
Ученички корпуси србофоних говорника енглеског језика у контрастивној анализи међујезика на примеру језичке неодређености

Научни рад

DOI: 10.18485/judig.2025.1.ch6

Јелена Марковић¹  0000-0003-4962-3754

Апстракт

У овом чланку говоримо о контрастивној анализи међујезика из угла србофоне академске заједнице, односно описујемо постојеће ученичке корпусе србофоних говорника енглеског језика: *ICLE-SE*, компоненту тренутно најзначајнијег ученичког корпуса енглеског језика, *The International Corpus of Learner English v3* (UCLouvain, Белгија), као и *Корпус студената англистике (КорСАНг)* при Центру за дигиталну хуманистику Филозофског факултета Пале, користећи их на примеру истраживања феномена језичке неодређености.

Феномен језичке неодређености посматрамо доминантно кроз неодређено *some*, уз неодређене заменичке облике *somebody*, *someone*, *something*, односно еквивалентне форме неодређеног *неки* у српском језику. Резултати, које добијамо након анализе и тумачења коришћења циљних експонената језичке неодређености у писању на енглеском као страном у корпусу *ICLE-SE*, указују на учестало и снажно ослабање србофоних говорника на поменуте експоненте, неретко на уштрб јасноће или природности исказа. Даљим поређењима уочавамо да србофони говорници енглеског значајно чешће посежу за наведеним облицима од изворних говорника других матерњих језика. Анализирање употреба еквивалентних облика у писању на српском као матерњем језику на примерима корпуса *КорССАНг* уверава нас да

¹ Универзитет у Источном Сарајеву, Филозофски факултет Пале, jelena.markovic@ff.ues.rs.ba

постоји таква тенденција и у овом корпусу, како по фреквентности тако и по снажном ослањању на еквивалентно *неки*.

Закључујемо да је тенденција коришћења неодређеног језика у енглеском као страном код србофоних говорника резултат како универзалних тенденција у писању на страном језику тако и културолошких и језичких особености писања на српском као матерњем језику. Додатно закључујемо да је за сродна методолошка истраживања контрастивна анализа међујезика незаобилазна методологија у истраживањима ученичких корпуса, а поменути србофони ученички корпуси ресурси значајног потенцијала за сродне теме у примењеној лингвистици на српском говорном подручју.

Кључне речи: контрастивна анализа међујезика, *International Corpus of Learner English v3*, *КорСАНг*, језичка неодређеност

1. Увод

Појављивање контрастивне анализе међујезика (енгл. *Contrastive Interlanguage Analysis*),² као новог типа контрастирања у примењеној лингвистици, резултат је редефинисања и иновирања контрастивне методологије у теорији учења и усвајања страног језика. У новонасталом дигиталном окружењу овај нови тип контрастирања пратило је формирање првих обимних електронских ученичких корпуса (енгл. *learner corpora*), као нове парадигме података у истраживањима у оквиру примењене лингвистике (Granger, 2002; Gilquin, Granger, & Raquot, 2007), за разлику од до тада коришћене оскудне парадигме језичке продукције.³ Појава ученичких корпуса стога представља много више од основе за иновирање контрастирања, јер заправо представља нову фазу у развоју теорије учења и усвајања страних језика, која је за кратко време свог постојања дала велики научни допринос широм света: објављивање бројних монографских и других публикација, осмишљавање нових универзитетских курсева на различитим нивоима, извођење докторских истраживања, покретање нових научних часописа, уз значајан пратећи теоријски допринос у

² Термин *Contrastive Interlanguage Analysis* први пут је употребила Силвијана Грејндер [Sylviane Granger] 1996. године у раду под називом “From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora”.

³ Изненађујуће је колико је језичка продукција неизворних говорника до појаве ученичких корпуса заправо каскала у периоду процвата примењене лингвистике, очигледно занемарена у великом броју како методолошких тако и теоријских неусаглашености и изазова с којима се примењена лингвистика сусретала у својим првим деценијама формалног постојања (Chaudron, 2003: 774–776; Gass & Polio, 2014: 168–169; Марковић, 2019).

комплементарним дисциплинама попут филозофије језика, прагматике, анализе дискурса, социолингвистике и стилистике.

Критеријуми формирања електронских корпуса у циљу сврсисходности морају да буду строги и експлицитни. Наиме, морају се бележити примери језичке продукције најпре заједно са пратећим и релевантним метаподацима, а одлука о томе који ће метаподаци бити укључени зависи од проницљивости и расуђивања у предвиђању које се варијабле могу сматрати значајним за будуће истраживаче, што јесте захтеван корак у формирању електронских база аутентичне језичке продукције. Подразумева се да се метаподаци, без обзира на то које се варијабле бирају, бележе систематично и доследно, будући да једино такво бележење одабраних метаподатака има научни потенцијал. Метаподаци могу бити класификовани у више типова метаподатака, попут варијабли самих говорника (енгл. *learner variables*) или варијабли контекста језичке продукције (енгл. *task variables*). Као примере варијабли самих говорника наводимо узраст, односно старосну доб, дужину учења страног језика или познавање других страних језика. Као примере варијабли контекста језичке продукције наводимо циљни жанр, фактор могућности коришћења извора или речника, фактор интегрисаности говорне ситуације у испитне задатке или формално образовање или фактор временског ограничења говорне ситуације.

Нагласићемо да се критеријум одређивања дискурса као аутентичног у језичкој продукцији изворних говорника неумитно редефинише у односу на одређивање аутентичности у језичкој продукцији изворних говорника. Наиме, за разлику од животног окружења у којем се свакодневно аутентичност говора изворног језика заправо подразумева, простор за апсолутну аутентичност у учењу страних језика у учионици веома је ограничен (Sinclair, 1996; Granger, 2002: 8). Стога се аутентичном продукцијом на страном језику може сматрати језичка продукција која има за циљ комуникацију на страном језику, а може се одвијати у условима који су делимично контролисани, како истиче и Елис (Ellis, 1994: 671). Другим речима референтност корпуса у истраживањима ученичких корпуса по критеријуму аутентичности постиже се у прихватљивом степену уколико се продукција може сматрати комуникацијом у одређеном контексту. С друге стране, пошто се намеће потреба контекстуализоване писане дискурсне продукције у ученичким корпусима, један од начина постизања циља у формирању корпуса јесте, на пример, писање састава, а не појединачних реченица.

