

Игре са полилексичким јединицама за боље повезивање научне заједнице

УДК 811.163.41'37:794

САЖЕТАК: У овом раду приказујемо игре у чијем фокусу су полилексичке јединице – комбинације речи које показују неочекивана лексичка, синтаксичка, семантичка, прагматичка или статистичка својства. Ове игре се заснивају на вишејезичној колекцији полилексичких јединица коју су створили чланови научне мреже, стручњаци за обраду природних језика. У раду описујемо процес прикупљања података и њихово коришћење у игри упаривања полилексичких јединица истог значења у разним језицима, кроз пантомиму и квиз који помињу идиоме у којима се помињу националности. У овим играма је коришћена језичка разноврсност и богатство чланова акције PARSEME да би се учврстила њихова међусобна повезаност.

КЉУЧНЕ РЕЧИ: полилексичке јединице, идиоме, игре, лексички ресурси.

РАД ПРИМЉЕН: 27. октобар 2017.

РАД ПРИХВАЋЕН: 25. новембар 2017.

Цветана Крстев

cvetana@matf.bg.ac.rs

Универзитет у Београду,
Филолошки факултет, Србија

Агата Савари

agata.savary@univ-tours.fr

Универзитет Франсоа Рабле,
Тур, Француска

1. Увод и мотивација

Учење кроз игру је познати приступ учењу, као средство за ефикасан развој друштвених, когнитивних и емоционалних вештина, посебно код деце. У научном истраживању се такозване „озбиљне игре“ или „игре са сврхом“, које се често заснивају на групној расподели рада, могу употребити да се искористи корисничково уживање у игри за обављање неког истраживачког задатка. У овом раду бавимо се играма заснованим на групној расподели рада које су употребљене на маргини научног истраживања, а за повећање повезаности научне заједнице. Извештавамо о напорима де се овакве игре организују у

оквиру COST акције PARSEME,¹ која представља европску научну мрежу у области обраде природних језика. Ова акција је посвећена истраживању интеракције између парсирања – аутоматске синтаксне или семантичке анализе природнојезичких исказа – и полилексичких јединица. Полилексичке јединице су комбинације речи које показују неочекивана лексичка, синтаксичка, семантичка, прагматичка или статистичка својства. Истакнути примери полилексичких јединица су идиоми, какви су у енглеском *by and large* (у целини, уопште), *a hot dog* (лит. ВРУЋЕ КУЧЕ – сендвич с виршлом), или *to pull one's leg* (лит. ВУЋИ НЕКОГА ЗА НОГУ – зафркавати некога). Њихово суштинско својство је семантичка некомпозитност, што значи да се њихово значење не може директно израчунати на основу значења речи из којих су састављене. Ово чини полилексичке јединице изазовним и за људе који нису изворни говорници неког језика и за рачунарске апликације у обради природних језика, какве су машинско превођење или екстракција информација. Показаћемо да се оне могу успешно користити и због својих забавних и изненађујућих ефеката.

PARSEME акција, која је трајала од 2013. до 2017, окупила је преко 200 истраживача из више од 30 земаља, који раде на исто толико различитих језика. За завршни догађај који се одиграо септембра 2016. у Дубровнику у Хрватској² учесници су прикупили примере полилексичких јединица који су претворени у квизове. Основна идеја је била да се искористи огромна језичка разноврсност и богатство чланова научне заједнице за пријатно дружење и учвршћивање заједништва. Успут је формирана вишејезична база података са неколико стотина полилексичких јединица која је слободно доступна.

2. Сродна истраживања

Игре са сврхом (енгл. Games With A Purpose – GWAPs) су се појавиле 2004. година са појавом игре за етикетирање слика ESP (Von Ahn and Dabbish, 2008) отприлике у исто време када је почела да се користи и групна расподела рада (Howe, 2006). У ствари, ова два начина за обављање озбиљних задатака имају пуно тога заједничког: задаци који су још увек изазовни за рачунаре, а тривијални (или бар оствариви) за људе деле се на мале задатке који се расподељују међу

¹ PARSEME – PARsing and Multi-word Expressions, www.parseme.eu

² Страница догађаја

многим „радницима“, који не морају бити стручњаци да би их решили. Разлика између ова два приступа у експлоатисању снаге људског ума лежи у начину мотивисања људи да обаве ове мале, обично досадне и заморне задатке: у случају групе расподеле рада мотивација може бити зарада, жеља да се научи нешто ново или да се допринесе неком заједничком циљу, док је у случају игара са сврхом главни мотив људи да се забаве. Потреба да се забаве „радници/играчи“ свакако утиче на организовање рада на подзадацима и одређује за које задатке су игре право решење.

Од самог настанка ова два концепта чули су се гласови који су из другог угла посматрали ову врсту организовања рада. Осим квалитета посла који се овако обавља, доводи се у питање и његова етичност. За групу расподелу рада поставило се питање: „Можемо ли да замислимо групно радно место на коме бисмо желели да раде наша деца?“ (Kittur et al., 2013), док Tuite (2014) указује да „игре чији је циљ да понуде забаву у замену за обављени посао, могу изгледати израбљивачке, иако то можда није била намера њихових твораца.“ Иста ауторка сматра да играчи игара са сврхом треба унапред да знају која је сврха тих игара и да у њима учествују по слободном избору – „забава“ ће им помоћи да обаве постављене задатке. Штавише, она сматра да људи не би требало да обављају само досадне и незанимљиве задатке по сценарију „људске јединице за обраду“ (енгл. human-processing unit – HPU), већ би игре требало конципирати за стручњаке који би обављали сложен и креативан посао, као што је већ постигнуто, рецимо у случају игре за протеинско савијање Foldit (Cooper et al., 2010).

