

Душица В. ТЕРЗИЋ\*  
Универзитет у Београду  
Филолошки факултет

Оригинални научни рад  
Примљен: 7. 2. 2023.  
Прихваћен: 22. 2. 2023.

## ОЗНАЧАВАЊЕ ГРЕШАКА У УЧЕНИЧКОМ КОРПУСУ СРПСКОГА ЈЕЗИКА\*\*

Међу доступним ученичким корпусима мали је број оних у којима су означене ученичке грешке. Највећи број означених корпуса садржи текстове на енглеском језику као страном. Међутим, расте број означених корпуса других језика. Они се могу искористити како би се убрзао процес израде нових корпуса. Да би се предности означених корпуса могле искористити у потпуности при означавању нових ученичких корпуса, неопходно је процес означавања претходно детаљно испланирати. Преузимање постојећих система означавања и скупова ознака олакшава и стандардизује поступак означавања грешака, те омогућује да резултати даљих корпусно заснованих истраживања буду упоредиви. Међутим, особености појединачних језика подразумевају да ће типови грешака које ученици праве зависити од циљнога језика. Стога није могуће у потпуности стандардизовати означавање ученичких корпуса. У овој раду поредимо системе који се користе за означавање ученичких корпуса. Због језичке сродности, издвајамо корпусе словенских језика. Пажњу пак поклањамо и ученичком корпусу француског језика будући да се систем означавања овога корпуса може применити на означавање корпуса српског језика. На основу ове анализе предлажемо скуп грешака за означавање корпуса српског језика као страног. Због специфичности српског језика, систем за означавање грешака који предлажемо подразумева означавање на више нивоа. Грешке за сада класификујемо према домену и типу остављајући могућност даљег проширења предложеног система.

**Кључне речи:** ученички корпуси, означавање грешака, таксономија грешака, ученички корпус српског језика, ученички корпус француског језика.

---

\* [dusica.terzic@fil.bg.ac.rs](mailto:dusica.terzic@fil.bg.ac.rs)

\*\* Рад је настао у оквиру Програма финансирања научноистраживачког рада НИО, према уговору који је Универзитет у Београду – Филолошки факултет склопио с Министарством просвете Републике Србије.

## 1. Увод

Касних 80-их и раних 90-их развија се дисциплина која се бави истраживањем ученичких<sup>1</sup> корпуса. Ученички корпуси су електронске збирке писаних или усмених узорака продукције на језику који говорницима није матерњи (Грејнцер 2008: 1). Ученичке корпусе од осталих збирки текстова разликује управо ученички језик који се испитује као посебан језички варијетет (Грејнцер и др. 2015: 3) и који одликују грешке које су ученици направили у усменој или писаној продукцији. Стога се у ученичким корпусима могу означити<sup>2</sup> како уобичајена својства текста (морфосинтаксичка, лематска, семантичка, дискурсна својства), тако и ученике грешке.

Седамдесетих година XX века анализа грешака (енг. *error analysis* – ЕА) постаје водећи приступ у теорији усвајања другог језика као алтернатива контрастивном приступу грешкама (енг. *contrastive analysis* – СА) (Дјулеј и др. 1982: 140). Иако овај приступ наилази на критике као ненаучни и непоуздан приступ ученичком језику (Грејнцер 2003: 466), развој технологије у последњих двадесетак година, а нарочито развој рачунарских ученичких корпуса допринео је поновном развоју интересовања за анализу грешака (Макдоналд 2016: 106). Без обзира на одређене мане, два основна значаја анализе ученичких грешака не могу се оспорити – из података који се добијају анализом грешака могуће је извући закључке о природи процеса усвајања језика; анализа грешака указује на елементе циљног језика које ученици најтеже производе, те који тип грешака највише умањује способност ученика да ефикасно комуницира на циљноме језику (Дјулеј и др. 1982: 138).