Ревидирани методолошки модел контрастивне анализе међујезика (Granger, 2015: 17) иновиран је у односу на претходно дефинисани првенствено истицањем варијабилности, односно дефинисањем варијанте језика која се истражује, како у случају изворног тако и ученичког корпуса, с тим што се под варијантом језика може подразумевати свако специфично раслојавање у употреби језика.

2. International Corpus of Learner English v3

Ученички корпус *International Corpus of Learner English v3* (*ICLE v3*), формиран је при Католичком универзитету у Новом Лувену (фр. Université catholique de Louvain, Louvain-la-Neuve, скр. UCLouvain), односно Центру за англистичку корпусну лингвистику (*Centre for English Corpus Linguistics*) као трећа верзија корпуса, публикована 2020. године (Granger et al., 2020).⁴ Ова верзија садржи око 5,7 милиона речи, односно 25 подкорпуса одређених матерњим језиком говорника, међу којима је и српски језик.

На платформи која подржава корпус,⁵ интегрисани су одговарајући корпусни алати, тако да је коришћење платформе додатно олакшано. Корпус је иначе и морфосинтаксички тагиран, а приручник садржи детаљна упутства о коришћеном сету ознака за основну и напредну претрагу. Додаћемо да напредна претрага није само именом напредна, већ омогућава задавање веома различитих упита, континуираних и дисконтинуираних, што, уз све претходно наведено, пружа бројне истраживачке могућности. При том истичемо да упити нису увек и искључиво засновани на формалним карактеристикама језичких јединица, већ су често у питању комбинације формалних и семантичких, односно функционалних карактеристика упита, уз висок степен тачности резултата.

3. ICLE-SE, ученички корпус србофоних студената англистике

Један од подкорпуса који је интегрисан у поменути корпус *ICLE* јесте корпус србофоних говорника енглеског језика као страног, *ICLE-SE*, а формиран је захваљујући учешћу четири универзитета, односно четири катедре за енглески језик на српском говорном подручју (Источно Сарајево, Бања Лука, Београд, Нови Сад). Корпус се састоји од 325 састава просечне дужине од 623 речи, доминантно аргументативног типа, које су писали углавном студенти завршних

⁴ Прва верзија публикована је 2002. (Granger et al., 2002), а друга верзија 2008. године (Granger et al., 2008).

⁵ <https://corpora.uclouvain.be/cecl/icle/home>

година првог циклуса студија. Сваки учесник припремио је по један састав, што је омогућило формирање веома репрезентативне базе података.⁶

Корпус *ICLE-SE* формиран је углавном без временског ограничења, такође углавном без коришћења речника, приручника или других референтних издања, а у условима који нису подразумевали испитне задатке.

Сликовит опис тематског оквира корпуса *ICLE-SE* даје набрајање неколико најчешће коришћених наслова састава, редом од најпопуларнијег, а то су били следећи:

- Is reading a book better than watching its movie adaptation?
- One should never judge a person by external appearances.
- Most university degrees are theoretical and do not prepare students for the real world.
- Are we too dependent on computers?
- In the words of the old song "Money is the root of all evil." Agree or disagree.
- Conflicts are necessary for healthy relationships.
- Does age matter in relationships?
- Discuss the advantages and disadvantages of banning smoking in restaurants.
- In our modern world, dominated by science, technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?
- Discuss the necessity of make-up.

4. Корпус КорСАНг

Корпус студената англистике (*КорСАНг*)⁷ формиран је при Филозофском факултету Универзитета у Источном Сарајеву, у периоду од 2016. до 2019. године (Радоња и Шућур, 2021). Настао је захваљујући сарадњи четири катедре за англистику на српском говорном подручју, у Источном Сарајеву, Бањој Луци, Београду и Новом Саду, у оквиру два национална пројекта при ресорном

⁶ Индикативно је истаћи и резултат који је остварен у оквиру оцењивања насумично одабраних 20 састава из сваког националног подкорпуса, а у складу с Заједничким европским оквиром (Common European Framework of Reference for Languages). Наиме, за корпус србофоних говорника, *ICLE-SE*, односно за његових 20 састава, 4 састава оцењена су као ниво В2, чак 11 састава оцењено је као ниво С1, а преосталих 5 састава као ниво С2, што је свеукупно резултат који вреди истаћи (Granger et al., 2020: 12).

⁷ <https://noske.rgf.rs/>

Министарству Владе Републике Српске.⁸ Корпус има око 140.000 речи, односно 327 текстова. У формирању корпуса укупно је учествовало 194 студената англистике, изворних говорника српског језика.

По саставу овај корпус је трокомпонентан. Садржи корпус превода са енглеског језика (*КорПСАнг1*), корпус превода на енглески језик (*КорПСАнг2*) и корпус аргументативних састава студената англистике писаних на српском језику као матерњем (*КорССАнг*). Прве две компоненте јесу паралелни ученички корпуси, док је трећа компонента, *КорССАнг*⁹ осмишљена као референтни корпус корпусу *ICLE-SE*, претходном описаном подкорпусу међународног ученичког корпуса аргументативних састава *ICLE* у трећој верзији.

Корпус *КорПСАнг1* састоји се од 127 превода са енглеског на српски језик (око 47.200 токена). Полазни текстови на енглеском језику укључују и књижевне и некњижевне текстове. Један учесник припремио је највише два превода (једног књижевног и једног некњижевног текста). Сваки превод има свој код који означава правац превођења, универзитет, врсту текста, годину студија учесника и редни број превода за дати текст на том универзитету. Укупно је коришћено шест различитих текстова, који су преузети из више извора, углавном часописних за некњижевне текстове.

