

Sveučilište u Zagrebu
Filozofski fakultet
Odsjek za psihologiju

**NORME SLOBODNIH ASOCIJACIJA ZA UZORAK RIJEČI NA HRVATSKOM
JEZIKU**

Diplomski rad

Nika Vezmar

Mentor: Doc. dr. sc. Andrea Vranić

Zagreb, 2017.

Sadržaj

Uvod.....	1
Podjela pamćenja.....	1
Slobodne asocijacije.....	1
Primjena slobodnih asocijacija u proučavanju lažnih sjećanja.....	2
Asocijativna povezanost.....	3
Metode procjene povezanosti među riječima.....	6
Postojeće baze slobodnih asocijacija.....	7
Ovisnost strukture mentalnih asocijacija o kulturi.....	8
Cilj.....	9
Problem.....	10
Metoda.....	10
Sudionici.....	
Instrumenti.....	
Sociodemografski upitnik.....	10
Skale asocijacija.....	11
Računalni program.....	11
Postupak.....	
Kreiranje početnog fonda riječi.....	12
Prikupljanje podataka.....	12
Priprema podataka za obradu.....	14
Obrada.....	16
Rezultati i rasprava.....	17
Mogućnost primjene normi u istraživanjima lažnih sjećanja.....	23
Metodološki nedostaci i smjernice za buduća istraživanja.....	24
Zaključak.....	26
Literatura.....	28

Norme slobodnih asocijacija za uzorak riječi na hrvatskom jeziku

Free association norms for a sample of Croatian words

Nika Vezmar

Ovim istraživanjem nastojali smo proučiti asocijativnu strukturu riječi kod govornika hrvatskog jezika kreiranjem normi slobodnih asocijacija za hrvatski jezik. Metodom diskretnih slobodnih asocijacija prikupljene su asocijacije za 700 hrvatskih riječi. U istraživanju je ukupno sudjelovao 971 govornik hrvatskog jezika, stariji od 18 godina. Svakom sudioniku prezentirano je 100 podražajnih riječi, a na svaku je riječ bilo potrebno dati jednu asocijaciju. Prikupljeni odgovori su lematizirani i uniformirani. Svaka je podražajna riječ u prosjeku normirana 137 puta. Za svaki par riječi podražaj - odgovor izračunata je asocijativna snaga unaprijed (FAS), a prikupljene riječi zajedno s izračunatim FAS-om tvore usmjereni težinski graf koji ocrta asocijativnu strukturu riječi. Graf sadrži 3113 različitih riječi te 11215 direktnih veza među riječima. Osim FAS-a, za parove riječi izračunate su i povratna (BAS), medijacijska (MAS) i preklapajuća (OAS) snaga. Za svaku riječ priloženi su i sljedeći podaci: oznaka je li riječ normirana, broj sudionika koji ju je normirao te ukupan broj različitih asocijacija na tu riječ. Prikupljeni podaci primjenjivi su u različitim granama psihologije – poglavito kognitivnoj psihologiji i psiholingvistici. Vjerojatnosti pojavljivanja pojedinih odgovora prikupljenih postupkom slobodnih asocijacija značajne su za objašnjavanje mehanizama pronalaženja informacija u pamćenju te za predviđanje izvedbe u raznim zadacima vezanima uz pamćenje, kao što su zadaci za proučavanje lažnog sjećanja. U tom kontekstu, norme mogu poslužiti za konstruiranje lista riječi u DRM paradigmi.

Ključne riječi: asocijativna struktura, metoda slobodnih asocijacija, asocijativna snaga, hrvatski jezik, norme

The aim of this paper is to study the associative structure of words for Croatian language speakers by creating free association norms for the Croatian language. We collected associations for 700 Croatian words using the discrete association method. A total of 971 Croatian language speakers, older than 18, participated in the study. Each participant was presented with 100 cue words and instructed to give one association for each cue word. The answers were lemmatized and normalized. On average, each stimulus was normed 137 times. For each stimulus-response pair, the forward (cue-to-target) associative strength (FAS) was calculated. The collected associations, combined with the calculated FAS, form a weighted directed graph that outlines the associative structure of the words. The graph contains 3113 unique words and 11215 direct edges between them. In addition to the FAS, the backward (BAS), mediated (MAS) and overlapping (OAS) strengths were calculated for each word pair. Additionally, the following information is included: whether the target word is normed, the number of participants that normed the cue, and the total number of different associations of that cue. The constructed norms are applicable in various subfields of psychology - primarily cognitive psychology and psycholinguistics. The probabilities of individual responses collected in the free association task are significant in explaining memory retrieval mechanisms and in predicting performance in various memory-related tasks, such as in false memory tasks. In this latter context, the norms can be used to construct word lists in the DRM paradigm.

Keywords: associative structure, free association task, associative strength, Croatian language, norms

Uvod

Podjela pamćenja

Pamćenje je mogućnost usvajanja, zadržavanja i korištenja informacija. Postoje tri skladišta pamćenja kroz koje prolazi svaka informacija: osjetno, kratkoročno i dugoročno. Najčešće korištena podjela dugoročnog pamćenja je Andersonova (1976), koja dijeli pamćenje na dvije komponente: deklarativno i nedeklarativno. Deklarativno pamćenje odnosi se na znanje o činjenicama, dok se nedeklarativno pamćenje odnosi na znanje o tome kako izvesti neku radnju. Deklarativno pamćenje dijeli se na semantičko i epizodičko. Pod epizodičkim pamćenjem podrazumijevamo sjećanja na događaje koje je netko osobno doživio ili koji su se dogodili u njegovoj blizini. Semantičko pamćenje nužno je za uporabu jezika, odnosi se na pamćenje značenja riječi i pojmova te gramatičkih pravila za slaganje rečenica (Zarevski, 1997).

Slobodne asocijacije

Jedan od načina proučavanja semantičkog pamćenja je putem procjene jačine povezanosti riječi. Ugrubo takve postupke možemo podijeliti na asocijativne (utemeljene na asocijacijama) i neasocijativne (primjerice, utemeljene na frekvenciji uzastopne pojave u tekstu). Paradigma utemeljena na asocijacijama datira još od 19. stoljeća kada su Galton i Wundt započeli sa psihometrijskim mjerenjem i sustavnim proučavanjem asocijacija (Boring, 1950). U tu svrhu koristili su zadatak slobodnih asocijacija u kojem su sudionici zamoljeni odgovoriti na prezentirani podražaj prvom asocijacijom koje se dosjete. Kasnije su zadaci tog tipa korišteni u kliničkim istraživanjima ljudske patologije i inteligencije, često su korišteni u području biheviorizma, a danas su najrašireniji u kognitivnoj psihologiji, u istraživanju jezika i pamćenja (Cramer, 1968). U području psihologije pamćenja i psiholingvistike postoji veliko zanimanje za asocijacije među riječima jer su, iz teoretskog gledišta, u središtu razumijevanja organizacije znanja i pojmova (De Deyne & Storms, 2008). Iako u kognitivnoj psihologiji i psihologiji općenito još nije u potpunosti razjašnjen mehanizam u podlozi asocijacija među riječima, trenutni konsenzus je da asocijacije

među riječima odražavaju naše leksičko znanje o njima te o njihovim odnosima, stečeno kroz svakodnevno iskustvo i interakciju sa svijetom (Nelson, McEvoy i Schreiber, 2004). Nadalje, struktura samih asocijacija (što je s čim asocirano i koliko jako) obuhvaća i važne aspekte značenja riječi, odnosno njihove semantičke reprezentacije (Nelson i sur., 2004). Takvu ukupnu strukturu među riječima, putem asocijacija, nazivamo asocijativna struktura. Zadatkom slobodnih asocijacija saznajemo koje su riječi međusobno asocirane, a ujedno i koliko je ta povezanost jaka, putem vjerojatnosti pojavljivanja pojedinog asocijativnog odgovora. Te vjerojatnosti govore nam nešto i o samim mehanizmima pronalaženja informacija (Nelson, McEvoy i Dennis, 2000) te o izvedbi u drugim zadacima vezanim uz sjećanje, npr. u zadacima dosjećanja u kojima uz zadanih par slova treba dovršiti riječ (Steyvers, Shifirin i Nelson, 2004).

Primjena slobodnih asocijacija u proučavanju lažnih sjećanja

Osim navedenog, asocijativna struktura važna je i za proučavanje lažnih sjećanja. To su sjećanja o informacijama i/ili događajima koje nismo doživjeli (Schwartz, 2014). Krajem 1950-ih, Deese (1959a, 1959b) je proučavao kako asocijativni faktori među riječima utječu na lažno dosjećanje. U tu svrhu konstruirao je liste sastavljene od 12 riječi odabranih tako da imaju najveću asocijativnu snagu prema nekoj riječi koja nije dio liste (kritična riječ ili riječ *mamac*). Asocijativnu snagu između dvije riječi definirao je kao povezanost tih riječi u zadatku slobodnih asocijacija, odnosno kao vjerojatnost da će prva riječ, prikazana kao podražaj, dovesti do dosjećanja druge riječi kao odgovora sudionika u takvom zadatku. Koristeći tako konstruirane liste, proveden je eksperiment u kojem je sudionicima prikazana lista riječi koju trebaju zapamtiti (faza učenja), a zatim je neposredno nakon učenja od njih zatraženo da se dosjete riječi koje su bile na listi (faza dosjećanja). Sudionici su se dosjećali kritične riječi u 44% slučajeva, ali se vjerojatnost dosjećanja razlikovala za različite liste, odnosno liste koje su bile jače asocirane s kritičnom riječi češće su dovodile do lažnog dosjećanja.

