2. Značajke korpusa

U nastavku ćemo se pozabaviti dimenzijama prema kojima se razlikuju računalni korpusi i računalno pretražive zbirke tekstova koje smo sami prikupili, a to su: modalitet podataka (npr. tekstualni ili audiovizualni podaci), postojanje anotacije i broj jezika (jednojezični, višejezični).

U istraživanjima jezika i prijevoda tradicionalno se koriste tekstualni podaci. Na primjer, za sintaktička istraživanja zanimat će nas rečenice, a najpristupačniji (i najmanje problematičan) izvor rečenica jest neki pisani tekst. Slično tome, istražujemo li pismene prijevode, jasno je da ćemo koristiti pisane tekstove. Međutim, u nekim slučajevima željet ćemo možda istraživati govorni jezik ili usmene prijevode, što znači da ćemo morati imati snimke govora. Snimke treba transkribirati jer je to način bilježenja govora u pismu, a omogućuje nam i pretraživanje snimljenog korpusa. Neki korpusi – poput Britanskog nacionalnog korpusa engleskog jezika – sadrže i pisanu i govornu komponentu.¹⁸ Osim govora, lingvisti se u posljednje vrijeme bave i istraživanjem gesta pa je u tom slučaju riječ o multimodalnom korpusu koji sadrži audiovizualnu komponentu: transkribiranu videosnimku osobe koja nešto govori (npr. dostupnu preko mreže Red Hen Lab). Za istraživanje prevođenja pomoću podslovljavanja potreban nam je audiovizualni korpus koji sadrži tekst (podslove) i izvornik (snimku s govorenim ili pisanim tekstom). Primjeri takvih korpusa su Pavia Corpus of Film Dialogue i Forlì Corpus of Screen Translation (Valentini i Linardi 2009). Primjer korpusa usmenih prijevoda je EPIC (European Parliament Interpretation Corpus), koji, nažalost, ne sadrži hrvatski jezik.

U velikim računalnim korpusima (poput hrWaC-a i Britanskog nacionalnog korpusa) podaci su obično anotirani. *Anotacija* je dodavanje (obično lingvističkih) informacija o nekim jedinicama u korpusu. Pogledajmo dvije rečenice:

- (1) Jabuke su bogate pektinom.
- (2) Priča o jabukama.

U primjeru (1) koristi se oblik *jabuke*, a u (2) *jabukama*. Oba prepoznajemo kao različite gramatičke oblike iste riječi koja u nominativu jednine glasi *jabuka*. Ako želimo analizirati sve gramatičke oblike riječi *jabuka*, bilo bi dobro kad bismo jednom pretragom mogli dobiti pristup svim oblicima, a ne da za

¹⁸ Za poveznice na korpose spomenute u poglavlju v. [Dodatni izvori](#).

različite oblike moramo upisivati pojedinačne pretrage (npr. sve padežne oblike riječi *jabuka*). Upravo je to jedan element anotacije: moguće je anotirati sve gramatičke oblike jedne riječi i pridružiti ih jednom citatnom obliku (lemi), tako da pretraživanjem citatnog oblika možemo dobiti sve anotirane oblike koji su mu pridruženi, što se zove lematizacija. Na taj se način mogu anotirati različiti elementi: vrste riječi, padeži, glagolska vremena, rečenične funkcije, ali i, primjerice, metafore i pragmatičke funkcije. U korpusima koji se koriste pri prevođenju, npr. višejezičnim paralelnim (usporednim) korpusima uz gramatičke elemente obično se navodi oznaka jezika i područje te je li riječ o izvornom ili ciljnom tekstu. U nekim specijaliziranim prijevodnim korpusima anotiran je i prijevodni postupak (npr. Zhai, Max i Vilnat 2018), pogreške u strojnom prijevodu i/ili njihovi ispravci (npr. Costa i sur. 2016). Anotacija korpuse čini vrlo zanimljivima za znanstveni rad, budući da omogućuje pristup različitim vrstama podataka. Čak je i u slučaju tekstova koje smo sami skupili osnovnu anotaciju po vrstama riječi, oblicima i lemmama moguće napraviti automatski, za što postoje alati za različite jezike (v. 9.3).

Anotacija je zapravo vrsta analize i u korpus unosi interpretaciju. Ako anotaciju nismo napravili sami (npr. u postojećim računalnim korpusima), onda moramo biti sigurni da se s njom u potpunosti slažemo. To nije velik problem kod najčešćih oblika vrsta riječi jer je razina slaganja među stručnjacima za njih razmjerno visoka. Međutim, budući da se anotacija vrsta riječi obično odvija automatski, moramo imati na umu da može doći do grešaka koje mogu utjecati na rezultate (to je obično slučaj kod rijetkih oblika, primjerice aorista u hrvatskom). Najzad, u slučaju anotacije elemenata poput pragmatičkih funkcija, metafora ili prijevodnih postupaka moramo biti sigurni da smo suglasni s definicijama anotiranih elemenata.