Корпус *КорПСАнг2* састоји се од 130 превода са српског на енглески језик (око 51.000 токена). Полазни текстови на српском језику имају и књижевни и некњижевни карактер. Један учесник припремио је највише два превода: један књижевни и један некњижевни текст. Сваки превод има свој код, формиран по принципу као и за *КорПСАнг1*. Укупно је коришћено шест различитих текстова из више извора, углавном часописних за некњижевне текстове.

Корпус *КорССАнг* састоји се од 70 аргументативних састава просечне дужине око 570 речи (око 45.000 токена). Састави су писани на опште теме, тако да чине референтни корпус ученичком корпусу србофоних говорника *ICLE-SE*. Наслови састава следе, заједно с бројем састава који су писани на појединачне теме (Прилог 2):

⁸ Називи пројеката су „Фразеолошка компетенција српских говорника енглеског кроз призму контрастивне анализе међујезика“ и „Научни потенцијал анотираних ученичких корпуса кроз призму контрастивне анализе међујезика“, оба остварена уз подршку ресорног Министарства Владе Републике Српске (редом, број 19/6-020/961-46/18, од 31. 12. 2018. године и број 19.032/961-135/19, од 31. 12. 2019.).

⁹ Комплетан корпус *КорСАнг* називамо ученичким у овом чланку, узимајући у обзир његов састав, као и могуће употребе у истраживањима. Прецизније би било описати компоненте превода ученичким корпусима, а компоненту *КорССАнг* корпусом састава изворних говорника.

- Породица или посао: шта је важније? (9)
- Како музика утиче на живот? (8)
- Како замишљам доброг родитеља (7)
- Од склапања брака до развода данас (6)
- Сиромаштво и здравље (5)
- Да ли друштвене мреже нарушавају односе међу људима? (4)
- Ја (не) бих у Европску Унију (3)
- Пластична хирургија: да или не? (3)
- Млади не треба да се баве политиком (3)
- Да ли је брак превазиђена институција? (3)
- Зашто се гријеши при избору професије? (2)
- Зашто (не) желим да будем наставник страног језика (2)
- Слобода медија данас (2)
- Будућност књиге (2)
- Пушење на јавним мјестима: да или не? (2)
- Како замишљам идеално друштво (2)
- Љепота је пролазна (2)
- Родитељи би требало да поштују личност своје дјеце (1)
- Шта решавају затворске казне? (1)
- Значај образовања у друштву у коме живимо (1)
- Да ли критичко мишљење има своје мјесто у нашем образовном систему? (1)
- Без муке нема науке (1)

Сваки учесник припремио је по један аргументативни састав, без изузетка. На тај начин обезбеђена је максимална репрезентативност овог корпуса, иначе наизглед релативно малог обима у поређењу с другим корпусима. Сваки састав обележен је својим кодом.

Комплетан корпус *КорСанг* анотиран је вишеструко. Најпре, урађена је библиографска анотација, која је повезала постојеће метаподатке по варијаблама са текстовима учесника. Даље, анотирани су сви текстови врстама речи (*PoS tagging*), а затим је урађено и поравнање паралелних корпуса превода на нивоу реченице (*sentence alignment*). Приликом обраде текста сачувано је писмо које су учесници користили у писању или куцању. У текстовима су отклоњене једино поједине очигледне техничке грешке у куцању, у циљу постизања ефективнијег претраживања корпуса. Изворни текстови представљени су у оригиналном облику, како су задати (без форматирања). Корисницима је омогућено претраживање корпуса путем корисничког интерфејса. Планирано је и проширење овог

корпуса укључивањем нових сродних текстова, по истим принципима формирања, посебно када је у питању трећа компонента, односно корпус референтан корпусу *ICLE-SE*.¹⁰

5. Феномен језичке неодређености

У оквиру овог чланка одабрали смо феномен језичке неодређености као студију случаја којом дајемо увид у могућности коришћења србофоних ученичких корпуса, *ICLE-SE* и *CorCCAng*, у примењенолингвистичким истраживањима.

Феномен језичке неодређености или неодређеног језика (енгл. *vague language*) представља једну од инхерентних особености природних језика, иако наизглед природни језици теже прецизности и одређености. Већ почетком двадесетог века Расел је забележио да је језик, уопштено говорећи, суштински мање или више нејасан или неодређен, управо због различите природе механизма природних језика и свеукупности значења која се изражавају (Russell, 1923). Нешто касније, Карл Попер [Carl Popper] је нагласио да се језик заправо суштински опире или отима потпуној прецизности, насупрот нашим очекивањима.¹¹

Феномену језичке неодређености у новијем периоду поклања се више пажње, првенствено у оквиру интердисциплинарних области анализе дискурса, примењене лингвистике и прагматике. Теорија језика, постало је извесно, уколико претендује да буде потпуна теорија, неизоставно укључује феномен језичке неодређености (Channell, 1994: 5). У најширем смислу, можемо рећи да је неодређени језик заправо резултат одступања од прецизности и јасноће у комуникацији (Jucker et al., 2003: 1737), али при том одмах треба истаћи да је коришћење неодређеног језика заправо типично у свакодневној употреби језика. Другим речима, језичка неодређеност само наизглед личи на непожељан феномен. У суштини, није у питању некакав недостатак или мањкавост, већ заправо незаобилазна компонента комуникативне, односно стратегијске и прагматичке компетенције којом се свакодневно служе изворни говорници (Carter & McCarthy 2006: 202; Parvaresh & Ahmadian, 2016: 437). Језичка

¹⁰ Користимо прилику да обавестимо научну јавност на српском говорном подручју да у предстојећем периоду планирамо проширивање корпуса састава на српском језику чији су аутори студенти, изворни говорници српског језика, примарно филолошке оријентације.