Polazeći od Deeseovog nalaza, Roediger i McDermott (1995) kasnije su modificirali paradigmu testirajući prepoznavanje umjesto dosjećanja. Po navedenim

autorima, opisani postupak za proučavanje lažnog sjećanja naziva se DRM paradigmom, a liste riječi koje su asocijativno snažno povezane s jednom kritičnom riječi nazivamo DRM listama. DRM paradigma predstavlja robusnu tehniku za istraživanje lažnih sjećanja. Zbog svoje jednostavne primjene, velikih mogućnosti eksperimentalne manipulacije, etičke neupitnosti i pouzdanosti trenutno je najraširenija tehnika za istraživanje lažnog sjećanja (Schwartz, 2014). Koristi se u preko 40% svih istraživanja na tu temu (Pezdek i Lam, 2007).

Asocijativna povezanost

Kako bismo pojam asocijativne povezanosti stavili u intuitivan okvir, modelirat ćemo asocijativnu strukturu u okvirima matematičke teorije grafova. U tom pogledu, semantičke mreže općenito možemo prikazati kao grafove čiji čvorovi (*engl. nodes*) predstavljaju riječ, a bridovi (*engl. edges*) neku vrstu veze među riječima. Bridovi mogu biti usmjereni i ponderirani, što nam prikazuje smjer i snagu povezanosti (Steyvers i Tenenbaum, 2005). Jedan mogući prirodni odabir za pondere je neka od mjera asocijativne povezanosti među riječima. Na taj način dobivenu semantičku mrežu možemo promatrati kao prikaz asocijativne strukture.

Jedna od osnovnih asocijativnih mjera povezanosti je asocijativna snaga unaprijed (*engl. forward associative strength - FAS*). FAS za par riječi (A,B) definiramo kao proporciju odgovora B pri prezentaciji podražaja A u metodi slobodnih asocijacija (Nelson, McEvoy i Schreiber, 1998). Srodan je i pojam povratne asocijativne snage (*engl. backward associative strength - BAS*) koji predstavlja asocijativnu snagu u drugom smjeru, odnosno $BAS(B,A) = FAS(A,B)$. Drugim riječima, povratna asocijativna snaga između riječi B i A , u kontekstu metode slobodnih asocijacija, govori nam koliko često će sudionik ponuditi riječ B kao asocijaciju kad mu se prezentira podražajna riječ A .

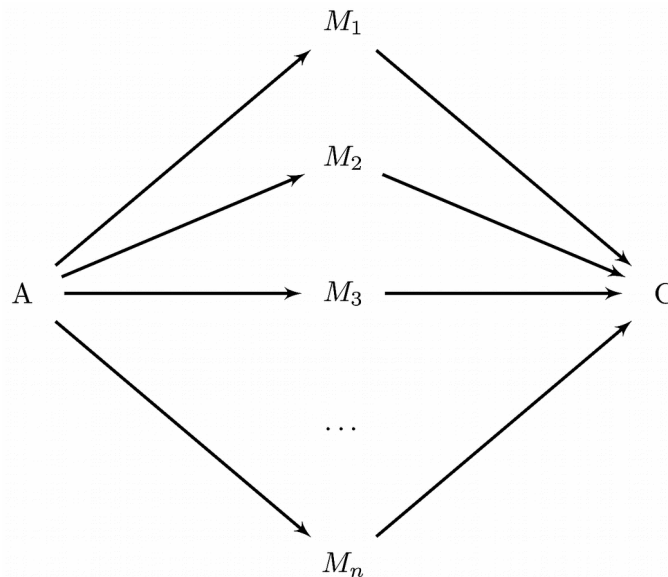
Budući da svi parovi riječi neće biti izravno povezani, ali je moguće da su povezani posredno putem drugih riječi, korisno je definirati i posrednu asocijativnu snagu (*engl. mediated associative strength - MAS*). U literaturi se MAS ponekad naziva dvostupanjskom snagom (*engl. 2-step*). Za par riječi (A,C) koje su u FAS grafu

povezane posredstvom jedne riječi B , MAS riječi A i C definiramo kao umnožak asocijativnih snaga unaprijed između A i B te B i C (Nelson i sur., 1998):

$$MAS(A,C) = FAS(A,B) \cdot FAS(B,C)$$

Općenito, ukoliko su riječi A i C povezane posredstvom n različitih riječi M_1, M_2, \dots, M_n , MAS riječi A i C definiramo kao zbroj snaga svake pojedine veze:

$$MAS(A,C) = \sum_{i=1}^n FAS(A, M_i) \cdot FAS(M_i, C)$$



Slika 1. Grafički prikaz općenitog slučaja za MAS riječi A i C

Analogno tome moguće je definirati i cijelu porodicu posrednih asocijativnih snaga za dulje puteve, odnosno n -stupanjsku posrednu asocijativnu snagu između riječi A i C , povezanih posredstvom puteva duljine $n-1$. Ako na sljedeći način definiramo n -stupanjsku snagu po jednom putu p duljine n , postavljajući $M_0 = A$ i $M_n = C$:

$$Put_p(A,C) = FAS(A, M_1) \cdot FAS(M_1, M_2) \cdot \dots \cdot FAS(M_{n-1}, C) = \prod_{i=0}^{n-1} FAS(M_i, M_{i+1})$$

Tad n -stupanjski MAS možemo definirati kao sumu takvih snaga po svim putevima p :

$$MAS_n(A, C) = \sum_p Put_p(A, C)$$

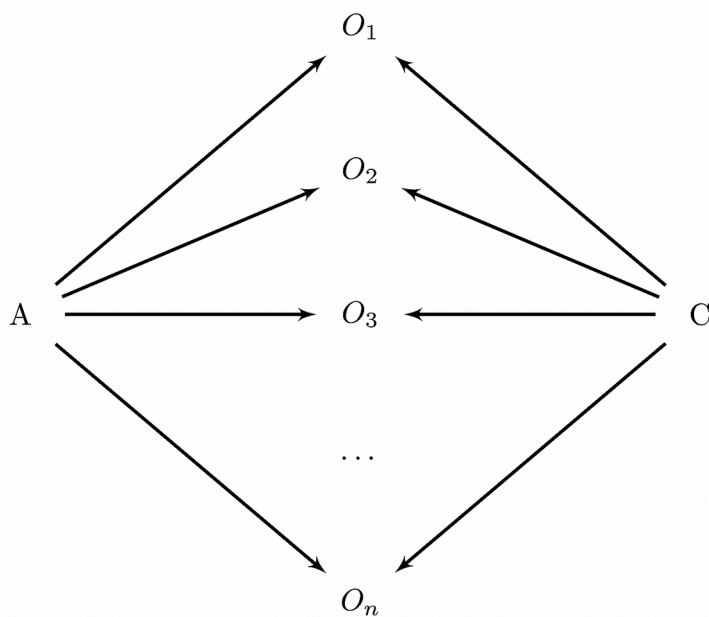
Grafički prikaz općenitog slučaja za n -stupanjski MAS dodan je kao prilog B.

Konačno, može se dogoditi da dvije riječi A i B u FAS grafu nisu povezane ni direktno, ni putem neke druge riječi, ali da obje asociraju na neku treću riječ C (odnosno, da i iz A, i iz B, postoji brid koji vodi u C). Makar između takvih riječi ne postoji kratak put, one su i dalje često semantički srodne jer dijele zajedničku asocijaciju. Takvu vrstu povezanosti možemo prirodno izraziti pomoću koncepta *preklapajuće asocijativne snage* (engl. overlapping associative strength - OAS), čija je formula slična formuli za MAS (Nelson i sur., 1998):

$$OAS(A, B) = FAS(A, C) \cdot FAS(B, C)$$

Slično kao i za MAS, možemo definirati općenitu formu, u kojoj riječi A i C dijele n različitih susjednih riječi O_1, O_2, \dots, O_n :

$$OAS(A, C) = \sum_{i=1}^n FAS(A, O_i) \cdot FAS(C, O_i)$$



Slika 2. Grafički prikaz općenitog slučaja za OAS riječi A i C

Metode procjene povezanosti među riječima

Procjena stupnja asocijacije između dvije riječi može se provesti na nekoliko načina. Neka istraživanja u tu svrhu koriste *vjerojatnost zajedničkog pojavljivanja* određenog para riječi u jeziku koju procjenjuju prebrojavanjem frekvencije pojave tog para u nekom korpusu. Takav postupak opravdan je pretpostavkom da stvaranje mentalne asocijacije između neka dva događaja ovisi o vremenskoj blizini pojave tih događaja (Spence i Owens, 1990) te pretpostavkom da asocijacije odražavaju način na koji koristimo jezik. Odnosno, ukoliko govornici nekog jezika često čuju ili vide dvije riječi zajedno, te riječi će u njihovom umu postati povezane (asocirane) (Spence i Owens, 1990).