Група расподела рада је успешно коришћена у многим пројектима обраде природних језика (Mitrović, 2013), пре свега за развој језичких ресурса, а посебно за језике са мање таквих ресурса чије је говорнике лакше мотивисати да допринесу развоју (Millour and Fort, 2017). Група расподела рада, игре са сврхом и други приступи засновани на колективној интелигенцији могу се користити не само за изградњу нових ресурса за обраду природних језика, већ се постојећи ресурси и алати за обраду природних језика могу користити за евалуацију ресурса изграђених колективним напором (Gurevych and Zesch, 2013).

У (Lafourcade et al., 2015) аутори групишу игре са сврхом у четири најшире категорије: биолошке игре, игре са медицинском сврхом, игре за обраду природних језика и остале игре које се не могу класификовати. Они истичу да су игре са сврхом посебно важне за развој језичких ресурса јер оне не омогућавају само прикупљање огромних ресурса,

већ могу узимати у обзир и динамику и променљивост самог језика. Такође, добијени ресурси могу да укључују само податке око којих се сагласило више играча што обезбеђује њихову исправност, а ово је тешко постићи код традиционалног прикупљања података. Према попису који је представљен у (Lafourcade et al., 2015) већина игара са сврхом за обраду природних језика се користи за задатке анотације (анафоре, узајамне референце, синтаксне анотације, разрешавање семантичких вишезначности) и прикупљања података (здраворазумске чињенице, парафразе, везе термина и њихових карактеристика, именовани ентитети, итд.), али и за изградњу нових ресурса (лексичка мрежа термина) и валидацију (везе између концепата). Недавна истраживања укључују игру за анотацију полилексичких јединица у француском.³ Коначно, постоје многе друге игре са речима и словима које наликују играма са сврхом али не спадају сасвим у ту категорију (нема консензуса, нема валидације одговора играча, итд.).

Игре које ћемо представити у овом раду су по својој природи специфичне. Оне нису игре са сврхом у пуном смислу речи, већ су пре свега замишљене за посебну сврху учвршћивања заједнице с намером да се играју само једном. Прецизније, њихов главни циљ је био да се успешан пројекат изградње мреже истарживача заврши у најбољем расположењу. Осим прикупљању ресурса (вишејезичног пописа полилексичких јединица), стварање и играње ових игара доринело је ближем упознавању шире публике са феноменом полилексичких јединица и откривању стереотипа везаних за разне националности који се исказују у идиоматској употреби језика.

3. Прикупљање података

Улазни подаци за наше игре су прикупљени углавном кроз групну расподелу рада. Чланове PARSEME акције смо позвали да попуне онлајн образац са полилексичким јединицама на њиховом матерњем језику. Сваку полилексичку јединицу је требало записати на оригиналном писму, у латиничној транскрипцији, ако је то било потребно, и требало је додати њено буквално и идиоматско значење. Образац за попуњавање се састојао од три одељка:

³ <https://zombilingo.org/game/upl>

- врло непрозирне полилексичке јединице, то јест оне чије значење странци тешко могу да докуче, као у примеру (1),⁴
 - полилексичке јединице у којима се помињу националности, посебно оне које одражавају позитивне или негативе стереотипе, као у примеру (2),
 - полилексичке јединице са општим значењем, то јест оне које имају једно од десет задатих значења: УМРЕТИ, БИТИ УМОРАН, БИТИ ГЛУП, БИТИ ЉУТ ИЛИ НЕРАСПОЛОЖЕН, БИТИ ЗАЉУБЉЕН ИЛИ ЗАЉУБИТИ СЕ, РАДИТИ НЕШТО НЕКОРИСНО, БИТИ ПИЈАН ИЛИ СКЛОН ПИЋУ, ОБРАЂАТИ ПАЖЊУ НА НЕБИТНЕ ДЕТАЉЕ, БИТИ ВРЕДАН ИЛИ РАДИТИ НАПОРНО, БИТИ ЛЕЊ; два прикупљена примера су (3)–(4).⁵
- (1) немати длаке на језику (SR) – *nemati dlake na jeziku*
'*to have no hair on one's tongue*' ⇒ TO EXPRESS ONE'S OPINION OPENLY
 - (2) *saoul comme un polonais* (FR)
'*as drunk as a Pole*' ⇒ VERY DRUNK
 - (3) *pójsć do Abrahama na piwo* (PL)
'*to go have a beer with Abraham*' ⇒ TO DIE
 - (4) *nach dem Bettzipfel schielen* (DE)
'*to squint at the corner of one's bed*' ⇒ TO BE TIRED

На овај начин смо прикупили укупно 455 полилексичких јединица од 38 учесника на 25 језика. Ови подаци су били улаз за три игре које ћемо представити у даљем тексту.⁶

4. Игре с полилексичким јединицама општег значења

Ове игре су се заснивале на улазним подацима које су корисници приложили о полилексичким јединицама за десет општих значења.

⁴ Надаље ће сви примери бити обележени језиком из кога потичу. Видети на страни 23 листу кодова језика свих примера.

