# Potencial množičenja v sodobni leksikografiji

*Darja Fišer in Jaka Čibej*

**Abstract**
Owing to increasing volumes of linguistic data and time constraints, the nature of lexicographic work has changed significantly in the past two decades. A number of steps in the dictionary production process have already been automated, but the algorithms developed are still far from perfect. Dictionary construction therefore still involves a number of routine but time-consuming and expensive manual procedures, for which experienced lexicographers are overqualified. This is why contemporary lexicography has started to explore options such as crowdsourcing, which save both time and financial resources without reducing the quality of the results by taking into account the key principles of microtask design and campaign management. This allows lexicographers to devote all their energy to expert work on the dictionary. This paper provides an overview of language resources that were successfully crowdsourced, the main aspects and characteristics of crowdsourcing, and the quality control mechanisms that ensure the success of this innovative method, which, if implemented correctly, could have a lasting impact on the overall workflow of lexicographic projects as well as the use and life cycle of lexicographic products.

**Keywords:** crowdsourcing, microtask design, crowd motivation, quality control, legal and ethical aspects of crowdsourcing

**Ključne besede**: množičenje, oblikovanje mikronalog, motivacija množičnikov, zagotavljanje kakovosti množičenja, pravno-etični vidiki množičenja

# 1 UVOD IN OPREDELITEV KLJUČNIH POJMOV

V zadnjem desetletju so se z razmahom spleta in ob vse večji meri digitalizacije dela pojavile številne oblike spletnega sodelovanja, pri katerih uporabniki na različne načine prispevajo k uresničitvi skupnega projekta. Poleg odprtokodnih projektov (npr. Linux) in kolaborativnih iniciativ (npr. Wikipedija) med nove oblike dela spada tudi množičenje[1] (angl. *crowdsourcing*), ki označuje postopek, pri katerem skupina ljudi (množica, angl. *crowd*) prispeva k doseganju določenega cilja, in sicer tako, da se delo razdeli med posameznike, od katerih vsak opravi manjši, obvladljiv del, ki ne zahteva veliko truda in časa, vsi prispevani deli, združeni v celoto, pa predstavljajo znaten dosežek (Howe 2008). Pri tem je pomembno izpostaviti, da množice ne sestavljajo nujno strokovnjaki z določenega področja, saj so številni projekti z uporabo množičenja pokazali, da ob ustrezni podpori in pripravi nalog tudi laiki zmorejo opravljati naloge, ki so bili doslej v izključni domeni ekspertov. Zaradi sodobne tehnologije in globalne razširjenosti spleta postaja izkoriščanje potenciala množičenja vse preprostejše, ugodnejše in učinkovitejše, o čemer pričajo uspehi številnih podjetij, ki uporabljajo moč množic za reševanje problemov, izboljšanje storitev ali ustvarjanje izdelkov (npr. Threadless, iStockPhoto in Procter & Gamble).

Angleški izraz crowdsourcing je vpeljal Jeff Howe leta 2006. Ker gre za relativno nov koncept dela, ki se pojavlja v več različicah in po mnenju nekaterih zajema tudi več oblik uporabniškega prispevanja,[2] še vedno ni povsem enotno definiran. Estellés-Arolas in González-Ladrón-de-Guevara (2012) izpostavita, da se različne definicije množičenja razlikujejo predvsem v obsegu dela, ki ga definirajo kot množičenje – nekatere namreč zajemajo praktično vso spletno sodelovanje, tudi soustvarjanje (angl. *co-creation*) in uporabniške inovacije (angl. *user innovation*). Zato z luščenjem definicij iz relevantne literature izdelata enotno definicijo, ki učinkovito loči množičenje od ostalih dejavnosti:

> Množičenje je vrsta spletnega sodelovanja, pri katerem posameznik, institucija, neprofitna organizacija ali podjetje s pomočjo javnega razpisa ali vabila skupini posameznikov z različnimi znanji, velikostjo in stopnjo heterogenosti predlaga prostovoljno opravljanje določene naloge. Opravljanje teh nalog, ki se lahko razlikujejo po težavnosti in načinu dela in pri katerih množica sodeluje z delom, denarjem, znanjem in/ali izkušnjami, vedno prinaša korist obema stranema. Uporabnik bo z opravljanjem naloge potešil določeno potrebo, npr. po zaslužku, družbenem odobravanju, dvigu samozavesti ali razvoju svojih sposobnosti, pobudnik množičenja pa bo lahko pridobil rezultat dela in ga izkoristil (ibid.: 9–10) [prevod: J. Č.].

1 V prispevku za angleški izraz crowdsourcing uporabljamo izraz množičenje, ki ponuja tudi številne druge izpeljanke, npr. množičiti (angl. to *crowdsource*), množičnik/množičnica (angl. *crowdsourcer*), množičeni projekti (angl. *crowdsourced projects*).

2 V angleščini obstajata npr. tudi izraza *crowdfunding* (množično financiranje) in *crowdvoting* (množično glasovanje), ki ju nekateri obravnavajo kot podpomenki množičenja. Obenem nekateri avtorji pojem *crowdsourcing* obravnavajo kot nadpomenko za vse vrste spletnega sodelovanja.

Vsaka vrsta spletnega sodelovanja torej ne spada nujno v kategorijo množičenja, za katerega je ključno, da je izpolnjenih več kriterijev: prisoten mora biti pobudnik (podjetje, organizacija ali posameznik), ki posameznike povabi k opravljanju določene (mikro)naloge, pri čemer imajo od tega korist tako posamezniki, ki so deležni bodisi finančne nagrade bodisi spodbude v drugi (lahko tudi nematerialni) obliki, kot seveda tudi pobudnik, ki lahko rezultat množičenja uporabi pri nadaljnjem delu.

Na področju leksikografije je od omenjenih oblik uporabniškega prispevanja najbolj razširjena kolaborativna leksikografija, pod katero spadajo znani projekti, kot so Wikislovar,[3] Urban Dictionary,[4] Folkets lexikon[5] in slovenski Razvezani jezik.[6] Uporabniški prispevki h gradnji slovarjev so danes večinoma omejeni na kolaborativne leksikografski projekte, kot so prispevanje morebitnih slovarskih iztočnic in primerov ali pa na popravljanje slovarja po izdaji, ugotavljata Abel in Meyer (2013). Hkrati pa so se okoliščine v sodobni leksikografiji tako spremenile, da se leksikografi pri svojem delu soočajo z vse večjimi časovnimi omejitvami in količinami podatkov, zato je vse več leksikografskih postopkov (pol)avtomatskih. Nekatere stopnje gradnje slovarjev se zato spreminjajo v rutinska opravila, za katera so leksikografi prekvalificirani. Ob tem se ponuja priložnost, da v leksikografsko delo vključimo uporabniške prispevke v obliki množičenja – ne kot glavno fazo izgradnje slovarja, temveč kot način za filtriranje, obdelavo in čiščenje (avtomatsko luščenih) podatkov, ki nato leksikografom omogočijo hitrejšo izdelavo geselskega članka.

Čeprav je zadnjih nekaj let v leksikografskih razpravah vse več govora o množičenju, metoda še ni bila zadostno preizkušena na obsežnih, raznolikih leksikografskih projektih. V prispevku zato predstavljamo domače in tuje primere dobrih praks, različne vidike uporabe množičenja in načine, kako zagotoviti uspešnost te inovativne metode, ki bi ob ustrezni implementaciji lahko trajno vplivala na izvedbo leksikografskih projektov ter na rabo in življenjsko dobo leksikografskih izdelkov.