Sljedeća, najrjeđe korištena metoda za procjenu je *metoda subjektivnog ocjenjivanja povezanosti* prikazanih parova riječi. Sudionici, na temelju osobnog iskustva procjenjuju koliko su često vidali par riječi unutar jednog teksta. Na temelju njihovih procjena računa se snaga povezanosti među prezentiranim parovima riječi (Spence i Owens, 1990).

Osim navedenih, važna paradigma za procjenu stupnja asocijacije riječi je spomenuta *metoda slobodnih asocijacija*. Upravo je ta metoda korištena i u ovom istraživanju jer brojna istraživanja potvrđuju njenu visoku pouzdanost za mjerenje snage povezanosti među riječima (Cramer, 1968; Deese, 1965). Prema Nelsonu, McEvoyu i Dennisu (2000), vjerojatnost odgovora B na podražajnu riječ A u kontekstu metode slobodnih asocijacija procjenjuje vjerojatnost da će podražajna riječ A potaknuti misao o riječi B i u kontekstima izvan te metode, i to s visokim prosječnim stupnjem pouzdanosti ($r = 0.89$).

Ova metoda ima nekoliko prednosti nad ostalim navedenim metodama. Naime, prvim dvjema navedenim metodama nemoguće je utvrditi smjer povezanosti između dvije riječi (povezanost postoji bilo da je riječ o povezanosti unaprijed ili povratnoj povezanosti riječi i ne možemo razlikovati ta dva slučaja). Također, visoka povezanost se može dogoditi jer među riječima postoji medijator, no nema načina da utvrdimo je li to slučaj. Procjenom povezanosti metodom subjektivnog ocjenjivanja utvrđeno je i kako procjena ovisi o relativnom poretku prikazanih parova riječi. Npr. parovi riječi umjerene

povezanosti bit će procijenjeni kao nisko povezani ukoliko su prikazani nakon parova riječi koji su procijenjeni kao visoko povezani (Nelson i sur., 1998).

Metodu slobodnih asocijacija možemo dalje podijeliti na *diskretnu* i *kontinuiranu*. Za razliku od diskretne metode, gdje sudionici daju samo jednu asocijaciju na podražaj, u kontinuiranoj metodi sudionici na svaku podražajnu riječ daju više od jedne asocijacije. Problem ove metode je to što je podložna pristranosti, uključujući ulančavanje odgovora i inhibiciju pronalaženja (*engl.* retrieval inhibition). Ulančavanje nastaje kada sudionici umjesto na podražajnu riječ daju asocijacije na prethodni odgovor. Inhibicija pronalaženja se javlja kada sudionici više puta daju asocijaciju na istu podražajnu riječ. Istraživanja na tu temu pokazuju da sudionici često otežano daju naknadne odgovore zbog blokade u pronalaženju (Nelson i McEvoy, 1982).

Korištenjem metode diskretnih slobodnih asocijacija izbjegavamo navedene probleme, no moramo biti svjesni i njenih nedostataka. Budući da se ovom metodom ispituje samo prva asocijacija riječi, možemo pretpostaviti da ovako konstruirane norme podcjenjuju asocijativnu snagu veza nisko povezanih riječi. Dakle, iako norme osiguravaju pouzdani indeks za najsnažnije asocijacije, podcjenjuju snagu vrlo slabih asocijacija (Nelson i sur., 1998).

Postojeće baze slobodnih asocijacija

U većini istraživanja, podatak o povratnoj asocijativnoj snazi (BAS) preuzet je iz normi slobodnih asocijacija koje su kreirali Nelson, McEvoy i Schreiber (2004). Osim BAS-a, njihova baza sadrži i nekolicinu drugih karakteristika parova riječi, uključujući i asocijativnu snagu unaprijed (FAS). U izradi ovih normi sudjelovalo je više od 6000 sudionika, a svaku podražajnu riječ procjenjivalo je od 94 do 206 sudionika. Baza slobodnih asocijacija Nelsona i sur. (1998) trenutno sadrži više od 5000 normiranih riječi. Pojam *normirane* riječi odnosi se na riječi koje su u nekom trenutku istraživanja služile kao podražajne riječi, odnosno riječi na koje su sudionici davali asocijacije (Nelson i sur., 1998). Taj podatak je bitan jer o snazi povratne asocijativne veze (BAS) možemo zaključivati samo za normirane riječi.

Osim baze Nelsona i sur. (1998), važno je spomenuti još jednu od najvećih svjetskih baza slobodnih asocijacija - Edinburgh Associative Thesaurus (EAT; Kiss, Armstrong, Milroy i Piper, 1973). Ona se razvijala u razdoblju između 1968. godine i 1971. godine, a sadržava više od 8000 podražajnih riječi. Također, razvijene su norme za 1424 riječi u nizozemskom jeziku (De Deyne i Storms, 2012) te norme za 1004 portugalskih riječi (Janczura, Castilho, Keller i Oliveira, 2016). Za razvoj nizozemskih normi korištena je metoda kontinuiranih slobodnih asocijacija, dok je za ostale norme korištena metoda diskretnih slobodnih asocijacija.

Ovisnost strukture mentalnih asocijacija o kulturi

S obzirom na postojanje asocijativnih normi na drugim jezicima, potrebno je promotriti problem primjenjivosti normi na druge jezike od onih za koje su konstruirane. Poznato je kako su jezik i kultura vrlo blisko povezani te da pojedinci iz različitih kultura imaju različita lingvistička iskustva. Jezik možemo promatrati kao verbalnu ekspresiju kulture. On je sredstvo za održavanje i prenošenje kulture te pruža mnoge kategorije koje koristimo za izražavanje misli. Po nekim istraživanjima Sapir-Whorf hipoteze (Ahearn, 2011), naše misli su barem u nekoj mjeri pod utjecajem jezika kojim govorimo. Također, pretpostavimo li da mentalne asocijacije između pojmova odražavaju pripadajuće događaje u vanjskom svijetu (npr. istodobnu pojavu dviju riječi u pisanom materijalu), slijedi da će pripadnici različitih kultura imati različite strukture mentalnih asocijacija (Lee, Chiang i Hung, 2008). Iz tog razloga možemo očekivati da će asocijativne mreže riječi imati drukčiju strukturu u različitim jezicima te da norme konstruirane za jedan jezik neće biti prenosive na drugi.

Rosenzweig i Miller (1966) usporedili su asocijacije na pojedine podražaje prikupljene u državama engleskog govornog područja (Australiji, SAD-u, Engleskoj) te u državama u kojima engleski nije primaran jezik (Francuska, Njemačka, Italija). Njihovi rezultati pokazali su da su asocijacije u državama engleskog govornog područja međusobno značajno sličnije (npr. Australija i Engleska dijele 72% zajedničkih asocijacija) od asocijacija engleskog i ne-engleskog govornog područja (npr. SAD i Francuska dijele 48% zajedničkih asocijacija). Također, asocijacije među državama

engleskog jezika značajno su sličnije i od asocijacija između država ne-engleskog govornog područja u kojima se glavni jezik razlikuje (npr. Njemačka i Francuska dijele 45% zajedničkih asocijacija). To nas upućuje na zaključak da je jezik važan faktor u strukturi mentalnih asocijacija.

Jedan od praktičnih problema prikupljanja asocijacija u pojedinim jezicima je infleksija, odnosno modifikacija osnovnog oblika riječi pri izražavanju kategorija poput lica, roda, broja, padeža, vremena i slično (npr. konjugacija glagola, *plešem/pleše/plesati*), što u suštini ne mijenja osnovno značenje riječi. Takve oblike ne želimo smatrati zasebnim riječima jer nas u ovom kontekstu zanima osnovno značenje. U nekim jezicima, poput engleskog, to nije velik problem budući da su relativno flektivno siromašni. S druge strane, hrvatski je primjer flektivno vrlo bogatog jezika (Katz, Rexer i Lukatela, 1991). Jedno od mogućih rješenja tog problema je provođenje postupka lematizacije nad dobivenim podacima prije stvaranja normi. Prema Manning i sur. (2104) lematizacija u lingvistici označava proces inverzan infleksiji, odnosno svođenje riječi na osnovni oblik, njihovu lemu (npr. nominativ jednine za imenice).

U hrvatskom jeziku ne postoji prikladna baza slobodnih asocijacija, kao niti normirane liste riječi za izazivanje lažnog sjećanja. To je bitan faktor koji ograničava istraživanja na ovim područjima. Cilj ovog rada je konstruirati normiranu bazu slobodnih asocijacija na hrvatskom jeziku u svrhu proučavanja asocijativne strukture govornika hrvatskog jezika. Takva baza će se u budućnosti moći primijeniti i u konstrukciji lista riječi u DRM paradigmi.

Cilj

Cilj ovog istraživanja je proučiti asocijativnu strukturu riječi kod govornika hrvatskog jezika što, između ostalog, može poslužiti u svrhu pripreme valjanih verbalnih podražaja za primjenu u područjima psiholingvistike i kognitivne psihologije, primjerice u istraživanjima pamćenja i kognicije.

