

## ЕЛЕКТРОНСКИ КОРПУС СРПСКИХ КЊИЖЕВНИХ ДЕЛА И ЊИХОВИХ ПРЕВОДА НА ИТАЛИЈАНСКИ ЈЕЗИК<sup>1</sup>

Израда двојезичног електронског корпуса који ће обухватити српска књижевна дела у прози и њихове преводе на италијански језик, паралелизоване на нивоу реченица или пасуса, већ у почетној фази реализације омогућује увид у начин на који су дела српске прозе превођена на италијански језик. Анализа двојезичног корпуса може се усредсредити на традуктолошке и стилистичке аспекте појединих превода, или на више превода истовремено. Могуће је пратити рад појединих преводаца и поредити преводе истог дела, настале као рад различитих преводаца. Језички материјал садржан у овако устројеном двојезичном електронском корпусу може се проучавати у оквиру контрастивне анализе, али има и широку примену како у лексикографији, као велика база усвојених преводачких еквиваленција и решења, тако и у дидактици како италијанског, тако и српског као L2.

**Кључне речи:** Електронски корпус, паралелизовани текст, српски језик, италијански језик, српска књижевност, преводи на италијански.

Овим радом желимо да представимо поступак дигитализације српских књижевних дела у прози и њихових превода на италијански језик. Идеја, нипошто оригинална, јесте да се на овај начин српска књижевна дела и њихови преводи на италијански чувају у форми електронских двојезичних паралелних текстова, која је погодна за брзе, прецизне и свеобухватне лексичке, морфолошке, синтаксичке, традуктолошке, контрастивне, стилистичке и друге анализе. Стварањем корпуса двојезичних књижевних текстова и ширењем тог

---

\* smoderc@fil.bg.ac.rs

1 Рад је изложен на научном скупу *Савремено изучавање српског језика и књижевности и словенских језика као мајерњих, инословенских и сираних*, који је организовао Савез славистичких друштава Србије у сарадњи с Филолошким факултетом Универзитета у Београду (29-31. август 2014).

корпуса новим дигитализованим текстовима олакшава се увид у то која су дела наше књижевности у одређеном тренутку сматрана интересантним за италијанску публику и италијанске издаваче. Такође, овим корпусом олакшава се приступ самим текстовима превода на италијански језик, јер они, у неким случајевима, нису доступни у библиотекама у Србији (тако је, на пример, с преводом романа „Нечиста крв“ Боре Станковића<sup>2</sup>), а у неким случајевима нису ни унесени у каталог Народне библиотеке (у поменутом каталогу, на пример, не постоји запис о преводу на италијански романа „Рани јади“, Данила Киша<sup>3</sup>, што није случај с другим преводима Кишових дела на италијански). Поред олакшане доступности информације о преводима и доступности самих текстова превода, израдом двојезичног корпуса отвара се могућност да се у анализи појединих превода провери да ли су исти или слични лексички еквиваленти и преводачки поступци примењивани доследно у оквиру истог дела, односно у другим преводима истог аутора, али и да се та решења и поступци упореде с решењима и поступцима других преводаца српских књижевних дела на италијански језик. Такође, овај корпус омогућује да се брзо и једноставно упореде различити преводи истога дела на италијански; до сада су, додуше, само значајнија Андрићева дела преведена два пута на италијански<sup>4</sup>; постоји и нови превод романа „Нечиста крв“.<sup>5</sup>

Поред рада на стварању корпуса, пожељно је да се дефинишу одговарајући критеријуми који ће се примењивати у анализи превода. Под анализом превода мислимо пре свега на практичне активности које проистичу из постојања оваквог корпуса: на пример, уочавање оних језичких особина текста превода која се могу применити у дидактици италијанског као страног језика, у виду илустративног материјала, затим, подстицање студената да кроз инструмент паралелних текстова уочавају кореспонденције и разлике између српског и италијанског текста, да уочено примене и да, уопште, стекну увид

2 Ово дело превео је Умберто Урбани 1928. године (“Sangue impuro“, издавач Тревес из Трста); због техничких грешака у слогу књиге, превод је поново објављен 1930. године код истог издавача. У каталогу Народне библиотеке Србије забележена су оба издања и оба су у статусу „дезидерата“.

3 “Dologi presoci”, превео Лионело Костантини, издавач Adelphi 1993.

4 За Андрићева дела види фусноте 14, 15 и 16.

5 Дело је превела Кјара Де Бернарди, издавач је Фалиђи Едиторе (2011.).