## 2 PREGLED UPORABE MNOŽIČENJA ZA PRIDOBIVANJE JEZIKOVNIH PODATKOV

V tem razdelku predstavljamo pregled sorodnih raziskav in projektov z različnih področij obdelave naravnih jezikov, ki so uspešno uporabili množičenje.

3 http://sl.wiktionary.org/ (dostop 8. 8. 2015).
4 http://www.urbandictionary.com/ (dostop 8. 8. 2015).
5 http://folkets-lexikon.csc.kth.se/ (dostop 8. 8. 2015).
6 http://razvezanijezik.org/ (dostop 8. 8. 2015).

## 2.1 Množičenje pri izdelavi in označevanju jezikovnih virov

Klubička in Ljubešić (2014) sta s pomočjo množičenja izdelala oblikoskladenjsko označen in lematiziran korpus hrvaščine, ki ga bo pozneje mogoče uporabiti kot učno množico. Evalvacija množičenja je pokazala, da je bila natančnost posameznega množičnika v povprečju 90 %, natančnost večinskega glasu treh množičnikov pa približno 97 %.

Venhuizen et al. (2013) so s pomočjo spletne aplikacije Wordrobe[7] množičnikom v reševanje ponudili naloge za določanje besedne vrste prekrivnih besednih oblik, kategorizacijo lastnih imen in označevanje pomena večpomenskih besed za razvoj in jezikoslovno označevanje angleškega korpusa Groningen Meaning Bank.[8] Rezultati množičenja so že ob majhnem številu množičenih podatkov dosegli visoko natančnost glede na zlati standard.

Rumshisky (2011) in Rumshisky et al. (2012) so z množičenjem na platformi Amazon Mechanical Turk pridobili pomensko označen korpus in pomenski leksikon za angleščino, ki so ju označili materni govorci – nestrokovnjaki. Rezultati kažejo, da tudi tak izdelek dosega kakovostni nivo, kakršnega bi dosegli strokovnjaki, označevanje pa je obenem zelo hitro in ekonomično, saj je kompleksno označevanje razdeljeno na več preprostejših korakov, ki jih zmorejo tudi naključni uporabniki platforme Amazon Mechanical Turk.

Fossati et al. (2013) so platformo CrowdFlower uporabili za označevanje semantičnih vlog v angleških besedilih. Metodo so primerjali s standardno metodologijo označevanja in ugotovili, da množičenje, ki temelji na označevanju po več (enostavnejših) korakih, dosega boljše rezultate kot standardno označevanje, saj je natančnejše in obenem tudi hitrejše.

## 2.2 Množičenje za jezikovne tehnologije

Eden najbolj tipičnih jezikovnotehnoloških nalog, ki jih znanstveniki razrešujejo s pomočjo množičenja, je strojno prevajanje, pri čemer množico uporabljajo za različne namene. Zaidan in Callison-Burch (2011) množičnike najemata za izbiro najboljšega strojnega prevoda med več ponujenimi kandidati in v eksperimentih pokažeta, da množičniki dosegajo kvaliteto, ki je primerljiva s profesionalnimi prevajalci. Z množičenjem je mogoče uspešno opraviti tudi evalvacijo strojnih prevajalnikov (Bentivogli et al. 2011; Denkowski in Lavie 2010), izvajati

7 http://wordrobe.housing.rug.nl/ (dostop 8. 8. 2015).
8 http://gmb.let.rug.nl/ (dostop 8. 8. 2015).

zanesljivo besedno poravnavo vzporednih korpusov (Gao in Vogel 2010) in izdelati kvalitetne učne korpuse za statistične strojne prevajalnike (Negri et al. 2011).

Chamberlain et al. (2008) z množičenjem uspešno zbirajo podatke za razreševanje sklicev v angleščini, in sicer s pomočjo spletne igre Phrase Detectives,[9] pri kateri množičniki označujejo besede ali besedne zveze, na katere se nanašajo zaimenske oblike. Na voljo je tudi različica na družbenem omrežju Facebook.

Platformo za množičenje Amazon Mechanical Turk so uporabili tudi Snow et al. (2008), med drugim za analizo sentimenta v naslovih angleških časopisov. Pri ocenjevanju podatkov, ki so jih označili nestrokovnjaki, so ugotovili, da je za vsako nalogo potrebno le majhno število odgovorov, da dosežejo enak rezultat, kot če bi jo reševal strokovnjak. V povprečju so za enako kakovost kot z eno oznako strokovnjaka potrebovali štiri oznake nestrokovnjakov.

## 2.3 Množičenje za slovenščino

Množičenje je bilo že uspešno uporabljeno tudi za izdelavo jezikovnih virov za slovenščino.

Fišer et al. (2014) so posebej razvito orodje za množičenje sloWCrowd (Tavčar et al. 2012) uporabili za odpravljanje napak v avtomatsko zgrajenem semantičnem leksikonu sloWNet. Množičniki so z orodjem glasovali, ali avtomatsko generirani kandidati sodijo v določeno množico sinonimov ali ne. Eksperiment je pokazal, da je sloWCrowd enostaven za uporabo tako za administratorje kot tudi za množičnike. Zbrani odgovori so glede na visoko stopnjo ujemanja med množičniki zanesljivi (povprečna natančnost dosega dobrih 80 %, kar je za kompleksne semantične naloge visoka vrednost), število dvoumnih rešitev pa zelo majhno.

Množičenje so za čiščenje avtomatsko luščenih kolokacij in zgledov iz korpusa Gigafida ter razvrščanje kolokacij in njihovih zgledov v (pod)pomene preizkusili tudi Kosem et al. (2013). Rezultati eksperimenta so pokazali, kako pomembna za doseganje zanesljivih rezultatov je jasna formulacija vprašanja, ki ne sme biti večdimenzionalno in subjektivno.

Na spletu je na voljo tudi Igra besed,[10] ki služi zbiranju kolokacij v slovenščini, v kateri igralci predlagajo po tri najbolj tipične pridevniške ali samostalniške kolokatorje za naključno izbran samostalnik ali pridevnik, za svoje odgovore pa prejemajo točke glede na seznam najmočnejših kolokacij, izluščen iz korpusa Gigafida (na podlagi

9 http://anawiki.essex.ac.uk/phrasedetectives/ (dostop 8. 8. 2015).
10 http://www.igra-besed.si (dostop 8. 8. 2015).

besednih skic v konkordančniku Sketch Engine). Igra ponuja način za enega igralca, igro z izbranim soigralcem ter igro z naključnim soigralcem, ki je ob istem času prijavljen na strežnik. Z igro se zbirajo podatki o besedah, ki jih igralci vpišejo, pa tudi kdo (uporabniško ime), kdaj in s kom igra. Ti podatki bodo uporabljeni za primerjavo s seznami kolokacij, da bo mogoče preveriti, katera kolokacijska mera najbolje zazna jezikovni čut ljudi. Rezultati imajo velik potencial tudi za čiščenje avtomatsko luščenih podatkov, kar pa bi bilo treba še analizirati.

# 3 MOTIVACIJSKI VIDIKI MNOŽIČENJA

Proces množičenja vedno vključuje pobudnika, ki od množice pričakuje opravljanje določene naloge, in množico posameznikov, ki v zameno za delo prejmejo plačilo oziroma nadomestilo. Kot pišeta Estellés-Arolas in González-Ladrón-de-Guevara (2012), plačilo oz. nadomestilo služi kot motivacija za množičnika, da delo opravi oziroma nadaljuje z njim.