¹¹ Видоусон (Widdowson, 1984: 58) преноси мишљење Карла Попера, који каже да је трагање за прецизношћу ("the quest for precision") аналогно трагању за сигурношћу ("the quest for certainty"), а да је у језику немогуће испунити такве тежње, јер је идеја о прецизности у језику погрешна. Језик се опире прецизности због своје суштинске недоречености.

неодређеност заправо представља „напор који користимо попут стратегије у постизању циљева комуникације“¹² (Parvareh & Tayebi, 2014: 597).

Неодређеност може бити изражена језичком формом која нема прецизно значење у датом контексту (Zhang, 2011). То може бити реч, синтагма или комплекснија језичка јединица. Неки од експонената у савременом енглеском језику као матерњем у свакодневној комуникацији јесу *thing, stuff, or so, like, something, anything, and so on, whatever, kind of, sort of* (Carter and McCarthy, 2006: 202).

Неодређеност језичких експонената реализује се у специфичном контексту, будући да значење истог експонента неодређености може бити другачије у другачијем контексту (Zhang, 2011: 572):

- There are *many* people in his room.
- There are *many* people in the park.

Некада је употреба неодређених језичких израза језика заправо једини избор говорника, јер се саопштава информација која је сама неодређена, а некада је то заправо намерни и вољни избор говорника који бира неодређени језик као релевантнији у датој говорној ситуацији, као што се истиче у савременим студијама (Jucker et al., 2003: 1737, McGee, 2018: 45):

- *Many* friends attended her birthday party. (Zhang, 2022: 274)

Many може имати различите интерпретације у зависности од више фактора: нпр. особе која слави (у смислу популарности слављенице, односно ширини круга пријатеља), површине простора за прославу или броја гостију које простор може да прими (Zhang, 2022: 274), што уједно показује да неодређени облик може бити прецизнији од навођења конкретног броја присутних у датој ситуацији.

Можемо резимирати наведено констатацијом да је коришћење неодређеног језика заправо манифестација прагматичке компетенције говорника. Употреба језичких експонената неодређеног значења у језику мотивисана је и сврсисходна, тако да представља индикатор вештине (изворног) говорника, односно промишљено језичко понашање у интеракцији (Jucker et al., 2003: 1739).

¹² “an endeavor strategically made to achieve certain communicative goals” (превод аутора)

6. Неодређеност у продукцији на страном језику

Једна од основних карактеристика међујезика јесте његова варијабилност, видљива у језичкој продукцији неизворних говорника (Saville-Troike, 2006). Паралелно с израженим степеном варијабилности, неизворни говорници у покушају да надоместе своје језичке недостатке, користе различите стратегије, односно стратегијску компетенцију, значајно више од изворних говорника, који имају на располагању језичку компетенцију и познавање релевантних социолингвистичких, културолошких и прагматичких феномена.

Међу препознатљивим стратегијама неизворних говорника налази се и когнитивна тенденција ка генерализацији, односно коришћењу уопштенијих, неодређених језичких форми како у превођењу на страни језик (Spajić & Suknović, 2019) тако и у писању на страном језику (Hasselgren, 1994; Lee & Chen, 2009). И само разумевање неодређеног језика битно је различито код изворних и неизворних говорника. Стога можемо истаћи да је веома захтевно за неизворне говорнике да овладају коришћењем различитих дискурских карактеристика страног језика, укључујући и прагматички оправдано коришћење језичке неодређености (Evison et al., 2007; Parvaresh & Ahmadian, 2016; Marković, 2019; Marković, 2021), посебно имајући у виду комплексност значења и употреба израза неодређености у изворном језику.

7. Неодређени језик у корпусу ICLE-SE

У наставку следе најпре примери употребе неодређеног језика из србофоног корпуса *ICLE-SE*, остварени коришћењем различитих језичких механизма:

- However, there are *some* cases in which this *kind of* an obstacle is certainly relevant. SEBJ1016¹³
- *Sometimes*, we all get tired, or there are *some* days when we are not pretty to ourselves, so why not put *some* make-up on. (SEBJ1004)
- Just because *someone* looks a *certain* way it doesn't mean they are good or bad. (SENS1021)
- There are *some* movies that all of us love and we can watch them over and over again, [...]. (SEBJ1011)
- It can happen to *anyone* to fall in love with *someone* who is older or younger, or *someone* who is different or *whatever*. (SEBJ1022)

¹³ Након појединачних примера, назначен је код састава из корпуса.

Понављање језичких јединица неодређеног значења, различитих или некада чак истих, веома је често у оквиру исте реченице код србофоних говорника, чиме се стиче утисак неприродности, односно неизворности дискурса:¹⁴

- Long, long time ago, *someone* very conservative and narrow-minded said that, in order to do a *certain* job you have to look a *certain* way. (SENS2001)
- Unlike *many*, *some* companies realised *how many* problems smoking can cause for a *large number of* people, so they put warning messages on every cigarette pack like <*> or <*> and so on. (SEBJ1003)

У фокусу овог чланка јесу првенствено неодређено *some*, уз неодређене заменичке облике *somebody*, *someone*, *something*, који су и одабрани као експоненти језичке неодређености у корпусу србофоних говорника *ICLE-SE*, а које називамо циљним неодређеним облицима у наставку текста.

У корпусу србофоних говорника учесталост наведених облика¹⁵ релативно је висока (0,87%), што је и очекивано и разумљиво будући да је у питању писање на енглеском као страном језику. Ипак, поређење с ученичким корпусима говорника других страних језика из истог међународног корпуса *ICLE* (укупно 18, видети Табелу 1., Прилог) показује да је код говорника српског језика, уз турски, учесталост већа од говорника других матерњих језика (бугарског, чешког, кинеског, холандског, финског, француског, немачког, грчког, мађарског, италијанског, литванског, македонског, персијског, пољског, руског, шпанског и шведског), као и од процентуалне заступљености на нивоу комплетног корпуса (0,53%) и референтног дела корпуса *LOCNESS* (0,39%), иначе корпуса изворних говорника (Granger, 1998).¹⁶ Другим речима, иако, са једне стране, постоји општа тенденција коришћења израза језичке неодређености код неизворних говорника, с друге стране учавамо да је присуство циљних неодређених облика учесталије управо у корпусу србофоних говорника, што је и предмет нашег истраживања.