у примењене преводачке технике и поступке. Овај функционални, „технички“ и традиционално усмерени приступ не мора нужно да носи предзнак прескриптивног приступа превођењу јер из увида у то како је неко место преведено не морају да се изведу обавезујућа традуктолошка правила. Корпус може да послужи и као полазиште за анализе спроведене из перспективе савремене традуктологије, у чијем фокусу не мора бити искључиво језичка димензија превода. Критеријуми за прву, у нашем виђењу основну анализу, захваљујући томе што су и оригинал и превод доступни у електронској форми и што су претраживи помоћу рачунара, могу да укључе разне параметре, почев од критеријума лексичке прецизности и доследности у преводачким решењима до критеријума синтаксичке (на местима где је она могућа) и семантичке кореспонденције, критеријума очуваности фразеолошке димензије текста итд. У анализи могу се узети у обзир и егзактно мерљиви елементи текста, какви су квантитативни односи оригинала и превода (установити у којој је мери текст превода дужи или краћи у односу на оригинал – тачније, да ли је разлика у дужини појединих делова текста настала због различите лексичке и фразеолошке структуре српског и италијанског језика или је она настала као резултат преводиоачеве субјективне интервенције за коју нису уочљиви разлози нити језичке, нити књижевне, нити културолошке природе). Електронска форма двојезичног корпуса омогућује да се изведу и друге анализе; могу се, на пример, измерити однос независних и зависних реченица у тексту оригинала и превода, испитати интерпункција оригинала и превода (јер промена интерпункције повлачи одређене промене у ритму саме прозе),<sup>6</sup> евидентирати парафразе и описни преводи, разликујући нужне изборе који проистичу из синтаксичких стега и ограничења једног језика од немотивисаних удаљавања од текста оригинала. Оваквим приступом могуће је стећи бољи увид у то како је одређени текст пренесен на италијански језик, да ли је преводац био прецизан у одабиру преводних еквивалената и доследан у њиховој примени у целом тексту (односно – у случају да је превод рађен у коауторству, могуће је проценити, на основу лексичких и стилских особина текста превода, који је део текста превео један преводац, а који део други, и у коликој мери је језик њихових текстова уједначен и лексички координиран).

6 Покушај анализе интерпункције у српском оригиналу и италијанском преводу изведен је у Moderc (2014a: 203-216).

Развој корпусне лингвистике у последњих двадесетак година допринео је стварању бројних електронских корпуса; обиман, али не и свеобухватан попис постојећих корпуса, нуди Barbera (2013). Захваљујући електронским корпусима лингвисти могу да брзо и једноставно прикупе материјал за широк спектар истраживања. Италијанска корпусна лингвистика, која нас у стручном и научном погледу посебно интересује, такође се може похвалити великим бројем остварених пројеката и значајним резултатима како у информатичкој обради једнојезичних корпуса, тако и у изучавању језичког материјала који ти корпуси садрже. Детаљнији подаци о резултатима италијанске корпусне лингвистике могу се наћи у поменутом делу Barbera (2013) и код Andorno-Rastelli (2009).

Упоредо с радом на једнојезичним корпусима расло је и интересовање за двојезичне и вишејезичне корпусе: најзначајнији резултати остварени су у области стручних језика, нарочито правне терминологије, која због природе и осетљивости језика правне науке захтева нарочиту терминолошку прецизност и доследност у раду преводилаца, посебно у државама или заједницама држава у којима су у службеној употреби два или више језика; за нас је најближи и најзанимљивији пример Европске уније и вишејезичних електронских докумената који садрже законе и одлуке политичких органа ЕУ; више о тим корпусима може се прочитати код Pouliquen (2006).

Ипак, нашу пажњу привукла је чињеница да у обиљу постојећих корпуса релативно мали број садржи књижевне текстове и њихове преводе на друге језике; за српски језик оваквих инструмената још је мање. Можемо поменути и неколико важнијих пројеката, попут корпуса GRALIS, који садржи Андрићева дела и њихове преводе на немачки и на словенске језике, развијен при Универзитету у Грацу,<sup>7</sup> или српско-енглеске корпусе SELFEN и SrpEngCor, односно српско-француски корпус SrpFranCor, који су доступни на сајту Математичког факултета у Београду.<sup>8</sup> Такво стање подстакло нас је да започнемо стварање корпуса српских књижевних дела и њихових превода на италијански језик: усмереност корпуса (српско-италијански) одговара нашим интересовањима (контрастивна анализа италијанског и српског, традуктологија, изуча-

---

7 [http://www-gewi.uni-graz.at/gralis/korpusarium/gralis\\_korpus.html](http://www-gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html).

8 <http://korpus.matf.bg.ac.rs/prezentacija/paralelni.html>.