Problem

Stvoriti norme slobodnih asocijacija za što veći broj hrvatskih riječi, što podrazumijeva izračun snage različitih vrsta povezanosti među riječima (FAS, BAS, MAS i OAS) te drugih značajnih podataka za pojedinu riječ kao što su podaci o tome je li riječ normirana, broj sudionika koji ju je normirao te ukupan broj različitih asocijacija na tu riječ.

Metoda

Sudionici

Istraživanje je provedeno na prigodnom uzorku govornika hrvatskog jezika, starijih od 18 godina. Od ukupnog broja sudionika koji su pristupili istraživanju ($N = 983$), valjane odgovore dalo je njih $N = 973$. Iz daljnje obrade isključeno je četvero sudionika čiji su odgovori bili međusobno gotovo identični (više od 90% istih asocijacija na iste podražaje) zbog sumnje na prepisivanje. Isključeno je i pet sudionika čiji su odgovori bili nevaljani (korištenje jednog vulgarizma kao odgovor na većinu podražaja, korištenje jednog slova kao većinu odgovora i sl.) te jedan sudionik koji je odustao od sudjelovanja. Također, iz daljnje obrade isključena su dva sudionika mlađa od 18 godina. Konačni uzorak sastoji se od 971 osobe u dobi od 18 do 85 godina, s prosjekom dobi od $M = 31.02$, $SD = 14.67$. Tablica s demografskim karakteristikama uzorka priložena je kao Prilog A.

Instrumenti

Sociodemografski upitnik

Od sudionika su prikupljeni osnovni sociodemografski podaci: spol, dob, stupanj obrazovanja, županija u kojoj su proveli najveći dio života, veličina mjesta u kojem su proveli najveći dio života.

Skale asocijacija

Kreirane su dvije verzije skala; skala za primjenu *papir-olovka* metodom te skala za *on-line* primjenu.

Svaka skala kreirana za prikupljanje podataka metodom *papir-olovka* sadržavala je uputu te 10 stranica s listama riječi. Na svakoj stranici nalazilo se po 10 riječi. Stranice u svakom primjerku skale bile su raspoređene nasumično kako bi se minimizirao mogući efekt poretka riječi na asocijacije.

Skala sastavljena za prikupljanje podataka *on-line* metodom sastojala se od upute i 20 cjelina s riječima. Svaka cjelina sadržavala je po 5 riječi.

Računalni program

Za potrebe istraživanja napisan je računalni program u svrhu pripreme podataka za obradu te same obrade podataka. Pomoću programa, svi podaci prikupljeni *on-line* metodom pretvoreni su u isti format u kojem su prethodno ručno uneseni i podaci prikupljeni metodom *papir-olovka*, kako bi se olakšala zajednička obrada. Osim te pretvorbe, program podržava i obavljanje drugih čestih radnji pri pripremi podataka za obradu, kao što su pretraživanje svih odgovora sudionika, ispis dokumenta i broja linije u kojima se tražena asocijacija nalazi, pretvaranje navedenih asocijacija u uniformirani oblik, pretvorbu malog početnog slova u veliko, itd. Konačno, program podržava i obradu podataka, odnosno izračun statistike za demografske podatke sudionika, izračun podataka vezanih za same norme (FAS, BAS, itd.) te vizualizaciju grafa asocijacija.

Program je napisan u programskom jeziku Python (korištena inačica 3.6.2). Sama obrada podataka oslanja se na programsku biblioteku za obradu podataka *pandas* (McKinney, 2010). Vizualizacija grafa asocijacija izvedena je pomoću programske biblioteke *igraph*, odnosno korištenjem pripadajuće Python pomoćne biblioteke *python-igraph* (Csardi i Nepusz, 2006). Graf odnosa ranga asocijacija i prosječnog FAS-a izveden je pomoću Python biblioteke *matplotlib* (Hunter, 2007).

Postupak

Kreiranje početnog fonda riječi

Budući da su u istraživanju Nelsona i sur. (1998) sudionici procjenjivali 100-120 riječi, određeno je da će u ovom istraživanju svaki sudionik procjenjivati 100 riječi kako ne bi došlo do kognitivnog zamora. Konstruiran je početni fond riječi za normiranje koji se sastojao od $N_r = 100$ riječi. Početni fond je, između ostalog, uključivao neke prevedene riječi s lista za lažno sjećanje koje su kreirali Stadler, Roediger i McDermott (1999) te riječi iz neobjavljene baze slobodnih asocijacija koja je kreirana na Odsjeku za psihologiju (2000). Engleske riječi su prevedene na hrvatski jezik od strane dva nezavisna prevoditelja. Prijevodi su se podudarali u svim riječima. Motivacija za ovakav odabir riječi bila je pokušati stvoriti što asociraniji početni skup riječi, s pretpostavkom da će se barem dio povezanosti riječi na engleskom jeziku prenijeti i na povezanost među ekvivalentnim riječima u hrvatskom jeziku. Početne riječi izabrane su iz različitih semantičkih kategorija kako bismo dobili što više različitih skupova međusobno semantički povezanih riječi. Svaka semantička kategorija sadržavala je skup od 5 riječi (*npr. pacijent, bolest, bolnica, kuta, bolesnik*). Biranje riječi iz različitih semantičkih kategorija učinjeno je kako bismo s dobivenim riječima bolje pokrili semantičku mrežu, odnosno dobili riječi što raznovrsnijih značenja. Vjerojatno je da bi se to dogodilo svakako, ali bi proces vjerojatno trajao dulje.

Prikupljanje podataka

Prikupljanje podataka provedeno je u razdoblju od ožujka do lipnja 2017. godine. Prvi dio prikupljanja podataka proveden je metodom papir-olovka, a drugi dio putem Interneta, *web* aplikacijom Google Forms (<https://forms.google.com/>). Svi sudionici bili su zamoljeni da u istraživanju sudjeluju samo jednom. S obzirom da je u upitniku zatraženo nekoliko osobnih podataka, u uputi je naglašeno da je osigurana anonimnost i povjerljivost dobivenih podataka.

Zadatak sudionika bio je za svaku riječ napisati prvu asocijaciju na hrvatskom jeziku. Napomenuto je i da se kao asocijacija može dati bilo koja vrsta riječi (neke od vrsta riječi naknadno smo izbacili, primjerice prijedloge, ali se htjelo izbjeći uvjetovanje

poznavanja naziva za pojedinu vrstu riječi) te da je moguće više puta dati istu asocijaciju. Također, za riječ koja je sudioniku nepoznata, dana je uputa da se podcrta (po uzoru na istraživanje Nelsona i sur, 2004. i De Deyne i Storms, 2008). Kako bismo izbjegli efekt ulančavanja, težilo se da se na svakoj stranici liste za procjenjivanje nalaze riječi koje pripadaju međusobno različitim semantičkim kategorijama. Sudionike se za sudjelovanje nastojalo motivirati naglašavanjem važnosti znanstvenog doprinosa istraživanju. Zadatak je ukupno trajao oko 15 minuta.

Prikupljanje podataka metodom papir-olovka odvijalo se i grupno i individualno pod nadzorom eksperimentatora. Grupno prikupljanje podataka odvijalo se u kontroliranim uvjetima u nekoliko srednjih škola, u jednoj umirovljeničkoj udruzi te u jednoj tvrtki. Prikupljanje ovom metodom odvijalo se u nekoliko etapa:

1. prvim sudionicima (njih 34) dana je na normiranje lista s riječima iz ranije spomenutog fonda od 100 riječi
 2. iz prikupljenih podataka odabrali smo riječi koje su barem dva sudionika ponudili kao asocijaciju te iz njih stvorili tri nove liste od 100 riječi
 3. nove liste su zatim dane na normiranje većem broju sudionika (njih 93)
 4. nad prikupljenim podacima još je jednom proveden sličan postupak odabira novih riječi, sve dok fond riječi za normiranje nije sadržavao više od 500 različitih riječi
- Motivacija za ovakav postupak bila je dobivanje baze od barem 500 normiranih riječi na kraju istraživanja. Ukupno su metodom papir-olovka prikupljeni odgovori 305 sudionika.

Drugi dio prikupljanja podataka proveden je *on-line*, tzv. metodom "snježne grude". Kreirane su četiri skale od kojih je svaka sadržavala 100 različitih podražajnih riječi. Izabrane riječi u ovoj fazi bile su, kao što je već navedeno, najčešće asocijacije sudionika prikupljene prethodnom metodom *papir-olovka*. Sudionicima je putem *e-maila* i socijalnih mreža nasumično poslana jedna od poveznica te su bili zamoljeni da, ukoliko žele, prosljede poveznicu poznanicima. Poveznice su podijeljene i na nekoliko

studijskih grupa te grupa studentskih domova na Facebooku. Program je svakom sudioniku prikazao riječi u nasumičnom redosljedu. Budući da je odaziv sudionika bio veći nego što je očekivano, podražaji u 3 skale naknadno su zamijenjeni s 300 novih podražaja, postupkom koji je ranije opisan (riječima koje su se do tada najčešće pojavljivale kao asocijacije, a nisu još normirane). Ukupno su *on-line* metodom prikupljeni odgovori od 678 sudionika.

