Katarina Meglič
Zgornje Duplje

Matevž Treven
Idrija

DOI: 10.4312/HNBP5974
1.22

SLOVENŠČINA KOT VEZ MED GOVORCI IN UMETNO INTELIGENCO: INTERVJU Z DR. SIMONOM KREKOM

Dr. Simon Krek je eden glavnih akterjev na področju digitalnega opremljanja slovenščine in jezikovnega razvoja v okviru tehnoloških novosti. Bil je koordinator projektov Sporazumevanje v slovenskem jeziku in Razvoj slovenščine v digitalnem okolju, znotraj katerih se je nadaljevala in nadgrajevala digitalna opremljenost slovenščine v obliki zbirk besedil ali korpusov (prim. Gigafida), baz podatkov (prim. Sloleks) in jezikovnih tehnologij (prim. prevajalnik, razpoznavalnik govora). Naslednji korak v njegovem delu je razvoj velikih jezikovnih modelov, ki vključuje povečanje nabora besedil za strojno obdelavo in posledično izdelavo tehnologij, ki omogočajo boljšo povezavo človeškega govora in umetne inteligence. Je vodja Centra za jezikovne vire in tehnologije pod okriljem Univerze v Ljubljani in raziskovalec na Institutu »Jožef Stefan«. Pogovarjali smo se o nekaterih dilemah ob trku jezikoslovja in sodobne tehnologije, o vplivu tehnoloških sprememb na jezikovno kompetenco, o nadgradnji tehnologij in o problemih govorcev slovenščine, s katerimi se soočamo trenutno, pa tudi o prihodnosti slovenščine skozi prizmo digitalnega razvoja.

Preden ste se usmerili v delo s slovenščino, ste študirali angleščino in filozofijo. Kako ste se odločili za raziskovanje slovenščine in za razvijanje jezikovnih tehnologij in orodij?

Zunanje okoliščine so k temu kar precej pripomogle. Za diplomo na angleščini sem namreč delal poskus, kako bi izgledal nov angleško-slovenski slovar … Če pa grem še malo dlje nazaj: na Državni založbi Slovenije (DZS) so snovali serijo

filmski romani in v Slovenijo je prihajal Spielbergov *Jurski park*. Zaradi dolgotrajnih dogovarjanj o avtorskih pravicah je ostalo samo še tri tedne do projekcije. Miha Kovač, takratni glavni urednik na DZS, je iskal nekoga, ki bi v manj kot treh tednih prevedel roman *Jurski park* Michaela Crichtona. To je nekako prišlo do mene, sprejel sem nalogo in noč in dan prevajal. Skupaj s še dvema pomagačema nam je to dejansko uspelo. Vmes je Kovač izvedel, da pripravljam koncept angleško-slovenskega slovarja, sam pa je iskal nekoga, ki bi na založbi delal nov angleško-slovenski slovar. Tako sem prišel na DZS. To je bilo leta 1994. Bilo je super. Ekipa je bila velika, skupaj s terminologi, fonetiki, slovenisti okrog 120 ljudi. Ko smo bili nekje na polovici, je postajalo očitno, da se založba usmerja v prodajo različnih stvari in da jih založništvo kot tako niti ne zanima več. Poleg tega je bil to čas korpusov, začetkov korpusnega jezikoslovja, ki se je z angleščino začelo že v 80. letih. Leta 1987 je bil izdan slovar COBUILD, ki je bil osnovan na podlagi tega, iz česar je kasneje nastal korpus Bank of English. Potem so v 90. zgradili British National Corpus (BNC), hkrati je nastajal češki nacionalni korpus. Postalo je jasno, da imamo precej podatkov za angleščino, nimamo pa dovolj sodobnih podatkov za slovenščino. Tako se je znotraj angleško-slovenskega slovarja začel podprojekt, katerega rezultat je bil korpus FIDA, ki sta ga v glavnem financirala DZS in Amebis. Od tod tudi ime FIDA: Filozofska (fakulteta), Institut (Jožef Stefan), DZS in Amebis. Sam sem ta projekt vodil in kmalu se je izkazalo, da si bo glede besedilnih korpusov marsikaj treba na novo izmisliti. Hkrati je založba spreminjala svojo naravo, zato sem se, ko se je leta 2004 angleško-slovenski slovar zaključeval, prestavil z DZS v akademsko sfero. Od tu naprej so se stvari nekako logično razvijale. Najprej smo delali novo različico korpusa, iz FIDE je 2006 nastal korpus FidaPLUS, potem je leta 2008 prišel nov velik projekt Sporazumevanje v slovenskem jeziku, ki se je zaključil 2013 – in tako sem prišel v slovenistične vode. Vmes je bilo treba napisati še doktorat – izvorno niti nisem imel teh načrtov, a se je izkazalo, da je za delovanje v akademskem svetu to pač treba narediti. Doktorat sem pisal v času, ko je potekal projekt Sporazumevanje v slovenskem jeziku.