вање италијанске фразеологије), али и потребама наставе италијанског као страног језика на Филолошком факултету у Београду, будући да овако устројен корпус може да послужи као репертоар из кога се може преузимати језички материјал како за потребе наставе језика, тако и за језичке вежбе. Неке могућности примене двојезичног материјала у учионици наводи Ристовић (2012.), а сам наставник може да искористи паралелне текстове за низ активности, одговарајући на дидактичке потребе, следећи своју инвентивност и радећи све време с електронским материјалом у познатом формату који не захтева посебна информатичка знања. Сваки наставник страног језика требало би да буде оспособљен да паралелизује текстове, будући да је често приморан да самостално набавља језички материјал и да од њега израђује дидактички материјал за своје потребе, чинећи разноврснијим своје дидактичке активности и материјале. Приближавање електронских корпуса учионици значајно је и због тога што ученик или студент могу већ у раним фазама учења језика да се поставе у улогу истраживача, да претражују паралелне текстове и да их користе као помоћно средство у усвајању страног језика и у усавршавању српског. Сматрамо да је изузетно важно да свако ко изучава страни језик буде упознат с електронским корпусима и упућен да их посматра као дидактички инструмент, уз уџбеник или речник.

На самом почетку рада на корпусу суочили смо се с дилемом која дела треба да уђу у састав корпуса. Определили смо се да применимо практичан приступ и да у корпус уврстимо што више превода на италијански, имајући у виду чињеницу да број српских књижевних дела преведених на италијански није претерано велик: успели смо да пребројимо осамдесетак наслова, мада их има извесно више, а прецизнији број дефинисаће се развојем корпуса и пажљивијим истраживањем. С друге стране, пошто не постоје ограничења када је реч о величини корпуса, односно о простору за складиштење, сматрамо да треба дигитализовати и паралелизовати и она дела која су преведена на италијански, а не спадају у књижевност у ужем смислу. Такав је случај с делима политичке и мемоарске садржине (Милован Ђилас) или с делима забавног карактера, намењеним млађој публици (Урош Ђурић). У вези с наведеним примерима наглашавамо да дајемо приоритет не толико књижевној вредности српског текста већ самом италијанском тексту, у уверењу да се у њему могу наћи примери преводилачких решења и поступака који су интересантни и у лингвистичком и у дидак-

тичком смислу. Поред оваквих дела, склони смо да у корпус уврстимо и књиге и преводе на италијански језик аутора попут Мирослава Крлеже, Мирка Ковача, Предрага Матвејевића или Августа Шеное. Сматрамо занимљивим и дела Доситеја Обрадовића, Светозара Марковића и српске народне приповетке, колико год да ова дела, због своје језичке архаичности, могу имати мању релевантност за синхронијске анализе. С друге стране, уколико желимо да нашем електронском корпусу припишемо и функцију централне дигиталне архиве превода са српског на италијански, онда преводи старијих дела на српском морају наћи своје место у корпусу. На пример, дело „Живот и прикљученија“ преведено је на савремени италијански језик и сусрет архаичног српског и савременог италијанског језика у сада доступном паралелизованом електронском тексту извесно отвара простор за бројне анализе. Наиме, само одустајање од иначе изводљиве архаизације текста италијанског превода овог дела можда треба сагледати као покушај да се Доситејева књига приближи савременом италијанском читаоцу.

Из онога што смо до сада изложили може се закључити да назив „електронски корпус српских књижевних дела...“ ваља схватити више као одраз доминантне тежње у избору преведених аутора и начелне предности коју дајемо српским писцима. С друге стране, флексибилност у селектовању књижевних дела која намеравамо да укључимо у наш корпус налази оправдање у самој природи електронских корпуса, јер корисник може да искључи из анализе оне ауторе, она дела или оне преводиоце које не сматра релевантним за своје истраживање, самостално одређујући оквире у којима намерава да се креће.

Радећи на дигитализацији корпуса покушали смо да сачинимо списак књижевних дела српских аутора која су преведена на италијански језик и установили да ни тај задатак није једноставан. Не постоји јединствена база података у којима су пописани сви преводи на италијански дела наше књижевности. Сваке године одређен број књига наших аутора бива преведен на италијански језик, тако да би се поменуте базе података морале стално ажурирати. У прикупљању информација од помоћи нам је био списак превода „српских, хрватских и босанских“<sup>9</sup> аутора, који је доступан на сајту римског Универ-

9 Наслов списка гласи *Autori di lingua serba, croata e bosniaca*, („Писци који пишу на српском, хрватском и босанском језику“ – превод наш). Ознака „босански“ у овом случају има искључиво географско значење; исто важи и за ознаке „српски“ и „хрватски“.