Priprema podataka za obradu

Nakon prikupljanja i unošenja svih podataka u računalo, sve asocijacije sudionika pretvorene su u uniformirani oblik, odnosno provedena je lematizacija. Djelomično je učinjena pomoću programa *reldi-tagger* (Ljubešić, 2016), a dio pravopisnih ispravaka (dodavanje dijakritičkih znakova tamo gdje su ih sudionici izostavili) pomoću programa *redi* (Ljubešić, 2016). Sve su riječi pretvorene u nominativ. Imenice su pretvorene u jedninu, osim zbirnih imenica, koje su ostavljene u izvornom obliku zbog svoje specifičnosti (npr. *perje*, *lišće*) te imenica koje postoje samo u množini (pluralia tantum, npr. *hlače*). Množina je zadržana i kod riječi koje se uobičajeno koriste samo u množini, npr. *kokice*, *brkovi* (prema Hrvatskom Jezičnom Portalu). Glagoli su pretvoreni u glagolske imenice. Ukoliko za neki glagol ne postoji pripadajuća glagolska imenica ili se takav oblik u govornom jeziku rijetko koristi, pretvoren je u infinitiv (npr. *tući*, *čuti*, *imati*, *rasti*). Svi su pridjevi pretvoreni u srednji rod, a komparativi i superlativi u pozitiv. Brojevi napisani znamenkama zamijenjeni su riječima. Od zamjenica su u obradu uvrštene samo osobne. Usklici, čestice, veznici, enklitike te prijedlozi nisu uvršteni u obradu. Također, hipokoristici i umanjenice su zamijenjeni s osnovnim oblikom riječi (lema). Iznimke su riječi *kućica*, *korice*, *pahuljice* i *štapići*. Budući da te riječi imaju drugačije značenje u takvom obliku, zadržane su u tom obliku. Vlastita imena, ako su se očigledno odnosila na poznatu ličnost, napisana su u punom obliku (ime i prezime). Iznimka su mononimna imena, odnosno imena osoba koje su prepoznatljive po jednom imenu (npr. Rembrandt), gdje je ime (ili prezime) ostavljeno u tom obliku. Ispravljene su i pravopisne pogreške. Ispravnost pravopisa je provjerena na Hrvatskom jezičnom portalu. Ukoliko je sudionik

kao asocijaciju napisao više riječi, u obzir je uzeta samo prva riječ. U slučaju kada se asocijacija sudionika sastojala od dvije riječi, a sadržavala u sebi i podražajnu riječ, kao odgovor je uzeta riječ koja nije podražajna riječ (npr. podražajna riječ: crno; asocijacija: crno nebo; ispravljeno: nebo). Također, ukoliko je sudionik kao asocijaciju napisao neku određenu boju i dodao riječ boja, odgovor je zbog redundancije ispravljen na naziv boje (npr. asocijacija: crvena boja, ispravljeno: crveno). U slučajevima kada je sudionik naznačio da ne prepoznaje značenje podražajne riječi, ali svejedno napisao asocijaciju na tu riječ, dana asocijacija je izbačena iz daljnje obrade. To je učinjeno jer se nastojalo da sudionik ima mogućnost stvoriti semantičku asocijaciju na zadani podražaj. Asocijacije stranog porijekla koje se uvrštene u standardni hrvatski jezik te asocijacije koje se izrazom razlikuju od standardnog hrvatskog jezika (*potpuni dijalektizmi*) ostavljene su u tom obliku (npr. *vanjkuš - jastuk*). Asocijacije napisane na nestandardnom narječju koje se samo fonološki razlikuju od standardnog jezika (*fonološki dijalektizmi*), pretvorene su u standardni oblik (npr. *lipo - lijepo*). U slučajevima kada je sudionik kao asocijaciju napisao prijedlog i imenicu, u daljnju obradu je uvrštena samo imenica (npr. *na snijegu - snijeg*), budući da prijedlog ne pridonosi semantičkom značenju asocijacije. Ukoliko je navedena asocijacija pojam koji se u hrvatskom jeziku pojavljuje samo u obliku sintagme, u obradu je uvršten u tom obliku, bez obzira na to što nije samo jedna riječ (npr. *dnevni boravak, medicinska sestra*).

U daljnju obradu nisu uvršteni ni idiosinkratski odgovori, koji su definirani kao asocijacije koje su se na zadani podražaj pojavile samo jednom. Nelson i sur. (1998) navode taj kriterij kao vrlo bitan u procjeni (ne)pouzdanosti danog odgovora. Riječi koje su dane kao odgovor na danu riječ barem dvaput, empirijski su se pokazale kao vrlo pouzdane, u smislu reproducibilnosti u novim normiranjima. Riječi koje se pojavljuju samo jednom često i dalje imaju smisla kao asocijacije, ali su slabo asociirane s podražajnom riječi pa ih je teško normirati, pogotovo metodom diskretnih asocijacija (Nelson i sur., 1998). Ujedno se korištenjem ovog kriterija s vrlo visokom vjerojatnošću isključuju i razni "besmisleni" odgovori. Pod tim primarno smatramo zabune, vandalizam, nerazumijevanje podražajne riječi makar to nije naznačeno, asocijacije vezane isključivo za specifična iskustva pojedinog sudionika, itd.

Ispravci su učinjeni nad ukupno 35 229 odgovora.

Obrada

Pri obradi podataka, dana je prednost direktnim vezama među riječima, odnosno par riječi je naveden u konačnoj bazi samo ako među riječima u paru postoji direktna veza.

Iz odgovora sudionika, prvo je konstruiran popis svih parova riječi koje su povezane direktno. Zatim je, za sve takve parove, izračunata asocijativna snaga unaprijed (FAS). Nakon toga su za svaki par dodatno izračunate povratna (BAS) asocijativna snaga, medijacijska asocijativna snaga (MAS) te preklapajuća snaga (OAS). U nekim slučajevima neke od ovih mjera nije moguće izračunati, npr. za par (A, B), ako riječ B nije normirana, nije moguće procijeniti $BAS(A, B)$. U tim je slučajevima polje za tu mjeru ostavljeno prazno.

MAS i OAS izračunati su po uputama Nelsona i suradnika (1998). Izračune ćemo objasniti konkretnim primjerima s riječima iz normi. Riječ KVAKA producira riječ VRATA kao asocijaciju s vjerojatnošću od 0.80, a riječ VRATA pak producira riječ ULAZ s vjerojatnošću od 0.26. Budući da ne postoje druge medijacijske riječi poveznice između KVAKA i ULAZ, medijacijska snaga za ove dvije riječi je $MAS(KVAKA, ULAZ) = FAS(KVAKA, VRATA) * FAS(VRATA, ULAZ) = 0.80 * 0.26 = 0.21$. Općenito, kao što je objašnjeno u uvodu, medijacijska snaga se računa kao zbroj svih ovakvih umnožaka za individualne poveznice između neke dvije riječi (Nelson i sur., 1998).

OAS predstavlja snagu asocijativne povezanosti dvaju riječi kroz zajedničke asocijacije. Drugim riječima, OAS nam daje informaciju o tome u kojoj mjeri dvije riječi proizvode iste asocijacije (Nelson i sur., 1998). Par riječi može imati jednu ili više zajedničkih asocijacija. Npr. i PROVALNIK i TAT asociiraju i na KRAĐA, i na LOPOV. Stoga je $OAS(PROVALNIK, TAT) = FAS(PROVALNIK, KRAĐA) * FAS(TAT, KRAĐA) + FAS(PROVALNIK, LOPOV) * FAS(TAT, LOPOV) = 0.15 * 0.10 + 0.41 * 0.54 = 0.24$.

Nadalje, za svaku je riječ navedeno: 1) je li normirana, 2) broj sudionika koji su normirali pojedinu riječ, 3) ukupan broj asocijacija danih na pojedinu riječ (isključujući idiosinkratske odgovore).

Rezultati i rasprava

Kao što je ranije navedeno, u istraživanju je normirano 700 riječi, a od sudionika je ukupno prikupljeno 94 330 odgovora. Isključimo li idiosinkratske odgovore, dobiveno je 3 113 različitih asocijacija te 11 215 direktnih veza među riječima. Prosječan broj direktnih veza po podražaju je 17.4 te su podražaji normirani u rasponu od 102 do 187 puta. Podražaji su u prosjeku normirani 137 puta.