⁵ Примери (1–4) су дати у облику у коме су оригинално приложени. Превод на српски буквалног превода и значења полилексичких јединица из примера (2–4) је: (2) 'пијан као Пољак' ⇒ ВЕОМА ПИЈАН; (3) 'Ићи на пиво са Аврамом' ⇒ УМРЕТИ; (4) 'чучати при углу кревета' ⇒ БИТИ УМОРАН.

⁶ Сви примери су у овом раду наведени онако како су прикупљени. Нисмо покушали да их хомогенизујемо нити су независни стручњаци проверавали њихова значења.

Идеја је била да се из прикупљеног улаза изаберу најнепрозирнији примери и да се дају играчима да погоде њихово значење. Истовремено смо желели да сви језици за које су корисници приложили неке примере у улазним подацима буду заступљени у игри. Да бисмо играчима дали полилексичке јединице које никада нису имали прилике да сретну, а под претпоставком да сродни језици (језици из истих фамилија) вероватно деле исте или сличне полилексичке јединице, груписали смо све прикупљене податке у четири групе: словенски језици, романски језици, германски језици и остали. На пример, за значење УМРЕТИ прикупљено је 27 полилексичких јединица: 10 за словенске језике (бугарски, македонски, пољски, руски, српски и хрватски), 7 за романске језике (бразилски португалски, португалски, румунски, француски и шпански), 4 за германске језике (дански, енглески, немачки и норвешки), и 10 за остале језике (грчки, летонски, литвански, мађарски, малтешки, персијски, свахили и хебрејски).

Играчи су на основу матерњег језика груписани у четири групе. Свака група је добила 12 примера изабраних међу онима који припадају другим трима групама, то јест, ниједна група није добила пример на матерњем језику неког играча из те групе. Задатак играча је био да повежу сваки пример са једним од 4 задата значења (за свако од ових значења су била по 3 примера). За сваки пример су биле дате следеће информације: језик из кога потиче, полилексичка јединица на оригиналном писму, транскрибована полилексичка јединица (ако оригинално писмо није латиница) и буквалан превод. На пример, романска група је добила следећа три примера за значење БИТИ ЗАЉУБЉЕН; ЗАЉУБИТИ СЕ:

- (5) *poszuć do kogoś mięte* (PL)
'осећати нану за некога'
- (6) *Schmetterlinge im Bauch haben* (DE)
'имати лептире у стомаку'
- (7) *belécsap a villám* (HU)
'муња га је погодила'

Нека од 10 општих значења за која је у улазним подацима било више непрозирних примера на неколико језика појавила су се као опције у више група: БИТИ ЉУТ, БИТИ ПИЈАН, БИТИ НЕКОРИСТАН, БИТИ УМОРАН, УМРЕТИ, БИТИ ГЛУП.

Као комплетан узорак ове игре наводимо скуп података који је дат словенској групи. Он се састојао од примера (8–19) којима је требало придружити значења а)–г). Решења овог узорка су дата у додатку 2.

Значења:

- а) УМРЕТИ (БИТИ МРТАВ)
- б) БИТИ УМОРАН
- в) БИТИ ГЛУП
- г) БИТИ ЉУТ

Изрази:

- (8) estar criando malvas (ES)
'узгајати слез'
- (9) estar na cara da gaita (PT)
'бити у кутији за хармонику'
- (10) lóg a bele (HU),
'његова црева висе'
- (11) πάω στα θυμαράκια (EL) – rao sta thimarakia
'ићи ка малим грмовима тимјана'
- (12) qabžitlu č-činga (MT)
'његов ремен се изглавио'
- (13) passer l'arme à gauche (FR)
'ставити оружје на леву страну'
- (14) har roterende fis i kasketten (DA)
'има ротирајући прдеж у капи'
- (15) באקלמר מחורר הכי העפרון לא (HE) – lo ha-iparon haki mexudad
ba-kalmar,
'није најоштрија оловка у перници'
- (16) βγάλω από τα ρούχα μου (EL) – vjaino apo ta rucha mu
'скидање моје одеће'
- (17) nincs ki a négy kereke (HU)
'он нема сва четири точка'
- (18) estar de saco cheio (PT)
'имати пуну врећу'
- (19) کردن کف کسی دهن (FA) – dahāné kasy kaf kardan
'нечија уста се пене'

На секцији „Игре се полилексичким јединицама“ седмог састанка PARSEME акције сви учесници, осим организатора игара, подељени су

четири групе и сви чланови једне су групе су заједно покушавали да дођу до решења. У групама је било од 8 до 13 учесника, а време за тражење решења је било ограничено на 15 минута. Све групе су се јако добро показале: германска група је решила свих 12 задатака, романска група 10, група осталих језика 9, а словенска група 8 задатака.

Игре за све четири групе, као и табела са улазним подацима свих прилагача за свих 10 општих значења може се наћи на веб страници догађаја.⁷

5. Пантомима с полилексичким јединицама

За ову игру смо користили само енглеске полилексичке јединице да би свако могао да учествује, претпостављајући да у публици сви познају енглески. Већина енглеских полилексичких јединица је изабрана из чувених почетних радова о полилексичким јединицама (Baldwin and Kim, 2010; Sag et al., 2002) које су сви учесници PARSEME COST акције сигурно прочитали. Неколико других примера је изабрано јер су опште познати, као *all of a sudden* (изненада), или су пуно пута поменути за време PARSEME састанака, као *it rains cats and dogs* (лит. кише мачке и пси; пада јака киша). Све изабране полилексичке јединице су рангиране као лаке, средње или тешке на основу два критеријума: (i) да ли је полилексичка јединица добро позната? (ii) да ли је лако описати полилексичку јединицу у игри пантомиме (на пример, прилошке полилексичке јединице је теже описати него именичке или глаголске полилексичке јединице)?