¹⁴ Приликом формирања коментара о наведеним примерима употребе из корпуса затражили смо и мишљење лектора, изворног говорника америчке варијанте енглеског језика.

¹⁵ Задата претрага укључила је *some* и све сложене неодређене облике чија је прва морфема *some**

¹⁶ Корпус састава изворних говорника који се типично користи као референтни корпус у истраживањима због сличних социолингвистичких особености говорника формиран при Центру за англистичку корпусну лингвистику у Новом Лувену (Centre for English Corpus Linguistics, Université catholique de Louvain, Louvain-la-Neuve).

(<https://www.learnercorpusassociation.org/resources/tools/locness-corpus/>).

Из конкорданци које су генерисане коришћењем интегрисаних корпусних алата на платформи корпуса *ICLE* издвојили смо неколико група примера употребе. Најпре наводимо примере који оправдано и прагматички јасно користе неодређени облик *some*, чак и поновљен у оквиру исте реченице:

- *Some* adaptations are successful, and *some* are not, mainly because *some* movies are successful, and *some* are not. (SEBG2021)
- *Some* people are against big age differences, *some* are not. (SEES2012)
- *Some* film adaptations are hits, and *some* are misses, but those hits hit hard. (SEBG2021)
- *Some* people that are under twenty seem to be more mature than *some* people that are over that age. (SEBJ1031)
- *Some* people tend to hide it, *some* are not ashamed to admit that it can't all be rainbows and butterflies, and *some* just don't give it too much thought. (SEBG2004)
- Пример употребе других неодређених облика следи:
- The question is how long we will wait until *someone* will do *something* serious to improve education system in Serbia. (SEBG2001)

Поједини примери нису неразумљиви, али је чињеница да коришћење неодређених облика одузима флуентност и захтева пажњу приликом читања:

- For instance, if *someone* sees a man with considerably younger woman the first thing that comes to their mind is that he is probably an extremely wealthy businessman and that finances are *the thing* that brought them together. (SEBG1026)¹⁷
- In the end we should not reject movie adaptations completely, because *sometimes* we can learn *something* new from them. (SEBJ1014)

За разлику од претходних, чести су примери где је неразумљивост евидентна, с тим што је управо коришћење неодређеног језика извор нејасноће. Такве примере било би потребно преформулисати у циљу постизања јасноће:

¹⁷ У корпусу *ICLE* задржана је потпуна аутентичност текстова неизворних говорника, укључујући и неправилности у формама речи (нпр. *financies).

- Even if *someone* is not trying to be *someone* they are not by their appearance, how they like to dress does not have to represent their personality. (SEBG2017)¹⁸
- Secondly, *everyone*, based on personal preferences and characteristics, is looking for *someone* who suits them, for *someone* it is an older person, and for *someone* younger. (SEBJ1016)¹⁹
- Generally, *some* people are more likely to judge *other* people immediately after they meet *someone* for the first time than *other* people. (SEBJ3005)²⁰

У следећим примерима непотребно оптерећење представља управо неодређено *some*. Стога, оправдано га је изоставити из примера који следе:

- But quite often movies that are adaptations of *some* books are not as good as the books. (SEBJ1027)
- In some cases, as *some* people are with their clothes, so they are with their habits. (SENS3014)
- Money is largely responsible for the enormous economic growth of *some* countries like America, Russia, China, Australia etc., where the utopian dream proposed by many a scholar has been realized. (SEBG2048)
- Native speakers are able to acquire some new words, or to add *some* new meanings to already known words. (SEES2026)
- What I remember best is that the invention of money immediately triggered *some* conflicts among people. (SENS1027)
- When *some* people hear where you're coming from, they make a picture of you and your country [...].(SEBJ1039)
- Yes, I still haven't lost hope in the movie industry and I hope there will be *some* better adaptations of this everlasting story in the future. (SEBG1028)

У примерима који следе уместо неодређеног *some* боље је користити неодређени члан, *a*:

¹⁸ Even if someone is not trying to express themselves through their appearance, how they like to dress does not have to represent their personality. (SEBG20172)

Напомена: Предлоге свих наведених парафраза у фуснотама, које су обележене индексом 2, припремио је изворни говорник енглеског језика америчке варијанте.

Конкретно за наведени пример изворни говорник је назначио да је извор нејасноће и језичка неодређеност и двострука негација у реченици.

¹⁹ Secondly, everyone, based on personal preferences and characteristics, is looking for someone who suits them; for one, it might be an older person, and for another, someone younger. (SEBJ10162)

²⁰ Generally, some people are more likely to judge other people immediately after they meet someone for the first time. SEBJ3005²

- Secondly, descriptions given in books never tell you something certain; for example, it is up to you how you will imagine *some* character's physical appearance, his/her facial expressions or the way he/she moves. (SEBG2030)
- Somebody says there is *some* boundary of five years difference and that should be ok. (SEBJ1031)
- All things considered, we can say that reading *some* book is always better than watching a movie. (SEES1008)

У примерима који следе уместо неодређеног *some* боље је користити неодређени придев *certain* или *various* (+pl.) у завршној реченици:

- Firstly, there are some universities that do not have the appropriate subjects for *some* degrees, so students are obliged to study subjects that are not so closely related to their major. (SEES2008)
- Age is not a crucial thing when you have to discuss *some* issues. (SEBJ1031)
- There are *some* social conventions that some people tend to follow. (SEBJ1050)
- [W]e learn to value different culture by being taught on different aspects from *some* country, such as literature, customs, its history, etc. we develop our vocabulary, listening, writing and speaking skills. (SEBG2051)²¹