Tablica 1

Usporedba nekih osnovnih karakteristika USF engleskih normi* te normi utvrđenih ovim istraživanjem

	engleske norme	hrvatske norme
broj normiranih riječi	5019	700
broj različitih asocijacija	8080	3113
broj direktnih veza	72176	11215
prosječan broj direktnih veza po podražaju	14.4	17.4
raspon normiranih podražaja	94 - 206	102 - 187
prosječan broj normiranja podražaja	149	137

Najjača veza dobivena je među riječima *kvaka* i *vrata* te njen FAS iznosi 0.802 (dok je snaga asocijacije iz riječi *vrata* na riječ *kvaka* = 0.067). Drugim riječima, sudionici su na podražaj *kvaka* kao odgovor dali *vrata* u 80% slučajeva, a obrnuto se dogodilo u samo 6.7% slučajeva. Minimalni izračunati FAS iznosi 0.011 (npr. između riječi *brak* - *čistoća*). Najveći MAS iznosi 0.246, a utvrđen je među riječima *metropola* i *grad*. Iz Tablice 2 možemo primijetiti kako minimalni MAS ne iznosi 0, što znači da za

* Nelson i sur. (1998)

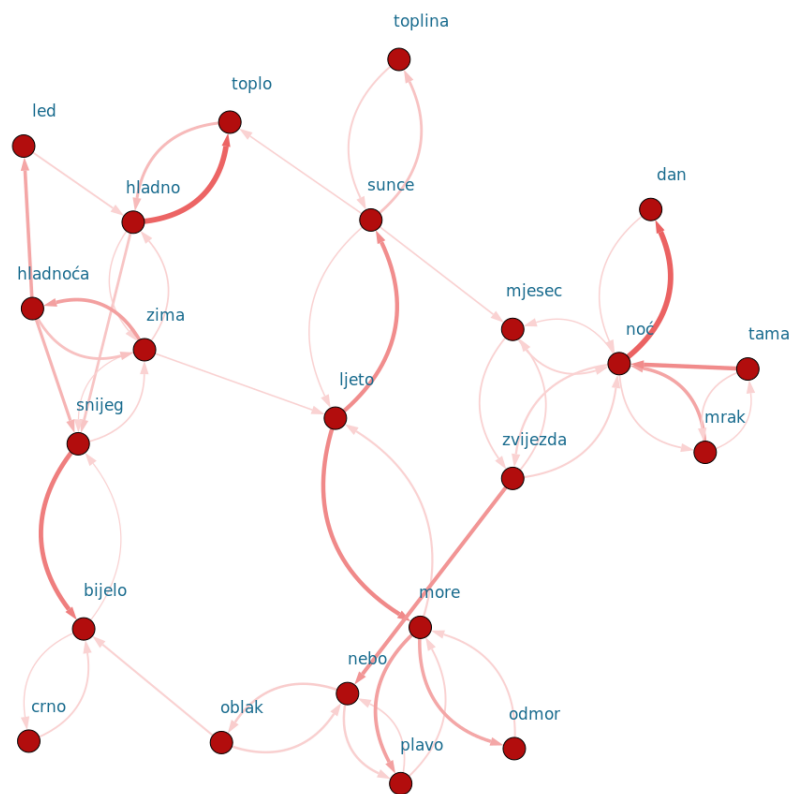
svaki neposredno povezani par riječi postoji barem jedna posredna veza (veza preko barem jedne normirane riječi). OAS je najjači među riječima *pruga* i *tračnica*, iznosi 0.416, a to je ujedno i par riječi koji se pojavio najveći broj puta, odnosno 120 puta.

Tablica 2

Deskriptivne karakteristike podataka utvrđenih u istraživanju slobodnih asocijacija za hrvatski jezik, ($N = 971$)

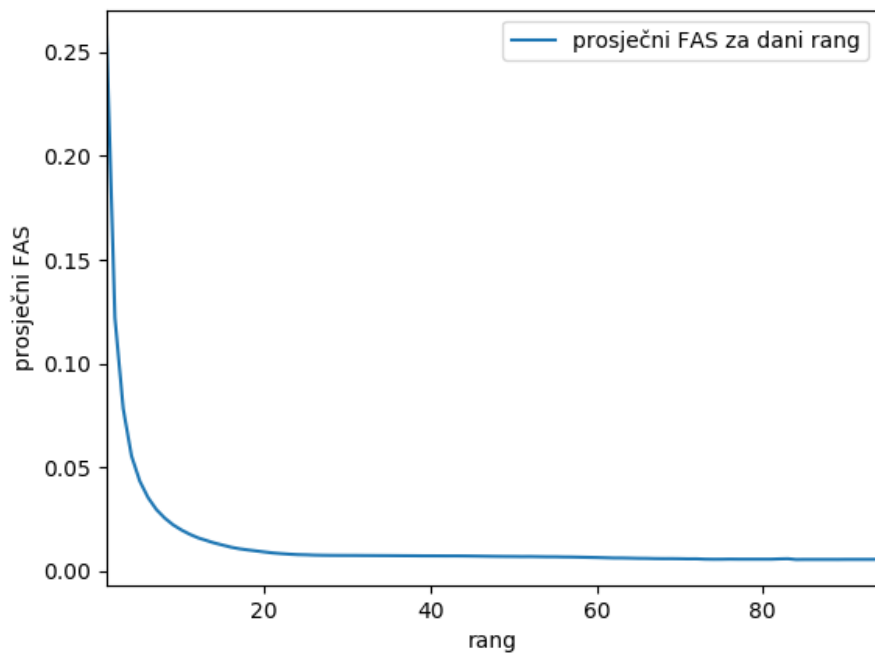
	<i>M</i>	<i>C</i>	<i>SD</i>	<i>min</i>	<i>max</i>
FAS	.049	.024	.071	.011	.802
BAS	.040	.007	.084	0	.802
MAS	.010	.003	.018	.00003	.246
OAS	.016	.006	.028	.00003	.416
broj asocijacija	17.436	17	4.826	3	34
broj normiranja	137.371	132	24.827	102	187

Prikupljene riječi skupa s izračunatim asocijativnim snagama unaprijed tvore usmjereni težinski graf koji ocrta asocijativnu strukturu riječi. Svaki krug u grafu na Slici 3 označava riječ. Crvenom bojom su označene normirane, a plavom bojom nenormirane riječi. Svaka strelica $A \rightarrow B$ prikazuje da je na podražaj A dan odgovor B kao asocijacija. Debljina strelice prikazuje snagu povezanosti među riječima. Ukupno, graf sadrži 3 113 različitih riječi te 11 215 direktnih veza među riječima.



Slika 4. Najveća povezana komponenta za bridove s FAS ≥ 0.1

Nadalje, promotren je odnos ranga asocijacija i prosječne asocijativne snage unaprijed. Iz Slike 5 možemo vidjeti kako dobivena krivulja ima oblik opadajućeg zakona potencije. Ovakav oblik karakterističan je za Zipfovnu distribuciju koju često empirijski pronalazimo u istraživanjima vezanim uz korpuse riječi. Drugim riječima, asocijacija ranga n je približno n puta rjeđa nego najjača asocijacija. Slična krivulja utvrđena je za portugalski (Janczura, Castilho, Keller i Oliveira, 2016) te engleski jezik te (Nelson i Schreiber, 1992).



Slika 5. Odnos ranga asocijacija i prosječne asocijativne snage unaprijed ($N_{\text{riječi}} = 700$)

Za usporedbu, u Tablici 3 navodimo prosječne asocijativne snage unaprijed za prvih 6 rangova u ovom istraživanju te istraživanju Janczura i suradnika (2016). Zanimljivo je kako su utvrđene vrijednosti vrlo slične, kao i standardne devijacije, što može ukazivati na neku univerzalnu strukturu, neovisnu o jeziku.

Tablica 3

Usporedba prosječnih asocijativnih snaga unaprijed prvih 6 rangova u istraživanju normi za portugalski jezik* te vrijednosti utvrđene ovim istraživanjem

rang	FAS	
	hrvatski jezik M(SD)	portugalski jezik M(SD)
1.	.26 (.13)	.23 (.13)
2.	.12 (.05)	.11 (.05)
3.	.08 (.03)	.07 (.03)
4.	.06 (.02)	.05 (.01)
5.	.04 (.01)	.04 (.01)
6.	.04 (.01)	.03 (.01)

* Janczura, Castilho, Keller, Oliveira (2016)

Nadalje, izračunate su korelacije između FAS, BAS, MAS, OAS te broja asocijacija. Intuitivno, očekivana je negativna korelacija između FAS i broja asocijacija jer što neka riječ ima više jakih asocijacija, to je teže proizvesti, a time i ispitati, slabije asocijacije. To pogotovo vrijedi za metodu diskretnih asocijacija (naspram metode kontinuiranih asocijacija) koja je korištena u ovom istraživanju. Jedini podatak o korelaciji među mjerama kojeg navode Nelson i sur. (1998) je korelacija između FAS i BAS, koja iznosi $r = .29$, dok je u našem istraživanju $r = .37$. Dakle, u oba slučaja korelacija je pozitivna, ali niska.

Tablica 4

Matrica interkorelacija FAS, BAS, MAS, OAS i broja asocijacija u istraživanju slobodnih asocijacija za hrvatski jezik ($N = 971$)

	FAS	BAS	MAS	OAS	broj asocijacija
FAS	1	.369	.226	.043	-.221
BAS	.369	1	-.035	.068	-.105
MAS	.226	-.035	1	.482	-.219
OAS	.043	.068	.482	1	-.210
broj asocijacija	-.221	-.105	-.219	-.210	1

Mogućnost primjene normi u istraživanjima lažnih sjećanja

U kontekstu primjene asocijativnih normi u istraživanjima lažnih sjećanja, Roediger, Watson, McDermott i Gallo (2001) navode BAS između podražaja i kritične riječi kao važan prediktor frekvencije lažnog dosjećanja DRM paradigmom. Također, eksperimentalno je utvrđeno kako je ukupna povratna asocijativna snaga (*engl.* total backward associative strength - TBAS) liste kreirane za lažno dosjećanje najvažniji prediktor proporcija lažnog sjećanja (Robinson i Roediger, 1997). Istraživanjem su utvrdili kako povećanjem broja asocijativnih riječi u listi (povećanje TBAS) raste broj lažnih sjećanja. Međutim, dodavanjem riječi koje nisu bile asocijativno povezane s kritičnom riječi (održavanje TBAS približno konstantnim) u ovom istraživanju nisu utjecali na proporciju lažnih sjećanja (Robinson i Roediger, 1997). Drugim riječima, povećanje broja asocijativno vezanih riječi u listi povećat će vjerojatnost lažnog sjećanja kritične riječi. U Tablici 5 navedeni su podaci 10 najjačih BAS povezanosti dobivenih našim istraživanjem. Smatramo kako bi neki od navedenih podražaja vrlo vjerojatno mogli predstavljati riječi mamce u izazivanju lažnih sjećanja DRM paradigmom.