Na katero temo ste pisali doktorat? Mentor je bil Marko Stabej?

Naslov je Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovnic in se nanaša na organizacijo leksikografskih podatkov na podlagi korpusov. Šlo je za strojno označevanje korpusov in ekstrakcijo podatkov na podlagi tega.

Na kakšen način, v kakšni meri se pri raziskovanju jezika lahko opremo na korpuse? Kako verodostojni so? Kje so njihove omejitve? Kje vidite možnosti za izboljšave?

Splošen odgovor je: empirično jezikoslovje. Poleg merjenja možganskih valov pri govoru je zame to bolj ali manj edino smiselno. Pomeni, da imate kot raziskovalec na voljo neke realne entitete, dele česarkoli že. To je mogoče prenesti na katerokoli znanost. Če denimo preučujete drevesa, je najbrž dobro, da pogledate

od blizu, kako drevo izgleda. Tako je tudi z jezikom. Korpusi so samo zbirke tega, kar jezik je. Imamo pa milijon različnih korpusov. Ne obstaja samo en korpus kot tak. Danes je seveda največji korpus internet. Temu se pridružuje še ena zanimiva stvar: strojno generiranje jezika, ki se je z velikimi jezikovnimi modeli pridružilo korpusom. Pri korpusih predpostavljamo, da je zadaj nek človeški akter ali proizvajalec vsebine. Stroj sicer včasih proizvaja nesmiselne, večinoma pa zelo smiselne dele jezika, stavke, izjave. Skratka, korpusno jezikoslovje je zame v bistvu jezikoslovje kot tako. Lahko seveda tudi proučujete različna stanja jezika v času, kar raziskujejo diahroni jezikoslovci, ampak še vedno se tudi oni opirajo na realne manifestacije jezika. Korpusi so to, kar je bilo kadarkoli zapisano. Od neke točke naprej oziroma nazaj pa zapisa nimamo in od tam naprej so v resnici samo ugibanja. Kako je izgledala indoevropščina, v resnici nihče ne ve.

O sodobnem jezikoslovju: kako razumete razmerje med terminoma knjižni jezik in standardni jezik?

To je samo opora za to, da lahko razmišljamo o dveh konceptih. Lahko bi imeli samo en izraz. Meni se je zdelo smiselno, da se uzavesti razlika, ker imamo neko tradicijo standardizacije, ki je po mojem mnenju naletela na svojo mejo. Gre za procese, s katerimi se ukvarja in jih proučuje sociolingvistika. Ponavadi je v izhodišču nek dominanten dialekt ali njegova varianta, iz česar se nato izvede standardizacijski proces, ki zajame večji ali manjši delež govorečih. Slovenci smo imeli v 19. stoletju – ali pa še prej, predvsem pa v 19. stoletju –, kar zahtevno situacijo, ker je večina tisto, kar je bilo zapisano, samo pisala, ne pa govorila. Od tod tudi izraz knjižni jezik, ker je bil zapisan. Menim, da na koncu 20. stoletja, predvsem pa na začetku 21. ta koncept ni več ustrezen. Treba je nekako povedati, da smo prišli do točke, ko je mogoče govoriti o tem, da ne gre več samo za zapisan standard, ki ga nihče ne govori, ampak ga dejansko uporabljamo tudi za govorno komuniciranje. Ni prisoten samo na RTV, ampak tudi to, kar sedaj govorim, je prilično standardna slovenščina. Koncept standardnega jezika je uporaben v povezavi z internetom in vsemi tehnologijami, ki so se temu pridružile. Imamo jezikovni standard, pri katerem lahko stroju rečemo: tole je nekaj, kar bo večina govorcev razumela in kar bo večina pričakovala. Imamo torej nek koncept tega, čemur rečemo standardna slovenščina, ki ni samo knjižna v smislu tega, kar je bilo treba vzpostaviti sredi 19. stoletja. Zato je to nepotrebna razlika, ki pa je hkrati uporabna za omenjeno razlikovanje.