U zbirkama tekstova koje prikupljamo sami, elementi koje analiziramo u osnovi su ujedno i anotacija. Primjerice, ako smo sami skupili paralelni prijevodni korpus i u njemu istražujemo prijevodne postupke, označavanje tih postupaka zapravo je anotacija, ali i temelj daljnje analize.

Prema broju jezika razlikujemo jednojezične i paralelne korpuse ili zbirke tekstova. Jednojezični računalni korpusi počeli su kao, iz današnje perspektive, razmjerno mali korpusi od milijun riječi (npr. poznati korpus Brown iz 1960-ih), dok danas postoje razmjerno veliki korpusi, kao što su korpusi iz obitelji TenTen, koji imaju više od deset milijardi riječi (<https://www>

sketchengine.eu/documentation/tenten-corpora/). Glavnina multimodalnih korpusa – koji sadrže i govorni ili videozapis – jest jednojezična. Većina velikih računalnih korpusa anotirana je i dostupna za pretraživanje pomoću već spomenutih specijaliziranih tražilica. Rezultati iz različitih jednojezičnih korpusa usporedivi su samo ako su korpusi sastavljeni na usporediv način, što valja imati na umu pri istraživanju. Paralelni korpusi zapravo su dva ili više jednojezičnih korpusa od kojih je jedan izvornik, a drugi su njegovi prijevodi. Kako bi se ti korpusi mogli dobro koristiti, nužno ih je sravniti (engl. *align*) tako da se dijelovi izvornog teksta upare s odgovarajućim dijelovima ciljnog teksta, što se obično radi na razini rečenice. Na taj način pretraživanjem izvornog teksta možemo dobiti i ciljni tekst i obrnuto. Velik broj dvo- i višejezičnih paralelnih korpusa pismenih prijevoda dostupan je preko repozitorija Europske koordinacije za jezične resurse, ELRC (https://elrc-share.eu/repository/search/?q=&selected_facets=resourceTypeFilter_exact%3Acorpus#). Primjer paralelnog korpusa s podslovima jest OPUS (<http://opus.nlpl.eu/>). U današnje se vrijeme za izradu paralelnih korpusa obično koriste prijevodne memorije, kao što su memorije Glavne uprave za pismeno prevođenje Europske komisije, koje se mogu besplatno preuzeti s interneta (<https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory>). Svaka se prijevodna memorija spremljena u formatu *.tmx* može prenijeti na Sketch Engine i tako pretvoriti u korpus koji se može pretraživati na način koji taj alat omogućava.

9.3. Stvaranje vlastitih korpusa ili zbirke tekstova

Stvaranje vlastitog korpusa danas je mnogo lakše nego ranije, pogotovo za jezike za koje su razvijeni različiti korpusni alati (među koje se ubraja i hrvatski). Vlastiti jednojezični korpus možemo dobiti i jednostavnim prikupljanjem elektronički dostupnih tekstova s neke internetske domene. Primjerice, ako nas zanimaju sve kolumne nekog autora na portalu index.hr, dovoljno je da njegove tekstove kopiramo u jedan ili više dokumenata, pregledamo tekst i uredimo ga (npr. izbacimo nepotrebne razmake ili dijelove koji nisu osnovni tekst) i unesemo u program za izradu korpusa kao što je Sketch Engine, AntConc

ili TextSTAT. Pritom valja provjeriti smijemo li koristiti navedene tekstove (zbog autorskih prava) i postoje li pritom neka etička pitanja (v. 17.). Tako dobivamo neanotirani korpus tekstova koji možemo pretraživati po riječima. Alati poput Sketch Enginea omogućuju i automatsko prikupljanje svih tekstova iz neke internetske domene pa čak i automatsko prikupljanje tekstova koji sadrže neke ključne riječi. Kako bismo stvorili paralelni korpus, tekstove koji nas zanimaju možemo prikupiti iz već postojeće prijevodne memorije ili sami napraviti prijevodnu memoriju od izvornika i prijevoda. To se radi pomoću samostalnih alata za sravnjivanje (npr. besplatni LF Aligner) ili alata u sklopu sustava prijevodnih memorija. Za neke jezike postoje alati koji nam omogućuju automatsku anotaciju vrsta riječi u korpusu. Naravno, to neće biti sasvim točno, ali se za neke jezike dobivaju rezultati koji su zadovoljavajući i koji olakšavaju daljnji rad. Ako je istraživač stvorio vlastiti korpus ili bilo koji drugi skup podataka, on može poslužiti i drugim istraživačima, što je u skladu s načelima otvorene znanosti, pa ga se može objaviti u specijaliziranim repozitorijima podataka ili uz sam rad (v. 14. i 16.).