зитета La Sapienza<sup>10</sup>, али је ажуриран само до 2011. године. Подаци о новијим преводима могу се наћи претрагом каталога Националне библиотеке Италије (претрагом имена аутора, или преводилаца – уколико су њихова имена позната истраживачу, или тако што се у образац за претрагу унесу ознаке Дјуијеве класификације за књижевна дела аутора из Србије или из бивше Југославије, односно претраживањем каталога и сајтова италијанских издавачких кућа, нарочито када је реч о најновијим преводима, чији подаци још нису унесени у каталоге библиотека.<sup>11</sup> У трагању за преводима књига српских аутора на италијански језик били су нам од помоћи и електронски каталог Народне библиотеке Србије, у којој се чува значајан број превода, и електронски каталог библиотеке Италијанског културног центра у Београду. И сама библиотека Катедре за италијанистику била је драгоцен извор, мада се у обнављању и ширењу њеног књижног фонда води рачуна у првом реду о научној литератури и о делима италијанске књижевности, док се преводи српских аутора на италијански набављају спорадично. Ипак, да су наведени инструменти од релативне помоћи може да илуструје чињеница да смо до једног превода дошли сасвим случајно, прегледајући полице катедарске библиотеке, а подаци о том преводу не постоје ни у евиденцији Народне библиотеке Србије ни Националне библиотеке у Риму.<sup>12</sup>

Дела која смо до тренутка писања овог рада паралелизовали и унели у корпус су: *На Дрини ћурија*, *Проклећа авлија* (два превода)<sup>13</sup> и *Травничка хроника* (два превода)<sup>14</sup>, *Госпођица* и више Андрићевих приповедака, у разним збиркама,<sup>15</sup> затим Селимовићев *Дервиш и*

10 <http://w3.uniroma1.it/sseuco>; до списка се долази кликом на линк Biblioteca.

11 У истраживању нарочито су нам од помоћи били каталози италијанске интернет књижаре IBS ([www.ibs.it](http://www.ibs.it)).

12 Реч је преводу књиге Арсена Диклића, „Салаш у Малом риту“, у преводу Анте Кнежевића (издавач превода је Едит, 1963.).

13 Корпус садржи преводе Јоланде Маркјори (1962.) и Лионела Костантинија (1988.).

14 Корпус садржи преводе Луиђија Салвинија (1962.) и Дуње Бадњевић (2001.).

15 *Ђоркан и Швабица*, *Мустафа Маџар*, *Анкина времена*, *Смрт у Синановој шекији* (два превода), *Миле и Прелац*, *Најаси*, *Свадба*, *Змија*, *Шала у Самсарином хану*, *Прича о везировом слону* (два превода), *Прича о кмету Симану*, *Осајичани*, *Жећ*, *Први сусрећи*, *Мост на Жейи*, *Исјовијег*, *Љубав у касаб*, *Писмо из 1920. године*, *Ћилим*, *Снојнићи*, *Свечаноси*, *Сарачи*, *Jedan dan u Sarajevu krajem jula 1878. godine*, *Разговор ђед вече*, *Ира*.

*смрѝ*, Кишова *Енциклопедија мртвих*, Тишмин роман *Упоишеба човека*, Албахаријеви романи *Цинк и Геџ и Мајер*, Басарина дела *Фама о бициклистима* и *Срце земље*, Кустуричина књига *Смрѝ је нероверена ласина*, Гаталичин роман *Век*, Петровићева *Сийничарница „Код срећне руке“*, Живковићев *Писац у најам* и Крлежин *Повраћак Филија Лајиновића*.<sup>16</sup> Корпус до сада паралелизованих текстова садржи 18 српских дела и 20 превода на италијански, с 2.920.489 речи, што да сада прикупљеном материјалу даје лингвистичку релевантност и одређену репрезентативност када је реч о ужој области превођења књижевних текстова са српског на италијански језик. Одредили смо се за ова дела вођени првенствено практичним критеријумом, односно физичком доступношћу превода на италијански језик, било у облику већ постојећег електронског текста или у облику папирног издања књиге. Број прозних књижевних дела наших аутора која су преведена на италијански језик, према нашој тренутној процени, превазилази бројку од осамдесет наслова. Имајући у виду тај податак, понављамо мишљење да двојезични српско-италијански електронски корпус књижевних дела треба да буде првенствено свеобухватан, а питање књижевне и естетске вредности, као што смо раније рекли, решаваће корисници корпуса, укључујући у претрагу и анализу она дела која буду сматрала релевантним за своје истраживање.