8. Неодређени језик у корпусу КорССАнг

Значај формирања корпуса *КорССАнг*, који садржи око 45.000 речи или 70 аргументативних састава писаних на српском језику чији су аутори изворни говорници српског, може се видети управо у поређењу које представљамо. Наиме, након запажања да су циљни неодређени облици учесталији код србофоних говорника, приступили смо анализи изворног корпуса *КорССАнг* и претражили неодређено *неки* у њему.²² Напомињемо да ниједан од наслова састава коришћених за формирање корпуса не садржи употребе циљних неодређених облика, што би у противном могло бити доведено у везу с учесталим употребама. Резултат обављене претраге јесу укупно 264

²¹ [W]e learn to value different cultures by being taught different aspects from various countries, such as literature, customs, its history, etc.; we develop our vocabulary, listening, writing and speaking skills. SEBG2051²

²² [lemma_lc="неки"]

примера употребе, односно учешће с 0,58% заступљености.²³ Одабрали смо неколико примера.

Најпре, има примера употребе, попут следећих, које можемо оценити као јасне и разумљиве изворним говорницима:

- Kada *neko* kaže „heroj“, прва асоцијација на ту ријеч је моја мајка. (*КорССАн232*)
- Као да је знанје *neki* колач кога има у одређеној количини, па што га више људи дели, то мање свако појединачно добија. (*КорССАн263*)
- Нпр. мени се често дешава да чујем *neku* пјесму која има неприкладан или totalно бесмислен текст, али veoma добар ритам. (*КорССАн234*)
- Нпр. *neka* особа слуша одређену пјесму кад рaskine са својим партнером. (*КорССАн234*)
- Кrv, novac, estrada, и *neke* екстремно битне информације попут, gdje спрва Cecin sin večeras. (*КорССАн23*)
- [I]мамо случај да *neke* земље (Нјемачка, Француска...) имају потписан споразум о безвизном уласку у неке од најразвијенијих земаља свијета (САД, Канада) [...]. (*КорССАн215*)

У наредним примерима умножавање циљних неодређених облика може представљати когнитивно оптерећење и за изворног говорника:

- Поготово када питате *неког* човјека за кога је гласао на изборима, а он вам одговори да није изашао на изборе уопште. Такви људи нису *никоме* јасни. Како *неко* може да са таквом лакоћом пусти *некога* да му сутра одлучује о судбини. (*КорССАн259*)
- [„] Banja Luka idealno mjesto za druženje, socijalizaciju“, и то jeste (donekle) tačno. На сваком koraku можеш видјети *некога*. Са *неким/а*. (*КорССАн236*)
- Уkoliko *neko* nije zadovoljan *неким* dijelom svog tijela, или želi *некакву* promjenu на себи како би се осјећао *zadovoljnije* [...]. (*КорССАн221*)
- Zašto *некога* osuđivati и zamjerati му *нешто*? (*КорССАн212*)
- [...] jer је sve у *некој* mjeri važno за *некога* и за *нешто*. (*КорССАн223*)
- Како ми реагујемо на *неку* пјесму много зависи од stanja нашег ума или *неких* prethodnih iskustava. Ако *neko* одређено *razdoblje* svog života повежемо с *неком* пјесмом [...]. (*КорССАн234*)

У примеру који следи у недоумици смо да ли су несигурни и они који не воле свој изглед иста група људи, или је ипак реч о две различите групе:

²³ Упит CQL [lemma_lc="неки|нечији|неки|нечији" & tag="PRO"] даје 273 резултата, односно 0,6%.

- *Neki* ljudi su nesigurni u sebe, *neki* ne vole svoj izgled, nešto im smeta i žele da se promene. (KopССАн230)
- Примере који следе радо бисмо парафразирали, опет због циљних неодређених облика:
- *Неки* од разлога зашто би требало бити забрањено је тај што дим цигарета може сметати *некоме*. (KopССАн258)
- Из тог разлога, можемо тврдити да музика делује искључиво терапеутски, као утеха у звуку који [је] створен од стране *некога* ко је имао иста или слична искуства, *неко* ко разуме. (KopССАн250)
- *Niko nikome* nije naudio što se zaljubio pa želi da se razvede. То је у redu. Међутим, *nije u redu ako taj neko* nije pošten pa ne prizna то и nastavi da živi u брачној заједници [...].(KopССАн212)

У примерима који следе можемо замислити исте реченице, али без неодређеног *неки*. На пример, у првој реченици јасно је да *неки* посао у просвети наставнички позив, односно да је довољно рећи „посао у просвјети“:

- Уколико је једини посао који могу да пронађем у својој струци *неки* посао у просвјети, онда биh се искрено радије запослила у *неком* кафићу или butiku него радила као наставник. (KopССАн218)
- Слично је и у бројним примерима који следе:
- [...] ljudi треба да посједују *неку* помоћ од ljudi који имају моћ да једним позивом riješe problem. (KopССАн25)
- Ако су већ у тим fazama да желе *неку* промјену, увијек могу промијенити frizuru, боју kose, начин одjevanja. (KopССАн27)
- [...] једноставно су пресебични да се одрекну нечега зарад друге osobe, зарад *неког* већег, општег добра. (KopССАн28)
- Улазак у брак uliva *неко* осећање сигурности па се ljudi polako опуштају и показују своје право lice или се једноставно ulenје [...]. (KopССАн29)
- У прошлости су овакви проблеми били мање изражени јер су се више njegovale *неке* vrijednosti попут solidarnosti и поштовања [...]. (KopССАн219)
- Требамо да се одлучимо, хоће ли то бити породица, посао или ће се пак покушати ostvariti *неки* баланс између то двоје, и опробати се у обе те sfere života. (KopССАн27)
- Понекад музички жанр који preferирамо може да каже много о нашој личности и то је разлог за *неке* smiješne али пак честе predrasude који ljudi имају. (KopССАн233)

Навођењем неколико група примера из корпуса *KopССАн2* нисмо исцрпили све маштовите употребе неодређеног *неки*.