Грешке се у контексту ученичких корпуса најчешће дефинишу као одступање од језичке норме или језик који не би произвео изворни говорник (Стритар 2012, Дјулеј и др. 1982, Сафари 2017, Грис/Берез 2017). Стога се оне у корпусима рачунарски могу означити приписивањем њиховог стандардног језичког облика у процесу који се назива нормализација. Реконструкција одговарајућег језикостандардног облика није увек очигледна јер се заснива само на текстуалном контексту, а не на учениковим сазнањима, схватањима или коментарима (Стритар 2012: 160). Другим речима, готово увек има више облика који би могли бити стандардни у датом текстуалном контексту, те, да би се одабрао одговарајући међу њима, мора се познавати говорникова намера, односно вањјезички контекст. М. Стритар закључује да је значајно записивати стандардне облике на часовима у току прегледања ученичких радова, те искористити прилику и питати ученике која им је била намера (Стритар 2012: 162).

Други начин да се грешке означе подразумева да се грешци припише податак о њеним особинама. Врсте грешака које се јављају у ученичким корпусима зависе како од природе језика који се усваја, тако и од природе ма-

<sup>1</sup> У овоме раду термин ученик означава све особе које усвајају одређени језик као страни независно од пола, старости и осталих социо-демографских одлика.

<sup>2</sup> О означавању в. Терзић 2020 и Терзић 2022.

терњег језика и других језика којима се ученик служи, али и бројних других фактора, као што је намена корпуса (Хана и др. 2010: 11–12). Велики број чинилаца који утичу на процес усвајања другог језика вишеструко умножавају број различитих типова грешака које ученици могу правити. Иако је тиме таксономија ученичких грешака отежана, она је неопходна да би се ученички корпуси што успешније могли искористити. Уколико су карактеристике грешака означене према прецизно дефинисаном систему, ученичке је корпуре лакше претраживати, обрађивати, те употребљавати у даљим истраживањима из области усвајања другог језика, нарочито при анализи грешака.

Анализа грешака одвија се у неколико фаза – дефинисање, детекција, означавање, анализа, те вредновање грешака, од којих су за корпуре усвајања страног језика кључни дефинисање, детекција и означавање (Стритар 2012: 118). У одељку 2 представљамо теоријску основу таксономије грешака, која омогућава први корак при анализи грешака – дефинисање. Затим у одељку 3 поредимо већ постојеће системе означавања грешака да бисмо указали на њихове предности и мане и тиме будућим истраживачима олакшали одабир одговарајућег система при изради новог ученичког корпуса. Посебну пажњу посвећујемо ученичким корпусима на језицима који су блиски српском будући да за ученички корпус српскога језика нисмо још увек пронашли скуп ознака. На основу ове анализе, у одељку 4 предлажемо систем означавања грешака у корпусу српског језика као страног, што је и главни циљ овога рада.

## 2. Начела означавања грешака у ученичким корпусима

Систем описивања грешака мора бити детаљан и свеобухватан, али у исто време једноставан, јасан и прилагођен како онима који корпус означавају, тако и онима који корпус користе (Стритар 2012: 120). С. Грејнцер (2003) наводи слична начела за израду система означавања грешака – систем треба да буде доследан, довољно детаљан да пружи корисне информације кориснику (али не толико да постане неупотребљив означивачу), довољно флексибилан да се ознаке могу додати у скуп или из њега избацити, док категорије треба да буду довољно опште да се могу употребити при означавању корпуса других језика (Грејнцер 2003: 467). У складу с тим потребно је радити на стандардизацији скупова ознака не само у оквиру једног језика, већ и, колико је то могуће, у оквиру више различитих језика (Дијаз Негриљо 2007).

Битан корак ка стандардизацији скупова ознака за означавање грешака представља дефинисање и класификација грешака. Творци означених корпуса често се позивају на теоријски модел који описује Х. Дјулеј са сарадницима (1982) (в. Грејнцер 2003; Стритар 2012; Сафари 2017) при дефинисању грешака које означавају у својим корпусима. Потребно је истаћи да се ова таксономија односи на опис грешака као производа језичког усвајања без објашњавања узрока тих грешака (Дјулеј и др. 1982: 141). Површински описи грешака олакшавају рачунарску обраду корпусног текста, класификацију