Motivacija je torej pri množičenjskih projektih ključnega pomena, zlasti v primeru manjših jezikov, ki nimajo na voljo velike baze množičnikov in je zato treba tiste, ki so na voljo, še dodatno motivirati za delo. Motivacija je lahko bodisi materialna bodisi nematerialna, vedno pa mora izpolniti eno ali več potreb množičnikov, kot so gmotna nagrada, družbena prepoznavnost, dvig samozavesti, razvoj posameznikovih sposobnosti. Nagrado zagotovi pobudnik množičenja kot poplačilo za delo množice. Lew (2013) v razpravi o motivaciji uporabnikov za dodajanje uporabniških vsebin na splet loči tri kategorije motivacije (psihološko, družbeno in ekonomsko), ki veljajo tudi za množičenje, zato jih podrobneje predstavljamo v nadaljevanju.

## 3.1 Družbena motivacija

Družbeni vidik motivacije črpa iz potrebe posameznikov, da se povezujejo z drugimi posamezniki, ki imajo podobne interese, s sodelovanjem pridobivajo nova znanja ali spretnosti in povečujejo svoj ugled v skupnosti.

### *3.1.1 Pripadnost skupnosti*

Večjo vlogo kot velikost skupnosti igra angažiranost njenih članov. Pomembno je torej, da se člani poistovetijo s skupnostjo in v njej navdušeno delujejo, ker želijo prispevati k njenemu uspehu, razvoju ali prepoznavnosti. Množičenje se v tem

primeru zanaša na pripravljenost posameznikov, da prispevajo k projektu, ki je v interesu in v korist vsej skupnosti.

Večina projektov kolaborativne leksikografije temelji prav na družbeni motivaciji, npr. že omenjeni Wikislovar, Urban Dictionary in Razvezani jezik, ki se je za slovenščino izkazal za uspešnega: v 10 letih trajanja projekta je okoli 1.600 anonimnih piscev prispevalo več kot 3.700 gesel in 2.300 člankov (Dolar 2014). To dokazuje, da so tudi v Sloveniji posamezniki pripravljeni sodelovati pri zbiranju leksikografskih podatkov. Da bi bili slovenski uporabniki družbenih omrežij v ustreznih okoliščinah pripravljeni prispevati tudi k izgradnji jezikovnih virov za slovenščino, lahko sklepamo iz aktivnih in konstruktivnih uporabniških skupin z jezikovno tematiko na Facebooku, npr. Prevajalci, na pomoč!, Za vsaj približno pravilno uporabo slovenščine, Skupina za ohranjanje rodilnika pri zanikanju, Društvo ljubiteljskih pravopisarjev in slovničarjev, Razgibane vejice ipd.

### *3.1.2 Izobraževalna motivacija*

Posebna podvrsta družbene motivacije je izobraževalna, pri kateri se množičniki z reševanjem nalog učijo določene vsebine ali spretnosti. Ustrezno pripravljene naloge bi torej lahko ponudili v reševanje v sklopu rednih učnih vsebin ali kot dodatno gradivo za vaje na različnih izobraževalnih stopnjah. Tovrstni motivacijski pristop ima npr. spletna stran za učenje jezikov Duolingo[11] (von Ahn 2013), ki uporabnikom ponuja brezplačne tečaje tujih jezikov. Tečaji sestojijo iz različnih nalog, med drugim tudi iz stavkov, ki jih uporabniki za vajo prevajajo v tuji jezik in hkrati pripomorejo k prevajanju spletnih vsebin v druge jezike.

### *3.1.3 Priznanja in nazivi*

Pod družbeno motivacijo spadajo tudi priznanja, ki jih množičnik prejme kot nagrado za delo v skupnosti. Lahko gre za priznanje v fizični obliki (npr. potrdilo o sodelovanju), prestižen naziv (npr. urednik Wikipedije) ali navedbo v dvorani slavnih projekta oz. skupnosti (angl. *hall of fame*).

## 3.2 Psihološka motivacija

Za mnoge uporabnike je dodajanje vsebin na splet psihološko izpolnjujoče, npr. ker radi delijo znanje z drugimi, ker tako izpolnjujejo potrebo po tem, da

11 https://www.duolingo.com/ (dostop 8. 8. 2015).

izražajo sami sebe, ali ker se jim sodelovanje zdi zabavno. Vidik zabave je bil podlaga za osnovanje t. i. iger z namenom, ki v zadnjem času postajajo ena najpopularnejših oblik sodela in množičenja, zato se jim v nadaljevanju razdelka podrobneje posvečamo.

### *3.2.1 Igre z namenom*

Igre z namenom (angl. *games with a purpose*) so igre, ki jih uporabniki primarno igrajo zaradi lastnega zadovoljstva, obenem pa z igranjem pomagajo pri zbiranju podatkov. Vse več ljudi ima dostop do spleta (veliko od teh igra tudi računalniške igre), nalog, ki jih računalniki ne zmorejo opraviti brez človeške pomoči, pa je kljub tehnološkemu napredku še vedno veliko. Kot piše von Ahn (2006), je igre z namenom zato mogoče uporabiti na različnih področjih, npr. za izboljšanje iskanja po internetu in za filtriranje vsebin, v številnih primerih pa je bil ta način zbiranja podatkov uporabljen tudi pri raziskavah z jezikovnimi podatki.

Uspešna primera iger z namenom sta ESP Game (von Ahn in Dabbish 2004) in Peekaboom (von Ahn 2006), s katerima je bilo dokazano, da lahko množica reši probleme, ki jih računalniki še ne zmorejo. Pri igri ESP Game se v paru znajdeta igralca, ki se med seboj še ne poznata, obema pa se prikaže slika. Cilj igre je uganiti, s katero besedo bo partner označil sliko. Igra je zelo uspešna, zato je bilo v kratkem času označeno veliko število slik, podatki pa so bili uporabni npr. za izboljšanje internetnih iskalnikov in za razvoj programske opreme za slabovidne. Na podoben način deluje tudi Peekaboom, le da igralci določajo, kje na sliki se nahaja določen predmet, podatki pa se na to uporabijo za strojno učenje računalniškega vida.

Med uspešnimi igrami z namenom so tudi JeuxDeMots (Joubert in Lafourcade 2012), igra za gradnjo leksikalne mreže francoščine; že omenjena igra Phrase Detectives (Chamberlain et al. 2008); Puzzle Racer (Jurgens in Navigli 2014), igra za označevanje slik s pomeni; in Verbosity (von Ahn et al. 2006), s katero so s pomočjo vprašanj ali dopolnjevanja stavkov zbirali splošno znane podatke (npr. izjave, kot je *mleko je belo*), ki so nato uporabni za gradnjo ontologij in izboljšanje inteligence računalniških sistemov.

Veliko število iger z namenom kaže, da je igrifikacija (angl. *gamification*, predstavitev oziroma oblikovanje orodij in aplikacij v obliki iger) danes pri zbiranju jezikoslovnih podatkov že precej pogosta praksa, ki prinaša dobre rezultate, uporabne na raznolikih področjih.

## 3.3 Ekonomska motivacija

Ekonomska motivacija temelji na denarnih plačilih za opravljanje nalog oziroma na drugih gmotnih nagradah.

### *3.3.1 Mikroplačila*

Denarno nadomestilo je pogosto pri velikih (zlasti komercialnih) projektih, pri katerih se od množičnikov pričakuje, da opravijo večjo količino dela, v katerega so vključeni dalj časa. Denarna nadomestila se najpogosteje izplačujejo v obliki t. i. mikroplačil (angl. *micropayments*), ki jih množičnik prejme za opravljeno nalogo oziroma za vnaprej določeno število opravljenih nalog. Na tak način delujejo številne znane platforme za množičenje, med drugim tudi Amazon Mechanical Turk,[12] CrowdFlower[13] in Clickworker.[14]

Postopek množičenja z mikroplačili je naslednji: pobudnik množičenja na platformo naloži projekt (sveženj mikronalog) in lastniku platforme vnaprej nakaže določeno količino denarja (odvisno od velikosti projekta, števila različnih nalog, zahtevnosti nalog ipd.). Določen del zneska pripada lastniku platforme za gostovanje projekta, ostalo pa lastnik platforme razdeli med množičnike glede na opravljeno delo.