Kad je riječ o stvaranju korpusa ili zbirke tekstova koja sadrži snimljeni govorni audio- ili videozapis, potrebno ga je transkribirati. Transkribiranje je razmjerno dugotrajan proces, a kako ćemo točno transkribirati tekst ovisi o potrebama istraživanja. Razvijeni su različiti sustavi transkripcije govora koji uključuju sasvim različite elemente: od onih najosnovnijih (izrečene riječi, govornike, pauze, nejasne riječi) do onih koji daju detaljan fonetski opis (naglašenost, duljinu itd.). Ako se transkribirani tekst koristi kao korpus, bilo bi dobro da snimka bude povezana s transkribiranim tekstom tako da je uz neku jedinicu transkribiranog teksta (npr. rečenicu) dostupan i audio- ili videozapis tog teksta.

9.4. Operacionalizacija i korištenje korpusa ili zbirke tekstova

Kao i pri korištenju svakog izvora podataka, operacionalizacija (v. 5.2.) rada sa zbirkama tekstova i korpusima ima svoje posebnosti. Ako radimo sa zbirkom tekstova koju nismo pretvorili u korpus, onda nam je kao potencijalni podatak

dostupno grafičko uređenje samog teksta, njegov smještaj na stranici, način na koji se tekst nastavlja na prethodni tekst i slično. Osim toga, dakako, dostupan nam je i sam tekst. To znači da bilo koju od navedenih kategorija pri operacionalizaciji možemo koristiti kao varijablu (v. 6.).

U slučaju tekstova koji su uređeni kao korpus, dostupni su nam, uz različite tekstualne podatke, i svi podaci koji su eventualno anotirani. Ako nas zanima neka kategorija koja nije anotirana (npr. koliko je neki tekst čitak ili je li neki prijevod ideološki obojen), moramo je operacionalizirati pomoću kategorija koje su nam dostupne u samom korpusu. Primjerice, težinu teksta možemo operacionalizirati kao broj slova u riječima ili broj riječi u rečenicama (više slova ili više riječi – kompliciraniji tekst),¹⁹ a ideološku obojenost prijevoda pojavom nekih riječi ili konstrukcija koje smo definirali kao ideološki obojene. Sve to znači da formalne ili anotirane tekstualne kategorije prilikom korištenja korpusa moramo koristiti kao način operacionalizacije (a kasnije i interpretacija) onoga što nas zanima.

Najzad, pri interpretaciji podataka iz zbirke tekstova ili korpusa moramo imati na umu da je neke podatke teško ili nemoguće interpretirati na odgovarajući način bez standarda koji bi ih situirao. Podatak da neki tekst prosječno ima 23.7 riječi po rečenici nije nam informativan ako nemamo standard s kojim bismo ga usporedili ili neki drugi podatak za drugu vrstu teksta. Isto vrijedi i za, primjerice, frekvenciju riječi ili konstrukcija u korpusu. Standard možemo dobiti korištenjem drugog dijela istog korpusa (primjerice, usporedbom frekvencije neke riječi u novinskim tekstovima i internetskim forumima u korpusu hrWaC koji nam dopušta da ga podijelimo na takav način) ili korištenjem drugog korpusa. Pritom valja imati na umu da korpus koji koristimo mora biti po nekom kriteriju usporediv s onim koji istražujemo.

¹⁹ Postoje različiti testovi kojima se može odrediti težina (čitkost) teksta. V. npr. <https://datayze.com/readability-analyzer>. Pritom valja imati na umu da se jezici razlikuju u pogledu takvih testova. U vezi s temom čitkosti na hrvatskom jeziku v. Brangan 2014.

9.5. Triangulacija

Ponekad je u istraživanju potrebno koristiti nekoliko načina prikupljanja podataka. Primjerice, ako se bavimo utjecajem korištenja različitih prijevodnih strategija na recepciju prijevoda, prvo moramo prikupiti primjere prijevodnih strategija, a tek nakon toga istraživati njihovu recepciju. Dakle, sam nacrt istraživanja zahtijeva inicijalno prikupljanje podataka pomoću tekstova, a zatim njihovu procjenu od strane sudionika.

Prikupljanje podataka iz više izvora ili različitim metodama može proizlaziti i iz istraživačeve želje za što potpunijim uvidom u istraživanu pojavu, što se naziva *triangulacija* (Denzin 1978). Na primjer, konceptualnu je metaforu uobičajeno istraživati tekstualno, tako da se promatra koji su njezini jezični oblici, koja joj je učestalost u tekstu i slično. Međutim, postoje i tvrdnje da konceptualna metafora utječe na to kako ljudi shvaćaju svijet, što također možemo ispitivati. Istraživanja metafora koja u nacrtu kombiniraju tekstove i sudionike takvom će triangulacijom pokušati utvrditi hoće li različiti izvori podataka i različite metode donijeti konzistentne rezultate, što je i opći cilj triangulacije.