Se vam zdi, da je pri normiranju pomembno, na podlagi česa je normirana tista najbolj standardna različica jezika? Če bi naredili vzporednico knjižnemu jeziku: ali je pomembno, katera besedila so v korpusu, da se dobi celostno podobo in neko srčiko tega standarda?

Treba je razumeti, da je izraz »so v korpusu« nekoliko problematičen, ker je korpusov veliko. V katerem korpusu? Če govorimo o korpusu standardnega jezika, potem seveda mora biti v njem standardni jezik. Če je to korpus spletnega jezika, v katerem imamo vse twitterje, facebooke itd., je to drug žanr, druga jezikovna zvrst,

ki jo je treba obvladovati tako strojno kot teoretično. Zato rabimo različne korpuse, med drugim tudi korpus standardne slovenščine. Moramo razumeti, kaj za nas v tem trenutku sploh pomeni standardna slovenščina, in ali je korpus standardne slovenščine enak kot korpus knjižne slovenščine. S svojega stališča lahko to razumem na merljiv način, ki nam pove, kdaj smo še znotraj in kdaj že izven standarda. To se da izmeriti. Povsem mogoče je, da se standard tudi premika v času, vendar empirično osnovo za to, da lahko rečemo, kaj je standard, nujno rabimo. Če me vprašate, kaj je knjižna slovenščina … Ne vem, če bi znal narediti korpus knjižne slovenščine. Mogoče bi bil to korpus najboljših piscev ali pa najboljših govorcev, pri čemer bi moral nekdo povedati, kdo to so in kako daleč pri tem gremo: ali Jančar še spada v to, Andrej Skubic pa ne več? To je dilema in vprašanje, ki je zame nepotrebno.

Kako gledate na problematiko neprisotnosti slovenščine pri Applu, Netflixu? Kaj bi rekli mladim, ki pravijo, da se na ta način učijo tujega jezika?

Odgovor je zelo preprost. Bolj ali manj vse, kar se uporablja v Sloveniji, bi moralo biti dostopno tudi v slovenščini, vključno z Applovimi napravami in Netflixom. Gre za vprašanje politične moči in organizacije. V bistvu se strinjam z vsem, kar pravi Lenart J. Kučić, ki se na Ministrstvu za kulturo sedaj ukvarja s tem. Mogoče se je treba zadeve lotiti z lobiranjem z drugimi manjšimi jeziki; če imate hrvaški vmesnik v Applu ... Treba je iskati zaveznike, vključno z Evropsko komisijo, ki ima precejšnjo moč, kar se tiče lobiranja. Gre za politično zadevo, ki jo je treba tudi reševati politično.

Kako digitalizacija vpliva na razvoj komunikacijske kompetence pri mladih?

Neposrednega odgovora na to nimam, ker je vsako stvar najprej treba znati izmeriti. Če sklepamo po zadnjih rezultatih raziskave PISA, je bralna pismenost precej upadla. To verjetno nekaj pomeni, če imamo primerjave za 30, 50 let nazaj. Očitno se je nekaj zgodilo. Če mladi, ki jih merijo, uspejo manj sprocesirati in manj razumeti zapisano besedilo kot njihovi vrstniki po svetu ali pa njihovi predhodniki izpred petih, desetih let, pomeni, da je treba nekaj narediti v zvezi s tem. V tem smislu niso koristne evforične akcije in tudi ne medijsko napihnjene ugotovitve. Po mojem mnenju je raziskava PISA relevantna, ker je precej stroga in primerna.