Изворни говорници енглеског нису учествовали у овој игри; уместо тога, они су имали улогу судија. Публика је била подељена у две групе. Свака група је добила скуп енглеских полилексичких јединица: 4 лаке, 4 средње и 1 тешку полилексичку јединицу. Једна особа из групе је пантомимом приказивала полилексичку јединицу, док је остатак групе погађао. Време за погађање је било ограничено на 1 минут за лаке полилексичке јединице, 2 минута за средње и 3 минута за тешке полилексичке јединице. Демонстратор је могао прстима да покаже од колико елемената се састоји полилексичка јединица, а иначе је могао да користи сва средства невербалне комуникације да би покушао да објасни значење сваког појединачног елемента.

На пример, полилексичке јединице једне од група су биле:

⁷ Игре на Вебу

- Лаке полилексичке јединице (1 минут):
 1. *pain in the neck* ЛИТ. БОЛ У ВРАТУ – извор гњаваже, неугодности;
 2. *take the bull by the horns* ЛИТ. УХВАТИТИ БИКА ЗА РОГОВЕ – ухватити се отворено у коштац са проблемом;
 3. *spill the beans* ЛИТ. ПРОСУТИ ПАСУЉ – открити тајну или изненађење;
 4. *strike while the iron is hot* ЛИТ. УДАРАТИ ДОК ЈЕ ГВОЖЂЕ ВРУЋЕ – урадити нешто на време, док за то постоји могућност.
- Средње тешке полилексичке јединице (2 минута):
 5. *cranberry word* ЛИТ. РЕЧ БРУСНИЦА – у лингвистичкој терминологији, морфема ‘брусница’ (или фосилизиран термин) је врста ограничене морфеме која се јавља само у једној или неколико полилексичких јединица и којој се не може придружити независно значење или граматичка функција, као у енглеском изразу *to go astray*, или у српском *ако хоће хоћо, онда неће нећо*;
 6. *red tape* ЛИТ. ЦРВЕНА ТРАКА – колекција или секвенција формулара и процедура који су неопходно да би се добило бирократско одобрење за нешто, посебно ако су сложени и одузимају пуно времена;
 7. *sell like hot cakes* ЛИТ. ПРОДАВАТИ СЕ КАО ВРУЋИ КОЛАЧИ – продавати се јако добро и брзо;
 8. *trip the light fantastic* ЛИТ. ЦУПКАТИ ЛАКИМ (КОРАЦИМА) ФАНТАСТИЧНО – плесати, посебно у балској дворани.
- Тешке полилексичке јединице (3 минута):
 9. *by and large* -- у целини, све у свему.

На крају је публика погодила све полилексичке јединице, иако морамо да признамо да судије нису строго контролисале време. Међутим, једна полилексичка јединица је била погођена и пре почетка демонстрације. Било је довољно да домаћин игре каже „Ово је полилексичка јединица коју сви знају“ и решење је одмах понуђено: *kick the bucket* ЛИТ. УДАРИТИ КОФУ – умрети.

Полилексичке јединице које су обе групе морале да погоде се могу наћи на веб страници догађаја.⁸

6. Квиз са националностима

Као што је речено у одељку 3., део полилексичких јединица које су корисници приложили односе се на националности и оне углавном

⁸ Доступно на Вебу.

одражавају позитивне или негативне стереотипе о њима. Прикупили смо од учесника из 23 земље укупно 80 таквих израза о 32 различите националности или националних група. Ови подаци су онда ручно преведени у графове који прекривају политичку карту Европе, тако да лук из земље А иде ка земљи Б ако на језику који се говори у земљи А постоји полилексичка јединица која се односи на националност земље Б. Оваква репрезентације превише поједностављује однос земаља и националности, али се може сматрати прихватљивом апроксимацијом за наше потребе. Боја сваког лука зависи од поларитета стереотипа који полилексичка јединица изражава:

- црвена: негативан стереотип,
- наранџаста: неутралан,
- зелена: позитиван стереотип.⁹

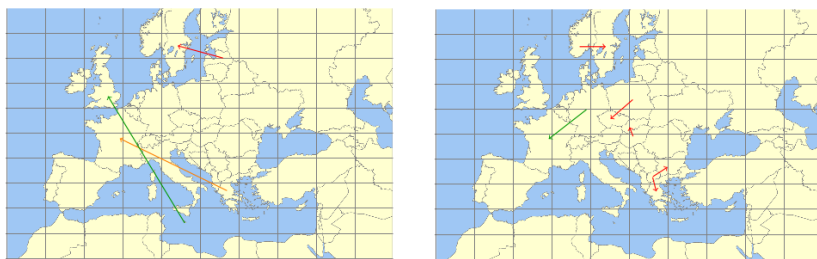
На пример, као што је приказано на слици 1, црвени лук који иде од Летоније ка Шведској представља израз (20), наранџасти лук од Грчке ка Француској одражава (21), док зелени лук од Малте ка Великој Британији одговара изразу (22). Лукови могу да буду шири или ужи у зависности од броја полилексичких јединица које имају исту изворну и циљну земљу као и поларитет.