Треба да напоменемо да наведени електронски материјал називамо корпусом иако он још нема неке техничке одлике савременог електронског корпуса: наиме, материјал није још дефинисан по питању квантитета (планирамо да се нови паралелни текстови непрекидно додају, онако како буду приређивани), самим тиме нема утврђен број токена; језички материјал није аотиран, то јест, није извршено његово машинско или ручно обележавање (речи или група речи, према њиховом морфолошким одликама, према синтаксичкој функцији или другим критеријумима). У техничком смислу, било би тачније казати да оно чиме сада располажемо јесте скуп паралелних текстова а не корпус у ужем смислу. Ипак, за почетна истраживања сакупили смо све паралелне текстове у један документ у програму Excel: сваки паралелни текст унели смо засебно у поменути програм (на пример, паралелни текст романа „Геџ и Мајер“ Давида Албахарија смештен је

16 Нека друга дела, која смо поменули у ранијем тексту, још су у фази обраде.



у колоне А и В, паралелни текст романа „Цинк“, истог аутора, смештен је у колоне С и D и тако даље за сваки паралелн текст; изузетак су дела која су преведена два пута: она заузимају три колоне). Ово решење нам омогућује да већ у овом облику истовремено претражимо садржај обрађених књижевних дела и њихових превода<sup>17</sup>; имајући у виду ову могућност, имамо основа да материјал којим тренутно располажемо ипак назовемо корпусом.

Израда електронског двојезичног текста и корпуса у наведеном облику, представља једноставан поступак, доступан просечном кориснику рачунара: сматрамо да су управо ова два елемента, једноставност израде и доступност просечном кориснику рачунара важни чиниоци јер се на тај начин у приређивање, али – касније – и у анализу паралелних текстова може укључити већи број особа. Сама израда обухвата три фазе: дигитализацију текстова (конверзију у електронски формат), паралелизацију и исправљање паралелног текста.

Дигитализација текста сканирањем и поступак препознавања слова представљају стандардан поступак који не треба посебно описивати јер зависи од уређаја и од програма који управља процесом сканирања. У овој фази довољно је имати на располагању стандардну информатичку опрему (рачунар и скенер) и неки од програма за оптичко препознавање слова (OCR), који се најчешће испоручују заједно са скенером. Након сканирања књиге програм OCR конвертује сканирани материјал у електронски текст који се може унети у неки од програма за обраду текста. Тако добијени текст треба очистити од недостатака и непрецизности насталих приликом оптичког препознавања текста: ти недостаци могу бити техничке природе (на пример, разни кодови, попут ознака за прелом секције или ручни прелом реда (у Word-у: Section Break, Manual Line Break), које треба уклонити, или цртице за дељење речи на слоге, које се отклањају пажљиво јер се цртица употребљава и у полусложеницама и за директни говор, уместо наводника), или правописне грешке, уз помоћ додатка за проверу правописа. На тај начин отклањају се грешке настале као производ програма за оптичко препознавање карактера. Грешке настају најчешће уколико је резолуција сканирања недовољна (у нашем

17 На основу оваквог корпуса могли смо да извршимо истраживање о томе како је превођен на италијански српски прилог „иначе“ (Moderc 2014b).

раду користимо резолуцију од 300dpi). Одређен број грешака настаје и због чињенице да програм за OCR који користимо, зависно од величине слова у штампаном издању и од њиховог графичког облика, у неким случајевима не препознаје ваљано ћирилична слова (дешава, се на пример, да се уместо „је“ у тексту добије „јс“, да се уместо „ново“ добије „пово“ итд.; грешака у препознавању има много више уколико је ћирилични текст штампан курзивом; грешака има и у италијанском тексту, где се релативно често група „m“ препознаје као „n“). Сканирани текст чува се и у формату pdf/a, тако да је могуће проверити у овом документу, који је нека врста „фотокопије“ оригиналног текста, да ли је грешка настала у фази препознавања и конверзије текста (OCR) и да ли и оригинални, штампани текст садржи штампарску грешку; такође, уз пажљиво читање може се утврдити и да ли је правописна неправилност резултат ауторове интенције – у том случају правописну неправилност не треба уклањати. Документ у формату pdf/a је неопходан уколико дело, штампано на ћирилици, садржи нетранскрибован текст на страном језику, јер се такав текст конвертује у електронски облик у виду неразумљивих симбола; такав текст треба унети ручно у електронски текст након увида у папирно издање или у документ у формату pdf/a. Неке грешке, опет, могу се уклонити само пажљивим читањем: на пример, програм за проверу правописа не може да препозна као грешку низ знакова попут „ви, дакле, иде те“ (уместо „ви, дакле, идете“, без размака): таквих грешака, међутим, релативно је мало. Приређивач електронског текста може да тражи одређене грешке циљано уколико примети да у грешкама постоји образац који се понавља. Поступак исправљања и дотеривања електронског текста не одузима много времена и може се обавити за неколико сати рада, зависно од дужине текста и квалитета сканираног материјала.