9. Дискусија и закључна разматрања

Чињеница је да је коришћење израза неодређености у писању на енглеском као страном у значајном степену универзална тенденција код неизворних говорника. Ипак, код србофоних говорника, као што смо видели, коришћење циљних форми, *some*, *somebody*, *someone*, има своје учестале, неоправдане или неприхватљиве реализације. Другим речима, постојећа навика ослањања на поменуте неодређене облике у писању на енглеском језику код србофоних говорника често има слабо прагматичко, односно комуникативно упориште.

Претрага и анализа еквивалентног неодређеног облика *неки* у корпусу *КорССАнг* показала је изненађујуће сличне тенденције, како квантитативно тако и квалитативно. Наиме, осим високе учесталости у коришћењу циљних форми у писању и на страном и на матерњем језику, србофони говорници снажно се ослањају на неодређене облике и у матерњем језику у ситуацијама када то није неопходно, некада више пута у оквиру једне реченице, а некада им радије прибегавају у поређењу с алтернативним формама сличног функционалног домена.

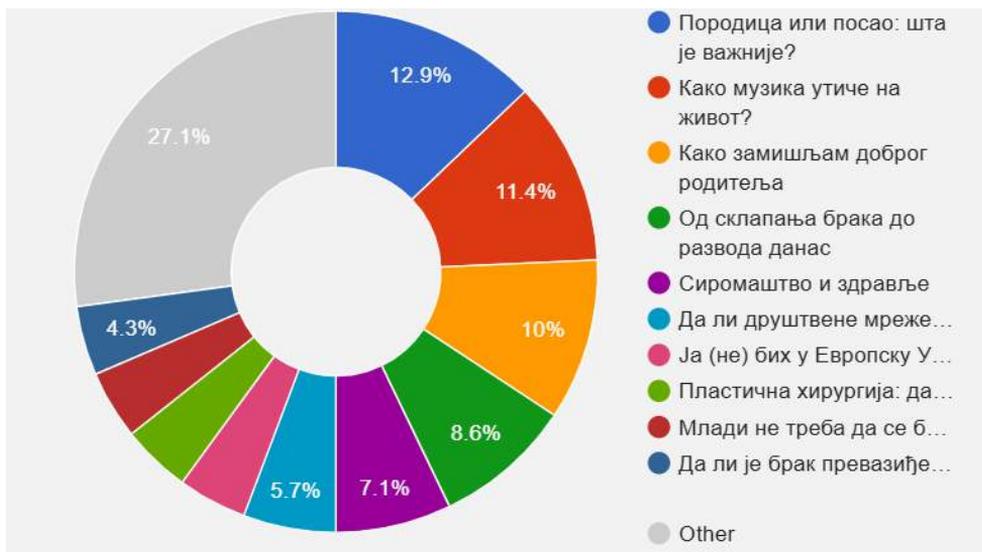
На основу наведеног, намеће се закључак да је коришћење неодређеног експонента *неки* код србофоних студената заправо најпре тенденција у писању на матерњем језику, самим тим и рефлексија друштвених и културолошких фактора. Поменута тенденција преноси се на продукцију на енглеском као страном, додатно оснажена универзалним карактеристикама често неоправданог или претерано присутног степена неодређености.

Извесно је да би истраживање требало даље проширити варијантама српског као матерњег код говорника различитих социолингвистичких варијабли (првенствено старосне доби и образовања), будући да се закључци до којих смо дошли не могу генерализовати на говорнике српског језика уопште. За такво проширивање истраживања неопходно би било користити и доступне корпусе српског језика као матерњег, уз формирање подкорпуса за одговарајуће варијанте језика.

Значај формирања два корпуса које смо користили у истраживању постаје јаснији на примеру изведене студије. Надамо се и даљем проширивању корпуса *КорССАнг*, будући да би свако проширење било драгоцено за откривање и објашњавање различитих дискурских феномена у међујезику србофоних говорника, што даље носи са собом научне и педагошке импликације.

Табела 1. Учесталост неодређених облика у различитим корпусима

	<i>some*</i> (%)	<i>some</i> (%)
<i>ICLE</i>	0.53	0.32
<i>LOCNESS</i>	0.39	0.23
Serbian	0.87	0.42
Turkish	0.87	0.43
Persian and Urdu	0.66	0.46
Czeck	0.62	0.32
Chinese	0.60	0.52
Bulgarian	0.58	0.32
Finnish	0.58	0.31
Russian	0.54	0.30
Macedonian	0.53	0.33
Swedish	0.48	0.24
Polish	0.47	0.31
Lithuanian	0.46	0.29
Spanish	0.43	0.24
German	0.42	0.21
Greek	0.42	0.20
Italian	0.42	0.21
Dutch	0.41	0.23
French	0.41	0.26
Hungarian	0.30	0.19



Графикон 1. Приказ учесталости најпопуларнијих наслова састава у корпусу *KorSSAng*

Извори

- [1] <https://noske.rgf.rs/>.
- [2] <https://www.learnercorpusassociation.org/resources/tools/locness-corpus/>.
- [3] <https://corpora.uclouvain.be/cecl/icle/home>.