- (20) *zviēdru gardīnes* (LV)
'шведске завесе' ⇒ ЗАТВОР
- (21) *γαλλικό κλειδί* (EL) – *jaliko klidi*
'француски кључ' ⇒ ПОДЕСИВ ОДВИЈАЧ И ЗАВИЈАЧ
- (22) *għandu ħin Ingliz* (MT)
'имати енглеско темпирање' ⇒ БИТИ ВЕОМА ТАЧАН

Оваква репрезентација олакшава анализу скупа података. На пример, очекивали смо да ће из историјских разлога бити велики број полилексичких јединица које се односе на суседне земље. Међутим, такви изрази нису били веома чести у нашем скупу – видети примере (23)–(28) и слику 2.

- (23) *leben wie Gott in Frankreich* (DE)
'живети као Бог у Француској' ⇒ ЖИВЕТИ ЛУКСУЗНО

⁹ Боје се не виде у штампаној верзији овог рада, али су видљиве у онлајн верзији.

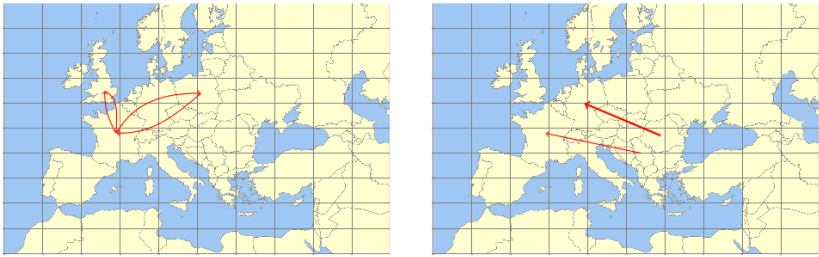


Слика 1. Репрезентација преко **Слика 2.** Полилексичке јединице из графа полилексичких јединица из примера (23)–(28) који се односе на примера (20)–(22) који се односи суседна земље. на националности и поларитет стереотипа.

- (24) tótágast áll (HU)
 ‘стоји као словачка вешалица’ ⇒ СТОЈИ НАОПАКО
- (25) бугарски чорапи (МК) – bugarski chorapi
 ‘бугарске чарапе’ ⇒ БИТИ БОС
- (26) солунски диреци (МК) – solunski direci
 ‘солунски стубови’ ⇒ ДЕБЕЛЕ НОГЕ, ОБИЧНО ЖЕНСКЕ
- (27) svenskemetoden (NO)
 ‘шведски метод’ ⇒ ЛАКО РЕШЕЊЕ
- (28) czeski film (PL)
 ‘чешки филм’ ⇒ СИТУАЦИЈА У КОЈОЈ НИКО НИШТА НЕ РАЗУМЕ

Други занимљив феномен се односи на минималне непразне цикличне путање, то јест, на ситуације у којој се у неколико земаља, обично две, узајамно користе полилексичке јединице негативног поларитета. Нашли смо два случаја оваквих цикличних путањи која се односе на примере (29)–(32) а приказани су на слици 3.

- (29) filer à l’anglaise (FR)
 ‘отићи у енглеском стилу’ ⇒ ОТИЋИ БЕЗ ЗБОГОМ
- (30) to take a French leave (EN)
 ‘отићи на француски начин’ ⇒ ОТИЋИ БЕЗ ЗБОГОМ
- (31) saoul comme un polonais (FR)
 ‘пијан као Пољак’ ⇒ ВЕОМА ПИЈАН



Слика 3. Полилексичке јединице из примера (29)–(32), који одражавају реципрочне негативне стереотипе. **Слика 4.** Полилексичке јединице из примера (23)–(28) у којима се користи антонимија за изражавање сарказма.

- (32) francuski piesek (PL)
 ‘француско кученце’ ⇒ НЕКО КО ЈЕ КАПРИЦИОЗАН И КОМЕ ЈЕ ТЕШКО УДОВОЉИТИ

Што се тиче лингвистичких феномена који се користе у полилексичким јединицама заснованим на стереотипима, антонимија је потврђена као начин за изражавање сарказма. Релевантни примери (33)–(35) су илустровани на слици 4.

- (33) наивна к’о француска собарица (SR) – naivna k’o francuska sobarica
 ЈАКО ИСКУСАН, АЛИ СЕ ПРИКАЗУЈЕ СУПРОТНО
- (34) cald nemțesc (RO)
 ‘немачка топлота’ ⇒ ХЛАДАН
- (35) a merge drept ca neamțul (RO)
 ‘ходати право као Немац’ ⇒ ТЕТУРАТИ СЕ КАО ПИЈАНАЦ

У случају полилексичких јединица с националностима уочен је сличан феномен као и код игре с полилексичким јединицама општег значења: исто значење се у различитим језицима изражава позивањем на друге националности, као у примерима (29)–(30). Уочена су два популарна значења ове врсте ПРАВИТИ СЕ ДА НЕ РАЗУМЕШ and БИТИ ПИЈАН/ПИТИ ПУНО. Она су у нашим подацима представљена са 4, односно 3, националности као што се види из примера (36)–(39), (31), (40)–(42) и на сликама 5–6.¹⁰

¹⁰ Обратите пажњу да када изворна, односно циљна, земља није у Европи, као Кина у примеру (36), почетна, односно, крајња тачка лука нису на мапи.