Преводи на италијански језик могу се набавити у електронском облику, директно од издавача, захваљујући чињеници да је електронско издаваштво у Италији развијеније него у Србији. Такви текстови су по правилу и поузданији и лакши за рад, мада понекада и они могу да садрже словне грешке, тако да је препорука да се и они обраде програмом за проверу правописа пре него што се приступи самој паралелизацији.

Треба да напоменемо да се у електронском тексту губи информација о пагинацији: управо због тога треба сачувати копију сканиране књиге у облику документа pdf/a да би се, у случају потребе, дошло до информације о пагинацији.

Након фазе дигитализације, оригинал и превод треба да се паралелизују. Овим поступком два текста се обједињују у један документ и распоређују се у табелу с две колоне, тако да сваком сегменту текста у колони лево одговора текст превода у колони десно, у истом реду. Табела омогућује да се распоред оригиналног текста и превода не мења с променом величине слова, ширине табеле или величине странице. Посао паралелизације текстова обавља се помоћу посебних програма; у нашем раду забележили смо позитивна искуства се програмом LF Aligner 4.05.<sup>18</sup> Овај програм сегментује текст оригинала и превода на нивоу реченица, а реченицом сматра текст садржан између две тачке (или текст од тачке до узвичника, односно до упитника). Уколико се текст оригинала и превода разликују у интерпункцији, та разлика утиче на прецизност паралелизације. Програм LF Aligner нуди могућност да се паралелизовани текст сачува као табела програма Excel: користећи ту опцију, довољно је селектовати колоне које садрже паралелни текст и копирати их у неки од програма за обраду текста. Овај поступак је двосмеран: табелу из текстуалног документа у Word-у, на пример, могуће је вратити неизмењену у Excel. Након тога прелази се на трећу фазу, исправљање погрешно паралелизованих места. Овај посао такође не захтева много времена и обавља се директно у неком од програма за обраду текста (Word, OpenOffice итд.). Делове оригинала или превода који нису смештени у ред ком нападају треба преместити у одговарајући ред.

Потешкоћу у раду представљају фусноте, нарочито у преводима на италијански: текст фуснота мора да се измести у сам текст - на пример, у угласте заграде које стоје одмах након речи код које је стајала ознака за фусноту.

Из наведеног закључује се да дигитализација текста, његова конверзија у електронски облик који је погодан за текстуалну обраду, затим исправљање евентуалних грешака у препознавању слова и сама паралелизација јесу технички задаци који не захтевају посебна информатичка знања.

Имајући у виду релативну једноставност поступка израде електронског корпуса, у процес стварања електронског корпуса српских књижевних дела и њихових превода на италијански може се

---

18 Аутор програма је Андраш Фаркаш, а сам програм доступан је на адреси <http://sourceforge.net/projects/aligner/>

укључити више извршилаца и за релативно кратко време, без посебне обуке и без материјалних улагања, могуће је проширити и обогатити поменути корпус. У тренутку писања овог чланка у обради је више десетина књига српских писаца, тако да до краја 2015. очекујемо да сакупимо седамдесетак паралелних српско-италијанских текстова, што би допринело важности корпуса и повећало поље истраживања како у оквирима двојезичних анализа, тако и у оквиру једнојезичних проучавања, што је, такође, једна од могућности које нуди инструмент паралелног корпуса. Следећи, планирани корак је сакупљање двојезичног електронског материјала за корпус обрнутог смера, који ће садржати италијанска књижевна дела преведена на српски, односно на хрватски (мислимо на дела италијанске књижевности која су од педесетих до осамдесетих година прошлог века преведена у хрватској редакцији а након распада Југославије нису преведена на српски).

Корпус ових димензија већ у постојећем облику може да послужи као нека врста српско-италијанског двосмерног речника књижевног језика, што је нарочито важно ако се узме у обзир чињеница да још увек не постоји велики српско-италијански речник и да би двојезични корпус састављен од више десетина, па можда и стотина оригиналних књижевних дела и њихових превода могао да послужи као значајна база за решавање разних лексикографских недоумица. Чињеница да језички материјал корпуса није лематизован и уподобљен стандардној форми речника одузима само део од његове практичне и информативне димензије: корисник који има основна знања српске и италијанске морфологије има у овом корпусу драгоцено помагало у свом раду.