Mikroplačila preko platform za množičenje so za motivacijo množičnikov uporabili že številni avtorji jezikoslovnih raziskav (Akkaya et al. 2010; Rumshisky 2011; Rumshisky et al. 2012; Fossati et al. 2013), a je treba omeniti, da imajo platforme, kot je npr. Amazon Mechanical Turk, svoj nabor množičnikov (registriranih uporabnikov, ki lahko rešujejo naloge), večinoma pa gre za angleške govorce (oziroma govorce večjih jezikov). Registriranih govorcev manjših jezikov na tovrstnih platformah ni dovolj, zaradi lokalne finančne in davčne zakonodaje pa se lahko zaplete tudi pri ustvarjanju računov za izvajanje množičenjskih projektov in nakazovanju mikroplačil.

### *3.3.2 Ostale nagrade*

Ekonomska motivacija vključuje tudi druga nadomestila, kot so npr. kuponi, vstopnice, licence za programsko opremo in druge predmetne nagrade (majice, priponke ipd.). Po plačilih v tej obliki najpogosteje posegajo manjši projekti z

12 https://www.mturk.com/ (dostop 8. 8. 2015).
13 http://www.crowdflower.com/ (dostop 8. 8. 2015).
14 http://www.clickworker.com/en/ (dostop 8. 8. 2015).

omejenim financiranjem. Primeri dobre prakse (El-Haj et al. 2014; Fišer et al. 2014) kažejo, da množičnike pritegnejo tudi tovrstne nagrade, pogosto v kombinaciji z družbeno in psihološko motivacijo.

# 4 PRAVNI, FINANČNI IN ETIČNI VIDIKI MNOŽIČENJA

V tem razdelku predstavljamo pravne, finančne in etične omejitve, na katere naletimo pri uporabi množičenja. Pravne in finančne omejitve so v veliki meri odvisne od lokalne zakonodaje in financiranja, in čeprav ne vplivajo neposredno na kakovost in vsebino projekta, pogosto predstavljajo veliko oviro pri uvedbi množičenja v raziskovalno delo, še zlasti na področju leksikografije. Večina raziskovalcev namreč ni seznanjena s pravnimi omejitvami na tem področju, pomoč pravnih strokovnjakov pa je redka. Ker je množičenje še vedno relativno nova oblika dela, ni izrecno predvideno v zakonodaji, zato marsikatero vprašanje ostaja odprto.

## 4.1 Plačevanje množičnikov

Kot pišejo Sabou et al. (2014), je etična dolžnost pobudnika množičenjskega projekta, da v primeru ekonomske motivacije množičnikov upošteva realne življenjske stroške in mikroplačila prilagodi tako, da zneski v povprečju presegajo lokalno minimalno plačo oziroma da odsevajo urno postavko, ustrezno za tovrsten način dela. Tudi Fort et al. (2014) opozarjajo, da množičenje kot nova oblika dela še vedno ni obravnavano v delovni zakonodaji, kar postavlja množičnike v kočljiv položaj z vidika višine plačila, varnosti pri delu, delavskih pravic ipd. Na platformi Amazon Mechanical Turk naj bi se tudi do 20 % delavcev preživljalo le z reševanjem množičenjskih nalog, zato je ključno, da se jim zagotovi ustrezen zaslužek. Silberman et al. (2010) izpostavijo tudi dejstvo, da pobudniki množičenjskih projektov pogosto zamujajo s plačili. Na platformi Amazon Mechanical Turk mora pobudnik odobriti, da je bilo delo ustrezno opravljeno, preden lahko množičnik prejme plačilo. Platforma naloge samodejno odobri po 30 dneh, če tega prej ne stori pobudnik, a to pomeni, da lahko množičnik čaka do konca izteka roka, nakar pobudnik njegovo delo zavrne, množičnik pa ne dobi plačila, ki ga je pričakoval. Tovrstnim praksam se je treba izogibati.

Sabou et al. (2014) priporočajo, da se pred množičenjem izvede pilotno reševanje nalog in predhodno določi, koliko časa naj bi delo trajalo. Nekatere mikronaloge so težje in bolj zapletene od drugih, zato od množičnika zahtevajo več truda in

časa. Pri takšnih nalogah je mikroplačilo ponavadi višje, da je tudi urna postavka primerljiva. Ta vidik že upoštevajo npr. Krek et al. (2013), ki za lažje množičenjske naloge predvidevajo mikroplačilo 0,02 € na odločitev (kar pri približno 350 odločitvah na uro znese 7 €), za težje naloge pa 0,05 € (odločitev na uro je v tem primeru nekoliko manj, postavka pa je podobna). Cena je vsekakor odvisna tudi od proračuna projekta in količine podatkov, ki jo je treba obdelati. Pri izplačevanju mikroplačil je treba upoštevati obstoječe načine plačevanja (v Sloveniji npr. avtorske pogodbe, plačilo preko s. p., plačilo po študentski napotnici) in morebitne omejitve v davčni zakonodaji.

Upoštevanje etičnih načel je še toliko pomembnejše, če bodo zbrani podatki uporabljeni v komercialne namene in bodo ponudniku projekta prinesli zaslužek. V takem primeru je sporno, da množičniki za delo ne prejmejo plačila ali da je plačilo nerazumno nizko.

## 4.2 Omejitve pri najemanju množičnikov

Pri izbiranju množičnikov za projekt je treba imeti v mislih, da morda pri tem obstajajo pravne omejitve. To še zlasti velja v primeru mladoletnih delavcev (npr. dijaki), pri katerih je treba pridobiti predhodno soglasje staršev.

## 4.3 Priznavanje avtorstva

Ker množičniki na projektu pogosto opravijo nezanemarljiv delež dela, je treba vnaprej določiti, kako in kje se jim pripiše zasluge (npr. ali so navedeni kot soavtorji). Čeprav za navedbo avtorstva v primerih množičenja ni na voljo jasnih smernic, nekateri prostovoljski projekti (npr. FoldIt,[15] Phylo[16]) množičnike navajajo na seznamu avtorjev.

## 4.4 Podpis soglasja in obveščanje množičnikov o projektu

Ponavadi množičniki pred začetkom dela podpišejo soglasje, s katerim jih pobudnik množičenja obvesti o naravi projekta in o namenih, za katere bodo podatki uporabljeni. Množičnikom mora biti jasno predstavljeno, da bodo podatki npr. uporabljeni v raziskovalne namene in ali bodo po koncu projekta dostopni tudi tretjim osebam (zlasti če gre za odprte licence, kot je Creative Commons).

15 https://fold.it/portal/ (dostop 8. 8. 2015).
16 http://phylo.cs.mcgill.ca/ (dostop 8. 8. 2015).

## 4.5 Dostopnost podatkovnih zbirk

V primeru, da bodo podatkovne zbirke, ki bodo z množičenjem nadgrajene, prosto dostopne, je treba zanje izbrati primerno licenco v skladu z lokalno zakonodajo o avtorskih pravicah in varstvu osebnih podatkov.