ZADACI:

1. Pronađite sažetke sljedećih radova i pročitajte ih:

Gumul, Ewa. 2011. „Translational Shifts of Syntactic and Lexical Markers of Ideology: Reporting the Iraqi Conflict in Polish Reprint Press.“ *Meta: Journal Des Traducteurs* 56 (4): 758–74. doi:10.7202/1011251ar.

Vanderbauwhede, Gudrun, Piet Desmet i Peter Lauwers. 2011. „The Shifting of the Demonstrative Determiner in French and Dutch in Parallel Corpora: From Translation Mechanisms to Structural Differences.“ *Meta: Journal Des Traducteurs* 56 (2): 443–64. doi:10.7202/1006186ar.

Merakchi, Khadidja i Margaret Rogers. 2013. „The Translation of Culturally Bound Metaphors in the Genre of Popular Science Articles: A Corpus-Based Case Study from Scientific American Translated into Arabic.“ *Intercultural Pragmatics* 10 (2): 341–72. doi:10.1515/ip-2013-0015.

Na temelju sažetaka odgovorite na sljedeća pitanja:

- Koja se vrsta korpusa koristi u radu (npr. paralelni, postojeći računalni, okazionalni...)?
- Koji će se podaci koristiti iz korpusa (npr. frekvencije, što još)?
- Koje moguće probleme vidite kod operacionalizacije (odredite što specifičnije)?
- Zašto navedeni korpus odgovara (ili ne odgovara) opisanom tipu istraživanja i zadanim ciljevima (v. 2.)?

2. Razmislite o temi vlastitog istraživanja i procijenite biste li mogli prikupiti podatke iz tekstova ili korpusa. Zašto biste ili zašto ne biste mogli s obzirom na posebnosti podataka koje možete prikupiti iz tekstova/korpusa? Postoji li potreba za triangulacijom?

3. Ako su u vašem istraživanju primarni izvor sudionici, koje biste (dodatne) podatke mogli dobiti iz tekstova/korpusa?

4. Ako su vam primarni izvor korpusi/tekstovi, planirate li koristiti tekstove koje ste sami skupili ili postojeće računalne korpuse? Zašto? Biste li mogli koristiti onu drugu vrstu korpusa?

5. Ako su vam primarni izvor korpusi/tekstovi, kako ćete operacionalizirati hipoteze? Je li vam za operacionalizaciju potrebna anotacija?

PROČITAJTE VIŠE:



McEnery i Hardie (2011: 1–21) daju opći uvod u lingvističke korpusne pristupe. Korpusima i zbirkama tekstova u istraživanjima prevođenja bave se Saldanha i O'Brien (2013: 57–108), dok se različiti primjeri korpusnih istraživanja u prevođenju mogu naći u Kruger, Wallmach i Munday (2011).

DODATNI IZVORI: ∞

Uz spomenute primjere korpusa, stranica europske infrastrukture CLARIN (Common Language Resources and Technology Infrastructure) sadrži kratke opise i poveznice na različite vrste korpusa: <https://www.clarin.eu/resource-families>.

Jednojezični tekstualni korpusi:

- Hrvatski nacionalni korpus, HNK (http://filip.ffzg.hr/cgi-bin/run.cgi/first_form)
- Hrvatski web korpus, hrWaC (<http://nlp.ffzg.hr/resources/corpora/hrwac/>)
- Corpus of Contemporary American English, COCA (<https://www.english-corpora.org/coca/>)
- British National Corpus, BNC (<https://www.english-corpora.org/bnc/>, <http://www.natcorp.ox.ac.uk/>)

Paralelni tekstualni/govorni korpusi i korpusi podslova:

- Parallel Corpus of Slavic and other languages, Parasol (<http://parasolcorpus.org/>)
- Hrvatsko-engleski paralelni korpus, hrenWaC (<http://nlp.ffzg.hr/resources/corpora/hrenwac/>)
- Open Parallel Corpus, OPUS (<http://opus.nlpl.eu/>) EPIC (European Parliament Interpretation Corpus) (<http://linghub.org/metashare/e33884de7611e2b1e400259011f6ea48ac8ce0b41f48e6be224bbee9d59cb9>)

Audiovizualni korpusi:

Red Hen Lab (<http://www.redhenlab.org/>)

Pavia Corpus of Film Dialogue (<https://studiumanistici.unipv.it/?pagina=p&titolo=pcfd>)

Primjeri računalnih korpusa i poveznice osvježeni su u lipnju 2020.