Tablica 5

Prikaz najjačih povratnih snaga među parovima riječi izračunatih u istraživanju slobodnih asocijacija za hrvatski jezik, ($N = 971$)

podražaj	asocijacija	povratna snaga
vrata	kvaka	.801
zatvor	rešetka	.747
olovka	šiljilo	.728
voda	žeđ	.695
zatvor	Remetinec	.671
olovka	kemijska	.665
jaje	kokoš	.657
kiša	kišobran	.656
vlak	tračnica	.652
miris	parfem	.648

U postupku kreiranja podražaja za normiranje, potkrala se pogreška te je, osim

riječi "sjedenje", normirana i riječ "sjediti". Pogrešku smo iskoristili za promatranje razlike u asocijacijama između te dvije riječi. Zanimljivo je da obje riječi dijele istu najjaču asocijaciju - "stolica" ($FAS(\text{sjedenje, stolica}) = 0,127$, $FAS(\text{sjediti, stolica}) = 0,150$). Prvih 10 najjačih asocijacija za riječ "sjediti", nalaze se i na listi za "sjedenje" (dane su kao asocijacija od najmanje dva sudionika), dok se prvih 5 najjačih asocijacija na "sjedenje", nalazi na listi od "sjediti".

U prilogu se nalaze bruto odgovori svih sudionika (i njihovi ispravci nakon lematizacije/uniformizacije), obrađeni podaci o svim parovima podražaj-asocijacija (cue-target) koji su se pojavili u odgovorima te podaci o prosječnom FAS-u prema rangu asocijacija. Podacima se može pristupiti na adresi: <https://dataverse.ffzg.unizg.hr/>.

Metodološki nedostaci i smjernice za buduća istraživanja

Potrebno je osvrnuti se na nedostatke ovog istraživanja. Prvi nedostatak vezan je uz samu metodu prikupljanja podataka, odnosno metodu diskretnih asocijacija. Iako su prethodno navedeni valjani razlozi za restrikciju sudionika na jednu asocijaciju po podražaju, bitno je naglasiti da je metoda kontinuiranih asocijacija značajna za prikupljanje slabijih asocijacija koje je teže ili nemoguće prikupiti korištenom metodom. Glavna prednost metode kontinuiranih asocijacija je ta što omogućuje veći varijabilitet asocijacija, odnosno ima veću osjetljivost. Prikupljanje višestrukih odgovora na podražaje omogućuje bolju procjenu vjerojatnosti povezanosti kod slabijih asocijacija. Osim toga, razlike u prvim i kasnijim asocijacijama na podražaj mogu pružiti dodatan uvid u konceptualnu strukturu i procese koji su u podlozi generiranja asocijacija (Barsalou, Santos, Simmons i Wilson, 2008). Zbog navedenih razloga, kako bi se dobio uvid u nove podatke, buduća bi istraživanja mogla koristiti metodu kontinuiranih asocijacija.

Potencijalni problem valjanosti u istraživanju je pretvaranje svih pridjeva različitog roda u isti rod. Pridjevi u hrvatskom jeziku razlikuju se od onih u engleskom po tome što se mogu nalaziti u muškom, ženskom te srednjem rodu, dok pridjevi u engleskom jeziku nemaju rodove. Postoji mogućnost da asocijacije sudionika češće budu istog roda kao podražajni pridjev, nego da se po rodu razlikuju. Dakle, korištenje

pridjeva u jednom rodu (u ovom slučaju srednjem) moglo je dovesti do toga da sudionici na pridjeve srednjeg roda daju asocijacije u srednjem rodu češće od asocijacija u ženskom ili muškom rodu. Kao što je već rečeno, pretvaranje u jedan rod učinjeno je isključivo zbog lakše provedbe obrade riječi. Zadržavanje sva tri roda pridjeva bilo bi nepraktično, budući da bismo zbog ograničenih resursa u tom slučaju mogli normirati manji broj različitih riječi. Srednji rod je odabran zbog toga što smo taj rod u hrvatskom jeziku smatrali najneutralnijim, odnosno smatrali smo da će pristranost u davanju asocijacija biti najmanja koristimo li taj rod. Budući da o tome, koliko nam je poznato, nema prijašnjih istraživanja, a rodovi su u hrvatskom jeziku vrlo zastupljeni, bilo bi zanimljivo to pitanje istražiti u budućnosti. Također, bilo bi zanimljivo provjeriti ovisi li skup dobivenih asocijacija nekog pridjeva o rodu u kojem je podražaj prezentiran sudioniku.

Nadalje, prilikom prikupljanja odgovora metodom *papir-olovka*, sudionicima je prikazano 10 podražaja po stranici, za razliku od 5 podražaja po stranici u sklopu *on-line* metode. Naveli bismo to kao potencijalni nedostatak, budući da je postojanje više od jednog podražaja unutar vidnog polja tijekom ispunjavanja skale teoretski moglo utjecati na asocijacije sudionika. Ipak, tako je učinjeno iz praktičnih razloga i potencijalnog pada motivacije kod sudionika suočenih s velikim brojem stranica. Kako bi se potencijalne posljedice ovog problema minimizirale, nastojalo se da su podražaji međusobno fizički razmaknuti što je više moguće. Također, nastojalo se da fizički bliski podražaji na jednoj stranici materijala budu odabrani iz različitih semantičkih kategorija, budući da je kod manje semantičke sličnosti podražaja manja šansa za izazivanje efekta ulančavanja (Xu, Regier i Malt, 2015).

Kao što je ranije spomenuto, po jednoj stranici papira ispisano je 10 podražaja. Kako bi se minimizirao efekt poretka podražaja na asocijacije, svaki primjerak skale je uvezan s nasumičnim rasporedom stranica. Zbog lakšeg umnožavanja papirnatih materijala, redosljed podražaja na pojedinoj stranici nije varirao. To je također jedan od nedostataka, budući da ne možemo sa sigurnošću reći da raspored podražaja na papiru nije utjecao na asocijacije sudionika. U *on-line* verziji skale navedeni je problem izbjegnuto budući da se za svakog sudionika nasumično odabrao raspored podražaja na stranici.

Jedan od propusta koji je učinjen tijekom prikupljanja podataka je što nije provjereno koliko dobro sudionici znaju hrvatski jezik. Budući da je samo 5 sudionika naznačilo da ne prepoznaje više od 5 riječi u listi, možemo pretpostaviti da u istraživanju nije sudjelovao značajan broj ljudi koji nemaju zadovoljavajuće znanje hrvatskog jezika. Ipak, u budućim istraživanjima ovog tipa bilo bi dobro kontrolirati i taj faktor.

Treba uzeti u obzir i potencijalne pristranosti u uzorku. Primjerice, u prikupljenom uzorku radi se pretežno o odgovorima mlađih ljudi ($M = 31.02$), s više žena nego muškaraca. Budući da asocijacije mogu varirati u ovisnosti o dobi ili spolu osobe, moguće je da norme imaju pristranost u tom pogledu. Međutim, na asocijacije mogu potencijalno utjecati i mnogi drugi faktori, poput stupnja obrazovanja, zanimanja, interesa, ličnosti, itd., što može biti nepraktično kontrolirati. Imajući to na umu, naglasak u ovom istraživanju bio je u prikupljanju što većeg broja podataka, ali pritom bilježeći najbitnije demografske podatke za svakog sudionika. Takav postupak bilježenja učinjen je kako bi naknadno postojeću bazu bilo lako proširiti s novim sudionicima, a i kako bi se olakšalo kreiranje specifičnih normi za pojedinu skupinu, ukoliko je to potrebno. U budućim istraživanjima, bilo bi zanimljivo istražiti koliko koji od navedenih faktora utječe na distribuciju asocijacija.

Također, nedostatak povezan s temporalnom valjanosti je provedba prikupljanja podataka unutar razdoblja od 3 mjeseca. Istraživanje Jenkinsa i Palerma (1965) pokazuje kako se asocijativne norme sustavno mijenjaju u funkciji vremena. Međutim, visoko povezane riječi najmanje su podložne promjenama. Stoga, budući da su neke asocijacije vremenski ovisne, za ovakav bi tip istraživanja bilo dobro periodički osvježavati podatke i prikupljati ih kroz dulje vremensko razdoblje kako bi se minimizirao utjecaj vremena.