# 5 MIKRONALOGE

Osnovna ideja množičenja je, da obsežen in kompleksen problem razdeli na manjše, obvladljive in enostavnejše dele. Celoten sklop dejavnosti, potrebnih za reševanje zastavljenega problema, imenujemo množičenjska kampanja, posamezne dele, ki jih v reševanje dobivajo množičniki, pa mikronaloge. Oblikovanje mikronalog je ključna stopnja v kateremkoli množičenjskem projektu. Zato v tem razdelku predstavljamo načela, ki jih je treba upoštevati pri izdelavi mikronalog, če želimo z množičenjem doseči kakovostne in predvsem uporabne rezultate, in navedemo še nekaj primerov uspešno oblikovanih mikronalog.

## 5.1 Načela oblikovanja mikronalog

**Nezahtevnost** – Ker pri reševanju mikronalog pogosto sodelujejo nestrokovnjaki, je pomembno, da so naloge kognitivno kar se da nezahtevne. Reševanje ene same naloge od množičnika ne sme zahtevati pretiranega razmisleka, saj je bistveno, da v čim krajšem času reši čim več nalog (prim. Rumshisky 2011; Snow et al. 2008; Lease in Alonso 2014).

**Ustrezna vprašanja** – Mikronaloge naj ne vsebujejo vprašanj, ki pri množičenju ne bodo dala dobrih rezultatov. Izključiti je treba predvsem nejasna ali dvoumna vprašanja in prekomerno subjektivne in nezanesljive ocene, saj se rezultati, pridobljeni iz takšnih nalog, pogosto izkažejo za nezanesljive in neuporabne (prim. Kosem et al. 2013). Zastavljena vprašanja morajo biti enodimenzionalna, zato je v primerih, ko gre za kompleksen, večplasten problem, priporočljivo, da se naloga razdeli na več preprostejših korakov (prim. Biemann in Nygaard 2010).

**Prilagojenost ciljni skupini** – Različne mikronaloge lahko od množičnikov zahtevajo različno stopnjo predznanja. Uvajanje množičnikov v postopek označevanja mora biti čim krajše, zato je treba za vsak skupek mikronalog izbrati ustrezno ciljno skupino glede na potrebno predznanje (npr. nestrokovnjaki, študenti ali strokovnjaki). Množičniki z nezadostnim predznanjem potrebujejo več uvajanja (kar je časovno neugodno), dajali pa bodo manj zanesljive in posledično manj

uporabne rezultate. Po drugi strani pa je strokovnjake, ki morajo reševati trivialne naloge, težje motivirati, medtem ko njihovo delo prav tako zahteva višje plačilo.

**Tehnična preprostost in uporabniku prijazen vmesnik** – Reševanje mora biti nezahtevno tudi z logističnega vidika, npr. da zahteva čim manjše število klikov z miško, čim manj premikanja po zaslonu in, če je le mogoče, čim manj tipkanja in vnašanja podatkov. Množičniki naloge najpogosteje rešujejo s pomočjo platform za množičenje (Amazon Mechanical Turk, Clickworker ipd.). V primeru, da se za projekt razvija lastna platforma, je treba zagotoviti, da ima uporabniku prijazen vmesnik, ki bo omogočal neteţavno registracijo, tekoče reševanje nalog in prehajanje med njimi. Množičniku je treba zagotoviti tudi možnost, da primer označi kot nejasen oz. ga preskoči, če npr. iz danega konteksta ne more jasno sklepati, kako bi ga označil, ali kot nerešljiv, če npr. nobena od danih oznak ni ustrezna. Pri igrah z namenom von Ahn in Dabbish (2008) izpostavita, da se mora igra končati v kratkem času, Jurgens in Navigli (2014) pa poudarita, da je vmesnik ključnega pomena – prednost je, če igra sploh ne vsebuje jezikoslovne terminologije, kadar večino podatkov zbira množica nestrokovnjakov, saj mora igra biti zanje čimbolj razumljiva in preprosta.

**Kratka navodila** – Navodila za reševanje mikronalog morajo pojasniti namen množičenjske kampanje, morajo biti jasna in kratka, priporočljivo je tudi, da vsebujejo ponazoritev na primeru.

**Povratna informacija** – Priporočljivo je, da množičnik za svoje odgovore dobi povratno informacijo. Na tak način se lahko nauči nekaj novega, obenem pa ga pravilni odgovori motivirajo, da z delom nadaljuje, saj lahko sproti preverja razumevanje vprašanj in pravilnost odgovorov ter se postopoma izboljšuje v reševanju problema.

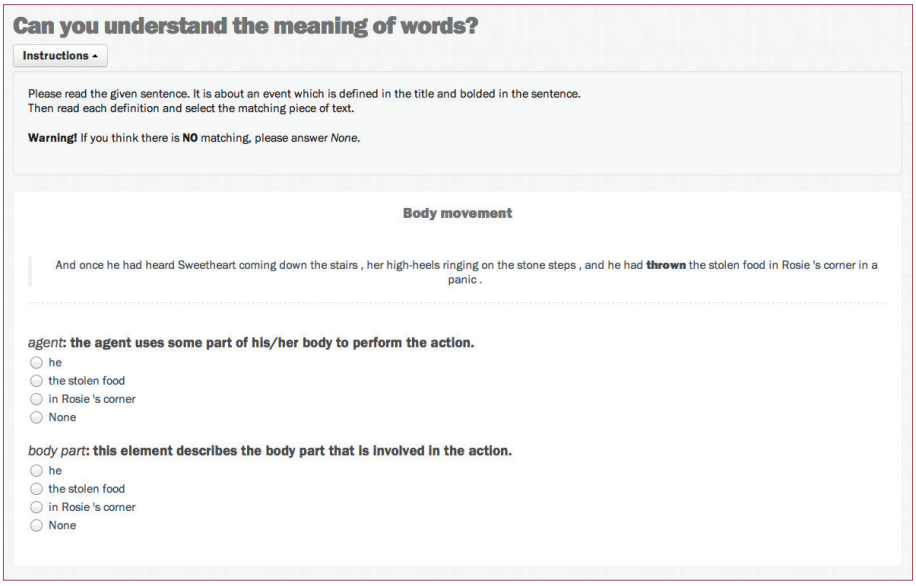
**Izziv, naključnost in časovna omejitev** – Nekaj vidikov, na katere je treba biti pozoren pri izdelavi iger z namenom, izpostavita tudi von Ahn in Dabbish (2008), ki so relevantne za vse množičenjske kampanje. Bolj kot je naloga zabavna, bolj je učinkovita. Zabavnost naloge zagotovimo tako, da jo zasnujemo kot izziv za igralca, npr. z uvedbo točkovnega sistema, časovne omejitve za reševanje naloge, lestvic množičnikov z najvišjim številom doseženih točk ipd. Ključno je tudi, da je število nalog, ki jih mora množičnik opraviti v določenem časovnem obdobju, tempirano tako, da mu je v izziv (da torej naloga ni niti preveč preprosta niti preveč težavna), časovna omejitev oz. preostali čas pa morata biti med reševanjem izpisana na zaslonu, da množičnika spodbujata. Pomembno je tudi, da naloga vsebuje elemente naključnosti, npr. da naključno združuje množičnike v pare, naključno izbira besede, ki jih morajo prepoznati ipd. S tem poskrbimo, da naloga ni predvidljiva, obenem pa se izognemo morebitnemu goljufanju množičnikov.

## 5.2 Primeri mikronalog

V tem razdelku predstavljamo nekaj primerov različnih načinov množičenja (tako klasičnih mikronalog kot iger z namenom), ki so se izkazali kot uspešni v sorodnih raziskavah.