грешака и даље квантитативне и квалитативне анализе које би могле довести до закључака о узроцима грешака. Грешке се могу класификовати на основу четири критеријума: 1) лингвистичке категорије; 2) површинске структуре; 3) компаративне анализе; 4) комуникативног ефекта (Дјулеј и др. 1982: 146). Грешке које се односе на лингвистичку категорију обухватају грешке на фонолошком, морфолошком, синтаксичком, семантичком, лексичком нивоу и стилске грешке (в. Дјулеј и др. 1982: 146–150). Подела грешака на основу површинске структуре подразумева уметање сувишних или изостављање неопходних елемената, погрешно обликовање или погрешан редослед елемената (в. Дјулеј и др. 1982: 150–163). Компаративна таксономија обухвата поређење структуре грешака које ученик прави у циљном језику са структуром грешака у неком другачијем контексту, најчешће структуром грешака код деце које тај језик усвајају као први (в. Дјулеј и др. 1982: 163–188). Категорија која се осврће на комуникативни ефекат испитује да ли грешке изазивају неразумевање код слушалаца, односно читалаца, или не изазивају, па издваја оне које су кључне за неразумевање (в. Дјулеј и др. 1982: 188–197). Велики број аутора који преузима ову таксономију за стварање система ознака за грешке у ученичким корпусима преузима само лингвистичке категорије и површинску структуру (уп. нпр. Сафари 2017: 71; Стритар 2012: 133; Грејндер 2003: 467).

Често се, међутим, једна грешка може сврстати у више категорија. Тако је, на пример, некад тешко одлучити да ли се ради о словној, правописној грешци или је у питању непознавање морфологије (нпр. *Они хоћу*). Тако М. Стритар (2012: 155) истиче да је за доследно означавање значајно обезбедити јасно упутство у коме ће бити описане грешке и наведена начела на основу којих би се могле разрешити потенцијалне дилеме при означавању. Најчешће се (уп. Николс 2003) предност даје оном типу грешке који доноси више информација у конкретном случају. Међутим, то није увек очигледно. Из ових се разлога неким корпусима додаје више слојева анотације који најчешће представљају дводимензионалну или тродимензионалну комбинацију грешака с нивоа лингвистичке категорије и површинске структуре (Грејндер 2003: 467). У следећем одељку поредимо конкретне системе означавања корпуса да бисмо издвојили системе погодне за обраду српског ученичког језика.

### 3. Означени ученички корпуси

Најисцрпнији списак ученичких корпуса који смо успели пронаћи налази се на интернетској страници<sup>3</sup> Центра за енглеску корпусну лингвистику Католичког универзитета у Лувену и броји 200 ученичких корпуса<sup>4</sup>. Бројни аутори су на основу овога списка анализирали ученичке корпусе према за-

<sup>3</sup> <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>. Последњи приступ 1. фебруара 2023. године.

<sup>4</sup> Број корпуса у табели 1. фебруара 2023. године.

ступљености енглеског и осталих језика у ученичким корпусима, уделу писаних и усмених текстова, односу слободних и комерцијалних корпуса итд. (в. Стритар Кучук 2022; Алфаифи 2015; Сафари 2017; Микелић Прерадовић 2020). Користан преглед 74 ученичка корпуса налази се и на платформи CLARIN (Хинрикс/Краувер 2014)<sup>5</sup>, која за одређене корпусе нуди могућност директног преузимања с платформе.

Највећи удео означених корпуса садржи узорке продукције ученика на енглеском језику (в. Николс 2003; Грејнцер 2014). М. Стритар (2012: 26–48, 136–138) даје детаљан преглед не само ових корпуса, већ и система означавања грешака у њима. Ознаке које се користе за означавање корпуса енглеског језика као страног углавном су довољно опште да се могу применити на друге језике, али недовољно детаљне да би обухватиле грешке које се у великом проценту јављају у корпусима високофлексивних језика као што је српски. Уз 8 корпуса ученичких текстова који нису на енглеском језику пронашли смо детаљан опис система означавања грешака. Међу њима су ALC (Алфаифи 2015: 159–160), FALCO (Лиделинг 2005), ASK (Тенфјорд и др. 2006: 1822), SFLC (Сафари 2017: 74). У овоме раду издвајамо само корпусе словенских језика као најсроднијих српском језику – CzeSL (Хана и др. 2010), CroLTeC (Микелић Прерадовић 2020) и KOST (Стритар Кучук 2022). Поред њих, представљамо и систем ознака у француском корпусу FRIDA (Грејнцер 2003: 468, 477–478) будући да садржи ознаке на више нивоа, као и детаљнији опис подтипова грешака у односу на остале несловенске корпусе.