- (36) κάνω τον Κινέζο (EL) – kano ton Kinezo
'правити се Кинез'
- (37) udać Greka (PL)
'претварати се да си Грк'
- (38) hacerse el sueco (ES)
'понашати се као Швеђанин'
- (39) правити се Енглеz (SR)
ПРАВИТИ СЕ ДА НЕ РАЗУМЕШ
- (40) pít jako Dán (CZ)
'пити као Данац'
- (41) пие како Рус (MK) – pie kako Rus
'пити као Рус'
- (42) пијан като казак (BG) – piyan kato kazak
'пијан као козак'
БИТИ ПИЈАН/ПИТИ ПУНО



Слика 5. Полилексичке јединице из примера (36)–(39) који исказују значење ПРАВИТИ СЕ ДА НЕ РАЗУМЕШ с различитим националностима. **Слика 6.** Полилексичке јединице из примера (40)–(42), који исказују значење БИТИ ПИЈАН/ПИТИ ПУНО с различитим националностима.

Одређене полилексичке јединице изражавају стереотипе који се не тичу одређених националности већ шире популације. Одлични примери долазе из свахилија и тичу се белих људи — видети примере (43)–(44).

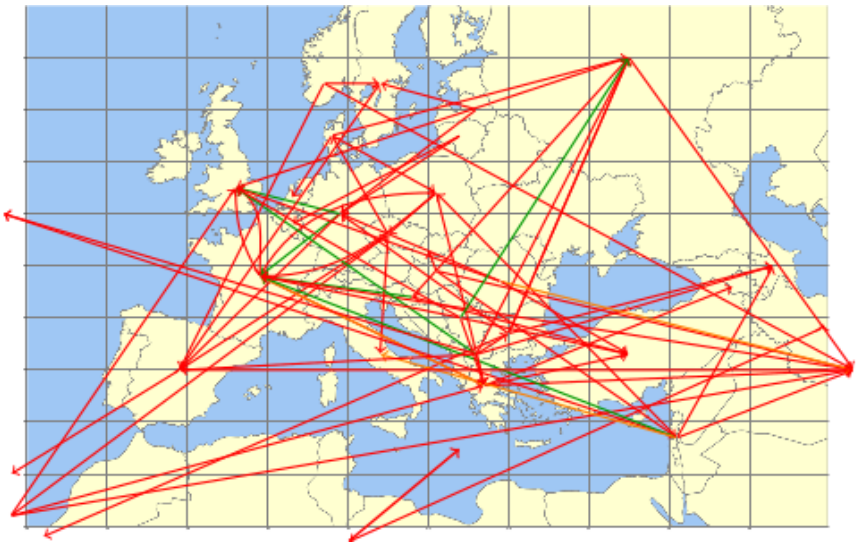
- (43) Mzungu wa reli (SW)
'бели железничар' ⇒ НЕЗНАЛИЦА

(44) Mzungu wa unga (SW)

‘белац од брашна’ ⇒ ДИЛЕР ДРОГЕ

Неки резултати анализе података су понуђени публици у виду једноставних квиз питања. На пример, три европске земље су прикупиле *ex aequo* највећи број улазних лукова, а публика је требало да погоди које су то земље. Сличан квиз се односио на земљу која се помиње у полилексичким јединицама на више језика датог геополитичког региона, конкретно Балкана. Примери који су релевантни за оба ова квиза и одговарајући графови дати су у додатку 2.

Граф који представља све прикупљене полилексичке јединице које се тичу националности приказан је на слици 7. Од укупно 80 полилексичких јединица које су њиме представљене, само 7 изражава позитивне стереотипе, 3 су неутралне, док преосталих 70 израза преноси негативне стереотипе. Цео скуп података је слободно [доступан](#).



Слика 7. Европски граф полилексичких јединица које се односе на националности.

7. Додатни исходи

Многе од прикупљених полилексичких јединица нису искоришћене за наше игре а представљају вредан извор за изучавање. Међу њима је више од 100 примера које су приложили истраживачи јер су изузетно непрозирни или смешни.¹¹ Они се могу анализирати из угла фигурације или транспарентности (Sheinfux et al., 2017). Фигурација се односи на степен до кога се полилексичкој јединици може придружити буквално значење. Већина наших примера из категорије непрозирних или смешних има фигуративно значење, а део забаваног ефекта потиче од неочекиваности слика које изазива њихово буквално тумачење, као у случају (45)–(48).

- (45) viaĵar na maionese (PT)
 ‘путовати у мајонезу’ ⇒ РЕЂИ ИЛИ МИСЛИТИ НЕШТО ГЛУПО ИЛИ ЛУДО
- (46) vziť nohy na ramena (CZ)
 ‘ставити ноге на своја рамена’ ⇒ ПОБЕЂИ
- (47) prendre des vessies pour des lanternes (FR)
 ‘побркати бешику с лампом’ ⇒ НЕ БИТИ У ПРАВУ
- (48) puust piilites (LV)
 ‘дувати патке’ ⇒ ПРИЧАТИ БАЈКЕ

Транспарентност се односи на то колико је лако да се реконструише мотивација за настанак неког идиома захваљујући вези између буквалног и идиоматског тумачења. Забавни ефекат може да потиче од транспарентности, захваљујући хиперболи или сарказму, као у случајевима (49)–(52) или, насупротив томе, од нетранспарентности, као код (53)–(57).