На крају, важно је да нагласимо да двојезични српско-италијански корпус паралелизованих књижевних текстова може да послужи као основа за стварање вишејезичног корпуса који би садржао преводе наших дела на више страних језика. Да би се ова идеја остварила неопходно је да се уз постојећи српски текст додају колоне с паралелизованим текстом на неком другом страном језику, као у примеру ниже:<sup>19</sup>

---

19 Вишејезични текст Макијавелијевог „Владара“ преузет је са сајта Андраша Фаркаша ([http://www.farkastranslations.com/bilingual\\_books.php](http://www.farkastranslations.com/bilingual_books.php)), на ком је доступно више сличних, вишејезичних текстова за која су ауторска права истекла.

Italian	English	Hungarian	German
Il Principe	The Prince	A fejedelem	Das Buch vom Fürsten
Niccolò Machiavelli	Niccolò Machiavelli	Niccolò Machiavelli	Niccolò Machiavelli
Cap. I Quot sint genera principatuum et quibus modis acquirantur. [Di quante ragioni sieno e' principati, e in che modo si acquistino]	CHAPTER I — HOW MANY KINDS OF PRINCIPALITIES THERE ARE, AND BY WHAT MEANS THEY ARE ACQUIRED	I. Hányféle az egyeduralom, és hogyan lehet megszerezni	1. Verschiedene Arten der Herrschaft, und Wege, zu ihr zu gelangen.
Tutti li stati, tutti e' dominii che hanno avuto et hanno imperio sopra li uomini, sono stati e sono o repubbliche o principati.	All states, all powers, that have held and hold rule over men have been and are either republics or principalities.	Az uralom és birtoklás minden fajtája, amely valaha is hatalmában tartotta az embereket, köztársaság vagy egyeduralom volt.	Alle Staaten und Gewalten, welche Herrschaft über die Menschen gehabt haben und noch haben, sind Republiken oder Fürstenthümer.
E' principati sono o ereditarii, de' quali el sangue del loro signore ne sia suto lungo tempo principe, o e' sono nuovi.	Principalities are either hereditary, in which the family has been long established; or they are new.	Az egyeduralom vagy örökletes, amikor a fejedelem vér szerint a már régóta uralkodó családhoz tartozik, vagy újkeletű.	Diese sind entweder ererbt, indem sie von dem Geschlechte des Herrschers schon lange regiert worden sind; oder sie sind neu errichtet.

Макијавелијев текст у горњој табели не подлеже прописима који регулишу ауторска права; друго је питање за преводе, који су свакако новијег датума и вероватно су заштићени наведеним прописима. Уопште узев, сматрамо да се у вези с нашим корпусом не би могло говорити о кршењу ауторских права уколико би увид у текстове оригинала и превода био ограничен на пет до десет речи лево и десно од тражене речи. Сличан приступ примењен је у корпусу GRALIS, у ком је увид у Андрићева дела ограничен на 60 словних места лево и десно од тражене речи<sup>20</sup>. За локално консултовање корпуса, унутар мреже Филолошког факултета, величина видљивог контекста може се повећати; у сваком случају, сам корисник не може ни у ком случају да преузме интегрални текст. На овај начин аутори и издавачи не би требало да буду оштећени; сматрамо да много већа штета по њих настаје организованим или индивидуалним (врло често, аматерским)

20 [http://www-gewi.uni-graz.at/gralis-alt/0.Projektarium/Gralis-Korpus/Andric-Korpus/Andric-Korpus\\_Koncepcija.html](http://www-gewi.uni-graz.at/gralis-alt/0.Projektarium/Gralis-Korpus/Andric-Korpus/Andric-Korpus_Koncepcija.html)

сканирањем и дистрибуирањем књига преко интернета, при чему за интегритет текста не постоје гаранције (тако, дешава се да се стави у оптицај само половина неког дела; нека страница је прескочена приликом сканирања, а могућа је и појава слободних интервенција на самом тексту<sup>21</sup>). Када је реч о италијанским делима, велики број књига може се преузети помоћу програма peer-to-peer (emule, на пример).

Накнадно ширење српско-италијанског корпуса текстовима превода на друге језике представљао би значајан подстицај за традицијологију, контрастивну анализу, лексикографију, науку о језику и науку о књижевности, али и за србистику. Имајући у виду једноставност поступка приређивања паралелних текстова и практично непостојеће трошкове поступка, не постоје препреке за израду мултилингвалног корпуса.