Литература

- [1] Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- [2] Channell, J. (1994). *Vague Language*. Oxford: Oxford University Press.
- [3] Chaudron, C. (2003). Data collection in SLA Research. In C. J. Doughty, & M. H. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 762–828). Oxford: Blackwell Publishing.
- [4] Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- [5] Evison, J., McCarthy, M., & O’Keeffe, A. (2007). 'Looking out for love and all the rest of it': Vague category markers as shared social space. In J. Cutting (Ed.), *Vague Language Explored* (pp. 138–157). Basingstoke: Palgrave MacMillan.
- [6] Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335.
- [7] Gass, S., & Polio, C. (2014). Methodological influences of "Interlanguage" (1972): Data then and data now. In Y. Han, & E. Tarone (Eds.), *Interlanguage: Forty years later* (pp. 147–172). Amsterdam and Philadelphia: John Benjamins.
- [8] Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast: Text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.
- [9] Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer*. London and New York: Addison Wesley Longman.
- [10] Granger, S. (2002). A Bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3–33). Amsterdam and Philadelphia: John Benjamins.
- [11] Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7–24.

- [12] Granger, S., Dagneaux, E., & Meunier, F. (Eds.). (2002). *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- [13] Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2008). *The International Corpus of Learner English. Handbook and CD-ROM (Version 2)*. Louvain-la-Neuve: Presses universitaires de Louvain.
- [14] Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (Eds.). (2020). *The International Corpus of Learner English v3*. Louvain-la-Neuve: Presses universitaires de Louvain.
<https://dial.uclouvain.be/pr/boreal/object/boreal:229877>
- [15] Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4, 237–258.
- [16] Jucker, A. H., Smith, S. W., & Ludge, T. (2003). Interactiobal strategy: Interactive aspects of vagueness in conversation. *Journal of Pragmatics*, 35, 1737–1739.
- [17] Lee, D. Y., & Chen, S. X. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18, 281–296.
- [18] Marković, J. (2019). “It is a thing that gives you...”: The lexeme thing(s) as the Serbian EFL ‘teddy bear’. *Komunikacija i kultura online*, 10, 19–37.
- [19] Marković, J. (2021). I, you, and we in Serbian EFL Argumentative Writing from the Essay Title Perspective. *Folia Linguistica et Litteraria*, 36, 271–290. doi:10.31902/fl.36.2021.16
- [20] Parvaresh, V., & Ahmadian, M. (2016). The impact of task structure on the use of vague expressions by EFL learners. *The Language Learning Journal*, 44, 436–450.
- [21] Parvaresh, V., & Tayebi, T. (2014). Vaguely speaking in Persian. *Discourse Processes*, 51, 565–600.
- [22] Russell, B. (1923). Vagueness. *Australian Journal of Philosophy and Psychology*, 1, 84–92.
- [23] Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- [24] Sinclair, J. (1996). EAGLES. Preliminary recommendations on Corpus Typology. Retrieved November 5, 2018, from www.ilc.pi.it/EAGLES/corpusstyp/corpusstyp.html

- [25] Spajić, S., & Suknović, M. (2019). The Choice of Lexemes According to their Frequency in Translation into L2. *Komunikacija i Kultura Online*, 10, 104–119. doi:10.18485/kkonline.2019.10.10.6
- [26] Widdowson, H. G. (1984). *Explorations in Applied Linguistics 2*. Oxford: Oxford University Press.
- [27] Zhang, G. (2011). Elasticity of vague language. *Intercultural Pragmatics*, 8(4), 571–599. doi:10.1515/IPRG.2011.026
- [28] Zhang, G. Q. (2022). Vague Language from a Pragmatic Perspective. In I. Kecskes (Ed.), *In The Cambridge Handbook of Intercultural Pragmatics* (pp. 274–300). Cambridge: Cambridge University Press.
- ***
- [29] Марковић, Ј. (2019). Кроз призму контрастивне анализе међујезика. Пале: Филозофски факултет.
- [30] Радоња, М. и Шућур, С. (2021). О Корпусу студената англистике (КорСАНГ) и могућностима његове софтверске експлоатације. *Infotheca – Journal for Digital Humanities*, 21(37–58).

Serbian EFL Learner Corpora in Contrastive Interlanguage Analysis Through The Example of Vague Language

Jelena Marković

Summary

This paper focuses on the two Serbian EFL learner corpora, i.e., ICLE-SE and KorSANG, exploring their potential for applied linguistics research in the field of Second Language Acquisition through the example of Vague Language exponents. The learner corpus ICLE-SE is a Serbian subcorpus of The International Corpus of Learner English, version 3 (ICLE, Université catholique de Louvain, Belgium), currently the most significant learner corpus of English. The KorSANG corpus consists of three parts: the two parallel corpora of English-Serbian and Serbian-English translations, and a subcorpus of argumentative essays in L1 Serbian, collected with the purpose to serve as a referential corpus for ICLE-SE.

The focus of the article is the use of Vague Language, as an inherent language feature exploited on a daily basis. In actual fact, Vague Language has proved to be a pragmatically justified means of language use, fulfilling various contextual purposes, only seemingly resembling a language

imperfectness or deficiency. The two learner corpora have been used in order to compare and contrast the use of the targeted indefinite forms, some→*, and neki.

The research is embedded within Contrastive Interlanguage Analysis, a renowned contrastive methodology in Second Language Acquisition. The research has shown that Serbian EFL writing exploits the targeted Vague Language more frequently than the referential native corpus; moreover, Serbian EFL writing contains the targeted forms more frequently than EFL writing originating from several other L1 backgrounds. The analysis of the KorSAng subcorpus serving as the ICLE-SE referential, has shown that native speakers of Serbian use the Vague Language form neki almost as frequently in their L1 Serbian as they use its English equivalents in their EFL writing. Moreover, there are striking similarities in both L1 and EFL writing, i.e., overreliance through repetition, optional or sometimes erroneous uses, and unclear utterances.

We conclude that the use of the targeted Vague Language forms in EFL writing results from both universal features of EFL writing and the L1 writing habits of EFL writers. Vague Language turns out to be challenging not only for Discourse Analysis, but even more for Applied Linguistics, since it is an extremely demanding feature of the foreign language system. We also conclude that the two learner corpora in question possess a huge potential in Second Language Acquisition research for cognate research questions regarding Serbian EFL written production.

Key words: Contrastive Interlanguage Analysis, International Corpus of Learner English v3, KorSAng, Vague Language