- (49) der er ingen ko på isen (DA)
 ‘нема краве на леду’ ⇒ НЕМА НИЧЕГ ЗАБРИЊАВАЈУЋЕГ
- (50) κόβω καρφιά (EL) – kovo karfia
 ‘сећи нокте’ ⇒ ОСЕЂАТИ ВЕЛИКУ ХЛАДНОЋУ/НЕКОМЕ ЈЕ ВЕОМА ХЛАДНО
- (51) itatja az egereket (HU)
 ‘он чини да мишеви пију’ ⇒ ОН ПЛАЧЕ
- (52) rope på elgen (NO)
 ‘посетити лоса’ ⇒ ПОВРАЋАТИ

¹¹ Цео скуп података је доступан на [страни догађаја](#)

- (53) вешать лапшу на уши (RU) – veshat' lapshu na ushi
'закачити резанце на нечије уши' ⇒ ПРЕВАРИТИ, ГОВОРИТИ
ЛАЖИ
- (54) си легна на брашното (МК) – si legna na brashnoto
'лећи на брашно' ⇒ ПРЕПУСТИТИ СЕ СУДБИНИ
- (55) tirar los tejos a alguien (ES)
'бацати некоме плочице' ⇒ ФЛЕРТОВАТИ
- (56) į jomarką joti (LT)
'возити се на пијацу' ⇒ БИТИ ТУП (О НОЖЕВИМА И МАКАЗАМА)
- (57) kuzunguka mbiyu (SW)
'ићи око дрвета баобаб' ⇒ ПОДМИТИТИ

Анализа података о полилексичким јединицама за десет општих значења које су прикупили корисници (видети оделљак 4.) је показала да се многе исте или сличне метафоре користе у различитим језицима за исто значење. Одабрали смо оне које су се појавиле у бар два језика и понудили нову табелу корисницима да је попуне са сличним полилексичким јединицама у њиховим језицима, ако постоје. Показало се да се најчешће користи стилска фигура поређења *радити као...* у којој се празно место може попунити са *нас* (EN, FR, EL, HR, SR), *коњ* (ET, HR, LV, SR), *магарац* (HE, МК), *мазга*, *стока*, и *црв* (HR, SR), и *во* (PL). Поређење са националностима и етничким групама укључује *црнац* (HR, МК, SR), *Турчин* (LV), and *Кинез* (ES).

За значење РАДИТИ НЕШТО БЕСКОРИСНО прикупљено је више полилексичких јединица:

- (58) מַיִם חַיִּים (HE) – Тахан mayim,
'млети воду'
- (59) толочь воду в ступе (RU) – toloch' vodu v stupe
'млети воду у малтеру'
- (60) a bate ara in piuă (RO)
'тући воду у млину'
- (61) pestare l'acqua nel mortaio (IT)
'млети воду у малтеру'
- (62) pretakati iz šupljeg u prazno (HR)
'пресипати из шупљега у празно'
- (63) کُوبِدَانِ هَشَنِ دَرِ آبِ (FA) – āb dar hāvan koobidan,
'ударати воду у малтеру'
РАДИТИ НЕШТО БЕСКОРИСНО

Табела сличних полилексичких јединица у разним језицима за 10 општих значења је **јавно доступна**.

8. Закључци и перспективе

Претставили смо наше напоре у креирању скупа игара које су засноване на примерима полилексичких јединица на многим језицима које су прикупили корисници. Ове игре, чији је основни циљ био изградња заједништва, показале су се успешне, ако судимо по ентузијазму учесника и позитивном одзиву чланова мреже. Главни разлог овог успеха лежи вероватно у самој природи полилексичких јединица, чија се семантичка некомпозитност може лако искористити за изненађујуће и забавне ефекте.

Ове игре су биле намењене стручњацима у области лингвистике и обраде природних језика, и посебно полилексичких јединица. Било би занимљиво да се истражи њихова примена на ширу публику, пре свега у контексту усвајања страних језика, новинарства, међународних односа и многим другим. Такође верујемо да би се проширивањем опсега и покривености овог истраживања новим изразима могле уочити многе занимљиве корелације међу историјским и лингвистичким феноменима.

Скраћенице

BG ‘бугарски’, CZ ‘чешки’, DA ‘дански’, DE ‘немачки’, EL ‘грчки’, EN ‘енглески’, ES ‘шпански’, EST ‘екстрасензорна перцепција’, FA ‘персијски’, FR ‘француски’, GWAP ‘игре са сврхом’, HE ‘хебрејски’, HR ‘хрватски’, HPU ‘људска јединица обраде’, HU ‘мађарски’, IT ‘италијански’, LT, ‘литвански’, LV ‘летонски’, MK ‘македонски’, MT ‘малтешки’, MWE ‘полилексичке јединице’, NLP ‘обрада природних језика’, NO ‘норвешки’, PL ‘пољски’, PT ‘бразилски португалски’, RO ‘румунски’, RU ‘руски’, SR ‘српски’, SW ‘свахили’

Додаци

1. Решења игре с полилексичким јединицама општег значења словенске групе из одељка 4.