Na kakšen način bi se to lahko reševalo? Ker sedaj so na voljo aplikacije kot npr. ChatGPT, mladi s pomočjo tega že pišejo tudi seminarske naloge.

Ja, to je povezano in hkrati tudi nepovezano. Za primer: ali kdo od vaju zna jahati?

Ne. Kvečjemu poskusila sva, znava pa ne.

Jahanje konja je bilo pred sto leti povsem običajna spretnost. Ni bilo človeka vajine starosti, ki ne bi znal jahati, ker je bilo to treba znati. V tem smislu smo izgubili marsikakšno spretnost, ki je prej obstajala, pridobili smo pa tudi veliko novih možnosti.

Na primer to, da pridemo iz kraja A v kraj B, ki je oddaljen 50 kilometrov, v pol ure z avtomobilom. V tem smislu je ChatGPT samo novo orodje, ki dela nekaj, kar prej ni bilo na voljo. Tako kot vsaka nova tehnološka rešitev. Realno in bolj smiselno je vprašanje, ali to povzroča učinke, ki so splošno negativni? Eden od teh je, da ljudje ne razumejo več, kaj berejo, ali pa, da so zmožni brati samo štiri minute, potem pa jim koncentracija upade. Razni katastrofični scenariji meni niso preveč pri srcu. Gre bolj za to, da razumemo, kaj se dogaja, in na drugi strani popravimo učinke, ki so negativni. Ko prihajamo v nov svet velikih modelov in ChatGPT-ja, je treba skrbeti, da bomo razumeli, kaj je generiral stroj in kaj človek. Primer je akt Evropske unije o umetni inteligenci, ki skuša te stvari urejati. Naj otroci, dijaki in ostali uporabljajo ChatGPT, ampak naj bo jasno, kaj je napisal ChatGPT in kaj dijak sam. Ne da se to skriva. Če se vrnem na vprašanje o vplivu digitalizacije: enoznačnega odgovora v smislu negativno ali pozitivno ni. Je oboje. Ponuja veliko priložnosti, hkrati pa ima negativne učinke, kot vsaka stvar. Tudi z avtom se lahko zaletiš.

V kakšno smer naj gre slovenščina v prihodnosti? Kaj še potrebuje od priročnikov, jezikovnih orodij ipd.?

Predvsem potrebuje smiselno upravljanje. To je največji problem, ker je jezik, glede na to da še vedno precej komuniciramo v človeškem jeziku, pomembna stvar v družbi. Tukaj smo deloma uspešni in deloma neuspešni. Če uporabimo SWOT analizo, pri kateri z različnih vidikov ovrednotimo potencialne nevarnosti, priložnosti itd. in s tem bolje razumemo, kaj bi bilo treba narediti, kaj manjka, je za jezike, kot je naš, tipična nevarnost ta, da imamo tako malo govorcev. Če na drugi strani govorimo o nemščini ali španščini, gre za popolnoma drugačne razmere. V tem smislu je treba pogledati, kaj se dogaja z jeziki, ki so primerljivi s slovenščino, kot recimo danščina (5 milijonov govorcev) ali pa češčina (10 milijonov). Že za poljščino veljajo povsem drugačne razmere, saj jo govori približno 40 milijonov govorcev. V tem smislu je treba upravljati jezik tak, kot ga imamo, z realnim številom govorcev. Nekatere stvari je treba bolj podpreti, vanje več vložiti, ker ne pridejo same od sebe, druge vendarle pridejo tudi same od sebe. Malteščina ali pa makedonščina sta denimo v še slabšem položaju kot slovenščina. Kaj se zgodi v takem primeru? Velika podjetja podprejo npr. razpoznavo govora, strojno prevajanje, sintezo govora ipd. naprej za tiste, ki so komercialno zanimivi, ostale pa zanemarijo, med drugim tudi slovenščino. To je treba pri upravljanju jezika upoštevati – kar se nekako v zadnjih letih tudi dogaja – da se vlaga v odprte vire. Če dam konkreten primer: znotraj projekta Razvoj slovenščine v digitalnem okolju, ki je potekal med letoma 2020 in 2023, smo zbrali tisoč ur govora. Transkripcija in avdio posnetek sta odprto dostopna, kar pomeni, da ju lahko vzame kdorkoli, recimo Siemens, Mercedes, Tesla, in iz tega naredi razpoznavalnik govora za slovenščino in ga vgradi v avto. To je mogoče, ker je nekdo zasnoval projekt, pri katerem je bilo treba na koncu objaviti popolnoma odprte vire, ki jih lahko uporabljajo tudi podjetja. Tega zavedanja 15 let nazaj še ni bilo. Takrat se je razmišljalo »tega ne sme nihče komercialno izkoriščati«, vmes pa je postalo jasno, da bo odprte vire vendarle treba omogočiti tudi za podjetja, saj so se ta sicer odločala v smislu: »Ok,