Zaključak

Ovim istraživanjem nastojali smo proučiti asocijativnu strukturu riječi kod govornika hrvatskog jezika, kreiranjem normi slobodnih asocijacija za hrvatski jezik. Početno očekivanje bilo je stvoriti norme za barem 500 hrvatskih riječi, no zbog velikog

odaziva sudionika stvorene su norme za 700 riječi. Kao rezultat, norme sadržavaju 3113 različitih asocijacija te 11215 direktnih veza među riječima. Može im se pristupiti na adresi: <https://dataverse.ffzg.unizg.hr/>. Utvrđene norme analizirane su vizualizacijom grafa kojeg tvore i osnovnom deskriptivnom statistikom te uspoređene s rezultatima iz srodnih istraživanja u drugim jezicima kako bi se uvjerali da su rezultati usporedivi. Smatramo da smo dobili zadovoljavajući broj različitih asocijacija te raznovrsnost riječi po semantičkim kategorijama. Smatramo da će norme biti prikladne za konstrukciju lista riječi za korištenje u DRM paradigmi. Na teoretskoj razini, norme mogu pridonijeti mapiranju dijeljenog asocijativnog leksičkog znanja stečenog kroz svakodnevni kontakt s riječima. Ovakvo mapiranje može biti korisno za istraživanja pamćenja, jezika, logičkog rezoniranja i ostalih kognitivnih procesa.

Literatura

Ahearn, L. M. (2011). *Living Language: An Introduction to Linguistic Anthropology*. John Wiley & Sons.

Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale, NJ: Erlbaum.

Barsalou, L. W., Santos, A., Simmons W., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, embodiment, and meaning: Debates on meaning and cognition*. Oxford, UK: Oxford University Press.

Boring E. G. (1950). *A History of Experimental Psychology*. New York, NY: Appleton-Century-Crofts.

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

Csardi G., Nepusz T. (2006). The igraph software package for complex network research. *InterJournal of Complex Systems*, 1695. <http://igraph.org>

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, 40, 198 - 205.

Deese, J. (1959a). Influence of interitem associative strength upon immediate free recall. *Psychological Reports*, 5, 235-241.

Deese, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17-22.

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, MD: Johns Hopkins Press.

- Hrvatski jezični portal (2017), <http://hjp.znanje.hr/> Pristupljeno: 18. srpanj 2017.
- Hunter, J. D. (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9, 90-95. DOI:10.1109/MCSE.2007.55
- Janczura, G. A., Castilho, G. M., Keller, V. N. & Oliveira, N. R. (2016). Free Association Norms for 1004 Brazilian Portuguese Words. *Psicologia: Teoria e Pesquisa*, 32 [on-line] (<http://dx.doi.org/10.1590/0102-3772e32ne23>)
- Jenkins, J. J., & Palermo, D. S. (1965). Further data on changes in word-association norms. *Journal of Personality & Social Psychology*, 1, 303–309.
- Katz, L. Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research*, 53, 25–32.
- Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. (1973). *An associative thesaurus of English and its computer analysis*. In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), *The Computer and Literary Studies*. Edinburgh, United Kingdom: University Press.
- Lee, Y., Chiang, W. C., & Hung, H. C. (2008). Lexical association and false memory for words in two cultures. *Journal of Psycholinguistic Research*, 37, 49-58.
- Ljubešić, N. (2016). redi (git commit 7769e644ca7ce223cae36482a7e525f6931dc60a) [računalni program]. Dostupno na: <https://github.com/clarinsi/redi>
- Ljubešić, N. (2016). reldi-tagger (git commit 7eb73061ba262c357a79f4cb6776925dfd3dd195) [računalni program]. Dostupno na: <https://github.com/clarinsi/reldi-tagger>

Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., & McClosky, David (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60.

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference, 51-56. <http://pandas.pydata.org/>

Nelson, D. L., McEvoy, C. L., & Friedrich, M. A. (1982). Extralist cuing and retrieval inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, 8, 89-105.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>. Pristupljeno: 4. srpnja 2017.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402-407.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, 28, 887-899.

Neobjavljena baza slobodnih asocijacija, Odsjek za psihologiju, 2000.

Pezdek, K., Lam, S. (2007). What research paradigms have cognitive psychologists used to study "False memory" and what are the implications of these choices? *Consciousness and Cognition*, 16, 2-17.

Python Software Foundation, Python (3.6.2) [računalni program]. Dostupno na:

<https://www.python.org>

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231–237.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803–814.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407.

Rosenzweig, M. & Miller, K. (1966). Comparisons of word association responses obtained in the United States, Australia, and England. *Journal of Verbal Learning and Verbal Behavior*, *5*, 35-41.

Schwartz, B. L. (2014). *Memory: Foundations and Applications*. Sage Publishers: Thousand Oaks, CA.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*, 317–330.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494–500.

Steyvers, M., Shiffrin, R. M. & Nelson, D. L. (2004). Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory. In A. Healy (Ed.) *Experimental Cognitive Psychology and its Applications: Decade of Behavior*. 237-249.

Steyvers, M., & Tenenbaum, J. (2005). The Large Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29, 41-78.

Xu, Y., Regier, T. & Malt, B. (2015). Historical Semantic Chaining and Efficient Communication: The Case of Container Names. *Cognitive science*, 40, 2081-2094.

Zarevski, P. (1997). *Psihologija pamćenja i učenja*. Jastrebarsko: Naklada Slap.

Prilog A

Tablica 6

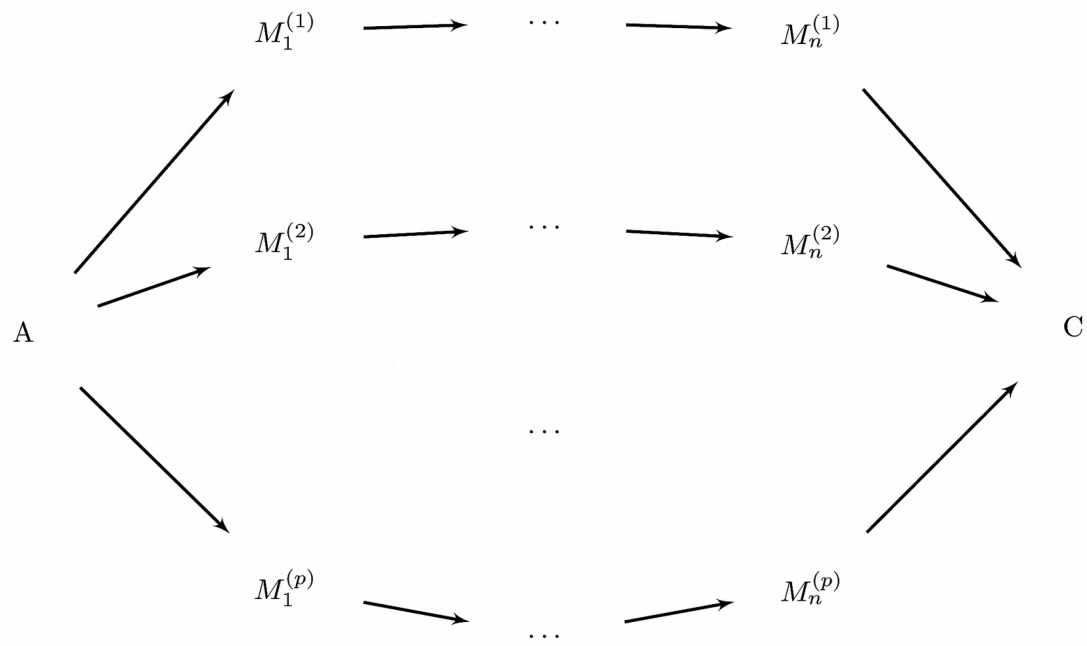
Demografske karakteristike uzorka u istraživanju slobodnih asocijacija
za hrvatski jezik, ($N = 971$)

	frekvencija	postotak
Spol		
M	249	25,64%
Ž	721	74,25%
nepoznato	1	0,1%
Dob		
18 – 27	609	62,72%
28 – 37	118	12,15%
38 – 47	85	8,75%
48 – 57	77	7,93%
58 – 67	43	4,43%
68 – 77	29	2,99%
78 – 87	7	0,72%
nepoznato	2	0,21%
Obrazovanje		
osnovna škola	40	4,12%
srednja škola	475	48,92%
viša stručna sprema/prvostupnik	248	25,54%
visoka stručna sprema	208	21,42%
Regija*		
Sjeverozapadna Hrvatska	386	39,75%
Srednja i Istočna Hrvatska	444	45,73%
Jadranska Hrvatska	141	14,52%
Veličina mjesta**		
selo	175	18,02%
manje mjesto (< 10 000 st.)	234	24,10%
manji grad (< 100 000 st.)	279	28,73%
grad (< 500 000 st.)	46	4,74%
veliki grad (> 500 000 st.)	237	24,41%

*Regija = područje u koje pripada županija u kojoj je sudionik proveo najveći dio svog života (u sirovim podacima je uz svakog sudionika navedena županija)

**Veličina mjesta = veličina mjesta u kojem je sudionik proveo najveći dio života; st. = stanovnika

Prilog B



Slika 6. Grafički prikaz općenitog slučaja za n -stupanjski MAS za riječi A i C