Чешки корпус CzeSL први је корпус неког словенског језика као страног с означеним грешкама (Микелић Прерадовић и др. 2015: 109). При означавању корпуса CzeSL грешке се означавају на три нивоа. На првом се нивоу исправља записивање, затим погрешне границе речи, интерпункција, као и морфолошке грешке, али се тип грешке овде не означава (Хана и др. 2010: 15–16). Исправке на другом нивоу подразумевају слагање, рекцију, прономиналну референцију, сложене глаголске облике, негацију, елементе који недостају, неогдговарајуће елементе, модалност, ред речи, лексику и фразеологију (Хана и др. 2010: 17). Наведено је и да неке од ових грешака имају и подтипове, али нисмо пронашли скуп свих ознака које су коришћене. Скуп грешака које се означавају није, пак, свеобухватан. Најављена је израда детаљнијег скупа ознака.

Систем за анотацију корпуса словеначког језика као страног прошао је кроз фазу евалуације. Првобитни систем анотације је искоришћен за означавање малог експерименталног дела корпуса и евалуиран је кроз анкету с изворним говорницима (в. Стритар 2012). Садржао је три нивоа анотације у складу с таксономијом предложеном у Дјулеј и др. (1982): према површинској структури (замена, додавање сувишних елемената, изостављање потребних елемената), према лингвистичком нивоу (правопис, морфологија, лексика, синтакса, ред речи) и према лингвистичкој категорији (именица,

<sup>5</sup> Број корпуса у табели 5. фебруара 2023. године када је и последњи пут приступљено страници <https://www.clarin.eu/resource-families/L2-corpora#publications-on-the-l2-learner-corpora>.

придев итд.) (Стритар 2012: 153). Евалуација је, међутим, показала да оваква таксономија није одговарајућа за словеначки ученички корпус будући да се први ниво као строго формални показао сувишним, док се јавила потреба да се означе детаљније карактеристике лингвистичких категорија. Стога је систем промењен у складу с начелима формалности, општости, објективности, описности и што мање интерпретативности (Стритар 2012: 154). У новом се систему грешке означавају у две равни – на првој је наведен лингвистички ниво на ком је направљена грешка (правопис, лексика, морфологија, синтакса), а на другој подтип грешке (Стритар Кучук 2022: 331). Примећујемо, међутим, да се као подтип лексичке и морфолошке грешке наводе лингвистичке категорије одређене речи (именица, глагол, придев итд.), а не која је врста грешке у питању. Сматрамо, дакле, да овај систем ознака није довољно детаљан за означавање грешака у српском ученичком корпусу без обзира на типолошку сродност.

Аутори корпуса CroLTeC (Микелић Прерадовић и др. 2015) наводе три слоја означавања, од којих први представља само дигитализован ученички текст, док се на другом нивоу исправљају грешке у једној речи, класификоване у три категорије – ортографске, лексичке и граматичке грешке (Микелић Прерадовић 2020: 906–908). Трећи слој је посвећен грешкама које се састоје од више речи као што су изостављање или уметање неодговарајућих синтагми, синтагме нејасног значења, синтаксичке грешке и погрешан ред речи (Микелић Прерадовић 2020: 909). Иако овај систем означавања уводи категорије којих нема у словеначком корпусу, а које су значајне за високофлексивне језике (погрешна конгруенција, вид глагола, нетачна флексија), сматрамо да категоризација грешака на другом нивоу може бити прецизнија и детаљнија да би се предности означеног ученичког корпуса могле искористити. Иако је једнослојна анотација с малим бројем ознака у скупу за означавање прецизнија и једноставнија, није довољна за високофлексивне језике код којих ученици често праве грешке на више нивоа истовремено (Хана и др. 2010: 11; Микелић Прерадовић и др. 2015).

Прецизнију категоризацију ученичких грешака од словенских корпуса нуди француски ученички корпус FRIDA (Грејнцер 2003). У овоме је корпусу најпре означено 9 домена грешака (правопис, морфологија, граматика, лексика, синтакса, регистар, стил, интерпункција, типографија) који се даље деле на укупно 36 подтипова (Грејнцер 2003: 468, 477–478). Морфосинтаксичко означавање представља трећи слој означавања и обухвата детаљни морфосинтаксички опис (нпр. за придеве је означен и степен компарације) (Грејнцер 2003: 479). Поређење постојећих система за означавање грешака у ученичким корпусима довело нас је до закључка да је за означавање српског ученичког корпуса најефикасније прилагодити систем примењен у француском ученичком корпусу FRIDA, што приказујемо у следећем одељку.