vir je nekomercialen, ne moremo ga uporabiti, torej ne bomo vključili slovenščine.« Preprosto. To pomeni upravljanje z jezikom. To lahko zagotovo rečemo sedaj, ne vemo pa, kaj bo čez pet let, ker se tehnologije hitro razvijajo. Ali bo takrat pomembno imeti tisoč ur transkribiranega govora ali ne? Kdo ve.

Kakšna je podpora države za tako delo?

Država ima problem s tem, da je slovenščina razpršena med štiri ministrstva. Služba za slovenski jezik deluje pod Ministrstvom za kulturo in vsaka stvar v zvezi s slovenščino gre avtomatično tja. Hkrati pa so infrastrukturne in raziskovalne zadeve, ki sem jih omenjal prej, pod Ministrstvom za visoko šolstvo, znanost in inovacije. Šolski del ostaja pod Ministrstvom za vzgojo in izobraževanje, Ministrstvo za digitalno preobrazbo pa pokriva jezikovnotehnološki del, umetno inteligenco itd. Vse je odvisno od tega, ali se bodo inštitucije uspele med seboj zmeniti ali ne – lahko vsak razmišlja po svoje in tudi vlaga po svoje in ni konsistentnega razvoja. Se pa to do neke mere uspešno rešuje z Resolucijo o nacionalnem programu za jezikovno politiko, ki vsebuje tudi tehnološki del. V tem smislu podpora je, ampak je odvisna od tega, ali se bodo omenjeni štirje akterji med seboj uskladili. Na drugi strani so tu še akademske inštitucije kot Univerza v Ljubljani in Inštitut za slovenski jezik, kjer se moramo tudi znati med seboj zmeniti v dobro vseh govorcev in govork.

Pri inštitucijah lahko pogledamo tudi Filozofsko fakulteto in študij slovenščine, slovenistiko. Kakšno je vaše mnenje o trenutnem nivoju študija? Kaj bi bilo potrebno spremeniti, dodati, odvzeti, mogoče spremeniti fokus glede na to, kar ste do sedaj povedali?

Študija slovenistike trenutno niti ne poznam dobro, ker nisem neposredno udeležen v njem, ampak moj odgovor je verjetno očiten. Pričakoval bi, da je bistveno več digitalnih vsebin znotraj študija, ker bi slovenisti na koncu morali biti popolnoma digitalno usposobljeni, glede na to da živimo v digitalnem svetu. Vprašanje je, koliko je taka usposobljenost res dosežena. Mislim, da ni.

Kaj pa vaši naslednji projekti? S čim se trenutno ukvarjate? Kaj, kar vam je še posebej ljubo?

