

Александра С. Трговац
 Универзитетска библиотека
 „Светозар Марковић”, Београд
 aleksandra@unilib.rs

УДК 81'33(049.32)
 004.822(049.32)
 02:004(049.32)

ДОМЕТИ СРПСКО-НЕМАЧКОГ ПАРАЛЕЛНОГ ЛИТЕРАРНОГ КОРПУСА

Андоновски, Јелена. *Паралелни српско-немачки корпус књижевних шекспирова: израда, проналажење информација и семантички веб*. Београд: Универзитетска библиотека „Светозар Марковић”, 2021.

Монографска публикација *Паралелни српско-немачки корпус књижевних шекспирова: израда, проналажење информација и семантички веб* представља приређени текст докторске дисертације *Мрежа ошворених повезаних података и језички ресурси у процесу израдање српско-немачког литеарног корпуса* коју је Јелена Андоновски одбранила 2020. године на Филолошком факултету Универзитета у Београду. Публикација припада едицији „Хомилије” коју чине научне монографије, а углавном су у њој заступљене приређене докторске дисертације.

Јелена Андоновски је своје знање и десетогодишње искуство рада у Универзитетској библиотеци „Светозар Марковић”, најпре на пословима каталогизације и предметне и стручне класификације монографских публикација, затим као део тима за опис дигиталних колекција и објеката, а потом и као члан Групе за језичке технологије Универзитета у Београду у оквиру које ради на припреми паралелних текстова и успостављању нових паралелних корпуса, објединила у овој изузетној научној монографији.

Монографска публикација Јелене Андоновски обима преко 300 страна, уобличена је у девет поглавља, а доноси преглед паралелних корпуса и њиховог значаја, потом развоја и значаја паралелних корпуса који укључују српски језик, методологију проналажења информација и веза корпуса са семантичким вебom и израдом првог српско-немачког паралелног корпуса, све до изузетно корисних прилога и веома обимне, актуелне и информативне библиографије са преко 200 библиографских јединица.

У самом уводу публикације истакнут је значај корпусне лингвистике и паралелних корпуса у изучавању књижевних текстова, уз повезивање корпуса са семантичким вебом, примену механизма за проналажење информација коришћењем постојећих језичких алата и ресурса и примену софтвера за паралелизацију. Нарочито се апострофира значај библиотека и библиотечких ресурса у погледу семантичког веба и наглашава се улога библиотекара који могу активно учествовати у најразличитијим истраживањима, па и у области рачунарске лингвистике.

Дефиниција, улога и значај паралелних корпуса изузетно су детаљно представљени, а аспект ауторских права и коришћења садржаја корпуса описан је и истакнут као незаобилазан почетни корак. Нарочит квалитет представља исцрпан преглед развоја постојећих паралелних корпуса у Србији који укључује и вишејезичне корпусе (Платонова *Рейублика*, Орвелова 1984, *Пуџ око свейџа* Жила Верна) и двојезичне корпусе (СрпФранКор, СрпЕнГКор, Библиша) као основу за настанак немачко-српског корпуса (СрпНемКор). Већину поменутих корпуса одржава и развија Група за језичке технологије. На информативан начин представљена је методологија рада на паралелном корпусу – прикупљање, дигитализација, обрада текстова, примена електронских морфолошких речника српског језика, анотација корпусног материјала, те паралелизација.

Посебну вредност ове научне монографије представљају поглавља 4 и 5 која се баве проналажењем информација и семантичким вебом као основом израде једног паралелног корпуса. Појмови и њихов значај су прецизно и обимно описани и представљени читаоцима, од појма метаподатак, стандарда за израду метаподатака, оптичког препознавања карактера, препознавања именованих ентитета до структуре семантичког веба, стандарда за описивање значења података – RDF¹, мреже отворених повезаних података – LOM², ресурса семантичког веба – DBpedia, Википодаци, система за организацију знања, те улоге библиотека и библиотекар у овим доменима. Сваки сегмент израде једног паралелног корпуса детаљно је описан. Тако конципиран садржај може послужити студентима библиотекарства и информатике, као и библиотекарима који улазе у свет дигиталних библиотека и проналажења информација у њима као незаобилазан извор неопходних сазнања о овим значајним темама.

¹ Resource Description Framework

² Linked Open Data

Нови научни допринос, како дисертације тако и приређене научне монографије, изражен је кроз свеобухватан и вишезначан немачко-српски паралелни литерарни корпус СрпНемКор. Садржински гледано, корпус се састоји од 14 романа – седам је изворно написано на немачком језику (немачки и аустријски аутори) и преведено на српски, а осталих седам је оригинално написано на српском језику и преведено на немачки. Дела су настала у другој половини 20. или почетком 21. века. Критеријуми за одабир дела за корпус били су: награђиваност аутора, доступност дела на оба језика у српским библиотекама, популарност дела код читалаца и обимност дела (из крајње практичних разлога). Ауторка истиче значај Аустријске библиотеке у Универзитетској библиотеци „Светозар Марковић”, координацију рада са Институтом Гете и Катедром за германистику Филолошког факултета у Београду, те сарадњу са Групом за језичке технологије као стубовима ослоња за одабир текстова, обраду истих и израду паралелног корпуса СрпНемКор.

Све описане фазе примењене су у изради овог паралелног корпуса. Процесу претходи дигитализација штампаног текста и обрада текста алатима за оптичко препознавање карактера. Следи израда структурних анотација применом XML³ етикета, графова за обраду реченица и регуларних израза, паралелизације програмским пакетом ACIDE који је практично представљен у свим корацима уз анализу и указивање на проблеме који се јављају приликом рада. Формална анотација текстова извршена је кроз оквир дигиталне библиотеке Библиша, а подразумева доследно унете метаподатке на оба језика: УДК број, име аутора, наслов, број страна, сажетак, место издавања, издавач, година издавања и име преводиоца. Сви метаподаци омогућавају услове за квалитетно проналажење информација у паралелном корпусу и двојезичну претрагу текста. Овако структуриран процес омогућио је и допуну постојећих лексичких ресурса, израду ограниченог немачко-српског речника директном ексцерпцијом из изграђеног корпуса. Као најзначајнији домет издваја се укључивање овог речника у мрежу отворених повезаних података из области лингвистике, а заснован је на семантичким везама између њих.

Будући да је ауторка дипломирала, одбранила мастер рад и докторирала из области библиотекарства и информатике, незаобилазна је њена намера да у тексту ове публикације доследно нагласи улогу библиотека и библиотекара и у истраживањима овог типа. Јасна улога огледа се у

³ eXtensible Markup Language

чињеници да библиотеке поседују грађу, да библиотекари у њима обављају процес дигитализације, доделу метаподатака дигиталним документима, као и да каталози библиотека и нормативне датотеке постају део мреже отворених повезаних података.