#### 4. Означавање грешака у корпусу српског језика као страног

Продукција ученика српског језика као страног, према нашим сазнањима, за сада се прикупља само у Корпус српског као страног језика КСКС (в. Миличевић 2016), који израђују Центар за српски као страни језик и Катедра за ошту лингвистику на Филолошком факултету Универзитета у Београду. Корпус чине, пре свега, ученички састави писани у оквиру испита, домаћих задатака и наставних активности у Центру за српски као страни језик, те радови настали у оквиру активности лектората за српски језик ван Србије (Миличевић 2016: 283–285). Радови се прикупљају од 2006. године уз одговарајуће метаподатке о ученицима (матерњи језик, старост, ниво знања), а процењује се да ће укупна величина корпуса износити 50.000 речи (Миличевић 2016: 283). Планирана лингвистичка обрада текста подразумева означавање како врста речи и основног облика, тако и означавање грешака и нормализацију. Имајући у виду значај детаљног планирања лингвистичке обраде (в. Лич 2004), а узимајући у обзир анализе из претходног поглавља, у даљем тексту предлажемо класификацију грешака и скуп ознака за њихово означавање на основу особености српскога језика.

Српски језик одликују велика сложеност флексије и деривације, слагање именских речи у роду броју и падежу, слагање предиката са субјектом, те слободан ред речи у реченици. Очекивано је да ученици који усвајају српски језик као страни праве више врста грешака истовремено, као што је Н. Микелић Прерадовић (2020: 904) приметила за ученике хрватског језика приликом означавања малог узорка текстова из хрватског ученичког корпуса. Неопходно је, дакле, применити вишеслојни тип означавања да би се обухватиле могуће вишеструке грешке на једној речи, што значи да се системи означавања већине корпуса енглеског језика или корпуса попут норвешког ASK-а не могу применити. У претходном одељку објаснили смо зашто сматрамо да је за српски језик најкорисније прилагодити систем означавања грешака у француском ученичком корпусу FRIDA. У табели 1 издвајамо домене и типове грешака из система означавања у описаном корпусу које сматрамо релевантним за српски језик. Уједно додајемо категорије које сматрамо релевантним за српски језик, а не налазе се у систему ознака за француски језик.

Први ниво (домен грешке)	Други ниво (тип грешке)
Правопис	Велика и мала слова Дијакритички знаци Спајање/раздвајање речи Интерпункција
Творба речи	Префиксација Суфиксација Инфиксација Флексија Композиција
Морфологија	Помоћни глагол Род Број Падеж Лице Време Вид
Лексика	Непостојећа реч Неодговарајућа реч
Синтакса	Ред речи Изостављање речи Уметање речи Кохезија Рекција
Регистар	Лексика Синтакса

Табела 1. Предлог система означавања грешака у корпусу српског језика као страног.

Пожељно је оставити место за једну недефинисану категорију у оквиру сваког домена пошто се не могу предвидети све грешке које ученици могу направити. Поред домена и типа грешке, сматрамо значајним означити граматичке категорије грешака. Морфосинтаксичко означавање српског језика је изузетно сложен процес за који постоје развијени системи (в. нпр. Милетић и др. 2016), али дубља разматрања превазилазе границе овога рада.

Постоји могућност да је вишеслојно означавање теже применити у аутоматској обради, те се у том контексту предност може дати једнослојном означавању. Ипак, корпуси се још увек означавају ручно јер не постоји довољно прецизан систем рачунарског означавања грешака. Даље, на основу закључака евалуације коју је спровела М. Стритар (2012), овакву класификацију сматрамо интуитивном и за означиваче корпуса и за потенцијалне кориснике. С друге стране, ова разматрања су заснована на прегледу доступне литературе, а не на конкретној анализи ученичких текстова. Да би се проверило да



ли предложени систем функционише у пракси, потребно је означити мањи део корпуса српскога језика као страног. На основу конкретног означавања, могу се утврдити проблеми предложеног система. Након тога се скуп ознака може једноставно модификовати и проширити означавањем грешака на другим нивоима (нпр. стилским и/или дискурсним).