Septembra letos smo začeli z novim raziskovalno-inovacijskim projektom, ki bo trajal tri leta. V projektu imamo pet podjetij in bomo delali slovenski ChatGPT oziroma velike jezikovne modele za slovenščino. Vzeli bomo odprto dostopni Facebookov oziroma Metin model Llama 2 in zbrali vse obstoječe podatke v slovenščini, do katerih bomo lahko prišli. Trenutno imamo v vseh korpusih, ki so na voljo, približno 10 milijard besed. Rabimo jih najmanj štirikrat, petkrat toliko in nismo popolnoma prepričani, če sploh obstajajo. Celoten NUK, digitalna knjižnica, transkribiranje celotnega arhiva RTV, vse, kar je na Arnesu in tako naprej. Bomo videli, do kod bomo prišli. Potem bo morda treba dodati še kaj od južnoslovanskih jezikov. Primer so Švedi, ki so se tega lotili takoj in naredili svoj model GPT-SW3. Sami niso mogli

nabrati dovolj švedskih tekstov in so morali vključiti še norveščino, danščino, islandščino in angleščino plus računalniško kodo, da so prišli do 300 milijard besed. Iz tega so potem učili model, ki pa sam na sebi ne pomeni dosti, ker je treba imeti še podatkovne množice, iz katerih se konkretna aplikacija uči za konkretno nalogo, recimo za konverzacijo. V OpenAI so za ChatGPT najprej naredili model, potem so zbrali ogromno, milijone in milijone dialogov, v katerih so ljudi spraševali, oni so odgovarjali. Potem se je aplikacija učila odgovarjati na podoben način in tako so prišli do ChatGPT-ja. Zato se tudi imenuje ChatGPT. Znotraj projekta PoVeJMo (prilagoditev velikih jezikovnih modelov za slovenščino) bo izdelana aplikacija za področje digitalne humanistike, s katero bo recimo mogoče generirati tekste za opis artefaktov v muzeju. Drugi primer je medicina in govorne tehnologije, razpoznava govora. Denimo, dialog med pacientom in zdravnikom je mogoče sproti transkribirati in hkrati prepoznati določene elemente v tekstu, kot so recimo zdravila, bolezni itd. To se poveže z bazo znanja, zato da lahko umetna inteligenca malo bolj pametno razmišlja o tem, o čemer sta se zdravnik in pacient pogovarjala. Tretji primer je industrijski, upravljanje skladišč z govorom. Četrti pa je računalniška koda, kar niti ni toliko povezano s slovenščino. To bomo delali tri leta. Hkrati prijavljamo še druge projekte. Vse pa se sedaj vrti okrog jezikovnih modelov in podobnih aplikacij. Korpusi, kot smo jih poznali, so na nek način pasé.

Glede umetne inteligence pri medicini: ali je kakšen problem glede varovanja osebnih podatkov?

Velik. To se rešuje z anonimizacijo in psevdoanonimizacijo, kar pomeni, da imen ne skrivamo, ampak jih zamenjamo in konkretne osebe postanejo razni Janezi Novaki. Potem načeloma ni več mogoče prepoznati, za koga je dejansko šlo. V ta namen se tudi sprejemajo različni akti, med drugim Zakon o varovanju osebnih podatkov. Nekatera okolja morajo biti zaprta. Prav to je problem, zaradi katerega moramo imeti jezikovne tehnologije sami, da nam podatkov ne bo treba pošiljati na neke strežnike v Severno Ameriko.

Kar pa je konec koncev velik izziv. Koliko je možnosti, da pride neko podjetje in na podlagi dostopnih tehnologij in baz podatkov razvije rešitev, ki jo za take vrste situacijo tudi ponudi?

Upam, da ta možnost vedno obstaja. Ključno sporočilo je – čim več. Ni rešitev v tem, da se vse zapira. Če obstaja samo en ponudnik, je to monopolistična situacija, ki se jo da izkoristiti na sto različnih načinov. Nekaj časa že govorimo o jezikovni suverenosti. Če pri jeziku nimamo digitalne komponente, ki jo lahko sami obvladujemo, potem smo odvisni od drugih. Saj v marsičem smo odvisni od drugih, nafte na primer nimamo. Ampak jezik je specifično naš.

Jezik je bil v bistvu to, na podlagi česar smo se Slovenci sploh lahko združil na tak način, kot smo sedaj združeni. Bil je velik faktor pri tem.