### *5.2.1 Reševanje mikronalog*

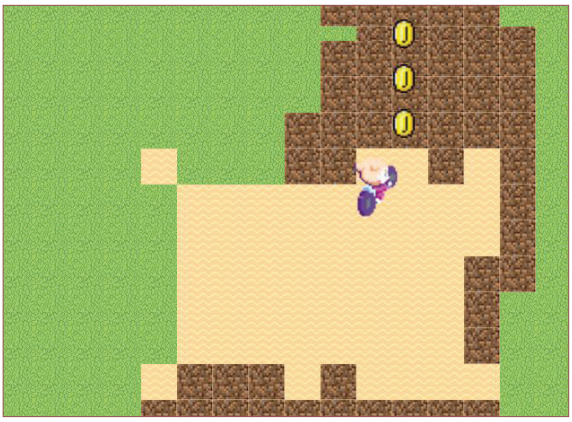
Slika 1 prikazuje primer mikronaloge za označevanje semantičnih vlog (Fossati et al. 2013). Naloga sestoji iz kratkega navodila, ki mu sledi poved, v kateri mora množičnik označiti vršilca dejanja (angl. *agent*) in del telesa (angl. *body part*). V tem primeru sta pravilna odgovora on (angl. *he*) in ni (angl. *none*).



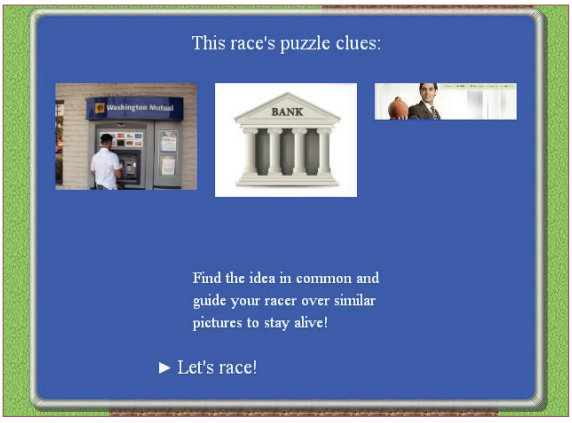

**Slika 1: Mikronaloga za označevanje semantičnih vlog.**

### *5.2.2 Reševanje mikronalog v igrah z namenom*

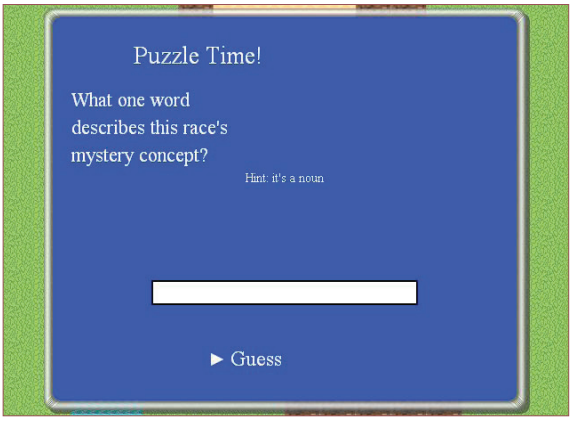
Slika 2 prikazuje vmesnik igre z namenom Puzzle Racer (Jurgens in Navigli 2014), pri kateri igralec tekmuje z avtomobilčkom ter pobira kovance in druge dobrine, ki prinašajo točke. Pred začetkom dirke igralec dobi namig v obliki treh slik (Slika 3), na podlagi katerih mora ugotoviti, kaj imajo skupnega, da lahko reši okvirček z uganko, ki se pojavi med dirkanjem (Slika 4). V tem primeru je pravilni odgovor denar (angl. *money*).



Slika 2: Igra z namenom Puzzle Racer.




Slika 3: Namig pri igri Puzzle Racer.




Slika 4: Uganka pri igri Puzzle Racer.

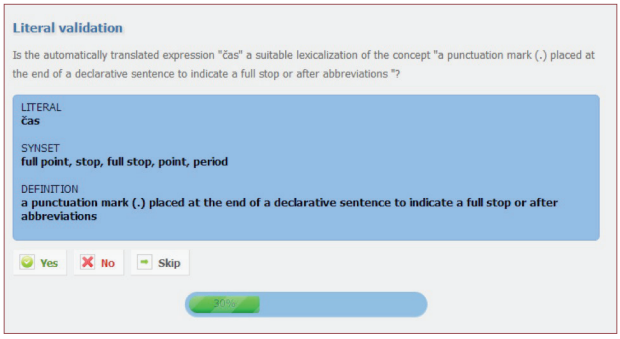
### 5.2.3 Reševanje mikronalog na družbenih omrežjih

Igre z namenom je mogoče vključiti tudi v družbena omrežja, kjer so prikladno dostopne velikemu številu uporabnikov, ki na njih preživljajo precej prostega časa in so dovzetni za tovrstne izzive. Na Sliki 5 je posnetek igre Phrase Detectives (Chamberlain et al. 2008) na družbenem omrežju Facebook. Igralec dobi zgled z dvema obarvanima besednima zvezama, od katerih se ena nanaša na drugo, označiti pa mora, ali se z oznakama strinja. Za pravilne odgovore (glede na ujemanje z drugimi igralci) prejme točke. V tem primeru je pravilen odgovor da (angl. *Agree*).


**Slika 5: Različica igre Phrase Detectives na omrežju Facebook.**

### 5.2.4 Primeri mikronalog za slovenščino

Na sliki 6 je prikaz mikronaloge v primeru množičenja za čiščenje sloWNeta (Fišer et al. 2014) z orodjem sloWCrowd (Tavčar et al. 2012). Množičnik mora pri nalogi označiti, ali dani literal (beseda ali besedna zveza) glede na angleške ustreznice in definicijo sodi v ta sinset (množico sinonimov). V tem primeru je pravilen odgovor ne (angl. *No*).




**Slika 6: Mikronaloga za potrjevanje literalov v orodju sloWCrowd.**

Slika 7 prikazuje vmesnik Igre besed. Igralec dobi iztočnico (v tem primeru pridevnik *gasilska*), za katero mora v 30 sekundah predlagati tri tipične kolokatorje. V ugibanju se lahko pomeri tudi z izbranim ali naključnim nasprotnikom. Odgovori se točkujejo glede na ujemanje s seznami kolokacij iz korpusa Gigafida.


**Slika 7: Vmesnik Igre besed.**

# 6 PREVERJANJE IN ZAGOTAVLJANJE KAKOVOSTI

V tem razdelku predstavljamo mehanizme, s katerimi lahko posredno ali neposredno zagotovimo kakovostne rezultate pri množičenju ter izločimo šumne prvine, ki se v množici rezultatov znajdejo npr. zaradi nejasnih navodil ali zaradi nezanesljivih množičnikov.

## 6.1 Zlati standard

Najpogostejša metoda nadziranja kakovosti se izvaja s pomočjo zlatega standarda (angl. *gold standard*), ročno označene množice podatkov, ki jo pri razvoju jezikovnih tehnologij uporabljamo za učno ali testno množico. Pri množičenju zlati standard vsebuje določeno število mikronalog, ki jih je vnaprej pravilno rešil strokovnjak. Naloge iz zlatega standarda so nato naključno vključene med mikronaloge, ki jih rešujejo množičniki, in služijo preverjanju njihove zanesljivosti – če množičnik ne odgovori pravilno na dovolj mikronalog iz zlatega standarda, se vsi njegovi odgovori izključijo iz končnih rezultatov.

Pri oblikovanju zlatega standarda je treba zagotoviti njegovo reprezentativnost, tako po obsegu kot težavnosti. S preveč preprostim zlatim standardom namreč

ne moremo učinkovito ločiti zanesljivih množičnikov od nezanesljivih, preveč težaven zlati standard pa bo izključil preveliko število množičnikov. Prav tako je treba pri reševanju mikronalog zagotoviti ustrezno razmerje vprašanj iz zlatega standarda in pravih vprašanj, saj s premajhnim številom vprašanj iz zlatega standarda ne moremo zanesljivo preveriti množičnikove zanesljivosti, s prevelikim številom vprašanj iz zlatega standarda pa namesto dragocenih odgovorov na nova vprašanja zbiramo že znane odgovore, kar je neekonomično.