## 6. Закључак

Развој ученичких корпуса значајан је и за области усвајања другог језика и за наставу страних језика. Иако број ових корпуса расте, малобројни су они у којима су означене ученичке грешке. Како би се олакшала израда ових корпуса и како би истраживања која се на њима врше била методолошки усклађенија, потребно је стандардизовати систем означавања колико је то могуће. Дефинисање и класификација грешака кључни су део планирања израде система означавања. У овоме смо раду анализирали корпусе с значеним грешкама, поредили документацију о систему означавања и издвојили систем који је најкориснији за планирање означавања грешака у ученичком корпусу српског језика. Највећи број (5) корпуси су енглеског језика као страног, који углавном садрже једнослојну анотацију грешака. Од 8 корпуса других страних језика, 3 су корпуси словенских језика као страних. Грешке у овим корпусима углавном су означене на више нивоа будући да природа ових језика захтева вишеслојну анотацију. Иако би било очекивано да систем означавања грешака у хрватском и словеначком корпусу буде најкориснији за означавање српског ученичког корпуса, анализа ових система означавања показала је да је за означавање српског ученичког корпуса најбоље применити систем за означавање француског ученичког корпуса FRIDA. У жељи да класификација грешака у корпусу српског језика као страног буде што ближа стандарду у означавању корпуса, предлажемо класификацију грешака на два нивоа – према лингвистичком домену и типу грешке. Предложени систем грешака ваљало би применити на корпусу српског језика као страног ради утврђивања потребних измена.

## ЛИТЕРАТУРА

- Алфаифи 2015:** A. Alfaihi, *Building the Arabic Learner Corpus and a System for Arabic Error Annotation*, PhD thesis, Leeds: University of Leeds.
- Грејнџер 2003:** S. Granger, Error-tagged learner corpora and CALL: A promising synergy, *CALICO Journal*, 20/3, 465–480.
- Грејнџер 2008:** S. Granger, Learner corpora, in: A. Lüdeling, M. Kytö, (eds.), *Corpus Linguistics. An International Handbook*, 1, Berlin, New York: Walter de Gruyter, 259–275.

- Грејнциер 2014:** Granger, S. *ICLE International Corpus of Learner English*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11372/LRT-859>. 01.02.2023.
- Грејнциер и др. 2015:** S. Granger, G. Gilquin, F. Meunier, Learner corpus research past, present and future, in: S. Granger, G. Gilquin, F. Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press, 1–5.
- Грис/Берез 2017:** S. T. Gries, A. L. Berez, Linguistic Annotation in/for Corpus Linguistics, in: N. Ide, J. Pustejovsky (eds.), *Handbook of Linguistic Annotation*, Dordrecht: Springer, 379–409. [https://doi.org/10.1007/978-94-024-0881-2\\_15](https://doi.org/10.1007/978-94-024-0881-2_15)
- Дијаз Негриљо 2007:** A. Díaz-Negrillo, *A fine-grained error tagger for English learner corpora*, Unpublished PhD thesis, University of Jaén.
- Дјулеј и др. 1982:** H. Dulay, M. Burt, S. Krashen, *Language Two*, New York: Oxford University Press.
- Лич 2004:** G. Leech, Adding linguistic annotation, in: M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*, Oxford: Oxbow Books, 17–29.
- Лиделинг 2005:** A. Lüdeling, M. Walter, E. Kroymann, P. Adolphs, Multi-level error annotation in learner corpora, in: *Proceedings of Corpus Linguistics 2005*, Birmingham.
- Макдоналд 2016:** P. MacDonald, “We All Make Mistakes!”. Analysing an Error-coded Corpus of Spanish University Students’ Written English, *Complutense Journal of English Studies*, 24, 103–129. <https://doi.org/10.5209/CJES.53273>
- Микелић Прерадовић и др. 2015:** N. Mikelić Preradović, M. Berać, D. Boras, Learner Corpus of Croatian as a Second and Foreign Language, in: K. Cergol Kovačević, S. L. Udier (eds.), *Multidisciplinary Approaches to Multilingualism*, Frankfurt am Main, Germany: Peter Lang, 107–126.
- Микелић Прерадовић 2020:** N. Mikelić Preradović, Označavanje pogrešaka u CroLTeC-u (računalnom učeničkom korpusu hrvatskog kao stranog jezika), u: *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, XLVI/2, 899–920.
- Милетић и др. 2016:** A. Miletic, C. Fabre, D. Stosic, Mise au point d’une méthode d’annotation morphosyntaxique fine du serbe, in : T. Hamon, L. Danlos (eds.) *Proceedings of Traitement Automatique des Langues Naturelles (TALN 2016)*, Paris, France: AFCP & ATALA, 506–514.
- Миличевић 2016:** M. Miličević, Korpus srpskog kao stranog jezika (KSKS): Opis građe i plan izrade, u: В. Крајишник (ред.), *Српски као страни језик у теорији и пракси III*, Београд: Филолошки факултет, 279–289.
- Николс 2003:** D. Nicholls, The Cambridge Learner Corpus: error coding and analysis for lexicography and ELT, in: *Proceedings of the Corpus Linguistics 2003 Conference*, Lancaster, United Kingdom, 572–581.