Seveda, zato je še toliko bolj pomembno, da smo glede tega suvereni. To, o čemer govorimo zdaj, je predvsem tehnološka suverenost, da imamo na voljo vse, kar imajo drugi. Prvi del te infrastrukture je bila v bistvu Biblija, če gledamo nazaj. Nemška Biblija, slovenska Biblija, vse ostale Biblije. Tako se je začelo. Potem je bila precejšnja časovna luknja, ko so drugi jeziki dobivali slovarje, literaturo itd. Mi smo bili v tem smislu podhranjeni. V 19. stoletju se je tudi za Slovence začela graditi jezikovna infrastruktura in ko je postalo jasno, da se oblikujejo nacionalne države, da je jezik tipično treba opisati v slovnici, v slovarju, smo se pri slovenščini s slovarjem ubadali precej predolgo. Slovensko-nemški slovar na koncu 19. stoletja je bil tako rekoč ves obstoječi slovarski opis. Potem je trajalo do leta 1990, da smo dobili celoten enojezični slovar. Logično bi bilo, da bi ga dobili recimo 1920. Je pa seveda tudi res, da so Švedi ravno končali svoj slovar po 120 letih (smeh). Situacije so različne, ampak bistvo je to, da jezik rabi infrastrukturo, ki pa se zelo spreminja v času. Sedaj bo definitivno na udaru digitalna infrastruktura.

To je verjetno povezano tudi s tem, kako se moč prerazporeja v družbi. Če ste prej omenili Biblijo, se to lahko poveže s tem, katere so bile takrat inštitucije družbene moči. V 19. stoletju se je družba izrazito spremenila in na nek način omogočila drugačno upravljanje z jezikom. Tudi zdaj smo v obdobju, ko smo zaradi hitrih in intenzivnih tehnoloških sprememb priča preobratu družbene moči in intenzivnemu spraševanju, kaj je najbolj pomembno, kar se jezika tiče. Konec koncev tudi to, kaj ima največji vpliv.

Ja, drži. So vzporednice, seveda. Če pogledamo zgodovinsko: večnacionalna država Avstro-Ogrska, stara Avstrija, ni bila najbolj prijazna do vzpostavljanja nacionalnih jezikov. Boljša situacija je bila potem kljub vsemu v stari Jugoslaviji in socialistični Jugoslaviji, ni pa bila idealna. Od leta 1991 smo v tako rekoč idealni poziciji v primerjavi s prej, ker je Evropska unija eksplicitno večjezična. Močno podpira vse uradne jezike in daje slovenščini in malim jezikom, kot so latvijščina, litvanščina itd. veliko moč, ki je prej ni bilo. Kar se je zgodilo novega, pa je najbolj povezano z nastankom Googla in s tem, da imamo izrazito močne akterje v Severni Ameriki. Pred sto leti nas ni brigalo, kaj se tam dogaja v smislu jezika, sedaj pa nam marsikaj določajo algoritmi. Ta del se je spremenil in vsakič je treba ravnati v skladu s tem, kar bo ljudem prineslo čim manj težav. Predvsem pa ne živeti v preteklosti. Če se vrnem nazaj na dilemo knjižno in standardno. Ukvarjati se z dilemami, ki so stare 30, 40, 50 let, ni smiselno, ker je dovolj težav, ki izhajajo iz zdajšnje situacije. Glede vprašanja standardnega in knjižnega jezika je moje sporočilo predvsem to: standardna slovenščina se ukvarja s problemi, ki so zdaj relevantni za govorce in govorke. Ljudje, profesorji, kdorkoli že ... marsikdaj razmišljajo o jeziku kot o entiteti, ki je ločena od govorcev in govork. Kot da obstaja izven nas kot neoprijemljiv duh. Jezik je v resnici samo to, kar delamo mi sedaj, ko govorimo, razmišljamo, sanjamo in tako naprej. Je v nas. Spoznanja, da nimamo Apple vmesnika, da ni razpoznavalnika, da nimamo tega, onega, tretjega, tudi slovarja pred 100 leti ... da nimamo nekega dela jezika, ki bi ga morali imeti zato, da bi lahko opravljali vse, kar hočemo, so realne in prave težave govork in govorcev slovenščine.