До сада, у овом раду, бавили смо се књижевним текстовима и њиховом паралелизацијом. Међутим, технички поступак који смо приказали има другу, заправо старију и практичнију димензију. На исти начин на који смо паралелизовали више књижевних текстова и објединили их у један документ-корпус у Excel-у могуће је паралелизовати и сакупити оригиналне стручне текстове (правне, економске, медицинске, машинске струке итд.) и њихове преводе на страни језик. Да таквај поступак није новост потврђује сам програм за паралелизацију LF Aligner 4.05, који је настао као инструмент за стручне преводиоце, алат за израду „преводиначких меморија“ (Translation Memory) које се употребљавају као база података уз програме за компјутерско превођење (CAT – Computer Assisted/Aided Translation). Сакупљени и „архивирани“ паралелизовани текстови оригинала и превода стручних текстова могу представљати у пракси верификовану терминолошку базу за разне области, у којој се чувају интегрални текстови а не само преводни еквиваленти (као у преводиначким меморијама). Подсећајући на изворну примену паралелних текстова и програма за паралелизацију, на неки начин затварамо круг и овим радом указујемо на могућност да се практична димензија наведених поступака, изворно везана за стручно превођење, прошири и на поље књижевног превода и лингвистике.

---

21 О таквим интервенцијама лично нас је обавестио проф. др Зоран Живковић говорећи о својим романима који се размењују преко интернета – наиме, учесници у „размени“ понекад мењају текст или га дописују према властитом укусу.

## Литература

- Andorno-Rastelli (ed.). 2009. *Corpora di Italiano L2: Tecnologie, metodi, spunti teorici*. Perugia, Guerra Edizioni.
- Barbera, Manuel. 2013. *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*. Qu.A.S.A.R. s.r.l., Milano.
- Gandin, Stefania. 2009. Linguistica dei corpora e traduzioni: definizioni, criteri di compilazione e implicazioni di ricerca dei corpora paralleli. *Annali della Facoltà di Lingue e Letterature Straniere*. Vol. 5, 2005 (2009), стр. 133-152.
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Rodopi, Amsterdam and New York.
- Laviosa, Sara. 1999. Come studiare e insegnare l'italiano attraverso i corpora. *Italica*, Vol. 76, No. 4. Linguistics and Pedagogy (Winter, 1999), pp. 443-453.
- Melamed, Dan. 2001. *Empirical Methods for Exploiting Parallel Texts*. The MIT Press. Cambridge, Massachusetts. London, England, 2001.
- Moderc, Saša. 2014a. I testi letterari paralleli e la valutazione della traduzione: il caso dell'interpunzione. *Наслеђе. Часопис за језик, књижевност, уметност и културу*. Година XI, Број 29. Филолошко-уметнички факултет. Крагујевац 2014, стр. 203-216.
- Moderc, Saša. 2014b. Su un modo di tradurre l'avverbio serbo "inače" in italiano: il caso dell'equivalente "altrimenti". *Italica Belgradensia* 1/2015, стр. 61-79.
- Модерц, Саша. (у штампи). Паралелизовани текстови и учење на даљину. *Научни скупи Дигиталне библиотеке и дигитални архиви*. Универзитет у Београду, Универзитет у Новом Саду. Београд-Нови Сад, 7-8. април 2014.
- Pouliquen et alii. 2006. (Pouliquen Bruno, Widiger Anna, Varga Daniel, Erjavec Tomaz, Ignat Camelia, Tufis Dan, Steinberger Ralf). The JRC-Acquis: A Multilingual Aligned Parallel Corpus with 20+ Languages. *The 5th International Conference on Language Resources and Evaluation – Proceedings*, стр. 2142-2147.
- Ристовић З. (2012). Од корпуса до учионице: примена паралелизованих текстова у настави енглеског језика у основној школи. *ИНФОШЕКА*, број 2, година XIII, децембар 2012.

- Toury, Gideon (1995). *Descriptive Translation Studies and beyond*. Benjamins 1995.
- Véronis, Jean (ed.). 2000. *Parallel Text Processing. Alignment and Use of Translation Corpora*. Springer Science+Business Media Dordrecht.

Saša Moderc

### Summary

#### AN ELECTRONIC BILINGUAL CORPUS OF SERBIAN LITERARY WORK AND THEIR TRANSLATION INTO ITALIAN

Building an electronic bilingual corpus containing Serbian literary texts and their translations in Italian, parallelized at sentence or paragraph level, allows a deep insight, even in the initial stages of digitalization, into the approach adopted in translating Serbian prose into the Italian language. The analysis of the linguistic material present in this Serbian-Italian corpus can be focused on translational and stylistic aspects of one or more translations at the same time; it is possible to study the work of one or more translators and to compare different translations of the same literary text, produced by different translators. Linguistic data contained in this bilingual corpus can be studied from the perspective of not only contrastive analysis but also lexicography, as a rich basis of translational equivalents and lexical solutions, or it can be applied in foreign language teaching, both for Italian as L2 and Serbian as L2.