- Сафари 2017:** S. Safari, *Constructing and analysing an error-tagged learner corpus of Persian*, Neobjavljena doktorska disertacija, Beograd: Filološki fakultet.
- Стритар 2012:** M. Stritar, *Korpusi usvajanja tujeга језика*, Ljubljana: Zveza društven Slavistično društvo Slovenije.
- Стритар Кучук 2022:** M. Stritar Kučuk, KOST med korpusi usvajanja tujeга језика, u: N. Pirih Svetina, I. Ferbežar (red.), *Obdobja 41: Na stičiču svetov: slovenščina kot drugi in tuji језик*, Ljubljana: Založba Univerze v Ljubljani, 323–334.
- Тенфјорд 2006:** K. Tenfjord, P. Meurer, K. Hofl, The ASK Corpus – a Language Learner Corpus of Norwegian as a Second Language, in: *LREC 2006 Proceedings*, 1821–24.
- Терзић 2020:** Д. Терзић, Анализа успешности парсера *Talismane* на нивоу синтаксичких етикета у корпусу ParCoJour, у: С. Гудурић (ред.), *Језици и културе у времену и простору*, 9.2, Нови Сад: Филозофски факултет, 265–279. <https://doi.org/10.18485/msc.2022.51.1.ch12>
- Терзић 2022:** Д. Терзић, Српски језик у контексту тенденција у рачунарској обради језика, *Научни састанак слависта у Вукове дане*, 51/1, 145–157.
- Хана и др. 2010:** J. Hana, S. Škodová, B. Štindlová, Error-tagged learner corpus of Czech, in: *Proceedings of the Fourth Linguistic Annotation Workshop at ACL 2010*, Association for Computational Linguistics, 11–19.
- Хинрикс/Краувер 2014:** E. Hinrichs, S. Krauwer, The CLARIN Research Infrastructure: Resources and Tools for E-Humanities Scholars, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, 1525–31.

Dusica Terzic

ANNOTATION DES ERREURS DANS LE CORPUS D'APPRENTISSAGE  
DU SERBE LANGUE ÉTRANGÈRE

Résumé

Le serbe étant une langue dont le corpus d'apprenants est au cours de développement, il est important de bien planifier le processus d'annotation de ce corpus. Annotation des corpus d'apprenants, et surtout l'annotation des erreurs, est une condition importante pour pouvoir exploiter au maximum les corpus d'apprenants dans les recherches de l'acquisition des langues étrangères. Dans cet article, nous avons d'abord présenté les principes théoriques qui se trouvent à la base de la plupart des systèmes d'annotation existants. Nous comparons ensuite des schémas d'annotation afin de répondre à la question quel système d'annotation des erreurs peut être adapté à l'annotation du corpus d'apprenants du serbe. Enfin, un système d'annotation des erreurs du corpus d'apprenants du serbe langue étrangère est créé en adaptant le schéma d'annotation du corpus d'apprenants du français FRIDA, ce qui représente le but du présent article.

*Mots clés:* corpus d'apprenants, annotation des erreurs, classification des erreurs, corpus d'apprenant du serbe langue étrangère, corpus d'apprenant du français langue étrangère.