## 6.2 Ujemanje med množičniki

Druga metoda za nadziranje kakovosti množičenja je izračun ujemanja med množičniki (angl. *inter-annotator agreement*). To storimo tako, da več različnim množičnikom ponudimo v reševanje iste mikronaloge. Na ta način pridobimo večje število odgovorov za vsako nalogo, odgovore pa lahko potem primerjamo in na podlagi primerjave odgovorov različnih množičnikov na isto vprašanje ugotovimo, kolikšno je razhajanje med njimi. Glede na porazdelitev odgovorov lahko izračunamo tudi stopnjo zanesljivosti (angl. *confidence score*) za posamezno mikronalogo ali množičnika (Oyama et al. 2013).

Če je veliko primerov, v katerih se odgovori množičnikov močno razhajajo, lahko to pomeni, da mikronaloge niso bile ustrezno zasnovane, da jih ni reševala skupina s pravo mero predznanja ali da smernice za označevanje niso bile dovolj jasno opredeljene, zaradi česar jih je treba izboljšati. V primerih, ko razhajanje ni pretirano, se lahko upošteva večinski glas (angl. *majority vote*), ki končno odločitev sprejme tako, da upošteva rešitev večine množičnikov.

Poudariti je treba, da je pomembno najti optimalno zgornjo mejo večkratnega postavljanja istega vprašanja različnim množičnikom, saj z vsakim ponovljenim vprašanjem ne dobimo nobenega novega odgovora, kar nezanemarljivo poveča stroške množičenja. Običajno so za eno odločitev potrebne 3 oznake, za bolj zapletene naloge pa 5 (tako predvidevajo tudi Krek et al. 2013b).

## 6.3 Razsojanje

Razsojanje (angl. *refereeing*) je proces, v katerem lahko težavne primere, pri katerih množičniki niso prišli do enotne rešitve, označi strokovnjak – razsodnik. Če je bila priprava na množičenje (z oblikovanjem mikronalog in smernic za označevanje) uspešna, je na tak način strokovnjakom na koncu prepuščen le manjši delež težavnih primerov, večina dela pa je še vedno množičena. V primeru označevanja

hrvaškega korpusa (Klubička in Ljubešić 2014) se je izkazalo, da tovrstni postopek količino dela strokovnjakov skoraj razpolovi.

## 6.4 Doslednost množičnika

Doslednost (angl. *consistency* ali *intra-annotator agreement*) je zadnja od priljubljenih metod, s katero iz rezultatov izločimo nezanesljive množičnike. Pri merjenju doslednosti namreč množičniku v reševanje ponudimo večkrat isto nalogo (Gut in Bayerl 2004), s čimer lahko preverimo, ali so njegovi odgovori konsistentni. Če se odgovori na iste naloge preveč razhajajo, množičnikovih rezultatov ne upoštevamo, saj bodisi ni dovolj samozavesten oziroma usposobljen bodisi izbira naključne odgovore, da bi rešil čimveč nalog.

# 7 SOOČANJE S PREDSODKI

Kljub veliki pozornosti, ki jo zadnje čase dobiva množičenje tudi med leksikografi, v razpravah o uporabi množičenja v leksikografiji še vedno naletimo na številne predsodke.[17] V tem razdelku naslavljamo poglavitne pomisleke in skušamo odpraviti najpogostejše dvome, ki so ovira pri uvedbi množičenja v leksikografske projekte.

## 7.1 Slovarja ne morejo pisati nestrokovnjaki

Danes je pri gradnji slovarskih priročnikov področje, na katerem so uporabniki slovarjev najbolj udeleženi, kolaborativna leksikografija, pri kateri uporabniki aktivno prispevajo iztočnice, razlage ipd. Ker so tudi nekateri avtorji z opredelitvijo množičenja nekoliko nedosledni (prim. Estellés-Arolas in González-Ladrón-de-Guevara 2012), se množičenje pogosto neupravičeno zamenjuje ali celo enači s kolaborativno leksikografijo.

Za razliko od številnih kolaborativnih projektov, pri katerih vse delo opravijo nestrokovnjaki, pri množičenju vedno aktivno sodeluje tudi pobudnik množičenja (izvajalec projekta), in sicer s pripravo podatkov, oblikovanjem mikronalog, preverjanjem kakovosti, zagotavljanjem motivacije med množičniki ipd. Prav tako drugače kot pri številnih kolaborativnih projektih, ki sicer dokazujejo, da lahko tudi uporabniki prispevajo h koristnim in široko uporabljanim slovarskim izdelkom (Meyer in Gurevych 2012), množičenje, kot ga predlagamo za vključitev v izdelavo slovarja sodobnega slovenskega jezika, zajema predvsem čiščenje

17 http://www.sssj.si/pogosta-vprasanja/ (dostop 8. 8. 2015).

avtomatsko luščenih podatkov pred gradnjo slovarja in ne predvideva kolaborativnega pristopa, pri katerem je takoj objavljen vsak uporabniški prispevek, množica pa neposredno nadzira tudi nabor besed, vključenih v slovar, vsebino in organizacijo informacij v geselskem članku (členitev pomenov, vrstni red definicij ipd.), kar je za končni izdelek lahko problematično. Meyer in Gurevych (2012) sicer ugotavljata, da kolaborativni projekti predstavljajo vsoto mnenj številnih avtorjev, ki geselske članke intenzivno popravljajo, vse dokler ni dosežen splošni konsenz tako o njihovi strukturi kot vsebini, zaradi česar kolaborativna leksikografija v številnih segmentih daje popolnoma primerljive rezultate resnim leksikografskim projektom. Kot največjo pomanjkljivost izpostavita učinkovit mehanizem za ločevanje med zrelimi, kakovostno oblikovanimi geselskimi članki in tistimi, ki še potrebujejo izboljšave. Podobno opozarja Lew (2013), ki opaža, da je pri določenih geslih vrstni red definicij v Wikislovarju precej naključen, pri čemer so lahko pri vrhu tudi povsem marginalni pomeni. Podobno je pri slovarju Urban Dictionary, pri katerem uporabniki z glasovanjem vplivajo na vrstni red definicij, sporno dejstvo, da lahko uporabniki izglasujejo tudi definicijo, ki se jim zdi najbolj zabavna oziroma ki najbolje odraža njihovo ideološko prepričanje, ni pa nujno najustreznejša.

Nasprotno je množičenje le ena od faz izdelave slovarja: najprej jezikovni tehnologi avtomatsko izluščijo podatke iz korpusa in drugih podatkovnih zbirk, ki jih množičniki s pomočjo ciljnih mikronalog očistijo, nakar jih leksikografi uporabijo pri ročnem leksikografskem delu. Množičenje je torej vmesni člen med avtomatsko in ekspertno obdelavo podatkov, saj z avtomatskim luščenjem in z delom množičnikov bistveno razbremeni leksikografa, hkrati pa vključuje ročni pristop pri postopkih, ki jih še ni mogoče zadovoljivo avtomatizirati. Ta metoda v obsežnem leksikografskem okolju še ni bila temeljito preizkušena, a na podlagi številnih drugih projektov, v katerih se je kot postopek za čiščenje avtomatsko generiranih podatkov izkazala kot učinkovita (Klubička in Ljubešić 2014; Fišer et al. 2014; Kosem et al. 2013), sklepamo, da bo uspešna tudi v leksikografiji.