(8) – а (УМРЕТИ/БИТИ МРТАВ)

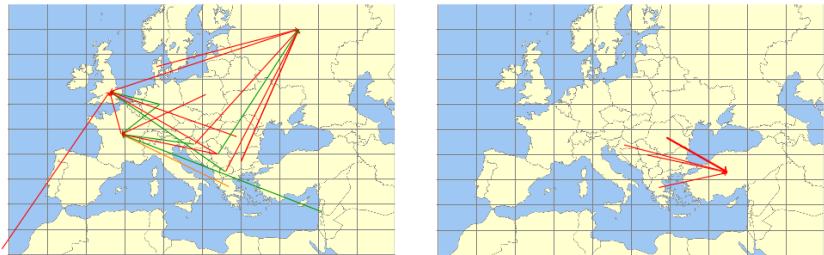
(9) – б (БИТИ УМОРАН)

(10) – б (БИТИ УМОРАН)

(11) – а (УМРЕТИ/БИТИ МРТАВ)

- (12) – г (БИТИ ЉУТ)
- (13) – а (УМРЕТИ/БИТИ МРТАВ)
- (14) – в (БИТИ ГЛУП)
- (15) – в (БИТИ ГЛУП)
- (16) – г (БИТИ ЉУТ)
- (17) – в (БИТИ ГЛУП)
- (18) – г (БИТИ ЉУТ)
- (19) – б (БИТИ УМОРАН)

2. Решења квиза о највећем броју улазних лукова (глобално и локално) из одељка 6.



Слика 8. Француска, Русија и **Слика 9.** Турска: земља која је Велика Британија: три европске привукла највећи број долазних земље с највећим бројем улазних лукова који потичу из балканских лукова. Узорак полилексичких земаља. Узорак полилексичких јединица је наведен у примерима јединица је наведен у примерима (67 (64)-(66))-(70)

- (64) (só) pra inglês ver (PT)
'(само) за Енглезе да виде' ⇒ (САМО) ЗА ПОКАЗИВАЊЕ
- (65) Francuski poljubac (HR)
ФРАНЦУСКИ ПОЉУБАЦ
- (66) en by i Rusland (DA)
'град у Русији' ⇒ НЕШТО ПОТПУНО НЕБИТНО

- (67) Puši k'o Turčin (HR)
a fuma sa un turc (RO)
'пушити као Турчин' ⇒ ПУШИТИ МНОГО
- (68) γίνομαι Τοῦρκος (EL) – jinome Turkos
'постати Турчин' ⇒ ПОСТАТИ БЕСАН
- (69) turcul plătește (RO)
'Турчин плаћа' ⇒ НЕКО ТРЕБА ДА ПРЕУЗМЕ ОДГОВОРНОСТ ЗА НЕШТО ШТО НИЈЕ УРАДИО
- (70) пролазити као поред турског гробља (SR)
ПРОЛАЗИТИ ПОРЕД НЕКОГ ИЛИ НЕЧЕГА НЕ ОБРАЂАЈУЋИ ПАЖЊУ

Литература

- Von Ahn, Luis and Laura Dabbish. "Designing Games with a Purpose". *Commun. ACM* Vol. 51, no. 8 (2008): 58–67. <http://doi.acm.org/10.1145/1378704.1378719>
- Baldwin, Timothy and Su Nam Kim. "Multiword Expressions.". *Handbook of natural language processing* Vol. 2 (2010): 267–292.
- Cooper, Seth, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite et al.. "The Challenge of Designing Scientific Discovery Games". In *Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG '10*, New York, NY, USA: ACM, 40–47, 2010. <http://doi.acm.org/10.1145/1822348.1822354>
- Gurevych, Iryna and Torsten Zesch. "Collective Intelligence and Language Resources: Introduction to the Special Issue on Collaboratively Constructed Language Resources". *Language Resources and Evaluation* Vol. 47, no. 1 (2013): 1–7.
- Howe, Jeff. "The rise of crowdsourcing". *Wired magazine* Vol. 14, no. 6 (2006): 1–4.
- Kittur, Aniket, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw et al.. "The Future of Crowd Work". In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13*, New York, NY, USA: ACM, 1301–1318, 2013. <http://doi.acm.org/10.1145/2441776.2441923>
- Lafourcade, Mathieu, Alain Joubert and Nathalie Le Brun. *Games with a Purpose (GWAPS)*. John Wiley & Sons, 2015.
- Millour, Alice and Karën Fort. "Why do we Need Games? Analysis of the Participation on a Crowdsourcing Annotation Platform". In *Games4NLP*, 2017.

- Mitrović, Jelena. “Crowdsourcing and its application”. *INFOtheca* Vol. 14, no. 1 (2013): 37–46.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. “Multiword Expressions: A Pain in the Neck for NLP”. *Computational Linguistics and Intelligent Text Processing* (2002): 189–206.
- Sheinfx, Livnat Herzig, Tali Arad Greshler, Nurit Melnik and Shuly Wintner. *Representation and Parsing of Multiword Expressions*, ed. Yannick Parmentier and Jakub Waszczuk, 5–38. Chapter Verbal MWEs: Idiomaticity and flexibility. Berlin: Language Science Press, 2017.
- Tuite, Kathleen. “GWAPs: Games with a Problem.”. In *FDG*, 2014.