## 7.2 Množičen slovar je nezanesljiv

Pogost predsodek zadeva tudi zanesljivost rezultatov množičenja, največkrat zato, ker so v delo (lahko) vključeni nestrokovnjaki. Na tem mestu je treba poudariti, da je serija mikronalog izdelana za določen profil množičnika glede na predznanje, ki je potrebno za reševanje naloge. Pri kakovostno zasnovanem postopku množičenja bodo bolj zapletene naloge reševali množičniki z več znanja (npr. študenti ali diplomanti jezikoslovnih smeri), preproste naloge pa bodo prepuščene tudi nestrokovnjakom.

V prispevku smo predstavili tudi vrsto mehanizmov, s katerimi lahko preverjamo in nadziramo kakovost pridobljenih rezulatov (npr. zlati standard, ujemanje med množičniki, večinski glas, doslednost, razsojanje) ter učinkovito izločimo tiste množičnike, ki dajejo nepravilne ali nezanesljive odgovore. Številni avtorji (prim. Rumshisky 2011; Fišer et al. 2014; Klubička in Ljubešić 2014; Fossati et al. 2013) so te mehanizme že preizkusili in ugotovili, da zagotavljajo visoko natančnost množičenjskih rezultatov, ki so enako kakovostni, kot če bi delo opravljali samo strokovnjaki (Snow et al. 2008).

## 7.3 Množičenje degradira leksikografski poklic

Množičenje kot nova oblika (prekarnega) dela, ki še ni izrecno predvidena v zakonodaji niti v tujini niti v Sloveniji, vzbuja tudi številne etične pomisleke, ki zadevajo predvsem plačilo množičnikov, pogoje dela in priznanje avtorstva. Pogosto uporabljane platforme za množičenje so največkrat le posrednice pobudnikov množičenja, ki ceno za opravljanje nalog na svojem projektu določijo sami. Množičnikom sicer ni treba sprejemati slabo plačanih nalog, a so v to pogosto prisiljeni, če želijo priti do zaslužka. Kot smo že omenili, na nizka plačila in izkoriščevalsko ravnanje z množičniki opozarjajo številni avtorji (Sabou et al. 2014; Silberman et al. 2010; Lease in Alonso 2014; Felstiner 2011), na tovrstne prakse nizkih plačil pa naletimo tudi pri sorodnih raziskavah: Snow et al. (2008) na platformi *Amazon Mechanical Turk* namreč plačajo skupno le 2 dolarja za 7.000 oznak nestrokovnjakov oziroma 1 dolar za 1.500 oznak strokovnjakov.

Dolžnost koordinatorjev vsakega leksikografskega projekta je torej, da množičnike kot vse ostale delavce obravnavajo korektno in jim zagotovijo ustrezno plačilo, kar je treba upoštevati že pri sami zasnovi projekta, ko se določa proračun. Obenem je treba poskrbeti, da je prispevek množičnikov na končnem izdelku tudi ustrezno priznan.

Poleg samega plačila in pogojev dela se pri množičenju pogosto soočamo tudi z mnenjem, da z izkoriščevalsko obliko prekarnega dela degradira poklic leksikografov in jezikoslovcev ter jim celo jemlje delo, ki ga preusmerja na (slabo plačano) nekvalificirano množico. Poudariti je treba, da je bistvo množičenja smiselno izkoriščanje virov – da se izurjenim leksikografom prihrani delo in dragoceni čas pri rutinskih opravilih, množičnikom pa omogoči, da po svojih močeh prispevajo h gradnji jezikovnih virov in v zameno dobijo motivacijo v različnih oblikah (denarno ali materialno plačilo, pridobivanje izkušenj in referenc, zabava ipd.).

## 7.4 Množičenje je sanjska rešitev

Za zagotavljanje ustrezne vloge množičenja v leksikografskih projektih je nujno prepoznati potencial, pa tudi omejitve množičenja, saj množičenje ni uporabno za vsako vrsto podatkov, vsako fazo leksikografskega dela in vsak leksikografski projekt. Množičenja recimo ne moremo uporabiti, če ne moremo zagotoviti rednega upravljanja s kampanjo (priprave mikronalog, preverjanja zbranih odgovorov, sprotne motivacije in plačevanja množičnikov). V vsebinskem smislu množičenje prav tako ni primerno za vprašanja odprtega tipa in vprašanja, ki zahtevajo podajanje subjektivnih ocen. Koristno je lahko le takrat, ko v leksikografskem projektu kljub vsem potrebnim pripravljalnim, vmesnim in naknadnim opravilom prihrani čas in/ali denar, pri tem pa še vedno zagotavlja zanesljive rezultate.

## 7.5 Ukrepi za zmanjšanje tveganj

Ker omenjeni predsodki pred množičenjskimi kampanjami niso prisotni samo v stroki, temveč nanje naletimo tudi pri splošni javnosti, je zelo pomembno, da ima pobudnik množičenja izdelano strategijo odnosov z javnostmi ter do potencialnih množičnikov pristopi pazljivo in premišljeno, hkrati pa od množice pričakuje vložek, ki je sorazmeren z vrsto predvidene motivacije. V primeru, ko množica za svoje delo ni plačana, ji na primer ni primerno zastavljati preveč ambicioznih nalog. Pomembno je tudi, da pobudnik množičenja skozi celotno množičenjsko kampanjo ohranja stik s skupnostjo množičnikov, jo obvešča o poteku projekta, vabi na javne predstavitve projekta in podobne dogodke, se jim javno zahvali za doprinos k projektu ipd.

# 8 ZAKLJUČEK

Številni jezikoslovni projekti so množičenje že uspešno uporabili na različnih področjih, kar kaže, da bi bila ta metoda lahko uporabna tudi v slovenski leksikografiji kot učinkovit način za obdelavo podatkov za gradnjo slovarja. Pri tem je treba vse potrebne vidike upoštevati že pri zasnovi leksikografskega projekta: od priprave podatkov, oblikovanja mikronalog in rekrutiranja množičnikov do zagotavljanja njihove motivacije in upoštevanja pravnih, finančnih ter etičnih omejitev projekta.

Da je množičenje mogoče uspešno uporabiti tudi v slovenskem okolju, dokazujejo dosedanje izkušnje pri sorodnih raziskavah, obenem pa je treba poudariti, da je tudi motiviranost slovenske javnosti za tovrstne projekte visoka. To kaže

npr. slovar pogovorne slovenščine Razvezani jezik, nezanemarljiv pa je tudi porast skupin z jezikoslovno tematiko na družbenih omrežjih, v katerih uporabniki zelo aktivno in redno sodelujejo.

Množičenje bo v okviru naslednje generacije leksikografskih projektov nedvomno postalo koristno orodje leksikografov, saj bo pripomoglo k hitrejšemu delu v obdobju, ko zahteve po hitri obdelavi vse večje količine jezikovnih podatkov naraščajo, in razbremenilo leksikografe pri rutinskih opravilih, zaradi česar jim bo ostalo več časa in energije za strokovno delo. Podatkovne baze, ki bodo rezultat množičenja, bodo imele dodano vrednost, saj jih bo mogoče uporabiti tudi za druge, neslovarske namene, npr. za izboljšanje jezikovnotehnoloških orodij z metodami strojnega učenja, pri čemer množičeni podatki služijo kot kakovostna učna množica. Slovar sodobnega slovenskega jezika je eden prvih leksikografskih projektov, ki ima v načrtu množičenje vključiti v celoten delotok, s čimer bo kot pionirski projekt začrtal smernice za številne prihodnje slovarje in jezikovne vire pri nas in po svetu.